

Robust Learned Quantum Experiment Design under Decoherence Distribution Shift

Changhao He

Abstract—Autonomous learned experiment-design policies can overfit an assumed decoherence model and lose performance when the dephasing rate changes between calibration episodes. We study two distinct robust formulations and keep them mathematically separate: a finite-radius order-1 Wasserstein DRO objective for analytic interrogation schedules, and support-based minimax or minimax-relative-regret objectives for learned policies. For the Wasserstein problem, strong duality converts the supremum over probability laws into a convex one-dimensional minimization in the transport multiplier whose function evaluations require a one-dimensional *global* maximization over the decoherence rate; the inner problem is not asserted to be convex. For online inference we use a univariate skew-normal assumed-density filter obtained by global moment matching, together with an optional rank-one expected-Fisher covariance proxy. Analytic robust schedules reduce the worst-case information-matrix risk from 0.265 to 0.139 and lower actual θ -RMSE by approximately 17–20% under the reported $3\times$ rate mismatch. In the learned-policy study, relative-regret training yields the lowest reported worst-case suboptimality across rates ($15\% \pm 3\%$, versus 19%–157% for rate-specialist policies), while single-rate RMSE differences remain statistically small. These results support robust objective design under episode-to-episode decoherence shift without conflating Wasserstein DRO, support minimax, and minimax regret.

Index Terms—Quantum metrology and sensing, adaptive Bayesian experimental design, Wasserstein distributionally robust optimization, minimax regret, distribution shift, decoherence.

I. INTRODUCTION

Learned agents for quantum experiment design. Choosing *which* measurement to make next, the interrogation time t of a qubit probe, is a sequential decision problem, and *learned* policies (amortized Bayesian optimal experimental design, reinforcement learning, evolutionary strategies) are increasingly used to automate it for quantum sensing and self-driving labs. Such an agent is trained against an *assumed* noise model. But in the operating regime of NISQ calibration (frequent re-estimation against a tight coherence budget and episode-to-episode variation in the dephasing rate) and single-spin / single-molecule sensing (a few coherent readouts before the probe decoheres or photobleaches), the decoherence rate is uncertain and may vary from one calibration episode to the next. An agent tuned to the *wrong* decoherence model suffers a *distribution shift*: it wastes its long interrogations on already-decohered signals (fig. 1).

Main contribution. We formulate robust experiment design under *episode-to-episode* decoherence shift. The analytic schedule is optimized with a finite-radius Wasserstein ambiguity

set, while learned policies are trained with either support worst-case risk or worst-case relative regret. The distinction matters: a finite Wasserstein ball, an unrestricted family of laws supported on Γ , and a regret-normalized objective are three different optimization problems. The experiments show that short-interrogation hedging can substantially reduce worst-case risk, whereas the learned-policy benefit is primarily a flatter cross-rate regret envelope rather than a statistically significant gain at every individual rate.

Supporting contributions.

- **Correct Wasserstein reduction.** Strong duality yields a convex scalar outer problem in λ and a scalar *global* inner maximization in γ . Compactness guarantees attainment; convexity or unimodality of the inner maximization is not assumed.
- **Fast approximate inference.** A univariate skew-normal ADF globally matches the first three posterior moments. Separately, an expected-Fisher rank-one precision update admits an exact Sherman–Morrison inverse for the resulting *approximate* covariance proxy.
- **Asymptotic calibration.** Under explicit adaptive-LAN and posterior-contraction conditions, the regular posterior has the usual Gaussian Bernstein–von Mises limit. A local cubic skew-symmetric expansion can improve higher-order approximation, but it does not identify the standard Azzalini skew-normal as a universal asymptotic family.

Scope. The robust-design conclusions do not depend on the skew-normal approximation: the reported 2×2 ablation shows less than a one-percent decision-level difference between Gaussian and skew-aware beliefs under the same robust schedule. The skew-normal component is therefore a finite-sample density approximation, not the mathematical basis of the robustness claim.

II. RELATED WORK

Quantum sensing and estimation. Quantum sensing turns a controllable quantum system into a high-precision probe [1], [2], with attainable precision governed by the quantum Fisher information and the quantum Cramér–Rao bound [3], [4]. Decoherence fundamentally limits the quantum enhancement [5], and quantum error correction can partially restore it [6]; fast, low-overhead recalibration of such probes is a core NISQ requirement [7].

Adaptive and learned experiment design. Adaptive Bayesian schemes for qubit-Hamiltonian and phase estimation [8], [9] and machine-learned measurement policies [10] predate the modern amortized view. Amortized / policy-based Bayesian

The author is with the CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia (e-mail: changhao.he@kaust.edu.sa).

