QMaxCal: Path-Space Regularization for Open Quantum Control via Girsanov's Theorem

Merijn Moody [*1 2 3] **Zier Mensch** [*2 4] **Miranda Cheng** [2 5 3] **Peter G. Bolhuis** [6] **Max Welling** [7]

Abstract

Reliable quantum control in the presence of decoherence requires policies that combat the effect of environmental noise on the controlled dynamics. Open quantum systems under continuous monitoring generate classical measurement records whose drift depends on the noise experienced by the system; the records of two evolutions sharing the same decoherence channels differ only in this drift, so Girsanov's theorem yields a closed-form, differentiable estimator of the KL divergence between their trajectory distributions. We instantiate this estimator with two physically motivated reference measures, yielding two regularizers that both drive the system toward states where the effects of decoherence are minimal: the *Wiener KL* ($\mathrm{KL_W}$), which is empirically more effective under certain conditions on the noise model, and the *drift-variance regularizer* (R_{DV}), which works for all noise models. Both are qualitatively distinct from existing penalties on control fluence or smoothness: they penalize the observable consequences of control on the decoherence channels rather than the control amplitude itself. The regularizers outperform unregularized gradient-based and reinforcement-learning baselines across a range of open quantum systems—including single- and multi-qubit benchmarks and a multi-qubit chain calibrated to a published snapshot of the IBM Kingston processor—along several axes of evaluation: final-state fidelity, robustness to mismatch in the assumed noise model (gains grow from $+17$ pp at training noise to $+27$ pp under $2.5\times$ noise mismatch), and occupation of forbidden states. The regularizers reduce infidelity by up to 50%, with $\sim$16% gains on the calibrated IBM Kingston chain.

[*]Equal contribution [1]Dutch Institute for Emergent Phenomena, University of Amsterdam, Netherlands [2]Institute of Physics, University of Amsterdam, Netherlands [3]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Netherlands [4]Department of Physics, National Taiwan University, Taiwan [5]Institute for Mathematics, Academia Sinica, Taiwan [6]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Netherlands [7]Amsterdam Machine Learning Lab, University of Amsterdam, Netherlands. Correspondence to: Merijn Moody <merijnmoody@gmail.com>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

1. Introduction

Quantum optimal control provides control policies that underlie a wide range of tasks in quantum technology, from implementing high-fidelity gates in superconducting and trapped-ion processors (Werninghaus et al., 2021; Motzoi et al., 2009) to preparing non-classical states for sensing and metrology (Pezzè et al., 2018), designing pulses for variational quantum algorithms (Magann et al., 2021), and realizing fast qubit reset and readout (Gautier et al., 2025). Because hardware parameters, available controls, and cost landscapes vary substantially across platforms, pulse design is typically cast as a numerical optimization problem (Koch et al., 2022; Glaser et al., 2015) and solved by methods such as GRAPE (Khaneja et al., 2005), Krotov's method (Reich et al., 2012), adjoint-state techniques (Gautier et al., 2025), or reinforcement learning (Ernst et al., 2025). Improvements in the quality of these solutions translate directly into deeper achievable circuits and better-performing quantum devices.

In any realistic setting, a central challenge facing quantum optimal control is decoherence: the irreversible coupling of the system to its environment, which degrades coherence on timescales comparable to the pulse duration and ultimately limits the fidelity of any protocol (Krantz et al., 2019; Abad et al., 2022). The evolution of a system subject to environmental noise can be modeled by a stochastic Schrödinger equation, a stochastic differential equation whose noise terms capture the system's coupling to each decoherence channel. Different control protocols can route the system through regions of state space that differ substantially in their exposure to decoherence (Lidar et al., 1998; Viola et al., 1999), so the trajectory through Hilbert space, not just the target state, is an object of optimization.

We propose QMaxCal, a framework for path-space regularization that acts directly on the trajectory distribution of an

open quantum system. The SSE generates, for each decoherence channel, a classical measurement record whose drift encodes the noise experienced by the system; the records of two evolutions sharing the same decoherence channels are Itô diffusions with shared noise structure and control-dependent drifts, so Girsanov's theorem (Girsanov, 1960) yields a closed-form, differentiable estimator of the KL divergence between their path distributions. The framework can be read as a quantum-control instance of the Maximum Caliber principle (Jaynes, 1980; 1985; Pressé et al., 2013); what is new is the choice of reference measure (Wiener noise or its closest constant-drift extension) and the use of Girsanov's theorem to obtain the resulting KL in closed form. The two resulting regularizers contrast with standard penalties on pulse *fluence* (Khaneja et al., 2005; Reich et al., 2012; Villanueva & Kappen, 2024) and evolution *smoothness* (Ernst et al., 2025): they penalize the observable consequences of control on the decoherence channels rather than the control amplitude or its derivatives. We make the following contributions:

1. We derive a **closed-form estimator for the KL divergence over quantum trajectory space** via Girsanov's theorem (Eq. 8), valid for any pair of evolutions sharing the same Lindblad operators.

2. We instantiate the general estimator with two physically motivated reference measures, yielding two regularizers that act on different structural features of the noise: the **Wiener KL regularizer** ($\mathrm{KL_W}$) and the **drift-variance regularizer** (R_{DV}). Both drive the system toward states where the effects of decoherence are minimal; $\mathrm{KL_W}$ is empirically more effective under certain conditions on the noise model, while R_{DV} works for any noise model.

3. We **demonstrate the framework on five open-system control benchmarks**, including amplitude damping, STIRAP, four-level shelving structures, a synthetic qubit chain with site-dependent noise, and a six-qubit chain calibrated to a published snapshot of the IBM Kingston processor, where the regularizers reduce infidelity by up to 50% across multi-qubit benchmarks and by $\sim$16% on the calibrated IBM Kingston chain over unregularized gradient-based and constrained-RL baselines.

2. Related Work

Gradient-based quantum optimal control. GRAPE (Khaneja et al., 2005) and its open-system variant Open GRAPE (Boutin et al., 2017) compute optimal pulses via piecewise evolution and gradient ascent over density matrices. Krotov's method (Reich et al., 2012) ensures monotonic convergence. Gautier et al. (2025) introduced a scalable adjoint-state method for large open systems. Abdelhafez et al. (2019) introduced trajectory-based gradient computation using automatic differentiation through the SSE, demonstrating the quadratic scaling advantage of state vectors over density matrices and enabling GPU parallelization. Our work builds on this computational paradigm but introduces a qualitatively different objective: rather than differentiating a fidelity cost with ad hoc control penalties, we regularize with a path-space KL divergence minimizing the effects of decoherence.

Reinforcement learning. Ernst et al. (2025) cast pulse design as a constrained RL problem and outperform previous gradient based methods. Their approach introduces penalties on the fluence and smoothness of the control pulse, as well as on the smoothness of the resulting quantum evolution, with the goal of producing experimentally realizable signals and improving computational efficiency. QMaxCal complements this line of work by introducing a regularizer that acts directly on the trajectory through Hilbert space, penalizing the controlled system's coupling to the environment rather than the shape of the drive waveform; the two types of penalizers are not interchangeable, as we verify by direct ablation in Appendix G.

Maximum Caliber and path integral control. The principle of Maximum Caliber (MaxCal) (Jaynes, 1980; 1985; Pressé et al., 2013) extends Maximum Entropy to dynamical settings, selecting a path distribution that minimizes a KL divergence to a reference subject to imposed constraints. MaxCal has found applications in classical nonequilibrium statistical mechanics (Ghosh et al., 2020) and molecular dynamics (Dixit et al., 2018; Bolhuis et al., 2021; Brotzakis et al., 2021; Bolhuis et al., 2023). Applied to control with the *uncontrolled* evolution as reference, the MaxCal paradigm reduces to path-integral control (Kappen, 2005; Kappen & Ruiz, 2016), which Villanueva & Kappen (2024) adapted to open quantum systems as the Quantum Diffusion Control (QDC) algorithm and Najafi et al. (2025) recently extended to quantum chemistry. Under additional structural conditions on the noise model, the resulting KL reduces to a fluence penalty on the controls. QMaxCal complements this approach by taking the Wiener measure as reference, yielding a KL that penalizes decoherence exposure rather than deviation from free dynamics, and by applying to arbitrary Lindblad systems without QDC's structural conditions on the noise model (Section 4.2, Appendix D).

3. Background

3.1. Open quantum systems and quantum trajectories

Open quantum systems are described by *density matrices* $\rho \in \mathbb{C}^{N\times N}$. When the system interacts with a Markovian

(memoryless) environment, ρ evolves according to the Lindblad master equation (Lindblad, 1976) (see Appendix A for a self-contained introduction):

$$\dot{\rho} = -i[H^{(\theta)}(t), \rho] + \sum_{k=1}^{K} \mathcal{D}_k(\rho), \tag{1}$$

where $H^{(\theta)}(t) = H_0 + \sum_a u_a^{(\theta)}(t) H_a$ is the system Hamiltonian with time-dependent control fields $u_a^{(\theta)}(t)$ depending on parameters θ, $\{L_k\}_{k=1}^K$ are Lindblad operators that model the irreversible interactions with the environment (e.g., energy decay, phase randomization), and $\mathcal{D}_k(\rho) = L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}$ is the dissipator associated with channel k.

The master equation describes the ensemble-averaged evolution, but it can be *unravelled* into quantum trajectories describing the evolution of individual *pure states* $|\psi(t)\rangle \in \mathbb{C}^N = \mathcal{H}$ under a given interaction with the environment. A given Lindblad evolution admits multiple unravellings; throughout this work we use the diffusive (homodyne) unravelling. Writing $\alpha_k(t) := \langle\psi(t)|(L_k + L_k^\dagger)|\psi(t)\rangle$, the conditional state obeys the stochastic Schrödinger equation (SSE) (Wiseman & Milburn, 2009; Percival, 1998)

$$\begin{aligned} d|\psi\rangle = &\underbrace{\Big[-iH^{(\theta)} + \sum_k \big(\tfrac{1}{2}\alpha_k L_k - \tfrac{1}{2}L_k^\dagger L_k - \tfrac{1}{8}\alpha_k^2\big)\Big]|\psi\rangle\, dt}_{\text{deterministic drift}} \\ &+ \underbrace{\sum_k \big(L_k - \tfrac{1}{2}\alpha_k\big)|\psi\rangle\, dW_k}_{\text{stochastic diffusion}}, \end{aligned} \tag{2}$$

where $dW_k(t)$ are independent Wiener increments satisfying $\mathbb{E}[dW_k(t)] = 0$ and $\mathbb{E}[dW_j(t)dW_k(t)] = \delta_{jk}\, dt$. The SSE induces a trajectory distribution P_θ on path space; averaging over P_θ recovers the density matrix evolution: $\rho(t) = \mathbb{E}_{P_\theta}[|\psi(t)\rangle\langle\psi(t)|]$ satisfies the Lindblad master equation (1). The trajectory representation also has a numerical advantage: state vectors carry $O(\dim)$ rather than $O(\dim^2)$ entries, so SSE integrators reach roughly two qubits beyond the largest system that the Lindblad reference solver fits on a single GPU when the noise levels are not too high (Appendix H).

In experimental platforms where the environment is itself measured (circuit QED, trapped ions, optical cavities), each individual trajectory $|\psi(t)\rangle$ corresponds to a specific realization of the classical *measurement record* $I_k(t)$ obtained by monitoring decoherence channel k. The same Wiener increments $\mathrm{d}W_k$ that drive the SSE (2) also drive the measurement record (Wiseman & Milburn, 2009), with the same drift α_k appearing in both:

$$dI_k(t) = \alpha_k(t)\,\mathrm{d}t + \mathrm{d}W_k, \tag{3}$$

and a unit diffusion coefficient that is independent of the control protocol.

Equation (3) is the central object of this paper: different controls produce different drifts α_k while sharing the same noise structure. This is precisely the setting in which Girsanov's theorem yields a tractable Radon–Nikodym derivative between path distributions, as we now make precise.

3.2. Girsanov's theorem

We state Girsanov's theorem informally here; a formal derivation with all conditions verified for the quantum setting is given in Appendix B. Consider two Itô diffusion processes sharing the same noise but with different drift processes:

$$dX_t^{(i)} = a_t^{(i)}\,\mathrm{d}t + \mathrm{d}W_t, \qquad i = 1, 2, \tag{4}$$

where each drift $a_t^{(i)}$ may depend on the entire trajectory history up to time t. Writing $\Delta a_t := a_t^{(1)} - a_t^{(2)}$ for the drift difference, Girsanov's theorem (Girsanov, 1960; Øksendal, 2003) states that the Radon–Nikodym derivative between the induced path measures $P^{(1)}, P^{(2)}$ is

$$\log \frac{dP^{(1)}}{dP^{(2)}} = \int_0^T \Delta a_t\, dB_t^{(2)} - \frac{1}{2}\int_0^T (\Delta a_t)^2\,\mathrm{d}t, \tag{5}$$

and the KL divergence takes the closed form

$$D_{\mathrm{KL}}[P^{(1)} \| P^{(2)}] = \frac{1}{2}\,\mathbb{E}_{P^{(1)}}\left[\int_0^T (\Delta a_t)^2\,\mathrm{d}t\right]. \tag{6}$$

3.3. Decoherence-free subspaces

For structured noise models there may exist subspaces on which the Lindblad operators $\{L_k\}_{k=1}^K$ act trivially. A subspace $\mathcal{S} \subseteq \mathcal{H}$ is a *decoherence-free subspace* (DFS) (Lidar et al., 1998) if each Lindblad operator acts as a scalar on it: $L_k|\phi\rangle = \ell_k|\phi\rangle$ for all $|\phi\rangle \in \mathcal{S}$ for all k, with constants $\ell_k \in \mathbb{C}$ depending only on the channel. On such a subspace the dissipator vanishes, and the stochastic back-action $(L_k - \frac{1}{2}\alpha_k)|\psi\rangle\,\mathrm{d}W_k$ in the SSE (2) reduces to a global stochastic phase $i\,\mathrm{Im}(\ell_k)|\psi\rangle\,\mathrm{d}W_k$ acting identically on every state in $\mathcal{S}$, and so is physically unobservable[1]; it is identically zero when $\ell_k \in \mathbb{R}$, as in all benchmarks of this paper. A trajectory that remains in $\mathcal{S}$ therefore undergoes unitary decoherence-free evolution, with drift

$$\alpha_k(t) = \langle\psi(t)|(L_k + L_k^\dagger)|\psi(t)\rangle = 2\,\mathrm{Re}(\ell_k), \tag{7}$$

[1] In quantum mechanics, two state vectors that differ only by a global phase $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ represent the same physical state: they yield identical probabilities for every measurement, the same density matrix $\rho = |\psi\rangle\langle\psi|$, and the same expectation values for all observables. A stochastic global phase therefore has no observable consequence on the trajectory.

constant in time and across realizations. Two cases arise: 1) when $\ell_k = 0$ for all k, the subspace lies in the joint kernel[2], $\bigcap_k \ker(L_k)$, and the drift vanishes and 2) when $\mathrm{Re}(\ell_k) \neq 0$ for some k, the state is decoherence-free but has nonzero constant drift.

4. Methods

4.1. Girsanov KL divergence over quantum trajectory space

Applying Girsanov's theorem (Section 3, Eq. (6)) to the measurement records of two evolutions sharing the same Lindblad operators yields the central result of this paper. Let $P^{(1)}$ and $P^{(2)}$ denote the path measures induced by the SSE (2) under two control protocols, two Hamiltonians, or a controlled-vs.-uncontrolled pair; the only requirement is that the Lindblad operators $\{L_k\}$ are shared, which is automatically satisfied for any pair of control protocols on the same physical system. Under each evolution $i \in \{1, 2\}$, the measurement record for channel k has drift $\alpha_k^{(i)}(t) = \langle\psi^{(i)}(t)|(L_k + L_k^\dagger)|\psi^{(i)}(t)\rangle$ and unit diffusion coefficient. Writing $\Delta\alpha_k(t) := \alpha_k^{(1)}(t) - \alpha_k^{(2)}(t)$ for the per-channel drift difference, Girsanov's theorem then yields:

$$D_{\mathrm{KL}}[P^{(1)}\|P^{(2)}] = \frac{1}{2}\sum_{k=1}^{K} \mathbb{E}_{P^{(1)}}\left[\int_0^T \Delta\alpha_k(t)^2\,\mathrm{d}t\right], \quad (8)$$

where the subscript denotes an average with respect to $P^{(1)}$.