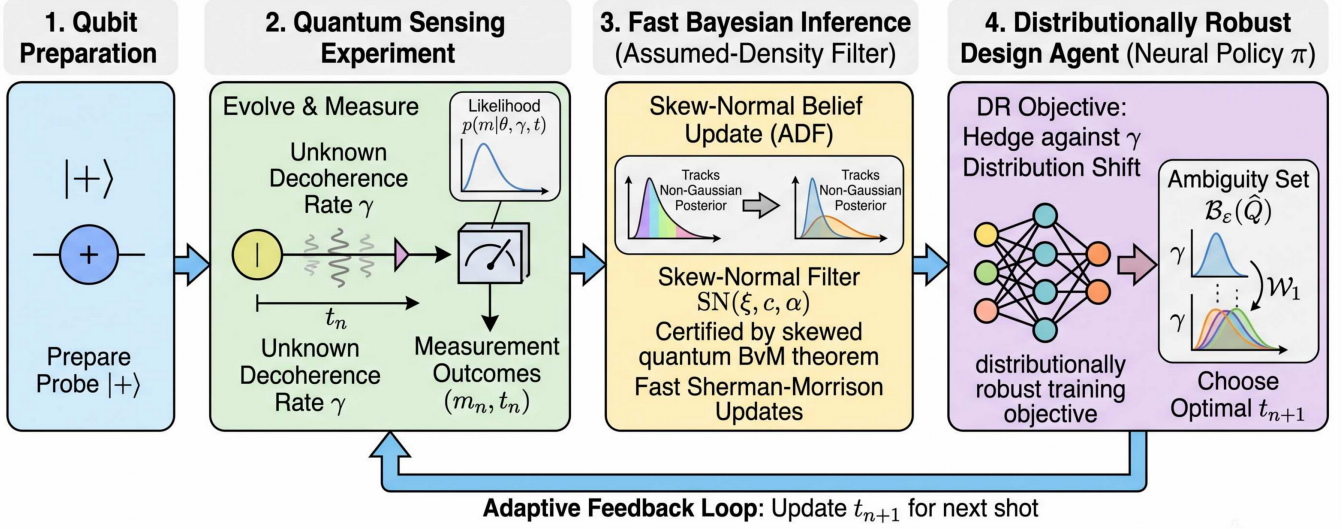


Fig. 1. **System model and adaptive sensing loop.** The four-stage pipeline of robust adaptive quantum sensing. **(1) Qubit preparation:** a probe is prepared in $|+\rangle$. **(2) Quantum sensing:** it evolves for a controllable interrogation time t_n under a dephasing rate γ that is constant within an episode but may change between episodes, and is measured with Born-rule likelihood $p(m | \theta, \gamma, t)$. **(3) Fast Bayesian inference:** a univariate skew-normal assumed-density filter tracks a potentially asymmetric posterior for θ by moment matching; an expected-Fisher rank-one update is used only as a covariance proxy. **(4) Robust design:** analytic schedules use a finite-radius Wasserstein ambiguity set, whereas learned policies use support-minimax or relative-regret training over a prescribed rate set.

optimal experimental design (Deep Adaptive Design and its implicit variant [11], [12]) and reinforcement-learning / evolutionary-strategy metrology policies [13], [14] now automate sequential measurement choice. We show these learned policies inherit a *distribution-shift* brittleness familiar from robust reinforcement learning, and import Wasserstein distributionally robust optimization for analytic schedules and minimax-regret training for learned policies [15], [16], [17], [18].

Robust quantum metrology and asymmetric posterior approximations. Robust quantum metrology guards a set of channels through protected quantum Fisher information [6] or worst-case measurement noise [24]. We instead optimize the distribution of an experimental design under an explicitly specified ambiguity model. Our finite-sample belief approximation uses the skew-normal family [19]. The higher-order skew-symmetric Bernstein–von Mises expansion of [20] motivates local cubic corrections, but it does not imply that a standard linear-shape skew-normal is the universal higher-order posterior limit.

III. SYSTEM MODEL, LIKELIHOOD, AND THE SEQUENTIAL PROBLEM

Table I fixes one symbol per quantity, used throughout.

Why this regime matters. Few measurements under an uncertain, episode-dependent decoherence rate is the *operating* regime of two important settings. (i) *NISQ calibration:* qubit frequencies and couplings must be re-estimated frequently

TABLE I
NOTATION: ONE SYMBOL PER QUANTITY, USED THROUGHOUT THE PAPER.

Symbol	Meaning
$\theta \in \mathbb{R}$	qubit frequency/phase (parameter of interest)
$\gamma \geq 0$	dephasing rate (nuisance, possibly unknown)
$\boldsymbol{\eta} = (\theta, \gamma)^\top \in \mathbb{R}^d$	full parameter vector ($d = 2$ minimal model)
$t > 0$	controllable evolution time (design variable)
$m \in \{+1, -1\}$	measurement outcome (σ_x eigenvalue)
$V(t) = e^{-\gamma t}$	fringe visibility
$p(m \theta, \gamma, t)$	Born-rule likelihood of one shot
$\mathcal{D}_N = \{(m_n, t_n)\}_{n=1}^N$	data after N shots
π_0, π_N	prior, posterior $\pi(\boldsymbol{\eta} \mathcal{D}_N)$
$\ell_N(\boldsymbol{\eta}) = \log \pi_N(\boldsymbol{\eta})$	log-posterior
$\mathbf{s}, \mathbf{J}, \boldsymbol{\Sigma}$	score, expected-information proxy, covariance proxy
$\phi(\cdot), \Phi(\cdot)$	standard normal pdf, cdf
$\text{SN}(\xi, c, \alpha)$	univariate skew-normal: location ξ , scale c , shape α
$\mathcal{W}_p, \varepsilon, \mathcal{B}_\varepsilon(\hat{Q})$	p -Wasserstein distance; radius ε ; ambiguity ball $\{Q \in \mathcal{P}(\Gamma) : \mathcal{W}_1(Q, \hat{Q}) \leq \varepsilon\}$ around $\hat{Q}$
$\hat{Q}, \Gamma$	nominal decoherence law; support of admissible rates
support minimax	$\sup_{\gamma \in \Gamma} L(\gamma)$, equivalently worst case over all laws supported on Γ
$L(t, \gamma)$	generic analytic design risk or information-matrix surrogate
$M(\mathbf{w}; \gamma)$	policy mean-squared error $\mathbb{E}[(\hat{\theta}_N - \theta)^2]$
$R(\mathbf{w}; \gamma)$	policy RMSE $\sqrt{M(\mathbf{w}; \gamma)}$
$r_R(\gamma)$	RMSE regret $R(\gamma)/R^*(\gamma) - 1$ for a stated comparator class

against a tight coherence budget, so a many-shot asymptotic design is unaffordable and the dephasing rate itself is uncertain

and may change between calibration episodes. (ii) *Single-spin / single-molecule sensing* (NV centers, single molecules): only a handful of coherent readouts are available before the probe decoheres or photobleaches, and the relaxation rate is a nuisance that may need to be co-estimated. Both demand accurate, *well-calibrated* estimates from few shots while remaining robust to a misspecified decoherence model, exactly the problem addressed here.

A. State, channel, and Born-rule likelihood

A qubit is prepared in $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, evolves for a controllable time t under a Hamiltonian $H = \frac{\theta}{2}\sigma_z$ while subject to pure dephasing at rate γ , and is read out in the σ_x basis. The dephasing channel shrinks the x - y Bloch components by $V(t) = e^{-\gamma t}$, so the pre-measurement state is

$$\rho_{\theta,\gamma,t} = \frac{1}{2}(\mathbb{I} + V(t)\cos(\theta t)\sigma_x + V(t)\sin(\theta t)\sigma_y). \quad (1)$$

Born's rule for the outcome $m \in \{+1, -1\}$ of σ_x gives the single-shot likelihood

$$p(m | \theta, \gamma, t) = \frac{1}{2}(1 + m V(t)\cos(\theta t)), \quad V(t) = e^{-\gamma t}. \quad (2)$$

B. Score and Fisher information

Differentiating $\log p$ in (2) w.r.t. θ (full steps in section B):

$$s(\theta; m, t) = \partial_\theta \log p(m | \theta, \gamma, t) = \frac{-m V t \sin(\theta t)}{1 + m V \cos(\theta t)}. \quad (3)$$

Because $m^2 = 1$, the single-shot (classical) Fisher information $I(\theta, t) = \mathbb{E}_m[s^2]$ collapses to the closed form

$$I(\theta, t) = \frac{V^2 t^2 \sin^2(\theta t)}{1 - V^2 \cos^2(\theta t)}. \quad (4)$$

For phase points with $\sin(\theta t) \neq 0$, the regular limit $V \rightarrow 1$ gives $I \rightarrow t^2$. At the likelihood-zero fold points of the perfectly coherent model, the usual regular Fisher formula must be interpreted by a limit. Any $\gamma > 0$ removes those zeros and attenuates the information. (3)–(4) are verified against finite differences and against $\mathbb{E}_m[s^2]$ to machine precision (section XII, Part 1).

a) *The joint (θ, γ) Fisher matrix.*: The nuisance analysis also needs the single-shot Fisher *matrix* of $\boldsymbol{\eta} = (\theta, \gamma)$. With $\partial_\gamma \log p = -m V t c / (1 + m V c)$ alongside (3) (and $c = \cos \theta t$, $s = \sin \theta t$), the same $m^2=1$ algebra gives

$$\mathbf{J}(\theta, \gamma; t) = \frac{V^2 t^2}{1 - V^2 c^2} \begin{pmatrix} s^2 & sc \\ sc & c^2 \end{pmatrix}, \quad (5)$$

whose (θ, θ) entry recovers (4). A *single* shot is rank-one ($\det \mathbf{J} = 0$): one interrogation cannot separate θ from γ . The accumulated matrix $\sum_n \mathbf{J}(\theta, \gamma; t_n)$ is full rank only when the schedule visits times with linearly independent (s_n, c_n) , which is exactly why a *varying- t schedule* (not one optimal t) is needed for joint (θ, γ) inference; (5) is checked against $\mathbb{E}_m[\text{ss}^\top]$ to machine precision (section XII).

C. The sequential design problem

Given a prior $\pi_0(\boldsymbol{\eta})$ and a loss (e.g. squared error on θ), an adaptive policy chooses each time t_n as a function of the history $\mathcal{D}_{n-1}$, observes $m_n \sim p(\cdot | \theta, \gamma, t_n)$, and updates the posterior. The goals are: (i) represent the few-shot posterior accurately (section V); (ii) update and design in (near) closed form (section VI); (iii) stay accurate when γ 's law is misspecified (section VII). Formally,

$$t_n^* = \arg \min_t \mathbb{E}_{m \sim p(\cdot | t, \mathcal{D}_{n-1})} [\mathcal{R}(\pi_n)], \quad (6)$$

$\mathcal{R}$ = posterior risk (section VI).