4.2. The Wiener KL and drift-variance regularizers

Consider the trajectory distribution P_θ introduced above in Section 3. The general estimator (8) requires a choice of reference measure. The simplest is the *Wiener measure* P_W (the law of pure Brownian motion with zero drift). Setting $\mathrm{KL_W} := D_{\mathrm{KL}}[P_\theta\|P_W]$, the estimator (8) reduces to

$$\mathrm{KL_W} = \frac{1}{2}\sum_{k=1}^{K} \mathbb{E}_{P_\theta}\left[\int_0^T \alpha_k^{(\theta)}(t)^2\,\mathrm{d}t\right], \quad (9)$$

and requires no reference SSE simulation. A more permissive reference is the closest constant-drift process: for $c \in \mathbb{R}^K$, let P_c denote the law of $dY_t = c\,\mathrm{d}t + \mathrm{d}W_t$. We define the *drift-variance regularizer* as $R_{\mathrm{DV}} := \min_{c\in\mathbb{R}^K} D_{\mathrm{KL}}[P_\theta\|P_c]$. The KL is quadratic in c, and minimization yields

$$R_{\mathrm{DV}} = \frac{1}{2}\sum_{k=1}^{K} \mathbb{E}_{P_\theta}\left[\int_0^T \left(\alpha_k^{(\theta)}(t) - \bar{\alpha}_k\right)^2 \mathrm{d}t\right], \quad (10)$$

with optimum at $c_k^\star = \bar{\alpha}_k := T^{-1}\mathbb{E}_{P_\theta}[\int_0^T \alpha_k^{(\theta)}(t)\,\mathrm{d}t]$, the time-and-ensemble drift mean (Appendix C).

Relating the two regularizers to the DFS. Both regularizers vanish on certain DFS evolutions, but they identify different subsets. The Wiener KL (9) vanishes on states with $\alpha_k = \langle L_k + L_k^\dagger\rangle = 0$, which includes the kernel component $\bigcap_k \ker(L_k)$ ($\ell_k = 0$) but not DFS evolutions with $\mathrm{Re}(\ell_k) \neq 0$. The condition $\alpha_k = 0$ can also hold on states outside $\ker(L_k)$ that incur full stochastic backaction $(L_k - \frac{1}{2}\alpha_k)|\psi\rangle\,\mathrm{d}W_k$ — Wiener KL does not penalize these.

The drift-variance regularizer (10) is the natural extension to arbitrary ℓ_k: it vanishes on any DFS, since (7) gives $\bar{\alpha}_k = \alpha_k(t) = 2\,\mathrm{Re}(\ell_k)$ identically. The converse also holds: a trajectory along which $\alpha_k(t)$ is constant in time and across noise realizations must lie in a DFS, since any deviation from a simultaneous eigenstate of the $\{L_k\}$ produces nonzero stochastic backaction, which generically yields different drifts across realizations. Constant drift therefore characterizes DFS evolutions exactly, and R_{DV} identifies them all. The two regularizers are complementary: $\mathrm{KL_W}$ has the stronger inductive bias when $\bigcap_k \ker(L_k)$ is nonempty and reachable, while R_{DV} is the appropriate choice for systems whose protected manifold has nonzero constant drift or where no joint kernel exists.

Relationship to QDC's fluence penalty. The fluence penalty of Villanueva & Kappen (2024) arises from a different instantiation of (8): with the *uncontrolled* evolution as reference and an additional structural condition on the noise model, the resulting KL reduces exactly to a noise-weighted control fluence (Appendix D). The structural condition is restrictive — none of our benchmarks satisfy it — and Villanueva & Kappen (2024) consequently use QDC primarily as an annealing framework for closed-system control rather than for combating decoherence.

4.3. The QMaxCal objective

We instantiate the framework on *state transfer* tasks: starting from an initial state $|\psi(0)\rangle$, the goal is to drive the system to a target state $|\phi_{\mathrm{target}}\rangle$ at a fixed final time T. The quantity of interest is the final-state fidelity $F(\psi(T)) = |\langle\phi_{\mathrm{target}}|\psi(T)\rangle|^2$, and combining either regularizer of Section 4.2 with this fidelity objective yields the QMaxCal loss

$$\begin{aligned}\mathcal{L}(\theta) = 1 - \mathbb{E}_{P_\theta}[F(\psi(T))] \\ + \lambda_{\mathrm{KL_W}}\,\mathrm{KL_W} + \lambda_{R_{\mathrm{DV}}}\,R_{\mathrm{DV}} + \lambda_{\mathrm{flu}}\,\mathrm{Flu},\end{aligned} \quad (11)$$

where $\mathrm{Flu} = \sum_a \int_0^T |u_a^{(\theta)}(t)|^2\,\mathrm{d}t$ is a standard control-fluence penalty (Khaneja et al., 2005; Reich et al., 2012; Villanueva & Kappen, 2024), included to bound the control amplitude and produce realistic control policies. The hyperparameters $\lambda_{\mathrm{KL_W}}, \lambda_{R_{\mathrm{DV}}} \geq 0$ control the trade-off between

[2]The kernel of an operator L is the subspace $\ker(L) = \{|\psi\rangle : L|\psi\rangle = 0\}$; the joint kernel $\bigcap_k \ker(L_k)$ consists of states annihilated by all Lindblad operators simultaneously.

path-space regularization and fidelity; setting both to zero recovers unconstrained gradient-based optimal control (Abdelhafez et al., 2019). The parameters θ may correspond to piecewise-constant pulse amplitudes, neural-network weights, or basis-function coefficients; optimization proceeds by gradient descent on θ with all terms evaluated on the same sampled SSE trajectory ensemble. We integrate the SSE with one of two diffusive schemes (Euler–Maruyama or an exponential split-step integrator); Appendix H describes both schemes, their trade-offs, and the per-experiment choice.

5. Experiments

We evaluate QMaxCal on five open-system control benchmarks across three structural families, comparing against an unregularized gradient baseline (the GRAPE-style trajectory optimizer of Abdelhafez et al. (2019)) and the PPO method of Ernst et al. (2025), with PPO hyperparameters tuned per-benchmark via a two-phase sweep (Appendix F). Fidelity differences are reported in percentage points (pp), $F_A - F_B$; relative infidelity reductions as percentages of $1 - F_B$. Figure 4 (Appendix A) shows the level structure of all five benchmarks for reference.

5.1. Single-qubit amplitude damping

Amplitude damping models a single qubit losing energy to its environment — the dominant noise process in superconducting, trapped-ion, and atomic platforms, where it sets the T_1 relaxation time (Krantz et al., 2019; Kubica et al., 2023). Population flows irreversibly from the excited state $|1\rangle$ to the ground state $|0\rangle$ at rate γ, via the Lindblad operator $L = \sqrt{\gamma}\,\sigma_-$ where $\sigma_- = |0\rangle\langle 1|$ is the qubit lowering operator. Since $\sigma_-|0\rangle = 0$, the kernel $\ker L = \mathrm{span}\{|0\rangle\}$ is one-dimensional and consists of the ground state alone (Figure 4a); the Wiener KL pulls the trajectory toward it. Regularized policies mirror the strategies of active qubit reset and just-in-time state preparation used in real hardware (Magnard et al., 2018; Egger et al., 2018), where ancillas are deliberately held at $|0\rangle$ until needed to limit T_1 exposure.

We consider the task $|+\rangle \to |Y\rangle$ — preparing the superposition $|Y\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$ from $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ under two control fields that drive rotations of the qubit (Pauli operators σ_x and σ_y, defined in Appendix A). Both endpoints sit on the Bloch-sphere equator with equal $|0\rangle$ and $|1\rangle$ populations, so a noise-naive geodesic stays on the equator and is fully exposed to decoherence; a decoherence-aware policy must detour through $|0\rangle \in \ker L$. The benchmark is therefore a clean test of whether $\mathrm{KL_W}$ routes the trajectory through the kernel. We sweep the dimensionless decoherence strength $\gamma T \in \{0.1, 0.5, 1, 2, 5\}$ at fixed $T = 1$ with $n_{\mathrm{trajs}} = 256$ stochastic trajectories per gradient step, training each policy for 5000 steps. We compare the unregularized baseline ($\lambda_{\mathrm{KL_W}} = \lambda_{R_{\mathrm{DV}}} = 0$), Wiener KL ($\lambda_{\mathrm{KL_W}} = 5$), drift-variance ($\lambda_{R_{\mathrm{DV}}} = 5$), and PPO on fidelity (Appendix E.2, Table 2) and SSE-trajectory population variance (Appendix E.2, Table 3), both computed using the same exact-Lindblad solver. Figure 1 confirms the predicted behavior. At $\gamma T = 2$, the Wiener KL policy routes the state through $|0\rangle \in \ker L$ where the SSE drift and noise both vanish, and the time-integrated population variance $\sum_k \int_0^T \mathrm{Var}[P_k(t)]\,dt$ drops $15\times$, from 0.0321 ± 0.0024 (baseline) to 0.0021 ± 0.0000 (Wiener KL). Because fidelity is an average over trajectories and the optimum is a delta on $|\phi_{\mathrm{target}}\rangle$ with zero variance, this contraction translates directly into fidelity gain. The Wiener KL gap over baseline is largest in the intermediate window $\gamma T \in [0.5, 2]$: at lower noise decoherence barely affects the final state, and at higher noise the bath alone drags every trajectory onto $|0\rangle$. Drift-variance trails the baseline by 0.25 to 4pp fidelity: Figure 5 shows that the Wiener KL integrand $\mathbb{E}_\psi[\alpha(t)^2]$ remains near the baseline level, confirming that drift-variance stabilizes α to a roughly constant non-zero value rather than driving the trajectory into $\ker L$. PPO trails Wiener KL at every γT, with the gap widening from 0.1% at low noise to over 2% at $\gamma T = 5$.

5.2. Lossy state avoidance: STIRAP and the diamond system

Both systems in this section share a structural pattern: the optimizer must move population between two states that are not directly coupled, and the available routes pass through one or more lossy intermediate states. The Wiener KL identifies the lossy states through their nonzero drift and pulls the trajectory toward routes that minimize their occupation.

STIRAP. Stimulated Raman adiabatic passage (STIRAP) is widely used for transferring population between two long-lived states across atomic, molecular, and solid-state platforms (Vitanov et al., 2017; Bergmann et al., 2019). The standard model is a three-level Λ system with metastable ground states $|g_1\rangle, |g_2\rangle$ coupled through a lossy excited state $|e\rangle$ that decays back to $|g_1\rangle$ at rate γ via $L = \sqrt{\gamma}|g_1\rangle\langle e|$ (Figure 4b). The two ground states are not directly coupled, so any protocol must route population through the $|e\rangle$ coupling; at finite protocol speed some $|e\rangle$ population is unavoidable, and a good protocol minimizes it. For this Lindblad operator the drift $\alpha = \langle L + L^\dagger\rangle$ vanishes on $|g_1\rangle$ and $|g_2\rangle$ but is nonzero on any state with $|e\rangle$ population, so the Wiener KL directly penalizes occupation of the lossy level — without the reward shaping that conventional methods require (Abdelhafez et al., 2019; Ernst et al., 2025).

At both noise levels we tested, the baseline and Wiener KL reach statistically indistinguishable fidelity. The regularizer's effect shows up instead in $|e\rangle$ occupation, which governs photon scattering and protocol viability outside the

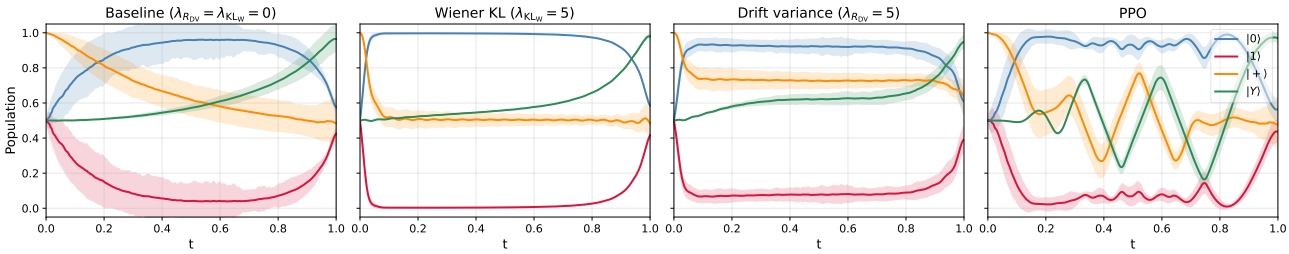

Figure 1. **SSE-trajectory population variance at $\gamma T = 2$.** Mean populations across 128 SSE samples (lines) with $\pm 1\sigma$ bands for the baseline ($\lambda_{\mathrm{KL_W}} = \lambda_{R_{\mathrm{DV}}} = 0$), Wiener KL ($\lambda_{\mathrm{KL_W}} = 5$), drift-variance ($\lambda_{R_{\mathrm{DV}}} = 5$), and PPO. Wiener KL contracts the time-integrated population variance from 0.0321 (baseline) to 0.0021, a $15\times$ reduction, consistent with the policy routing the state along $\ker L$ where both the SSE drift and noise vanish. Drift-variance reduces the variance relative to baseline but stabilizes outside $\ker L$.

strict adiabatic limit (Vitanov et al., 2017; Bergmann et al., 2019). At $\gamma T = 10$, Wiener KL reduces peak $|e\rangle$ population by 56% (from 0.097 to 0.043) and time-integrated $|e\rangle$ exposure by 35% (from 0.026 to 0.017). At $\gamma T = 1$ both metrics are already low under the baseline and the regularizer's effect is correspondingly smaller (Appendix E.3, Table 5).

Diamond system. A complementary structural pattern arises when the computational states themselves leak slowly but a nearby stable state is available as a refuge. This setup is common across qubit architectures: metastable trapped-ion qubits (Allcock et al., 2021; Yang et al., 2022), erasure-conversion qubits (Wu et al., 2022; Levine et al., 2024), and DFS gate designs (Ivanov et al., 2010) all route control through stable subspaces to avoid lossy ones. We model it as a four-level system with two *lossy* computational states $|b\rangle, |t\rangle$ (the source and target of the protocol), a *safe* auxiliary state $|d\rangle$, and a *dump* state $|\mathrm{dump}\rangle$ that collects population leaked from $|b\rangle$ or $|t\rangle$ (Figure 4c). The leak acts identically on both computational states, with rate γ: $L_b = \sqrt{\gamma}\,|\mathrm{dump}\rangle\langle b|$ and $L_t = \sqrt{\gamma}\,|\mathrm{dump}\rangle\langle t|$. Three controls couple adjacent levels: Ω_{bt} on $|b\rangle \leftrightarrow |t\rangle$ (the direct but lossy route), and Ω_{bd}, Ω_{dt} that together provide an indirect route via $|d\rangle$. The task is $|b\rangle \to |t\rangle$.

The unregularized baseline gets trapped in a local minimum that routes population through the lossy direct coupling, reaching only $F = 0.665$ with $\sim 30\%$ accumulating in $|\mathrm{dump}\rangle$ (Figure 2, left). Wiener KL forces the optimizer to discover the route through $|d\rangle$, improving fidelity to $F = 0.834$ (+17 pp, Appendix E.4, Table 6; Figure 2, middle). PPO trails at $F = 0.605$.

Wiener KL policies are also robust to noise-model mismatch in a striking way: a policy trained at $\gamma_{\mathrm{train}} = 2$ and evaluated at $\gamma_{\mathrm{test}} = 5$ — a $2.5\times$ stronger noise level than seen at training — reaches $F = 0.665$, matching the *baseline's training-noise fidelity* (+27 pp over the baseline at the same test noise). The advantage grows monotonically with γ_{test}: from +10 pp at the easiest setting to +27 pp at the hardest, with the baseline and PPO collapsing toward $F \sim 0.3$ while Wiener KL stays above 0.65 (Figure 2, right; full statistics in Appendix E.4, Table 7).

5.3. Drift-variance on superconducting qubit chains

Excitation transfer along a chain of nearest-neighbor qubits is a canonical primitive for short-range quantum communication (Bose, 2003; Lorenzo et al., 2015; Lyakhov & Bruder, 2005). The task is to move a single-qubit excitation from one end of an N-qubit chain to the other, using $N-1$ tunable nearest-neighbor couplings as controls: when active, control H_a swaps the excitation between qubits $a-1$ and a, leaving the rest of the chain untouched. Each site is subject to two local noise channels — dephasing (which destroys $|0\rangle/|1\rangle$ superpositions without changing populations, at rate κ_i) and amplitude damping (the T_1 process from §5.1, rate γ_i) — with site-specific rates that vary substantially across real hardware (Krantz et al., 2019). An effective protocol must therefore route the excitation around the noisier sites rather than treating the chain uniformly. Lindblad operators and parameter values are given in Appendices E.5 and E.6.

Asymmetry sweep. On a four-qubit chain ($\dim = 16$, eight simultaneous Lindblad channels; Figure 4d) we sweep the edge-to-interior dephasing ratio $\rho = \kappa_{\mathrm{edge}}/\kappa_{\mathrm{int}} \in \{1, 1.5, 2, 4, 8\}$, where $\kappa_{\mathrm{edge}} = \kappa_0 = \kappa_3$ are the dephasing rates on the chain endpoints and $\kappa_{\mathrm{int}} = \kappa_1 = \kappa_2$ those on the interior sites. Because of the high noise levels in this regime, training is unstable, so we use a conservative $\lambda_{R_{\mathrm{DV}}} = 0.02$. All training results are reported in Appendix E.5. Asymmetry opens a gap in favor of drift-variance: for $\rho \in \{1, 1.5, 2\}$ the two policies are statistically indistinguishable; at $\rho = 4$ drift-variance gains +4.5pp (0.52 ± 0.13 vs. 0.48 ± 0.02); at $\rho = 8$ the gap widens to +20pp (0.42 ± 0.11 vs. 0.22 ± 0.04), where the baseline collapses. The mechanism is visible in the populations (Figure 3): drift-variance parks an increasing fraction of its time-integrated excitation on a single low-noise interior site as the edges become noisier, while the baseline follows a less optimal population distribution.

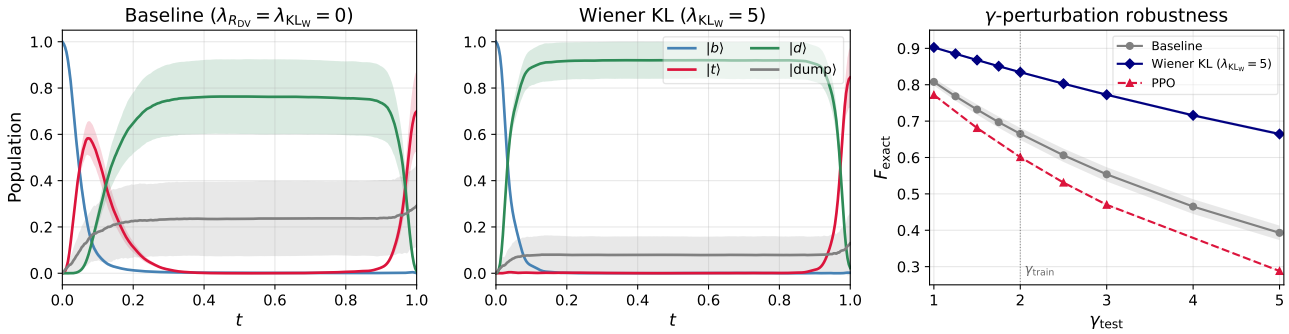

Figure 2. **Diamond system,** $\gamma_{\text{train}} = 2$**,** $T = 1$**.** *Left, middle:* SSE-trajectory populations (mean across 128 samples, $\pm 1\sigma$) for baseline ($F = 0.665$) and Wiener KL ($\lambda_{\text{KL}_\text{W}} = 5$, $F = 0.834$); the regularizer routes population through $|d\rangle \in \ker L$ rather than the lossy direct coupling. *Right:* fidelity vs. γ_{test} for policies trained at $\gamma = 2$; full statistics in Appendix E.4.

Hardware-calibrated case study. We further show that drift-variance regularization improves transfer fidelity on a six-qubit chain whose per-site T_1, T_2 rates are taken from a calibration snapshot of the `ibm_kingston` processor retrieved on 3 May 2026 from the IBM Quantum Platform (IBM Quantum, 2026). The chain has one site that is substantially noisier than the others (Figure 4e), placing it in the strong-asymmetry regime ($\rho \gg 1$) where the asymmetry sweep predicts drift-variance to help most. Table 1 reports fidelities; per-site excitation evolutions for all four policies are shown in Appendix E.6, Figure 7.

6. Conclusion

We derived a closed-form, differentiable estimator for the KL divergence between the trajectory distributions of any two open quantum system evolutions sharing the same decoherence channels. Instantiated with two physically motivated reference measures, this estimator yields the QMaxCal framework: a pair of regularizers that act on different structural features of the noise. The Wiener KL regularizer KL_W pulls the trajectory toward the joint kernel $\bigcap_k \ker(L_k)$ and is most effective when this manifold is reachable by control, while the drift-variance regularizer R_{DV} identifies arbitrary decoherence-free subspaces and applies to any noise model. Both are qualitatively distinct from existing penalties on control fluence or smoothness: they penalize the observable consequences of control on the decoherence channels rather than the control amplitude itself. Empirically, the two regularizers improve final-state fidelity by up to $+17$ pp on single-channel systems (rising to $+27$ pp under a $2.5\times$ noise-model mismatch) and $+20$ pp on multi-channel chains over unregularized gradient-based and constrained-RL baselines, with a $\sim$16% infidelity reduction on a six-qubit chain calibrated to a published snapshot of the IBM Kingston processor, and discover decoherence-avoiding policies without system-specific reward engineering. Because QMaxCal acts on quantum trajectories rather than density matrices, it inherits a potential quadratic memory advantage over density-matrix-based approaches: our SSE solvers reach 12-qubit ($\dim = 4096$) systems on a 40 GB A100 GPU, whereas the density-matrix Lindblad solver used by our PPO baseline runs out of memory at this size (Appendix H).

The regularizers help most when the noise model has noise structure that the policy can exploit: a reachable joint kernel (amplitude damping, diamond), a strongly lossy intermediate state (STIRAP), or pronounced site-to-site asymmetry in the decoherence rates (four qubit chain at $\rho \geq 4$, IBM Kingston). When the noise is small or if there is no DFS to exploit (four qubit chain at $\rho \in \{1, 1.5, 2\}$) the gains are small or absent. This is consistent with the DFS analysis of Section 4.2: the regularizers vanish exactly on decoherence-free evolutions and if these don't exist or are not reachable the regularization is not effective.

Limitations. Three points to note: (i) the Girsanov estimator is derived for the diffusive unravelling of the SSE; jump unravellings such as photon counting produce piecewise-deterministic trajectories, and Eq. (8) does not apply directly. (ii) The Lindblad assumption excludes non-Markovian environments. (iii) The regularizers act as a structural prior: they vanish on decoherence-free evolutions, so when no DFS-routed protocol exists between the initial and target states under the available controls the gains are small or absent.

Future directions. The diffusive Girsanov estimator can be evaluated on quantum hardware without access to $|\psi(t)\rangle$ via *contrastive density-ratio estimation*: a neural discriminator trained to separate experimental measurement records from synthetic reference samples gives a variational lower bound on $D_{\text{KL}}[P_\theta \| P_W]$ for KL_W or $D_{\text{KL}}[P_\theta \| P_{\bar{\alpha}}]$ for R_{DV} (Belghazi et al., 2018), and updating it online yields a feedback policy adaptive to noise drift. Extensions to jump unravellings (via marked-point-process change of measure) and non-Markovian settings (via system-plus-environment

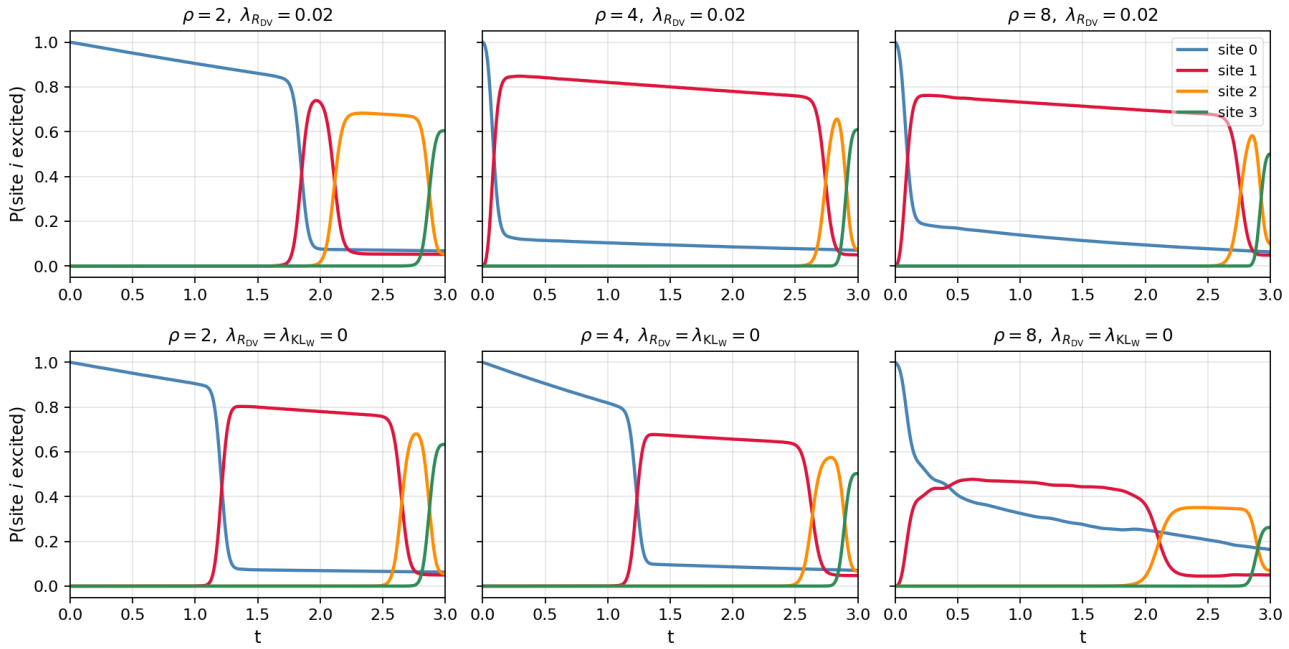

Figure 3. **Site populations across the asymmetry sweep.** Columns: $\rho \in \{2, 4, 8\}$. Rows: drift-variance ($\lambda_{R_{\mathrm{DV}}} = 0.02$, top), baseline (bottom). Each panel shows the single-site excitation population $\langle\psi(t)|(I - \sigma_z^{(i)})/2|\psi(t)\rangle$ at sites $i \in \{0, 1, 2, 3\}$ (i.e., the probability that site i holds the excitation); these four observables are the natural readout because the task is excitation transfer. Curves are drawn from a representative seed at each ρ cell (selection rule and per-seed fidelities in Appendix E.5, Table 8).

Table 1. Six-qubit excitation transfer on IBM Kingston, $T = 30\,\mu$s. Mean fidelity (three seeds). Bold marks the best in each row.

Excitation transfer	$\lambda_{R_{\mathrm{DV}}}$=$\lambda_{\mathrm{KL_W}}$=0 F	$\lambda_{R_{\mathrm{DV}}}$=16 F	$\lambda_{\mathrm{KL_W}} = 0.1$ F	PPO F
q14→q9	0.837 ± 0.004	$\mathbf{0.863 \pm 0.002}$	0.831 ± 0.004	0.604 ± 0.041

embeddings) broaden the framework's applicability.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning, with applications to quantum optimal control. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abad, T., Fernández-Pendás, J., Frisk Kockum, A., and Johansson, G. Universal fidelity reduction of quantum operations from weak dissipation. *Phys. Rev. Lett.*, 129:150504, Oct 2022. doi: 10.1103/PhysRevLett.129.150504. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.150504.

Abdelhafez, M., Schuster, D. I., and Koch, J. Gradient-based optimal control of open quantum systems using quantum trajectories and automatic differentiation. *Phys. Rev. A*, 99:052327, May 2019. doi: 10.1103/PhysRevA.99.052327. URL https://link.aps.org/doi/10.1103/PhysRevA.99.052327.

Allcock, D. T. C., Campbell, W. C., Chiaverini, J., Chuang, I. L., Hudson, E. R., Moore, I. D., Ransford, A., Roman, C., Sage, J. M., and Wineland, D. J. Omg blueprint for trapped ion quantum computing with metastable states. 119(21):214002, 2021. ISSN 0003-6951. doi: 10.1063/5.0069544. URL https://doi.org/10.1063/5.0069544.

Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, D. Mutual information neural estimation. In *International conference on machine learning*, pp. 531–540. PMLR, 2018.

Bergmann, K., Nägerl, H.-C., Panda, C., Gabrielse, G., Miloglyadov, E., Quack, M., Seyfang, G., Wichmann, G., Ospelkaus, S., Kuhn, A., Longhi, S., Szameit, A., Pirro, P., Hillebrands, B., Zhu, X.-F., Zhu, J., Drewsen, M., Hensinger, W. K., Weidt, S., Halfmann, T., Wang, H.-L., Paraoanu, G. S., Vitanov, N. V., Mompart, J., Busch, T., Barnum, T. J., Grimes, D. D., Field, R. W., Raizen,

M. G., Narevicius, E., Auzinsh, M., Budker, D., Pálffy, A., and Keitel, C. H. Roadmap on STIRAP applications. 52(20):202001, 2019. doi: 10.1088/1361-6455/ab3995. URL https://doi.org/10.1088/1361-6455/ab3995.

Bolhuis, P. G., Brotzakis, Z. F., and Vendruscolo, M. A maximum caliber approach for continuum path ensembles. 94(9):188, 2021. ISSN 1434-6036. doi: 10.1140/epjb/s10051-021-00154-3. URL https://doi.org/10.1140/epjb/s10051-021-00154-3.

Bolhuis, P. G., Brotzakis, Z. F., and Keller, B. G. Optimizing molecular potential models by imposing kinetic constraints with path reweighting. 159(7):074102, 2023. ISSN 0021-9606. doi: 10.1063/5.0151166. URL https://doi.org/10.1063/5.0151166.

Bose, S. Quantum communication through an unmodulated spin chain. *Phys. Rev. Lett.*, 91:207901, Nov 2003. doi: 10.1103/PhysRevLett.91.207901. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.207901.

Boutin, S., Andersen, C. K., Venkatraman, J., Ferris, A. J., and Blais, A. Resonator reset in circuit qed by optimal control for large open quantum systems. *Phys. Rev. A*, 96:042315, Oct 2017. doi: 10.1103/PhysRevA.96.042315. URL https://link.aps.org/doi/10.1103/PhysRevA.96.042315.

Brotzakis, Z. F., Vendruscolo, M., and Bolhuis, P. G. A method of incorporating rate constants as kinetic constraints in molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 118(2):e2012423118, 2021. doi: 10.1073/pnas.2012423118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2012423118.

Dixit, P. D., Wagoner, J., Weistuch, C., Pressé, S., Ghosh, K., and Dill, K. A. Perspective: Maximum caliber is a general variational principle for dynamical systems. *The Journal of Chemical Physics*, 148(1):010901, January 2018. doi: 10.1063/1.5012990.

Egger, D., Werninghaus, M., Ganzhorn, M., Salis, G., Fuhrer, A., Müller, P., and Filipp, S. Pulsed reset protocol for fixed-frequency superconducting qubits. *Phys. Rev. Appl.*, 10:044030, Oct 2018. doi: 10.1103/PhysRevApplied.10.044030. URL https://link.aps.org/doi/10.1103/PhysRevApplied.10.044030.

Ernst, J. O., Chatterjee, A., Franzmeyer, T., and Kuhn, A. Reinforcement learning for quantum control under physical constraints. *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

Gautier, R., Genois, E., and Blais, A. Optimal control in large open quantum systems: The case of transmon readout and reset. *Phys. Rev. Lett.*, 134:070802, Feb 2025. doi: 10.1103/PhysRevLett.134.070802. URL https://link.aps.org/doi/10.1103/PhysRevLett.134.070802.