IV. MOTIVATION: WHY THE CRAMÉR–RAO BOUND IS INSUFFICIENT BY ITSELF

A. Local vs. global error

The classical Cramér–Rao inequality is a local lower bound for unbiased estimators under regularity conditions; its usual attainability argument is asymptotic. In a few-shot, periodic likelihood, the posterior may be asymmetric or multimodal, so the scalar quantity $1/(NI)$ can be a poor approximation to the actual Bayes risk. It also conditions on a single data-generating model and therefore cannot, by itself, represent ambiguity in the probability law of γ . We consequently use information-matrix quantities only as transparent design surrogates and validate selected schedules with posterior Monte Carlo.

B. Bayesian bounds that are tight in the few-shot regime

The quantum Ziv–Zakai bound (QZZB) [21] and the quantum Weiss–Weinstein bound (QWWB) [22] are *Bayesian* bounds that, by converting estimation into a sequence of binary state-discrimination problems, remain informative exactly where the QCRB is loose: the non-Gaussian, few-measurement regime. They are used here as *motivation*, not as a claimed result.

C. The few-shot posterior can be measurably skewed

With a Gaussian prior $\theta \sim \mathcal{N}(1.0, 0.6^2)$, $\gamma = 0.15$, and $N = 3$ shots at $t \in \{0.9, 1.7, 2.3\}$, the exact grid posterior in the reported example has standardized third central moment approximately -0.46 . This is a finite-sample diagnostic showing that a symmetric Gaussian fit can miss posterior shape. It does *not* imply that the skewness remains $O(1)$ asymptotically: under a regular full-dimensional Bernstein–von Mises limit, every fixed standardized marginal converges to a Gaussian and its skewness vanishes.

V. THE FEW-SHOT POSTERIOR BELIEF: A GLOBAL SKEW-NORMAL FILTER

A. The skew-normal family and its moment map

The skew-normal density is

$$\text{SN}(x; \xi, c, \alpha) = \frac{2}{c} \phi\left(\frac{x - \xi}{c}\right) \Phi\left(\alpha \frac{x - \xi}{c}\right), \quad x \in \mathbb{R}, \quad (7)$$

with location ξ , scale $c > 0$, shape $\alpha \in \mathbb{R}$ ($\alpha = 0$ recovers the Gaussian). Writing $\delta = \alpha/\sqrt{1+\alpha^2}$ and $b = \sqrt{2/\pi}$, and setting $r := \delta b$, its first three moments are

$$\text{mean } \mu = \xi + cr, \quad (8)$$

$$\text{variance } \sigma^2 = c^2(1-r^2), \quad (9)$$

$$\text{skewness } \gamma_1 = \frac{4-\pi}{2} \frac{r^3}{(1-r^2)^{3/2}}, \quad (10)$$

$$|\gamma_1| < \gamma_1^{\max} \approx 0.9953.$$

Proposition 1 (Moment inversion). *Given a feasible triple $(\mu, \sigma^2, \gamma_1)$ with $|\gamma_1| < \gamma_1^{\max}$, the parameters are recovered uniquely by solving (10) for $r \in (-b, b)$; the map is strictly increasing on this interval. Then $c = \sigma/\sqrt{1-r^2}$, $\xi = \mu - cr$, $\delta = r/b \in (-1, 1)$, and $\alpha = \delta/\sqrt{1-\delta^2}$.*

Proposition 1 is the projection map used below; it is verified numerically (section XII, Part 3). The bound $\gamma_1^{\max}$ in (10) is a *limitation*: a single skew-normal cannot represent $|\gamma_1| \gtrsim 1$ (discussed in section XIII).

B. A decisive negative result: local skew correction is not enough

A formal local skew correction is the third-order Laplace (“skew-modal”) expansion of Durante et al.: expanding $\ell_N(\theta)$ about its mode $\hat{\theta}$ with $j = -\ell_N''(\hat{\theta})$ and $\ell_3 = \ell_N'''(\hat{\theta})$,

$$\pi_N(\theta) \approx 2\phi_j(\theta - \hat{\theta})\Phi\left(\frac{\ell_3}{6}\sqrt{\frac{\pi}{2}}(\theta - \hat{\theta})^3\right), \quad (11)$$

$$\phi_j(u) = \sqrt{\frac{j}{2\pi}} e^{-\frac{1}{2}ju^2},$$

derived in section C. **Numerically, (11) does not improve on the Gaussian–Laplace fit in the few-shot regime** (it wins the KL comparison in only 46.6% of trials; section XII, Part 3b). The reason is structural: few-shot skewness is a *global* property of the posterior, whereas (11) only reshapes the tail through the *local* third derivative at the mode. This negative result motivates, but does not mathematically force, a global moment-matched construction.