Ghosh, K., Dixit, P. D., Agozzino, L., and Dill, K. A. The Maximum Caliber Variational Principle for Nonequilibria. 71:213–238, 2020. ISSN 0066-426X. doi: 10.1146/annurev-physchem-071119-040206. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9827727/.

Girsanov, I. V. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960. doi: 10.1137/1105027. URL https://doi.org/10.1137/1105027.

Glaser, S. J., Boscain, U., Calarco, T., Koch, C. P., Köckenberger, W., Kosloff, R., Kuprov, I., Luy, B., Schirmer, S., Schulte-Herbrüggen, T., Sugny, D., and Wilhelm, F. K. Training Schrödinger's cat: Quantum optimal control. 69(12):279, 2015. ISSN 1434-6079. doi: 10.1140/epjd/e2015-60464-1. URL https://doi.org/10.1140/epjd/e2015-60464-1.

Gorini, V., Kossakowski, A., and Sudarshan, E. C. G. Completely positive dynamical semigroups of N-level systems. 17(5):821–825, 1976. ISSN 0022-2488. doi: 10.1063/1.522979. URL https://doi.org/10.1063/1.522979.

Guilmin, P., Bocquet, A., Genois, É., Weiss, D., and Gautier, R. Dynamiqs: an open-source python library for gpu-accelerated and differentiable simulation of quantum systems. 2025. URL https://github.com/dynamiqs/dynamiqs.

IBM Quantum. IBM Quantum Platform, 2026. URL https://quantum.ibm.com/. Calibration snapshot for `ibm_kingston` retrieved 3 May 2026.

Ivanov, P. A., Poschinger, U. G., Singer, K., and Schmidt-Kaler, F. Quantum gate in the decoherence-free subspace of trapped-ion qubits. *Europhysics Letters*, 92(3):30006, nov 2010. doi: 10.1209/0295-5075/92/30006. URL https://doi.org/10.1209/0295-5075/92/30006.

Jaynes, E. T. The minimum entropy production principle. *Annual Review of Physical Chemistry*, 31(1):579–601, 1980.

Jaynes, E. T. Macroscopic prediction. In *Complex Systems—Operational Approaches in Neurobiology, Physics,*

and Computers: Proceedings of the International Symposium on Synergetics at Schloß Elmau, Bavaria, May 6–11, 1985, pp. 254–269. Springer, 1985.

Kappen, H. J. Linear theory for control of nonlinear stochastic systems. *Phys. Rev. Lett.*, 95: 200201, Nov 2005. doi: 10.1103/PhysRevLett.95.200201. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.200201.

Kappen, H. J. and Ruiz, H. C. Adaptive Importance Sampling for Control and Inference. 162(5): 1244–1266, 2016. ISSN 1572-9613. doi: 10.1007/s10955-016-1446-7. URL https://doi.org/10.1007/s10955-016-1446-7.

Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T., and Glaser, S. J. Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms. 172(2):296–305, 2005. ISSN 1090-7807. doi: 10.1016/j.jmr.2004.11.004. URL https://www.sciencedirect.com/science/article/pii/S1090780704003696.

Koch, C. P., Boscain, U., Calarco, T., Dirr, G., Filipp, S., Glaser, S. J., Kosloff, R., Montangero, S., Schulte-Herbrüggen, T., Sugny, D., and Wilhelm, F. K. Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe. 9(1):19, 2022. ISSN 2196-0763. doi: 10.1140/epjqt/s40507-022-00138-x. URL https://doi.org/10.1140/epjqt/s40507-022-00138-x.

Krantz, P., Kjaergaard, M., Yan, F., Orlando, T. P., Gustavsson, S., and Oliver, W. D. A quantum engineer's guide to superconducting qubits. 6(2):021318, 2019. ISSN 1931-9401. doi: 10.1063/1.5089550. URL https://doi.org/10.1063/1.5089550.

Kubica, A., Haim, A., Vaknin, Y., Levine, H., Brandão, F., and Retzker, A. Erasure qubits: Overcoming the T_1 limit in superconducting circuits. *Phys. Rev. X*, 13:041022, Nov 2023. doi: 10.1103/PhysRevX.13.041022. URL https://link.aps.org/doi/10.1103/PhysRevX.13.041022.

Levine, H., Haim, A., Hung, J. S. C., Alidoust, N., Kalaee, M., DeLorenzo, L., Wollack, E. A., Arrangoiz-Arriola, P., Khalajhedayati, A., Sanil, R., Moradinejad, H., Vaknin, Y., Kubica, A., Hover, D., Aghaeimeibodi, S., Alcid, J. A., Baek, C., Barnett, J., Bawdekar, K., Bienias, P., Carson, H. A., Chen, C., Chen, L., Chinkezian, H., Chisholm, E. M., Clifford, A., Cosmic, R., Crisosto, N., Dalzell, A. M., Davis, E., D'Ewart, J. M., Diez, S., D'Souza, N., Dumitrescu, P. T., Elkhouly, E., Fang, M. T., Fang, Y., Flammia, S., Fling, M. J., Garcia, G., Gharzai, M. K., Gorshkov, A. V., Gray, M. J., Grimberg, S., Grimsmo, A. L., Hann, C. T., He, Y., Heidel, S., Howell, S., Hunt, M., Iverson, J., Jarrige, I., Jiang, L., Jones, W. M., Karabalin, R., Karalekas, P. J., Keller, A. J., Lasi, D., Lee, M., Ly, V., MacCabe, G., Mahuli, N., Marcaud, G., Matheny, M. H., McArdle, S., McCabe, G., Merton, G., Miles, C., Milsted, A., Mishra, A., Moncelsi, L., Naghiloo, M., Noh, K., Oblepias, E., Ortuno, G., Owens, J. C., Pagdilao, J., Panduro, A., Paquette, J.-P., Patel, R. N., Peairs, G., Perello, D. J., Peterson, E. C., Ponte, S., Putterman, H., Refael, G., Reinhold, P., Resnick, R., Reyna, O. A., Rodriguez, R., Rose, J., Rubin, A. H., Runyan, M., Ryan, C. A., Sahmoud, A., Scaffidi, T., Shah, B., Siavoshi, S., Sivarajah, P., Skogland, T., Su, C.-J., Swenson, L. J., Sylvia, J., Teo, S. M., Tomada, A., Torlai, G., Wistrom, M., Zhang, K., Zuk, I., Clerk, A. A., Brandão, F. G. S. L., Retzker, A., and Painter, O. Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons. *Phys. Rev. X*, 14:011051, Mar 2024. doi: 10.1103/PhysRevX.14.011051. URL https://link.aps.org/doi/10.1103/PhysRevX.14.011051.

Lidar, D. A., Chuang, I. L., and Whaley, K. B. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.*, 81:2594–2597, Sep 1998. doi: 10.1103/PhysRevLett.81.2594. URL https://link.aps.org/doi/10.1103/PhysRevLett.81.2594.

Lindblad, G. On the generators of quantum dynamical semigroups. 48(2):119–130, 1976. ISSN 1432-0916. doi: 10.1007/BF01608499. URL https://doi.org/10.1007/BF01608499.

Lorenzo, S., Apollaro, T. J. G., Paganelli, S., Palma, G. M., and Plastina, F. Transfer of arbitrary two-qubit states via a spin chain. *Phys. Rev. A*, 91: 042321, Apr 2015. doi: 10.1103/PhysRevA.91.042321. URL https://link.aps.org/doi/10.1103/PhysRevA.91.042321.

Lyakhov, A. and Bruder, C. Quantum state transfer in arrays of flux qubits. 7(1):181, 2005. doi: 10.1088/1367-2630/7/1/181. URL https://doi.org/10.1088/1367-2630/7/1/181.

Magann, A. B., Arenz, C., Grace, M. D., Ho, T.-S., Kosut, R. L., McClean, J. R., Rabitz, H. A., and Sarovar, M. From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms. *PRX Quantum*, 2:010101, Jan 2021. doi: 10.1103/PRXQuantum.2.010101. URL https://link.aps.org/doi/10.1103/PRXQuantum.2.010101.

Magnard, P., Kurpiers, P., Royer, B., Walter, T., Besse, J.-C., Gasparinetti, S., Pechal, M., Heinsoo, J., Storz, S., Blais, A., and Wallraff, A. Fast and unconditional all-microwave reset of a superconducting qubit. *Phys. Rev.*

Lett., 121:060502, Aug 2018. doi: 10.1103/PhysRevLett.121.060502. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.060502.

Manzano, G. and Zambrini, R. Quantum thermodynamics under continuous monitoring: A general framework. 4(2): 025302, 2022. ISSN 2639-0213. doi: 10.1116/5.0079886. URL https://doi.org/10.1116/5.0079886.

Motzoi, F., Gambetta, J. M., Rebentrost, P., and Wilhelm, F. K. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett.*, 103: 110501, Sep 2009. doi: 10.1103/PhysRevLett.103.110501. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.110501.

Najafi, P., Villanueva, A., and Kappen, H. Path integral quantum control for quantum chemistry applications. *arXiv preprint arXiv:2509.24104*, 2025.

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

Percival, I. *Quantum State Diffusion*. Cambridge University Press, 1998.

Pezzè, L., Smerzi, A., Oberthaler, M. K., Schmied, R., and Treutlein, P. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90:035005, Sep 2018. doi: 10.1103/RevModPhys.90.035005. URL https://link.aps.org/doi/10.1103/RevModPhys.90.035005.

Pressé, S., Ghosh, K., Lee, J., and Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.*, 85:1115–1141, Jul 2013. doi: 10.1103/RevModPhys.85.1115. URL https://link.aps.org/doi/10.1103/RevModPhys.85.1115.

Reich, D. M., Ndong, M., and Koch, C. P. Monotonically convergent optimization in quantum control using Krotov's method. 136(10):104103, 2012. ISSN 0021-9606. doi: 10.1063/1.3691827. URL https://doi.org/10.1063/1.3691827.

Villanueva, A. and Kappen, H. Stochastic optimal control of open quantum systems. *arXiv preprint arXiv:2410.18635*, 2024.

Viola, L., Knill, E., and Lloyd, S. Dynamical decoupling of open quantum systems. *Phys. Rev. Lett.*, 82: 2417–2421, Mar 1999. doi: 10.1103/PhysRevLett.82.2417. URL https://link.aps.org/doi/10.1103/PhysRevLett.82.2417.

Vitanov, N. V., Rangelov, A. A., Shore, B. W., and Bergmann, K. Stimulated raman adiabatic passage in physics, chemistry, and beyond. *Rev. Mod. Phys.*, 89:015006, Mar 2017. doi: 10.1103/RevModPhys.89.015006. URL https://link.aps.org/doi/10.1103/RevModPhys.89.015006.

Werninghaus, M., Egger, D. J., Roy, F., Machnes, S., Wilhelm, F. K., and Filipp, S. Leakage reduction in fast superconducting qubit gates via optimal control. 7(1):14, 2021. ISSN 2056-6387. doi: 10.1038/s41534-020-00346-2. URL https://doi.org/10.1038/s41534-020-00346-2.

Wiseman, H. M. and Milburn, G. J. *Quantum Measurement and Control*. Cambridge University Press, 2009.

Wu, Y., Kolkowitz, S., Puri, S., and Thompson, J. D. Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays. 13 (1):4657, 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32094-6. URL https://doi.org/10.1038/s41467-022-32094-6.

Yang, H.-X., Ma, J.-Y., Wu, Y.-K., Wang, Y., Cao, M.-M., Guo, W.-X., Huang, Y.-Y., Feng, L., Zhou, Z.-C., and Duan, L.-M. Realizing coherently convertible dual-type qubits with the same ion species. 18 (9):1058–1061, 2022. ISSN 1745-2481. doi: 10.1038/s41567-022-01661-5. URL https://doi.org/10.1038/s41567-022-01661-5.

A. Open quantum systems and stochastic unravellings

This appendix gives the physical background for the objects used in the main text. It is self-contained for a reader familiar with stochastic differential equations but not with quantum mechanics. For comprehensive treatments, see Wiseman & Milburn (2009) for open quantum systems and continuous measurement, Percival (1998) for quantum state diffusion, Manzano & Zambrini (2022) for a review of continuous monitoring frameworks, and Koch et al. (2022) for quantum optimal control. Figure 4 shows the level structure and noise topology of all five benchmarks at a glance, for reference throughout the paper.

Closed quantum systems. An isolated quantum system of dimension N is described by a unit vector $|\psi\rangle \in \mathbb{C}^N$ called the *state vector*, evolving under the Schrödinger equation $|\dot{\psi}\rangle = -iH|\psi\rangle$ for some Hermitian operator H called the *Hamiltonian*. The evolution is unitary: norms are preserved and the dynamics are reversible.

Open quantum systems and the density matrix. In practice, every quantum system couples to an external environment (electromagnetic field modes, phonons, nearby spins), and this coupling introduces irreversible dynamics that cannot be described by a state vector alone. The standard framework represents the state as a *density matrix* $\rho \in \mathbb{C}^{N\times N}$: a positive semidefinite, trace-one, Hermitian matrix. Pure states correspond to rank-one projectors $\rho = |\psi\rangle\langle\psi|$; mixed states to convex combinations $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ representing classical uncertainty over which pure state the system is in.

We illustrate with a single qubit ($N = 2$), with computational basis $|0\rangle = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $|1\rangle = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$. The single-qubit operators we use throughout are the Pauli matrices

$$\sigma_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad \sigma_y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \sigma_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{12}$$

together with the raising and lowering operators $\sigma_\pm = \frac{1}{2}(\sigma_x \mp i\sigma_y)$, so that $\sigma_- = |0\rangle\langle 1|$ and $\sigma_+ = |1\rangle\langle 0|$. A qubit density matrix

$$\rho = \begin{pmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{pmatrix} \tag{13}$$

has diagonal entries $\rho_{00}, \rho_{11} \geq 0$ giving the probabilities of measuring $|0\rangle$ or $|1\rangle$ (the *populations*, with $\rho_{00} + \rho_{11} = 1$), and off-diagonal entries $\rho_{01} = \overline{\rho_{10}}$ encoding quantum superposition (the *coherences*).

The Lindblad master equation. The Lindblad master equation (1) is the most general Markovian, trace-preserving, completely positive evolution for a density matrix (Lindblad, 1976; Gorini et al., 1976). It consists of two parts. The commutator $-i[H, \rho]$ generates unitary evolution under the Hamiltonian $H = H_0 + \sum_a u_a(t)H_a$, where H_0 is the free Hamiltonian and $u_a(t)$ are time-dependent *control fields* chosen by the experimenter and coupled through fixed control Hamiltonians H_a. The dissipator $\sum_k \mathcal{D}_k(\rho)$ with $\mathcal{D}_k(\rho) = L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}$ encodes the irreversible effect of the environment, with one *Lindblad operator* L_k per *decoherence channel*: L_k specifies what the environment does to the system. The five benchmarks of Section 5 use Lindblad operators of two basic forms — depopulating transitions like σ_- that move probability mass between levels, and dephasing operators like σ_z that scramble superpositions without changing populations. Figure 4 shows where each appears.

We illustrate with two channels that recur throughout the paper, both acting on a single qubit (Figure 4a uses amplitude damping; the chain panels d, e use both).

- *Dephasing*, $L = \sqrt{\kappa}\,\sigma_z$, models an environment that continuously probes whether the qubit is in $|0\rangle$ or $|1\rangle$ without exchanging energy with it. The populations ρ_{00}, ρ_{11} are unchanged, while the coherences ρ_{01}, ρ_{10} decay exponentially at rate 2κ. It is the dominant noise channel in many superconducting and spin-qubit platforms and sets the T_2 coherence time.
- *Amplitude damping*, $L = \sqrt{\gamma}\,\sigma_-$, models the qubit losing a quantum of energy to its environment, for example a photon emitted into the electromagnetic field. Population flows irreversibly from $|1\rangle$ to $|0\rangle$ at rate γ, and the ground state is a fixed point of the dissipation. This dominates in optical and atomic systems and sets the T_1 relaxation time in superconducting qubits.

For an n-qubit system, $N = 2^n$, and the density matrix has $N^2 = 4^n$ real parameters; propagating the Lindblad equation directly therefore costs $\mathcal{O}(N^2) = \mathcal{O}(4^n)$ per time step.

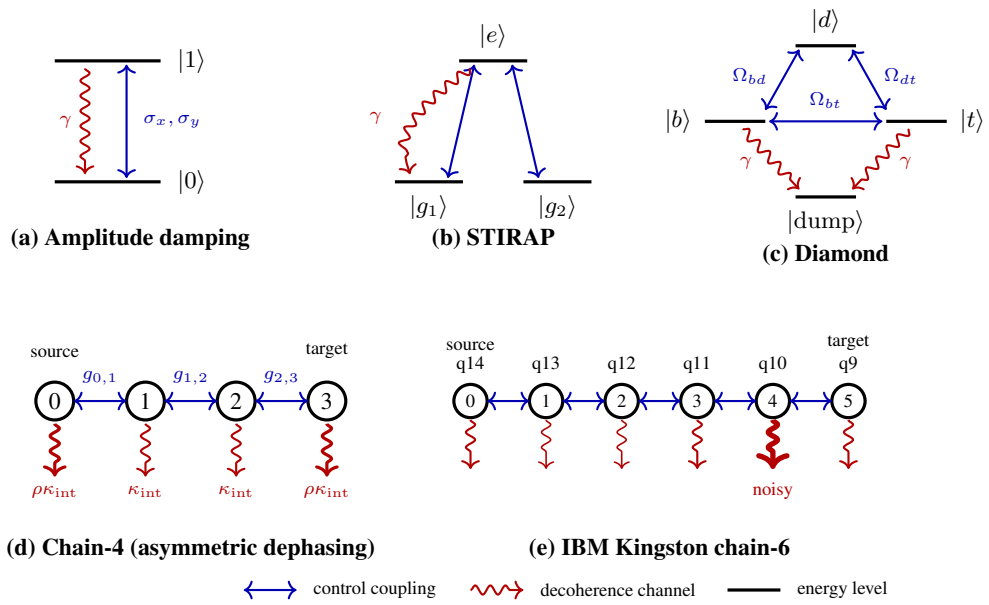

Figure 4. **Level structure and noise structure of the five benchmarks.** Blue double arrows mark coherent control couplings; red squiggly arrows mark decoherence channels (each acts on the density matrix as a Lindblad operator L_k). **(a)** Amplitude damping: $|1\rangle$ decays to $|0\rangle$, two controls drive arbitrary single-qubit rotations. **(b)** STIRAP: two stable ground states $|g_1\rangle, |g_2\rangle$ coupled through a lossy excited state $|e\rangle$ that decays back to $|g_1\rangle$. **(c)** Diamond: two lossy computational states $|b\rangle, |t\rangle$ each decay to $|\mathrm{dump}\rangle$; the auxiliary state $|d\rangle$ is in $\ker L_b \cap \ker L_t$ and provides a safe indirect route. **(d)** Chain-4: four qubits with site-specific dephasing; edge sites carry rates $\rho\kappa_{\mathrm{int}}$, interior sites κ_{int} (amplitude damping not shown for clarity but acts on every site). **(e)** IBM Kingston chain-6: six qubits with per-site T_1, T_2 from the calibration snapshot; site 4 (q10) is markedly noisier ($T_2 = 17\,\mu$s, indicated by the thicker decay arrow).

Unravelling the master equation. The Lindblad equation describes the ensemble-averaged dynamics. In many experimental platforms (circuit QED, trapped ions, optical cavities) the environment is itself *continuously monitored*, and each run of the experiment produces a specific measurement record and a corresponding conditional pure state $|\psi(t)\rangle \in \mathbb{C}^N$. The decomposition of the density-matrix evolution into a distribution over such pure-state trajectories is called an *unravelling*; averaging over trajectories always recovers the density matrix,

$$\rho(t) = \mathbb{E}\big[|\psi(t)\rangle\langle\psi(t)|\big], \tag{14}$$

regardless of which unravelling is chosen.

The choice of unravelling is determined by the type of measurement performed on the environment. Two principal classes exist:

- *Jump unravelling* (photon counting / direct detection). The environment is monitored by counting detection events. Between events, $|\psi\rangle$ evolves smoothly under a non-Hermitian effective Hamiltonian; at each event, the state jumps as $|\psi\rangle \to L_k|\psi\rangle/\|L_k|\psi\rangle\|$. Trajectories are piecewise deterministic with jump times governed by Poisson processes.
- *Diffusive unravelling* (homodyne / heterodyne detection). The output field is mixed with a strong local oscillator before detection, producing a continuous photocurrent. The conditional state evolves under the stochastic Schrödinger equation (2), driven by independent Wiener increments $\mathrm{d}W_k$.

Why diffusive unravelling. The argument of Section 3 relies on the measurement record being an Itô diffusion with state-dependent drift and unit, control-independent diffusion coefficient (Eq. 3); this is the structure required by Girsanov's theorem. Under the jump unravelling the measurement record consists of point processes (click times) rather than diffusions, and the change of measure for marked point processes takes a different form, involving products of intensity ratios at jump times rather than Itô integrals. Extending the framework in this direction is discussed in Section 6.

State vectors versus density matrices. The trajectory representation underlies the computational strategy adopted throughout this paper. The conditional state vector $|\psi(t)\rangle \in \mathbb{C}^N$ has N complex entries while the density matrix $\rho(t) \in$

$\mathbb{C}^{N\times N}$ has N^2, so the state space of a single trajectory is quadratically smaller than that of the master equation. For n-qubit systems with $N = 2^n$, this is the difference between $\mathcal{O}(2^n)$ and $\mathcal{O}(4^n)$ memory, and the gap matters in particular for autodiff-based control: gradient computation through the solver requires storing the entire forward trajectory, and the smaller state space lets a single GPU handle larger systems than direct integration of Eq. (1) would allow (Abdelhafez et al., 2019). Trajectories are also independent across the ensemble and parallelize accordingly. We adopt this framework, augmented with the path-space regularizer derived in the main text.

B. Formal derivation of the Girsanov KL on path space

We give a rigorous path-space derivation of Eq. (8). The argument is standard stochastic analysis (Øksendal, 2003); the state-dependent drift requires the path-space framing to be handled rigorously. We construct the relevant exponential martingale (Lemma B.1), use it to relate $P^{(1)}$ and $P^{(2)}$ as a Girsanov change of measure (Theorem B.2), and read off the KL divergence (Corollary B.3).

Setup. Let $\Omega := C([0,\tau];\mathbb{R}^K)$ with canonical filtration $\mathcal{F}_t$, and let $\mathbb{P}_W$ denote K-dimensional Wiener measure on Ω, under which the coordinate process $W_k(t) := \omega_k(t)$ is a standard Brownian motion. Let $|\psi_t^{(i)}(\omega)\rangle$ denote the SSE solution under protocol $i \in \{1,2\}$, and define the drift functionals

$$A_k^{(i)}(\omega,t) := \langle\psi_t^{(i)}(\omega)|(L_k + L_k^\dagger)|\psi_t^{(i)}(\omega)\rangle, \tag{15}$$

which are $\mathcal{F}_t$-adapted and uniformly bounded by $C_k := \|L_k + L_k^\dagger\|_{\mathrm{op}}$. Define $P^{(i)}$ as the pushforward of $\mathbb{P}_W$ under the map $\omega \mapsto I^{(i)}(\omega)$ with $I_k^{(i)}(t) := \int_0^t A_k^{(i)}(\omega,s)\,ds + W_k(t)$. By construction, under $P^{(i)}$ the coordinate process $\omega_k(t)$ has the law of $\int_0^t A_k^{(i)}\,ds + W_k(t)$ under $\mathbb{P}_W$, so setting $B_k^{(i)}(t) := \omega_k(t) - \int_0^t A_k^{(i)}(\omega,s)\,ds$ gives a standard K-dimensional Brownian motion under $P^{(i)}$ and

$$d\omega_k(t) = A_k^{(i)}(\omega,t)\,\mathrm{d}t + dB_k^{(i)}. \tag{16}$$

Set $\Delta_k := A_k^{(1)} - A_k^{(2)}$, so $|\Delta_k| \le 2C_k$.

The exponential martingale. With $B^{(2)}$ the standard Brownian motion under $P^{(2)}$ from (16), define

$$Z_t := \exp\left(\sum_k \int_0^t \Delta_k\,dB_k^{(2)} - \frac{1}{2}\sum_k \int_0^t \Delta_k^2\,ds\right). \tag{17}$$

Lemma B.1. *Z_t is a strictly positive uniformly integrable $(\mathcal{F}_t, P^{(2)})$-martingale with $Z_0 = 1$ and $\mathbb{E}_{P^{(2)}}[Z_\tau] = 1$.*

Proof. Apply Itô's formula to $Z_t = e^{M_t}$ with $M_t := \sum_k \int_0^t \Delta_k\,dB_k^{(2)} - \frac{1}{2}\sum_k \int_0^t \Delta_k^2\,ds$. Since $d\langle M\rangle_t = \sum_k \Delta_k^2\,dt$ exactly cancels the drift in dM_t,

$$dZ_t = Z_t \sum_k \Delta_k\,dB_k^{(2)}, \tag{18}$$

with $Z_0 = 1$. The bound $\|\Delta(t)\|^2 \le 4\sum_k C_k^2$ gives the deterministic estimate

$$\mathbb{E}_{P^{(2)}}\left[\exp\left(\frac{1}{2}\int_0^\tau \|\Delta\|^2\,\mathrm{d}t\right)\right] \le \exp\left(2\tau\sum_k C_k^2\right) < \infty, \tag{19}$$

verifying Novikov's condition, so Z_t is a true uniformly integrable martingale on $[0,\tau]$ (Øksendal, 2003, Thm. 8.6.5) with $\mathbb{E}_{P^{(2)}}[Z_\tau] = Z_0 = 1$. Strict positivity is immediate from the exponential form. □

Theorem B.2 (Girsanov transform on path space)**.** *$P^{(1)}$ and $P^{(2)}$ are mutually absolutely continuous, with*

$$\frac{dP^{(1)}}{dP^{(2)}}(\omega) = Z_\tau(\omega). \tag{20}$$

Proof. By Lemma B.1, $\mathcal{Q} := Z_\tau \, dP^{(2)}$ defines a probability measure on $(\Omega, \mathcal{F}_\tau)$. Girsanov's theorem (Øksendal, 2003, Thm. 8.6.6), applied to the $P^{(2)}$-Brownian motion $B^{(2)}$ with adapted bounded shift Δ, gives that

$$B_k^{(1)}(t) := B_k^{(2)}(t) - \int_0^t \Delta_k(\omega, s)\, ds = \omega_k(t) - \int_0^t A_k^{(1)}(\omega, s)\, ds \tag{21}$$

is a standard K-dimensional Brownian motion under $\mathcal{Q}$. Hence under $\mathcal{Q}$ the coordinate process has the law of $\int_0^\cdot A_k^{(1)}\, ds + B_k^{(1)}$ with $B^{(1)}$ Brownian, which is precisely the law of the pushforward $\omega \mapsto I^{(1)}(\omega)$ of $\mathbb{P}_W$, i.e. $P^{(1)}$. Therefore $\mathcal{Q} = P^{(1)}$, giving $dP^{(1)}/dP^{(2)} = Z_\tau$. The reverse direction is immediate from $Z_\tau > 0$ $P^{(2)}$-a.s., so $dP^{(2)}/dP^{(1)} = Z_\tau^{-1}$. □

Corollary B.3 (Path KL divergence)**.**

$$D_{\mathrm{KL}}[P^{(1)} \| P^{(2)}] = \frac{1}{2} \sum_k \mathbb{E}_{P^{(1)}} \left[\int_0^\tau \Delta_k(\omega, t)^2 \, \mathrm{d}t \right]. \tag{22}$$

Proof. By Theorem B.2,

$$D_{\mathrm{KL}}[P^{(1)} \| P^{(2)}] = \mathbb{E}_{P^{(1)}} \left[\log \frac{dP^{(1)}}{dP^{(2)}} \right] = \mathbb{E}_{P^{(1)}}[\log Z_\tau] = \mathbb{E}_{P^{(1)}} \left[\sum_k \int_0^\tau \Delta_k \, dB_k^{(2)} - \frac{1}{2} \int_0^\tau \|\Delta\|^2 \, \mathrm{d}t \right]. \tag{23}$$

Under $P^{(1)}$, $B_k^{(1)}$ is Brownian, so $dB_k^{(2)} = dB_k^{(1)} + \Delta_k \, \mathrm{d}t$. Substituting,

$$D_{\mathrm{KL}}[P^{(1)} \| P^{(2)}] = \mathbb{E}_{P^{(1)}} \left[\sum_k \int_0^\tau \Delta_k \, dB_k^{(1)} \right] + \frac{1}{2} \mathbb{E}_{P^{(1)}} \left[\int_0^\tau \|\Delta\|^2 \, \mathrm{d}t \right]. \tag{24}$$

Since $|\Delta_k| \le 2C_k$ is bounded, $\int_0^\cdot \Delta_k \, dB_k^{(1)}$ is a true $P^{(1)}$-martingale, so the first term vanishes. □

The Wiener KL of Section 4.2 is the special case $A^{(2)} \equiv 0$, where the pushforward map is the identity, so $P^{(2)} = \mathbb{P}_W$ and $\Delta_k = A_k^{(1)}$, recovering (9).

C. Derivation of the drift-variance regularizer

We derive the closed form of the drift-variance regularizer (10) by minimizing $D_{\mathrm{KL}}[P_\theta \| P_c]$ over the constant drift $c \in \mathbb{R}^K$, where P_c denotes the law of $dY_t = c \, \mathrm{d}t + dW_t$ as defined in Section 4.2. Substituting $\alpha_k^{(1)}(t) = \alpha_k^{(\theta)}(t)$ and $\alpha_k^{(2)}(t) = c_k$ into the multi-channel Girsanov KL (8) (suppressing the (θ) superscript and t argument hereafter for readability),

$$D_{\mathrm{KL}}[P_\theta \| P_c] = \frac{1}{2} \sum_k \mathbb{E}_{P_\theta} \left[\int_0^T (\alpha_k - c_k)^2 \, \mathrm{d}t \right]. \tag{25}$$

The KL decomposes as a sum over channels, each depending only on its own c_k, so we minimize per channel. Expanding the square,

$$\mathbb{E}_{P_\theta} \left[\int_0^T (\alpha_k - c_k)^2 \, \mathrm{d}t \right] = \mathbb{E}_{P_\theta} \left[\int_0^T \alpha_k^2 \, \mathrm{d}t \right] - 2c_k \, \mathbb{E}_{P_\theta} \left[\int_0^T \alpha_k \, \mathrm{d}t \right] + c_k^2 T, \tag{26}$$

which is quadratic in c_k with positive leading coefficient T. Differentiating and setting to zero gives the optimal constant drift

$$c_k^\star = \frac{1}{T} \mathbb{E}_{P_\theta} \left[\int_0^T \alpha_k \, \mathrm{d}t \right] =: \bar{\alpha}_k, \tag{27}$$

the time-and-ensemble drift mean. Substituting (27) back into (25),

$$R_{\mathrm{DV}} := \min_{c \in \mathbb{R}^K} D_{\mathrm{KL}}[P_\theta \| P_c] = \frac{1}{2} \sum_k \mathbb{E}_{P_\theta} \left[\int_0^T (\alpha_k - \bar{\alpha}_k)^2 \, \mathrm{d}t \right], \tag{28}$$

which is (10). The integrand is the squared deviation of the instantaneous drift from its time-and-ensemble mean: R_{DV} vanishes if and only if $\alpha_k^{(\theta)}(t) = \bar{\alpha}_k$ for $(P_\theta \otimes \mathrm{d}t)$-almost every (ω, t) and every channel k, i.e. the trajectory drift is constant in time and across realizations.