C. The skew-normal assumed-density filter (ADF)

Maintain a belief $b_n = \text{SN}(\xi_n, c_n, \alpha_n)$, summarized by its moments $(\mu_n, \sigma_n^2, \gamma_{1,n})$. On receiving (m_n, t_n) :

- 1) **Tilt** the belief by the likelihood factor, $\tilde{p}_n(\theta) \propto b_{n-1}(\theta)p(m_n | \theta, \gamma, t_n)$;
- 2) **Project** back by computing $(\mu_n, \sigma_n^2, \gamma_{1,n})$ from $\tilde{p}_n$, applying the inverse moment map of proposition 1 to obtain (ξ_n, c_n, α_n) , and setting $b_n = \text{SN}(\xi_n, c_n, \alpha_n)$.

The filter is initialized at the Gaussian prior ($\alpha_0 = 0$). The first likelihood tilt can already create nonzero skewness. If the empirical third moment lies outside the feasible skew-normal range, the implementation clips it to a value strictly inside $|\gamma_1| < 0.9953$ before inversion; this clipping is an additional approximation. This is a parametric assumed-density filter, not an exact Bayesian recursion. The tilted moments are one-dimensional integrals evaluated by Gauss–Hermite quadrature in the implementation. The Sherman–Morrison update in section VI is a separate expected-information covariance proxy

used for fast screening and should not be confused with the moment projection itself.

a) Empirical ADF tracking. On unimodal in-range posteriors the sequential ADF tracks the exact batch posterior with mean KL = 0.012 (median 0.0075) and attains lower KL than the same-moment Gaussian in 74% of trials (section XII, Part 4). The *batch moment-matched reference* skew-normal wins in 97% of trials, with KL improvement ratio rising from $1.2\times$ (low skew) to $1.7\times$ (moderate skew). These are empirical summaries, not a theorem about all posteriors.

VI. FAST EXPECTED-INFORMATION UPDATES AND ACQUISITION

A. Rank-one expected-Fisher covariance proxy

For the joint parameter $\boldsymbol{\eta} = (\theta, \gamma)^\top$, define the conditional mean of the binary outcome

$$\mu(\boldsymbol{\eta}, t) := \mathbb{E}[m | \boldsymbol{\eta}, t] = V(t) \cos(\theta t),$$

$$\mathbf{g}(\boldsymbol{\eta}, t) := \nabla_{\boldsymbol{\eta}} \mu = -Vt \begin{pmatrix} \sin(\theta t) \\ \cos(\theta t) \end{pmatrix}. \quad (12)$$

Because $m \in \{-1, +1\}$,

$$\text{Var}(m | \boldsymbol{\eta}, t) = 1 - \mu(\boldsymbol{\eta}, t)^2, \quad (13)$$

not $\mathbb{E}[m^2] = 1$. The expected single-shot Fisher matrix is therefore

$$\mathbf{I}(\boldsymbol{\eta}, t) = \frac{\mathbf{g}(\boldsymbol{\eta}, t)\mathbf{g}(\boldsymbol{\eta}, t)^\top}{1 - \mu(\boldsymbol{\eta}, t)^2} = \mathbf{u}(\boldsymbol{\eta}, t)\mathbf{u}(\boldsymbol{\eta}, t)^\top, \quad (14)$$

$$\mathbf{u} := \frac{\mathbf{g}}{\sqrt{1 - \mu^2}},$$

which is identical to (5). A common Laplace/EKF-style covariance proxy evaluates this matrix at the current location estimate $\hat{\boldsymbol{\eta}}_{n-1}$ and updates

$$\mathbf{J}_n^{\text{prox}} = \mathbf{J}_{n-1}^{\text{prox}} + \mathbf{I}(\hat{\boldsymbol{\eta}}_{n-1}, t_n) = \mathbf{J}_{n-1}^{\text{prox}} + \mathbf{u}_n \mathbf{u}_n^\top. \quad (15)$$

Equation (15) is an *approximate expected-information update*; it is neither the observed Hessian $-\nabla^2 \log p(m_n | \boldsymbol{\eta}, t_n)$ nor the exact posterior precision.

Lemma 2 (Sherman–Morrison inverse of the proxy update). *If $\mathbf{J}_{n-1}^{\text{prox}} = \Sigma_{n-1}^{-1} \succ 0$, then the inverse of (15) is exactly*

$$\Sigma_n = \Sigma_{n-1} - \frac{\Sigma_{n-1} \mathbf{u}_n \mathbf{u}_n^\top \Sigma_{n-1}}{1 + \mathbf{u}_n^\top \Sigma_{n-1} \mathbf{u}_n}, \quad (16)$$

at arithmetic cost $O(d^2)$.

Proof: see section A. The algebraic identity has been checked numerically against direct inversion. This check validates only the matrix inversion, not the statistical accuracy of the proxy. The univariate skew-normal ADF still computes its tilted moments by quadrature; no generic $O(d^2)$ complexity claim is made for a multivariate skew-normal moment projection.

B. Acquisition with exact moment projection or a variance proxy

For a candidate time t , the exact ADF acquisition used in the scalar- θ implementation is

$$a_n(t) = \sum_{m=\pm 1} \bar{p}(m | t) \mathcal{R}(b_n^{(m,t)}), \quad t_n^* \in \arg \min_{t \in \mathcal{T}} a_n(t), \quad (17)$$

where $\bar{p}(m | t)$ is the posterior predictive probability and $b_n^{(m,t)}$ is obtained by the moment-matching step of section V. For squared-error loss one may take $\mathcal{R}(b) = \text{Var}_b(\theta)$. A quantile width can also be used, but it requires the skew-normal moment projection and is not produced by (16) alone. When many candidates must be screened, $[\Sigma_n]_{\theta\theta}$ from (16) provides a closed-form variance proxy; shortlisted candidates can then be evaluated with the full quadrature-based ADF.