D. Proof: Girsanov KL reduces to fluence in the QDC-compatible case

Following Villanueva & Kappen (2024), consider a Lindblad equation with n_c Lindblad operators $\{C_a\}$ and real symmetric noise matrix D. The unravelling admits a gauge freedom: an invertible $A \in \mathrm{GL}(n_c, \mathbb{C})$ transforms $\tilde{C}_b = C_a A_{ab}$, $\tilde{D} = ADA^\dagger$ while leaving the Lindbladian dissipator invariant; the diffusive SSE in the transformed basis takes the form

$$d\psi = -iH\psi\,\mathrm{d}t - \frac{1}{2}\sum_{a,b}\tilde{D}_{ab}\left(\tilde{C}_b^\dagger\tilde{C}_a - 2\tilde{c}_a\tilde{C}_b + \tilde{c}_a\tilde{c}_b\right)\psi\,\mathrm{d}t + \sum_a(\tilde{C}_a - \tilde{c}_a)\psi\,d\tilde{W}_a, \tag{29}$$

where $\tilde{c}_a = \psi^\dagger\tilde{C}_a^{(h)}\psi$ and $\tilde{C}_a^{(h)} = \frac{1}{2}(\tilde{C}_a + \tilde{C}_a^\dagger)$ is the Hermitian part. Suppose there exists a choice of A such that $\tilde{D}$ remains real symmetric and $\tilde{C}_a = -iH_a$, identifying the transformed Lindblad operators with $-i$ times the control Hamiltonians. Anti-Hermiticity of $\tilde{C}_a$ implies $\tilde{C}_a^{(h)} = 0$, hence $\tilde{c}_a = 0$ identically: the nonlinear backaction terms in (29) drop out and the SSE becomes linear in ψ. Combining with the Hamiltonian control $H = H_0 + u_a H_a$, the diffusion and Hamiltonian terms acting along H_a combine into $-iH_a\psi(u_a\,\mathrm{d}t + d\tilde{W}_a)$, so the controls u_a enter as deterministic shifts of the Brownian increments (Villanueva & Kappen, 2024, Eq. 20). The corresponding measurement record in the transformed basis is

$$dY_a^{(u)} = u_a\,\mathrm{d}t + d\tilde{W}_a, \tag{30}$$

with drift u_a deterministic and state-independent.

Applying Corollary B.3 to the controlled record $dY_a^{(u)}$ versus the uncontrolled record $dY_a^{(0)} = d\tilde{W}_a$ (drifts u_a and 0, both deterministic), the expectation in the KL collapses and

$$D_{\mathrm{KL}}[P_u\|P_0] = \frac{1}{2}\int_0^T\sum_a u_a(t)^2\,\mathrm{d}t, \tag{31}$$

which is the control fluence. The structural condition $\tilde{C}_a = -iH_a$ is restrictive: it requires that an invertible A exists mapping the original Lindblad operators $\{C_a\}$ to $\{-iH_a\}$, which forces the noise channels (after mixing) to act in the same algebraic directions as the controls. The Wiener KL of Section 4.2 is a different quantity, computed in the original physical basis where the measurement-record drift is $\langle L_k + L_k^\dagger\rangle = 2\langle L_k^{(h)}\rangle$, twice the expectation of the Hermitian part. This drift is state-dependent and generically unrelated to u_a; the two KLs coincide only when the QDC conditions hold and the change of basis A brings the original $\{C_a\}$ to anti-Hermitian form.

E. Additional Experimental Details

This appendix collects multi-seed statistics and robustness experiments referenced in Section 5, organized to mirror the body's experiment ordering. Section E.1 fixes the common training setup; the per-experiment subsections then list only the values that differ.

E.1. Common training setup

We have the following common architectural choices for all our experiments:

- **Controller**: each control channel $u_a^{(\theta)}(t)$ is a truncated Fourier series in time with 16 Fourier modes (i.e. 16 cosine and 16 sine coefficients per channel plus a DC offset).
- **Optimizer**: Adam with learning rate 10^{-3}.
- **Solver**: Euler-Maruyama (EM) or Exponential-Split (ExpSplit), see Appendix H
- **Fidelity reported**: at every snapshot we evaluate the exact-Lindblad fidelity by integrating the master equation under the current controller and computing $F = \langle\psi_{\mathrm{tgt}}|\rho(T)|\psi_{\mathrm{tgt}}\rangle$.
- **Selection rule**: the reported F for any given run is the maximum over snapshots of the exact-Lindblad fidelity.
- **Hardware**: single NVIDIA A100 GPU (40 GB); total project compute reported in Appendix I.

Each subsection below lists T, the total number of optimiser steps, the regulariser weights actually swept, the fluence-warm-up length, and the number of training seeds.

Table 2. **Amplitude damping (full statistics):** Fidelity F for $|+\rangle \to |Y\rangle$, mean $\pm$ std across 4 seeds. Bold marks the best in each row.

γT	$\lambda_{\mathrm{KL_w}}{=}\lambda_{R_{\mathrm{DV}}}{=}0$	$\lambda_{\mathrm{KL_w}}{=}5$	$\lambda_{R_{\mathrm{DV}}}{=}5$	PPO
0.1	0.9942 ± 0.0001	$\mathbf{0.9952 \pm 0.0001}$	0.9917 ± 0.0001	0.9943 ± 0.0001
0.5	0.9831 ± 0.0001	$\mathbf{0.9900 \pm 0.0001}$	0.9698 ± 0.0001	0.9856 ± 0.0003
1.0	0.9772 ± 0.0002	$\mathbf{0.9863 \pm 0.0001}$	0.9555 ± 0.0001	0.9773 ± 0.0007
2.0	0.9715 ± 0.0004	$\mathbf{0.9810 \pm 0.0001}$	0.9458 ± 0.0004	0.9684 ± 0.0017
5.0	0.9633 ± 0.0001	$\mathbf{0.9673 \pm 0.0001}$	0.9222 ± 0.0027	0.9451 ± 0.0002

Table 3. **Amplitude damping (full statistics)**. SSE-trajectory population variance $\sum_k \int_0^T \mathrm{Var}[P_k(t)]\, dt$ for $|+\rangle \to |Y\rangle$, mean $\pm$ std across 4 seeds (128 SSE samples per seed). Bold marks the best in each row.

γT	$\lambda_{\mathrm{KL_w}}{=}\lambda_{R_{\mathrm{DV}}}{=}0$	$\lambda_{\mathrm{KL_w}}{=}5$	$\lambda_{R_{\mathrm{DV}}}{=}5$	PPO
0.1	0.0032 ± 0.0000	$\mathbf{0.0013 \pm 0.0000}$	0.0019 ± 0.0000	0.0036 ± 0.0002
0.5	0.0106 ± 0.0001	$\mathbf{0.0021 \pm 0.0000}$	0.0055 ± 0.0000	0.0090 ± 0.0001
1.0	0.0166 ± 0.0002	$\mathbf{0.0022 \pm 0.0000}$	0.0077 ± 0.0001	0.0134 ± 0.0005
2.0	0.0321 ± 0.0024	$\mathbf{0.0021 \pm 0.0000}$	0.0077 ± 0.0001	0.0192 ± 0.0019
5.0	0.0411 ± 0.0061	$\mathbf{0.0016 \pm 0.0000}$	0.0065 ± 0.0002	0.0243 ± 0.0040

E.2. Single-qubit amplitude damping

Tables 2 and 3 report the multi-seed statistics for the fidelity and trajectory-variance comparison underlying Section 5.1.

Experiment-specific parameters.

- $T = 1.0$ in dimensionless units, so $\gamma T = \gamma$.
- Decay rate sweep: $\gamma T \in \{0.1, 0.5, 1, 2, 5\}$.
- Total optimiser steps: 5000.
- Solver: ExpSplit
- Discretised time grid: 256 points per trajectory, snapshot every 500 steps; 256 SSE samples per gradient step.
- Regulariser weights: $\lambda_{\mathrm{KL_w}} \in \{0, 5\}$, $\lambda_{R_{\mathrm{DV}}} \in \{0, 1, 5\}$, $\lambda_{\mathrm{flu}} = 0.01$.
- Fourier controller: $n_{\mathrm{modes}} = 20$, init_scale $= 0.1$.
- Seeds: 4 per cell.

E.3. STIRAP

Table 4 reports multi-seed fidelity and $|e\rangle$-exposure statistics for the two-noise-level comparison in Section 5.2, and Table 5 reports the γ-perturbation robustness experiment, in which protocols trained at $(\gamma T)_{\mathrm{train}} = 10$ are evaluated at five test noise levels via exact Lindblad integration.

Experiment-specific parameters.

- $T = 2.0$ in dimensionless units.
- Decay rate sweep: $\gamma \in \{0.5, 5\}$ (cleanup, $\gamma T \in \{1, 10\}$); robustness evaluated at $\gamma_{\mathrm{test}} \in \{1, 2, 5, 10, 15\}$ from policies trained at $\gamma = 5$.
- Total optimiser steps: 10 000.
- Solver: ExpSplit

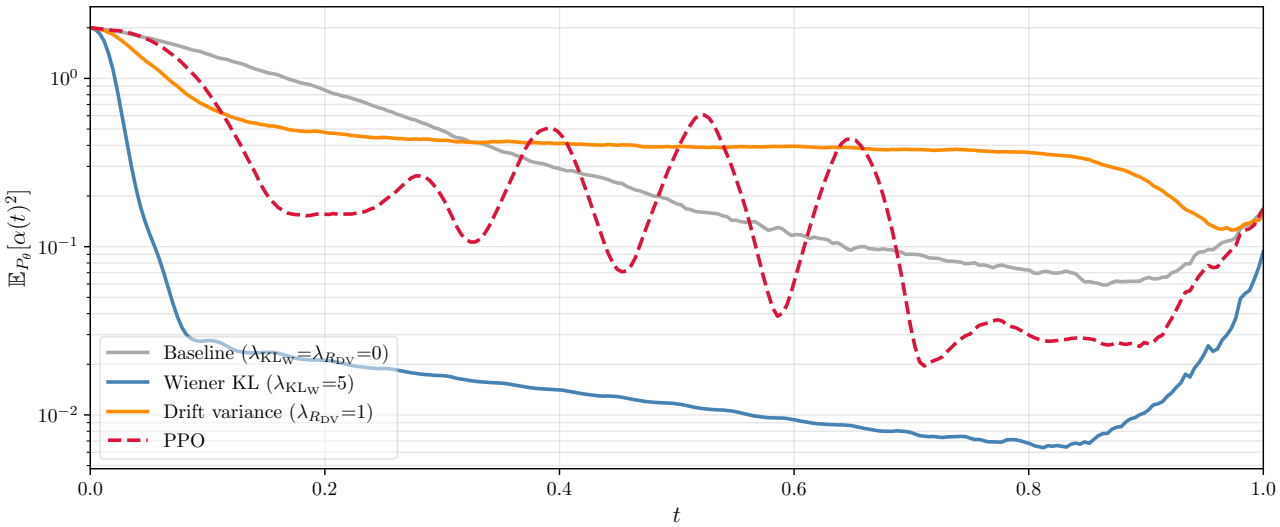

Figure 5. **Wiener KL integrand $\mathbb{E}_\psi[\alpha(t)^2]$, $\gamma T = 2$.** Trajectory-averaged drift squared on a logarithmic axis. Wiener KL suppresses α toward $\ker L$, reducing the time-integrated value $\frac{1}{2}\int\langle\alpha^2\rangle\,dt$ from 0.230 (baseline) to 0.031 — an order-of-magnitude reduction consistent with the trajectory routing seen in Figure 1. Drift-variance remains near the baseline level (0.237), while PPO achieves an intermediate reduction (0.160).

Table 4. **STIRAP results.** Fidelity F, peak $|e\rangle$ population, and time-integrated $|e\rangle$ exposure for the transfer $|g_1\rangle \to |g_2\rangle$, mean $\pm$ std across 3 seeds. Bold marks the best in each column per γT block; baseline and Wiener KL are statistically indistinguishable in fidelity at both noise levels.

γT	method	F_{exact}	peak $\lvert e\rangle$	$\int_0^T \lvert e\rangle\, dt$
1.0	baseline ($\lambda_{\mathrm{KL_W}}$=$\lambda_{R_{\mathrm{DV}}}$=0)	0.9959 ± 0.0003	0.056 ± 0.016	0.024 ± 0.000
1.0	Wiener KL ($\lambda_{\mathrm{KL_W}}$=1)	$\mathbf{0.9961 \pm 0.0005}$	$\mathbf{0.054 \pm 0.011}$	$\mathbf{0.022 \pm 0.001}$
1.0	PPO	0.9614 ± 0.0001	0.160 ± 0.003	0.164 ± 0.001
10.0	baseline ($\lambda_{\mathrm{KL_W}}$=$\lambda_{R_{\mathrm{DV}}}$=0)	$\mathbf{0.9797 \pm 0.0010}$	0.097 ± 0.028	0.026 ± 0.001
10.0	Wiener KL ($\lambda_{\mathrm{KL_W}}$=1)	0.9790 ± 0.0005	$\mathbf{0.043 \pm 0.012}$	$\mathbf{0.017 \pm 0.001}$
10.0	PPO	0.7289 ± 0.0000	0.163 ± 0.010	0.162 ± 0.001

- Discretised time grid: 256 points per trajectory, snapshot every 500 steps; 64 SSE samples per gradient step.
- Regulariser weights: $\lambda_{\mathrm{KL_W}} \in \{0, 1\}$, $\lambda_{R_{\mathrm{DV}}} = 0$, $\lambda_{\mathrm{flu}} = 0$.
- Fourier controller: $n_{\text{modes}} = 16$, init_scale $= 1.0$.
- Seeds: 3 per cell.

E.4. Diamond system

Table 6 reports the multi-seed fidelity statistics for the Wiener KL ablation at $\gamma = 2$ referenced in Section 5.2, and Table 7 reports the γ-perturbation robustness experiment, with protocols trained at $\gamma_{\text{train}} = 2$ and evaluated at six test noise levels spanning $\gamma_{\text{test}} \in \{1, 1.5, 2, 2.5, 3, 5\}$.

Experiment-specific parameters.

- $T = 1.0$ in dimensionless units.
- Training $\gamma = 2.0$; robustness evaluated at $\gamma_{\text{test}} \in \{1, 1.5, 2, 2.5, 3, 5\}$ from policies trained at $\gamma = 2$.
- Total optimiser steps: 5000.

Table 5. **STIRAP γ-perturbation robustness with forbidden-state exposure.** Each protocol trained at $(\gamma T)_{\text{train}} = 10$ and evaluated at the indicated test value. Mean $\pm$ std across 3 seeds.

$(\gamma T)_{\text{test}}$	method	F_{exact}	peak $\lvert e\rangle$	$\int_0^T \lvert e\rangle\, dt$
2.0	baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_\text{DV}}$=0)	0.9945 ± 0.0006	0.113 ± 0.028	0.023 ± 0.001
2.0	Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1)	0.9945 ± 0.0010	$\mathbf{0.042 \pm 0.007}$	$\mathbf{0.015 \pm 0.001}$
2.0	PPO	0.733 ± 0.001	0.282 ± 0.011	0.215 ± 0.002
4.0	baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_\text{DV}}$=0)	0.9904 ± 0.0006	0.108 ± 0.028	0.024 ± 0.001
4.0	Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1)	0.9906 ± 0.0006	$\mathbf{0.042 \pm 0.007}$	$\mathbf{0.016 \pm 0.001}$
4.0	PPO	0.784 ± 0.001	0.239 ± 0.012	0.199 ± 0.001
10.0 (train)	baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_\text{DV}}$=0)	0.9797 ± 0.0009	0.097 ± 0.023	0.026 ± 0.001
10.0 (train)	Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1)	0.9792 ± 0.0004	$\mathbf{0.043 \pm 0.009}$	$\mathbf{0.017 \pm 0.000}$
10.0 (train)	PPO	0.729 ± 0.000	0.163 ± 0.010	0.162 ± 0.001
20.0	baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_\text{DV}}$=0)	0.9626 ± 0.0017	0.084 ± 0.016	0.026 ± 0.001
20.0	Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1)	0.9603 ± 0.0008	$\mathbf{0.047 \pm 0.011}$	$\mathbf{0.018 \pm 0.000}$
20.0	PPO	0.521 ± 0.000	0.095 ± 0.004	0.110 ± 0.001
30.0	baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_\text{DV}}$=0)	0.9446 ± 0.0028	0.071 ± 0.013	0.026 ± 0.001
30.0	Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1)	0.9410 ± 0.0013	$\mathbf{0.050 \pm 0.014}$	$\mathbf{0.019 \pm 0.000}$
30.0	PPO	0.369 ± 0.000	0.061 ± 0.002	0.075 ± 0.000

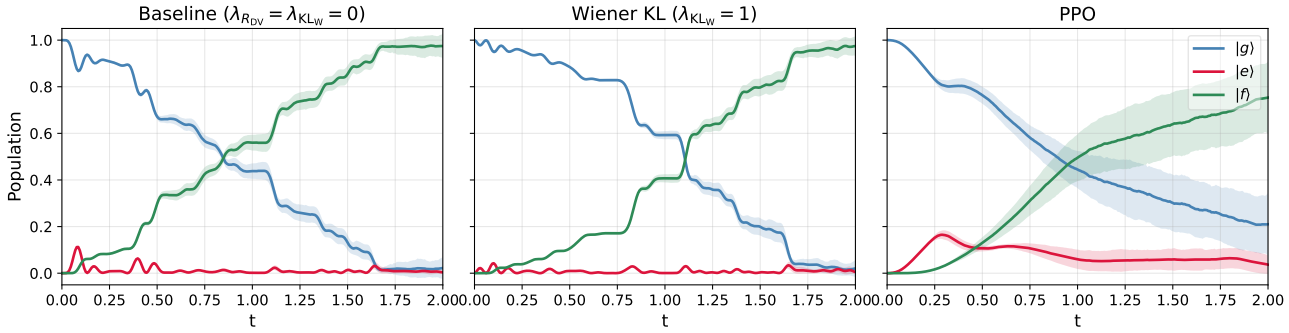

Figure 6. **STIRAP**, $\gamma T = 10$. SSE-trajectory populations (mean across 128 samples with $\pm 1\sigma$ bands) for baseline (λ=0), Wiener KL ($\lambda_{\text{KL}_\text{w}}$=1), and PPO.

- Solver: EM
- Discretised time grid: 256 points per trajectory, snapshot every 500 steps; 256 SSE samples per gradient step.
- Regulariser weights: $\lambda_{\text{KL}_\text{w}} \in \{0, 0.5, 5\}$, $\lambda_{R_\text{DV}} = 0$, $\lambda_{\text{flu}} = 0.001$.
- Fourier controller: $n_{\text{modes}} = 20$, init_scale $= 0.1$.
- Seeds: 3 per cell.

E.5. 4 qubit chain asymmetry sweep

Table 8 reports the complete five-point asymmetry sweep ($\rho \in \{1, 1.5, 2, 4, 8\}$) summarised in Section 5.3. Means and standard deviations are taken over three seeds for $\rho \in \{2, 4, 8\}$; $\rho = 1$ and $\rho = 1.5$ are reported on a single seed as control points. The single representative trajectories shown in Figure 3 are taken from the seed with the highest *final-step* exact-Lindblad fidelity at each ρ; per-seed best fidelities are reported in the lower half of Table 8. Due to high noise values, training was unstable so we implemented a fluence warm-up of 2000 steps.

The wide drift-variance spread at $\rho = 4$ and $\rho = 8$ (std $\approx$ 0.11–0.13) is concentrated in seed 2, which finds a qualitatively different basin in both cases. The mean improvement over the baseline holds at all asymmetry levels $\rho \geq 4$ even including

Table 6. **Diamond system: Wiener KL ablation at $\gamma = 2$, $T = 1$.** Mean $\pm$ std across 3 seeds.

method	F_{exact}
baseline ($\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_{\text{DV}}}$=0)	0.665 ± 0.017
Wiener KL $\lambda_{\text{KL}_\text{w}}$=0.5	0.810 ± 0.004
Wiener KL $\lambda_{\text{KL}_\text{w}}$=5	$\mathbf{0.834 \pm 0.003}$
PPO	0.605 ± 0.005

Table 7. **Diamond system: γ-perturbation robustness.** Mean $\pm$ std across 3 seeds.

γ_{test}	$\lambda_{\text{KL}_\text{w}}$=$\lambda_{R_{\text{DV}}}$=0	$\lambda_{\text{KL}_\text{w}}$=0.5	$\lambda_{\text{KL}_\text{w}}$=5	PPO
1.0	0.808 ± 0.010	0.895 ± 0.002	$\mathbf{0.903 \pm 0.002}$	0.772 ± 0.001
1.5	0.732 ± 0.014	0.851 ± 0.003	$\mathbf{0.868 \pm 0.003}$	0.681 ± 0.002
2.0 (train)	0.665 ± 0.017	0.810 ± 0.004	$\mathbf{0.834 \pm 0.003}$	0.601 ± 0.002
2.5	0.606 ± 0.018	0.771 ± 0.004	$\mathbf{0.803 \pm 0.003}$	0.531 ± 0.002
3.0	0.554 ± 0.020	0.734 ± 0.004	$\mathbf{0.772 \pm 0.004}$	0.470 ± 0.002
5.0	0.393 ± 0.020	0.607 ± 0.006	$\mathbf{0.665 \pm 0.004}$	0.288 ± 0.001

this seed.

Experiment-specific parameters.

- $T = 3.0$ in dimensionless units; rates are in inverse-time units of the same scale.
- Per-site decoherence rates $\kappa_{\text{int}} = 0.3$, $\gamma_{\text{int}} = 0.05$ at all interior sites, scaled by ρ on the two edge sites ($\kappa_0 = \kappa_3 = \rho\,\kappa_{\text{int}}$ and similarly for γ). The scale is chosen so that $T\,\kappa_{\text{int}} \approx 1$, placing decoherence on the same timescale as the protocol.
- Asymmetry sweep: $\rho \in \{1, 1.5, 2, 4, 8\}$.
- Total optimiser steps: 10 000.
- Solver: ExpSplit
- Discretised time grid: 256 points per trajectory, snapshot every 500 steps; 64 SSE samples per gradient step.
- Regulariser weights: $\lambda_{\text{KL}_\text{w}} = 0$, $\lambda_{R_{\text{DV}}} \in \{0, 0.02\}$, $\lambda_{\text{flu}} = 0.005$.
- Fluence warm-up: 2000 steps (penalty held at zero, then enabled).
- Fourier controller: $n_{\text{modes}} = 16$, init_scale $= 1.0$.
- Seeds: 1 at $\rho \in \{1, 1.5\}$, 3 at $\rho \in \{2, 4, 8\}$.

E.6. Hardware-calibrated 6 qubit chain experiment

Tables 9 and 10 report the per-site coherence times and the multi-seed fidelity statistics for the six-qubit excitation-transfer experiment of Section 5.3. The calibration source is retrieved from the IBM Quantum Platform at 11:33:48 UTC on 3 May 2026 for processor `ibm_kingston` (IBM Quantum, 2026). IBM Quantum calibrations refresh approximately daily; the snapshot used here is fixed, so the per-site rates in Table 9 are reproducible exactly from the specific timestamp but will differ from any newer snapshot. Per-qubit T_1 and T_2 are converted to Lindblad rates via

$$\gamma_i = 1/T_1^{(i)}, \qquad \kappa_i = \tfrac{1}{2}\max\Big(0,\ 1/T_2^{(i)} - 1/(2T_1^{(i)})\Big),$$

yielding two per-site Lindblad operators per qubit ($\sqrt{\gamma_i}\,\sigma_-^{(i)}$ amplitude damping and $\sqrt{\kappa_i}\,\sigma_z^{(i)}$ dephasing). The drive is a sum of five nearest-neighbor exchange couplings, $H_{\text{ctrl}}(t) = \sum_{i=0}^{4} g_{i,i+1}(t)\,\frac{1}{2}(\sigma_x^{(i)}\sigma_x^{(i+1)} + \sigma_y^{(i)}\sigma_y^{(i+1)})$, with no fixed exchange or local drive. Training used a fluence warm-up for stability.

Table 8. **Chain-4 asymmetry sweep:** maximum exact-Lindblad fidelity for excitation transfer site-0 to site-3, by edge-to-interior dephasing ratio ρ. *Top:* mean±std over three seeds where applicable. *Bottom:* per-seed best fidelities for the multi-seed rows.

ρ	Baseline F	Drift-variance ($\lambda_{R_{\mathrm{DV}}} = 0.02$) F	Δ
1.0 (uniform)	0.71	0.71	0.0 pp
1.5	0.69	0.69	0.0 pp
2.0	0.64 ± 0.01	0.62 ± 0.03	-2.4 pp
4.0	0.48 ± 0.02	0.52 ± 0.13	$+4.5$ pp
8.0	0.22 ± 0.04	0.42 ± 0.11	$+19.9$ pp

ρ	policy	seed 0	seed 1	seed 2
2.0	baseline	0.637	0.636	0.651
2.0	drift-variance	0.606	0.591	0.654
4.0	baseline	0.482	0.503	0.454
4.0	drift-variance	0.609	0.589	0.374
8.0	baseline	0.193	0.209	0.262
8.0	drift-variance	0.501	0.456	0.306

Table 9. Per-site coherence times for the IBM Kingston q14→q9 chain, with derived $\gamma + \kappa$ rate. Site index runs from source (left, q14) to target (right, q9).

Site	Qubit	T_1 (μs)	T_2 (μs)	$\gamma + \kappa$ ($\mu\mathrm{s}^{-1}$)
0	q14	147.5	191.2	0.0077
1	q13	402.9	352.0	0.0033
2	q12	355.1	272.8	0.0039
3	q11	402.7	93.3	0.0072
4	q10	346.5	17.0	0.0316
5	q9	137.8	61.8	0.0135

Experiment-specific parameters.

- $T = 30\,\mu$s in physical units; rates above are in $\mu\mathrm{s}^{-1}$.
- Initial state $|100000\rangle$ (excitation at site 0, q14), target $|000001\rangle$ (excitation at site 5, q9).
- Total optimiser steps: 3000.
- Solver: ExpSplit
- Discretised time grid: 1024 points per trajectory, snapshot every 250 steps; 64 SSE samples per gradient step.
- Regulariser weights: $\lambda_{\mathrm{flu}} = 0.02$; $\lambda_{R_{\mathrm{DV}}} \in \{0, 16\}$; $\lambda_{\mathrm{KL_w}} \in \{0, 0.1\}$, with the non-zero value chosen so that the average loss contribution of the KL term matches that of the drift-variance term at $\lambda_{R_{\mathrm{DV}}} = 16$ (we also tried $\lambda_{\mathrm{KL_w}} = 16$, reported as a single-seed control).
- Fluence warm-up: 500 steps.
- Fourier controller: $n_{\mathrm{modes}} = 16$, init_scale $= 1.0$.
- Seeds: 3 for each of baseline, $\lambda_{R_{\mathrm{DV}}} = 16$, and $\lambda_{\mathrm{KL_w}} = 0.1$.

F. PPO baseline configuration

We use the constrained-RL formulation of Ernst et al. (2025) as our trajectory-based RL baseline.

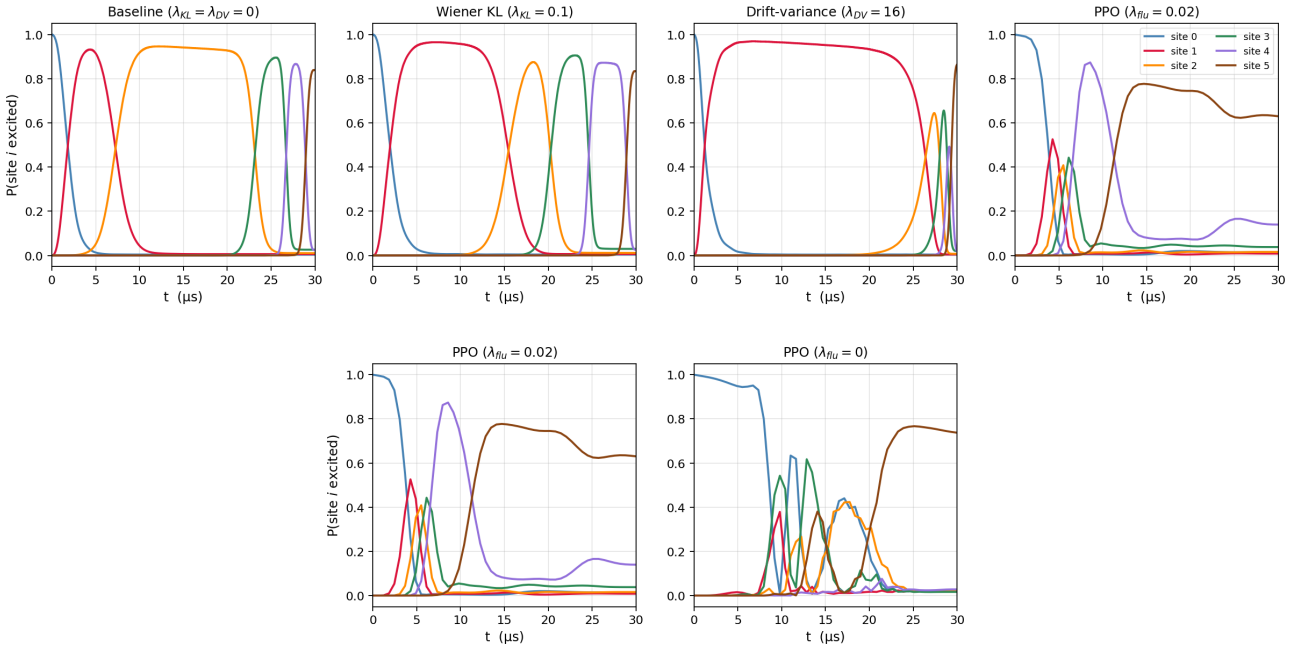

Figure 7. **IBM Kingston q14→q9 evolutions.** Per-site excitation populations $\langle\psi(t)|(I - \sigma_z^{(i)})/2|\psi(t)\rangle$ (i.e., the probability that site i holds the excitation). *Top row*: the four policies of Table 10 at matched squared-fluence penalty $\lambda_{\mathrm{flu}} = 0.02$ — baseline ($\lambda_{\mathrm{KL_W}}$=$\lambda_{R_{\mathrm{DV}}}$=0), Wiener KL ($\lambda_{\mathrm{KL_W}}$=0.1), drift-variance ($\lambda_{R_{\mathrm{DV}}}$=16), and PPO. Drift-variance reaches the highest final-time excitation on q9. *Bottom row*: PPO compared at matched fluence penalty ($\lambda_{\mathrm{flu}} = 0.02$) and with the fluence penalty removed entirely ($\lambda_{\mathrm{flu}} = 0$).

Table 10. **IBM Kingston q14→q9:** maximum exact-Lindblad fidelity for excitation transfer, $T = 30\,\mu$s, linear fluence penalty $\lambda_{\mathrm{flu}} = 0.02$. Mean ± std across 3 seeds where applicable. Bold marks the best. The bottom block reports two control experiments (overdriven KL, unconstrained PPO) discussed in Section 5.3.

method	F_{exact}
baseline ($\lambda_{\mathrm{KL_W}}$=$\lambda_{R_{\mathrm{DV}}}$=0, $\lambda_{\mathrm{flu}} = 0.02$)	0.837 ± 0.004
Wiener KL ($\lambda_{\mathrm{KL_W}}$=0.1, $\lambda_{\mathrm{flu}} = 0.02$)	0.831 ± 0.004
Drift-variance ($\lambda_{R_{\mathrm{DV}}}$=16, $\lambda_{\mathrm{flu}} = 0.02$)	$\mathbf{0.863 \pm 0.002}$
PPO ($\lambda_{\mathrm{flu}} = 0.02$)	0.604 ± 0.041
Wiener KL ($\lambda_{\mathrm{KL_W}}$=16, single seed)	0.427
PPO ($\lambda_{\mathrm{flu}} = 0$)	0.737

Reference implementation. We use the authors' `train_ppo.py` from the released codebase (Ernst et al., 2025) together with their PPO implementation, wrapping our QMaxCal system modules through Ernst's generic environment interface. The reward function follows Eq. 2 of Ernst et al. (2025).