VII. CORE METHOD: DISTRIBUTIONALLY ROBUST DECOHERENCE DESIGN

A. Why distributional, not worst-case

Existing robust quantum metrology guards against a *set* of channels (worst-case / QFI-protection) or worst-case measurement noise. We instead place an *ambiguity set over the probability law* of the decoherence rate γ , capturing misspecification of the *shape* of that law, not merely its support.

Definition 3 (Wasserstein ambiguity set). For a nominal law $\hat{Q}$ on γ and radius $\varepsilon \geq 0$, $\mathcal{B}_\varepsilon(\hat{Q}) = \{Q \in \mathcal{P}(\Gamma) : \mathcal{W}_1(Q, \hat{Q}) \leq \varepsilon\}$, with $\mathcal{W}_1$ the order-1 Wasserstein distance under ground metric $|\gamma - \gamma'|$.

The robust sequential design solves, at each step,

$$\min_t \sup_{Q \in \mathcal{B}_\varepsilon(\hat{Q})} \mathbb{E}_{\gamma \sim Q} [L(t, \gamma)], \quad (18)$$

$L(t, \gamma) = \text{expected posterior risk at rate } \gamma.$

B. Strong-duality reduction

Theorem 4 (Order-1 Wasserstein DRO duality). *Let Γ be a compact interval, let $L : \Gamma \rightarrow \mathbb{R}$ be upper semicontinuous, and let $\hat{Q}$ be a probability law on Γ . Then*

$$\begin{aligned} & \sup_{Q \in \mathcal{P}(\Gamma) : \mathcal{W}_1(Q, \hat{Q}) \leq \varepsilon} \mathbb{E}_Q[L(\gamma)] \\ &= \inf_{\lambda \geq 0} \left\{ \lambda \varepsilon + \mathbb{E}_{\hat{\gamma} \sim \hat{Q}} \left[\sup_{\gamma \in \Gamma} (L(\gamma) - \lambda |\gamma - \hat{\gamma}|) \right] \right\}. \end{aligned} \quad (19)$$

The inner supremum is attained. The outer objective is convex in λ .

This is the order-1 specialization of optimal-transport DRO duality. The inner problem is a one-dimensional *global maximization*; it need not be concave and cannot in general be solved by golden-section search with a global guarantee.

Proposition 5 (Lipschitz certificate). *If L is Λ -Lipschitz on Γ , then*

$$\mathbb{E}_{\hat{Q}}[L] \leq \sup_{Q \in \mathcal{P}(\Gamma) : \mathcal{W}_1(Q, \hat{Q}) \leq \varepsilon} \mathbb{E}_Q[L] \leq \mathbb{E}_{\hat{Q}}[L] + \Lambda \varepsilon. \quad (20)$$

Moreover, for every $\lambda \geq \Lambda$ the inner transform equals $L(\hat{\gamma})$, so an outer minimizer may be sought on $[0, \Lambda]$ when $\varepsilon > 0$. Equality in the upper bound is not asserted without additional structure.

The primal and dual are compared numerically in section XII. This numerical check supports the implementation of (19); it does not turn the inner maximization into a convex problem.

C. Continuous- γ ambiguity sets: regularity, duality, tractability

We consider a continuous decoherence law on the compact interval $\Gamma = [\gamma_{\min}, \gamma_{\max}]$. The exact finite-sample risk is

$$L_N(t, \gamma) := \mathbb{E}_{\mathcal{D} \sim P_\gamma} [(\hat{\theta}_N(\mathcal{D}; t) - \theta)^2], \quad (21)$$

where the expectation includes all measurement outcomes and any prior draw used by the experiment. Two information surrogates must be distinguished. In the primary analytic schedule experiment, γ is an uncertain environment variable but is not estimated within the episode. With a scalar prior precision $J_0 \geq 0$, the surrogate is

$$\tilde{L}_{\text{env}}(t, \gamma) := \frac{1}{J_0 + \sum_{n=1}^N I_{\theta\theta}(\theta_0, \gamma; t_n)}. \quad (22)$$

If γ is instead jointly estimated as a nuisance parameter, the correct local surrogate is

$$\begin{aligned} \tilde{L}_{\text{joint}}(t, \gamma) &:= [\mathbf{K}(t, \gamma)^{-1}]_{\theta\theta}, \\ \mathbf{K}(t, \gamma) &= \mathbf{K}_0 + \sum_{n=1}^N \mathbf{I}(\theta_0, \gamma; t_n), \end{aligned} \quad (23)$$

which automatically accounts for the nuisance coupling through the matrix inverse (equivalently, a Schur complement). Both are prior-regularized local approximations, not exact finite- N posterior risks; selected schedules are therefore checked separately by Monte Carlo.

Lemma 6 (Regularity of the information surrogates). *Let $j(t, \gamma) = J_0 + \sum_n I_{\theta\theta}(\theta_0, \gamma; t_n)$. If $j(t, \gamma) \geq \kappa > 0$ on Γ , then $\tilde{L}_{\text{env}}$ is continuously differentiable and Lipschitz, with*

$$\partial_\gamma \tilde{L}_{\text{env}} = -\frac{\partial_\gamma j}{j^2}. \quad (24)$$

If, in addition, $\mathbf{K}(t, \gamma) \succeq \kappa \mathbb{I} \succ 0$, then $\tilde{L}_{\text{joint}}$ is continuously differentiable and Lipschitz, with

$$\begin{aligned} \partial_\gamma \tilde{L}_{\text{joint}} &= -[\mathbf{K}^{-1}(\partial_\gamma \mathbf{K})\mathbf{K}^{-1}]_{\theta\theta}, \\ \partial_\gamma \mathbf{K} &= \sum_n \partial_\gamma \mathbf{I}(\theta_0, \gamma; t_n). \end{aligned} \quad (25)$$

Proof: see section A.

Theorem 7 (Continuous-support specialization). *For any upper-semicontinuous, integrable risk L on the compact metric space $(\Gamma, |\cdot|)$,*

$$\begin{aligned} \sup_{Q \in \mathcal{B}_\varepsilon(\hat{Q})} \mathbb{E}_Q[L] &= \inf_{\lambda \geq 0} \left\{ \lambda \varepsilon + \mathbb{E}_{\hat{\gamma} \sim \hat{Q}} [L_\lambda(\hat{\gamma})] \right\}, \\ L_\lambda(\hat{\gamma}) &:= \max_{\gamma \in \Gamma} (L(\gamma) - \lambda |\gamma - \hat{\gamma}|). \end{aligned} \quad (26)$$

Corollary 8 (Computational structure). *For fixed λ and $\hat{\gamma}$, evaluating $L_\lambda(\hat{\gamma})$ is a scalar global optimization on a compact interval. The outer objective is convex in λ . If L is Λ -Lipschitz and $\varepsilon > 0$, it suffices to search $\lambda \in [0, \Lambda]$. A grid with certified refinement, interval branch-and-bound, or a problem-specific global solver is valid; a unimodal method such as golden-section search requires an independently verified unimodality condition.*

Proof: see section A.

Proposition 9 (Radius-dependent upper bound). *For a Λ -Lipschitz risk,*

$$0 \leq \sup_{Q \in \mathcal{B}_\varepsilon(\hat{Q})} \mathbb{E}_Q[L] - \mathbb{E}_{\hat{Q}}[L] \leq \Lambda\varepsilon. \quad (27)$$

A first-order equality with a particular “effective slope” requires additional directional regularity and attainability assumptions and is not used in this paper.

The reported continuous-grid experiment verifies primal–dual agreement for the surrogate (22) and estimates a numerical Lipschitz constant on the specified compact interval.

VIII. LEARNED EXPERIMENT-DESIGN AGENTS UNDER DECOHERENCE SHIFT

We cast adaptive quantum sensing as an *autonomous experiment-design agent*: a policy mapping the current belief to the next interrogation time t , trained to minimize estimation error, the quantum instance of (amortized / reinforcement-learning) Bayesian optimal experimental design. We parametrize the policy as a small neural network and train it with an evolutionary strategy (cross-entropy method), as is standard in learned quantum metrology.

a) The agent.: The policy π_w maps a two-feature belief state $s_n = (\sigma_{\theta,n}, n/N)$ (the current posterior width and the fraction of the shot budget used) to the next interrogation time $t_n = \pi_w(s_n) \in [t_{\min}, t_{\max}]$, and is a small two-layer perceptron (policy weights w): a 4-unit tanh hidden layer, 17 weights in total, with a sigmoid output rescaled to $[t_{\min}, t_{\max}] = [0.3, 11]$, trained gradient-free by the cross-entropy method (26 iterations, population 50, elite 12, exploration floor 0.03; full architecture, hyper-parameters and seeds in section G). The architecture is *intentionally minimal*: our contribution is the robust *objective*, not a new policy class, and a small gradient-free policy isolates the effect of the training distribution from network capacity (we claim no new reinforcement-learning algorithm). Rolled out for N shots with the skew-normal assumed-density estimator of section V, it returns a terminal estimate $\hat{\theta}_N$; the episode conditional mean-squared error at true rate γ is $M(w; \gamma) = \mathbb{E}[(\hat{\theta}_N - \theta)^2]$, and the reported RMSE is $R(w; \gamma) = \sqrt{M(w; \gamma)}$.

b) Nominal, support-minimax, and regret-minimax training.: The nominal agent minimizes $M(w; \hat{\gamma})$. Two robust scoring rules are used in separate experiments. The raw support-minimax objective is

$$\mathcal{J}_{\text{raw}}(w) := \max_{\gamma \in \Gamma_{\text{train}}} M(w; \gamma). \quad (28)$$

trained agent	RMSE @ $\hat{\gamma}$ (in-dist.)	RMSE @ $3\times$ shift	worst-case subopt. (wide shift)
nominal specialist	0.11 ± 0.00	0.147 ± 0.00	19% to 157% (rate-dependent)
robust (ours)	0.12 ± 0.01	0.147 ± 0.00	15% $\pm$ 3% (lowest)