Architecture and core hyperparameters. Separate actor-critic MLPs with hidden width 256 and ReLU6 activations; Gaussian policy. PPO with clip $\epsilon = 0.2$, GAE $\lambda = 0.95$, discount $\gamma_{\mathrm{RL}} = 0.99$, max gradient norm 0.5, 4 minibatches per update, no entropy bonus. Adam optimizer with linear learning-rate annealing. Training runs 50,000 update steps with 16 parallel environments and 50 piecewise-constant action samples per episode (4×10^7 environment interactions per run), smoothed by a Gaussian kernel of length 25. The constrained-RL maximum ODE-solver step is $N_{\max}^{\mathrm{sim}} = 1000$ with penalty 25 when exceeded. For the IBM Kingston chain-6 system only, we chunk training into blocks of 5000 updates (saving an intermediate snapshot at each chunk boundary) and cap each run at 3 h of GPU wall time; if the cap is hit, the most recent snapshot is used. For reference, the QMaxCal training runs for the 6 qubit chain completed in under 2 h of GPU wall time.

Hyperparameter sweep. We swept the PPO hyperparameters in two sequential phases per benchmark.

Phase A: $N_{\max}^{\mathrm{sim}} \times \lambda_{\mathrm{flu}}$ on a $\{500, 1000, 2000, 4000\} \times \{0.001, 0.01, 0.1\}$ grid for amplitude damping, where $N_{\max}^{\mathrm{sim}}$ is

the headline knob from Fig. 1 of Ernst et al. (2025) and λ_{flu} is a quadratic fluence penalty $\lambda_{\text{flu}}\int_0^T |\Omega|^2 dt$ that we add to Ernst's reward (distinct from Ernst's linear pulse-area penalty $w_A \int |\Omega| dt$ in Eq. 2; the quadratic form matches QMaxCal's regulariser and places PPO on the same fluence–fidelity Pareto frontier; we set Ernst's linear $w_A = 0$). For STIRAP and the diamond system we used the reduced Phase A grid $\{1000, 4000\} \times \{0.001, 0.01, 0.1\}$ (6 cells), since the amplitude-damping sweep showed $N_{\max}^{\text{sim}}$ to be flat in the relevant range. For the IBM Kingston chain-6 we omit Phase A entirely: λ_{flu} is fixed at 0.02 to match the QMaxCal regulariser strength, and $N_{\max}^{\text{sim}} = 1000$ is inherited from the converged Phase-A choice on the smaller systems.

Phase B: at the best Phase-A cell, the constraint-weight ablation from Fig. 14 of Ernst et al. (2025): smoothness penalty $\in \{0, 0.001, 0.01\}$, Gaussian-kernel std $\in \{2, 4, 8\}$, and fidelity weight $w_F \in \{1, 5\}$ (18 cells). For the IBM Kingston chain-6 we further restrict Phase B to a 2×2 grid over smoothness $\in \{0, 0.001\}$ and kernel std $\in \{2, 4\}$ with w_F fixed at 1, motivated by GPU-budget constraints (each matched-fluence chain-6 PPO run takes ~ 2.5 h vs. ~ 1 h on the smaller systems).

This yields $12 + 18 = 30$ configurations on amplitude damping, $6 + 18 = 24$ on STIRAP and the diamond system, and $0 + 4 = 4$ on the IBM Kingston chain-6.

Seeds and selection. Sweep runs used a single seed; the final reported numbers come from 4 seeds on amplitude damping, 3 seeds on STIRAP, the diamond system, and the IBM Kingston chain-6 (per Tables 2, 4, 6, 10). Selection criterion: highest exact-Lindblad fidelity F on the deterministic-rollout snapshot.

Per-system overrides. Phase A converged to $(N_{\max}^{\text{sim}}, \lambda_{\text{flu}}) = (1000, 0.001)$ for amplitude damping, STIRAP, and the diamond system; for IBM Kingston, λ_{flu} was set to 0.02 a priori. Phase B winners differed per benchmark:

benchmark	smoothness	kernel std	w_F	λ_{flu}
amp. damping	0	2	5	0.001
STIRAP	0.001	2	5	0.001
diamond	0.001	2	1	0.001
IBM Kingston q14→q9	0.001	4	1	0.02

Fluence-penalty sanity check. To confirm that the quadratic fluence penalty $\text{Flu} = \sum_a \int_0^T |u_a^{(\theta)}(t)|^2\, dt$ is not handicapping PPO, we re-trained the best Phase-B configuration on all four benchmarks with $\lambda_{\text{flu}} = 0$ (single seed each). Removing the penalty did not improve PPO on amplitude damping and STIRAP: fidelity was unchanged within seed noise. On the diamond system, pulse fluence grew by $26\times$ ($\sum_a \int_0^T |u_a^{(\theta)}(t)|^2\, dt : 119 \to 3116$) while fidelity dropped slightly ($F : 0.603 \to 0.559$). On the IBM Kingston 6 qubit chain, removing the penalty raised fidelity from 0.604 to 0.737 (Table 10, single seed), while the fluence increase $66\times$ ($7.84 \to 519.89$). We nonetheless report the matched-fluence number as the headline PPO result to maintain a fair comparison with QMaxCal, which uses $\lambda_{\text{flu}} = 0.02$.

G. Smoothness regularization is not Wiener KL

We rule out the hypothesis that Wiener KL acts as an implicit smoothness penalty on the controls $u_a^{(\theta)}(t)$. The derivative penalties of Abdelhafez et al. (2019) (C4 and C5) take the form

$$\lambda_{d_1} \int_0^T |u'(t)|^2\, dt \;+\; \lambda_{d_2} \int_0^T |u''(t)|^2\, dt$$

we add them to QMaxCal at two weight settings on amplitude damping ($\gamma T = 2$) and STIRAP ($\gamma T = 10$), with smoothness terms warmed up over the first 1500 optimiser steps.

Table 11 shows that adding C4/C5 strictly costs fidelity in every cell and never recovers the Wiener-KL-only result. On STIRAP the heavier weight collapses fidelity by 16 pp; combining it with Wiener KL adds at most 1 pp over smoothness alone, confirming that Wiener KL cannot rescue a trajectory already constrained by pulse-derivative penalties. Reported numbers are single-seed; the effect size (16pp collapse on STIRAP) is large enough that seed variance is unlikely to overturn the conclusion.

The two regularizers act on physically distinct objects: Wiener KL acts on the dissipation-channel drift $\alpha_k(t) =$

Table 11. **Smoothness ablation.** Single seed; smoothness warmed up over 1500 steps. *Top:* amplitude damping ($\gamma T = 2$, $\lambda_{\mathrm{KL_W}} = 5$). *Bottom:* STIRAP ($\gamma T = 10$, $\lambda_{\mathrm{KL_W}} = 1$). Bold marks the highest fidelity per block.

system	method	F_{exact}
amp. damp.	baseline	0.972
amp. damp.	Wiener KL	**0.981**
amp. damp.	smooth only ($10^{-3}, 10^{-5}$)	0.932
amp. damp.	smooth + Wiener KL ($10^{-3}, 10^{-5}$)	0.941
STIRAP	baseline	0.978
STIRAP	Wiener KL	**0.979**
STIRAP	smooth only ($10^{-3}, 10^{-5}$)	0.819
STIRAP	smooth + Wiener KL ($10^{-3}, 10^{-5}$)	0.822
STIRAP	smooth only ($10^{-4}, 10^{-6}$)	0.914
STIRAP	smooth + Wiener KL ($10^{-4}, 10^{-6}$)	0.919

$2\,\mathrm{Re}\langle\psi|L_k|\psi\rangle$ in trajectory space, driving it toward $\ker L$; C4/C5 act on derivatives of the controls in pulse space, agnostic to the trajectory's relationship to the noise kernel. They cannot substitute for each other.

Parameters. Amplitude damping: $T = 1$, $\gamma = 2$, 5000 steps, 256 SSE samples, $n_{\mathrm{modes}} = 20$. STIRAP: $T = 2$, $\gamma = 5$, 10 000 steps, 64 SSE samples, $n_{\mathrm{modes}} = 16$.

H. SSE solver schemes and scaling behaviour

This appendix documents the two diffusive-SSE integrators used to optimise QMaxCal protocols, characterises how each one scales with system size N, and contrasts both with the Lindblad master-equation solver (`qiskit-dynamics` on top of `diffrax` Dopri5) used as ground truth. Our SSE-integrator implementations build on `dynamiqs` (Guilmin et al., 2025), a JAX-based library for differentiable open-quantum-system simulation; we extended its diffusive-SSE solvers with the Girsanov KL and drift-variance loss terms of Section 4.

H.1. The two SSE integrators

We solve the diffusive Stochastic Schrödinger Equation

$$\begin{aligned} d|\psi\rangle = \Big[-iH\,dt - \tfrac{1}{2}\sum_k L_k^\dagger L_k\,dt + \tfrac{1}{2}\sum_k \langle L_k + L_k^\dagger\rangle L_k\,dt - \tfrac{1}{8}\sum_k \langle L_k + L_k^\dagger\rangle^2\,dt\Big]|\psi\rangle \\ + \sum_k \big(L_k - \tfrac{1}{2}\langle L_k + L_k^\dagger\rangle\big)|\psi\rangle\,dW_k \end{aligned} \tag{32}$$

with two integrators, summarised in Table 12.

Euler–Maruyama (EM). A first-order weak-order-1 stochastic Euler step that applies H and the $\{L_k\}$ to $|\psi\rangle$ as matrix–vector products. EM never instantiates a dense propagator, so memory stays linear in $\dim$. Local truncation error is $O(dt \cdot \|H_{\mathrm{eff}}\|^2)$, so large-norm Hamiltonians require proportionally finer time discretisation.

Exponential split-step (ExpSplit). Splits each timestep into a deterministic half, solved exactly via $\exp(-iH_{\mathrm{eff}}\,dt)|\psi\rangle$ with $H_{\mathrm{eff}} = H - i\frac{1}{2}\sum_k L_k^\dagger L_k$, and an Euler-style stochastic correction. The matrix exponential makes the deterministic half Hamiltonian-exact and removes the $\|H\|$-dependence of EM's truncation error, but instantiates a dense $\dim \times \dim$ propagator at every step, costing $O(\dim^3)$ compute and $O(\dim^2)$ memory.

The Lindblad reference solver propagates the full density matrix $\rho(t)$ via adaptive Dopri5 on the Liouvillian, costing $O(\dim^3)$ compute and $O(\dim^2)$ memory per adaptive step.

Table 12. Per-step cost and error scaling. n_{trajs} is the number of stochastic samples; $\dim = 2^N$.

solver	per-step compute	per-step memory	Hamiltonian error
EM	$O(\dim^2 \cdot n_{\text{trajs}})$	$O(\dim \cdot n_{\text{trajs}})$	$O(dt^2 \|H\|^2)$
ExpSplit	$O(\dim^3)$	$O(\dim^2)$	exact

H.2. Scaling with N

We benchmarked all three solvers on random N-qubit XX+YY chains (per-site σ_- and σ_z Lindblad channels, random nearest-neighbour exchange, $T_{\text{final}} = 1$, Fourier control pulses with $n_{\text{modes}} = 4$). All runs used a single 40 GB A100. Wall-clock and peak GPU memory are reported in Table 13.

Table 13. Cached forward-pass wall-clock and peak GPU memory for the three solvers on random N-qubit chains; same controller, same Lindblad operators, $n_{\text{trajs}} = 256$, $n_{\text{time-steps}} = 2048$, single 40 GB A100.

N	dim	Lindblad call	Lindblad peak GPU	EM call	ExpSplit call
4	16	—	—	0.1 s	1.2 s
6	64	3.9 s	82 MiB	0.2 s	1.4 s
8	256	4.3 s	130 MiB	0.5 s	2.7 s
10	1024	21.8 s	1.85 GiB	3.6 s	15.1 s
12	4096	**GPU-OOM**		**45.8** s	JIT did not complete in 2 h

Two qualitative results stand out. First, the two $O(\dim^3)$ solvers hit a wall at the same scale: by $N = 12$, the dense $\dim \times \dim$ matrices needed for density-matrix propagation (Lindblad) and for the matrix exponential (ExpSplit) exceed practical GPU memory and JIT-compile budgets — the Lindblad solver runs out of GPU memory and the ExpSplit solver fails to finish JIT-lowering within a 2 h budget. EM's $O(\dim \cdot n_{\text{trajs}})$ state representation pushes the ceiling roughly two qubits further, with rapidly growing wall-clock. Second, at the system sizes that are reachable for all three (up through $N = 10$), EM and ExpSplit run several times faster than the Lindblad reference on identical hardware — consistent with the quadratic state-space advantage of trajectory representations discussed in Appendix A.

H.3. When EM is too coarse: Hamiltonian-amplitude sensitivity

EM's local error scales as $O(dt \cdot \|H_{\text{eff}}\|^2)$, so EM becomes inaccurate when $\|H_{\text{eff}}\|$ is large at fixed time-grid resolution. We illustrate with a controlled experiment at $N = 10$: fixed seed, $n_{\text{trajs}} = 1024$ (Monte Carlo noise floor ~ 0.031 in Frobenius distance to Lindblad), random-Fourier pulse amplitude varied over $\{1, 3, 10\}$ (peak $\|H\|_{\text{op}} \approx 17, 35, 105$), and $n_{\text{time-steps}}$ swept over a $256\times$ range.

Table 14. EM Frobenius distance to the Lindblad ground truth, $N = 10$. Rows: pulse amplitude. Columns: time-discretisation. Bold cells indicate bias dominating the Monte Carlo floor of ~ 0.031.

pulse amp	$\|H\|_{\text{op}}$	nts=256	1024	4096	16384	65536
1	17	0.039	0.028	0.029	0.024	0.032
3	35	**0.121**	0.038	0.031	0.028	0.031
10	105	**0.629**	**0.204**	**0.052**	0.033	0.030

Three observations: (i) at amplitude 1 EM is converged at $n_{\text{time-steps}} = 256$ and refining the grid 256-fold does not change the result; (ii) at amplitude 3 a $4\times$ refinement reaches the noise-limited regime; (iii) at amplitude 10 EM bias dominates the Frobenius error until $n_{\text{time-steps}} \geq 16384$, with the 256-step run producing a Frobenius error of 0.63 — twenty times the MC floor, making the SSE estimate effectively meaningless. The scaling tracks the theoretical prediction bias $\sim dt \cdot \|H_{\text{eff}}\|^2 \cdot T$: a $6\times$ growth in $\|H\|_{\text{op}}$ requires $\sim 36\times$ more time-steps for matched bias, and the empirical $64\times$ refinement required to recover convergence is within a factor of two of the prediction.

Solver choice for the experiments in this paper. We use ExpSplit for all training and evaluation runs reported in the body of the paper, except for the diamond system, which uses Euler–Maruyama. The amplitude-sensitivity experiment of

Section H.3 shows that EM accuracy at fixed $n_{\text{time-steps}}$ degrades quickly once peak $\|H_{\text{eff}}\|$ rises above ~ 30, requiring roughly $\|H\|^2$-scaled refinement of the time grid to recover convergence. ExpSplit's deterministic half is Hamiltonian-exact and removes this sensitivity entirely, so we do not have to revisit the time-grid resolution when sweeping decoherence rates, control amplitudes, or the number of qubits.

For the diamond system EM is comfortably within its accurate regime: the four-level Hilbert space and the small-amplitude controls ($\lambda_{\text{flu}} = 0.001$ keeps the controller amplitude well below the threshold of Table 14) place it firmly in the EM-converged zone, and at $\dim = 4$ the $O(\dim^2)$ trajectory representation is strictly faster than ExpSplit's $O(\dim^3)$ matrix exponential. The larger benchmarks (chain-4 at $\dim = 16$, chain-6 at $\dim = 64$, plus the amplitude-damping and STIRAP sweeps where we wanted insurance against the strong-noise corner of the parameter sweep) use ExpSplit. For systems beyond the dimensions in this paper EM remains the appropriate choice, with $n_{\text{time-steps}}$ chosen to match the expected operator norm.

I. Compute budget

All experiments ran on a single NVIDIA A100 GPU (40 GB, `gpu_a100`). The compute consumed by the experiments reported in this paper breaks down approximately as $\sim$55 h on amplitude-damping, STIRAP, and diamond system training (including their PPO baselines); $\sim$25 h on the chain-4 ρ-sweep; $\sim$30 h on the IBM Kingston q14$\rightarrow$q9 chain (including the PPO hyperparameter sweep and multi-seed extension); and $\sim$10 h on the scaling benchmarks of Appendix H, for a reported total of approximately **120 A100-hours**. The full project, including hyperparameter sweeps, alternative system parameterisations, and exploratory variants that did not make it into the final paper, consumed approximately **240 A100-hours** of GPU compute.