Equivalently, $\mathcal{J}_{\text{raw}}$ is the worst mean-squared error over all probability laws supported on Γ_{train} ; it is not the same as a finite-radius Wasserstein ball. For cross-rate generalization we instead use relative regret

$$\mathcal{J}_{\text{reg}}(w) := \max_{\gamma \in \Gamma_{\text{train}}} \left\{ \frac{R(w; \gamma)}{R^*(\gamma)} - 1 \right\}, \quad (29)$$

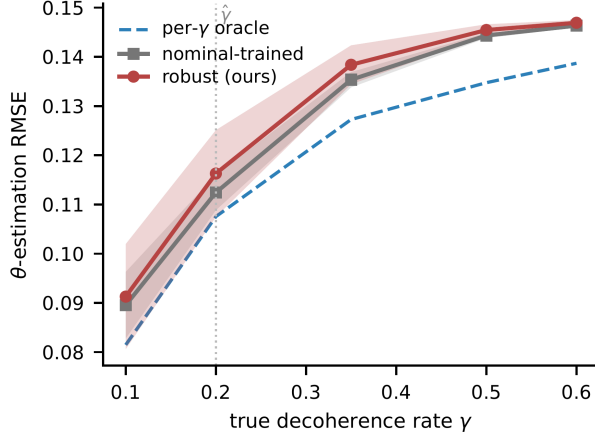
where $R^*(\gamma) > 0$ is the RMSE of the explicitly stated comparator class (in the reported experiment, the best constant-interrogation-time non-adaptive policy with γ known). Thus the reported percentage is suboptimality relative to that comparator, not relative to the globally optimal adaptive policy. Both objectives are optimized by the same cross-entropy method; only the candidate scoring rule changes.

c) Statistical protocol.: Unless stated otherwise, every reported estimation error is a Monte-Carlo average over independent rollouts, and every *learned-agent* quantity is reported as the mean ± 1 standard deviation over independent training seeds; error bands and bars in all agent figures are ± 1 s.d. Design-level Cramér–Rao surrogate quantities (section IX) are exact deterministic computations and carry no seed variance.

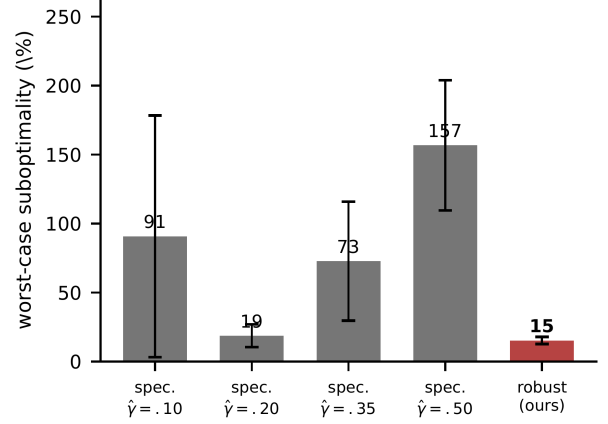
Figure 2 reports the agent-level phenomenon, with a nuance we state plainly. At a *single* operating rate an adaptive learned policy is *not* brittle: because it re-plans each interrogation from its current belief, a nominal-trained agent and the corresponding robust agent also tracks the per- γ constant-time reference to within one standard deviation across the shift, and are statistically indistinguishable at any single rate (panel a; mean ± 1 s.d. over training seeds). Adaptivity supplies its own partial robustness. The distribution-shift failure is instead a *specialist* effect: when a policy trained at one rate is *evaluated across a wide shift*, its worst-case suboptimality explodes, up to $157\% \pm 47\%$ for a high- $\hat{\gamma}$ specialist, whereas the *same* architecture trained with our relative-regret minimax objective keeps the lowest and most stable worst case of all ($15\% \pm 3\%$; panel b). This is the trustworthy-ML lesson, learned *specialists* overfit their training distribution, made concrete for autonomous quantum experiment design, and it motivates the principled robust *design* analyzed next (section IX).

Means ± 1 s.d. over independent training seeds. At any single rate the two agents are statistically identical, adaptivity already self-corrects; the robust objective’s payoff is its uniformly low *worst-case* across the shift, where rate-specialists range from 19% (incidentally near the test grid) to 157% (high- $\hat{\gamma}$). The *learned* robust agent rediscovers the short-interrogation hedging that our *analytic* robust design independently prescribes (section IX): the learned agent’s mean interrogation time is $t \approx 1.75$ (5–95% range 0.9–2.8), close to the analytic robust schedule ($T_{\text{rob}}=2.10$, mean 1.2) and far below the nominal long-interrogation design ($T_{\text{nom}}=8.36$, mean 4.3); the data-driven and principled routes agree. Full architecture, hyper-parameters, and seeds are in section G.

d) Generalization across the shift: specialists overfit, the robust agent does not.: To probe how the brittleness scales, we



(a) both near-oracle: adaptivity self-corrects



(b) specialists overfit; robust lowest worst-case

Fig. 2. **A learned experiment-design agent under decoherence shift (multi-seed).** (a) θ -estimation RMSE versus the true decoherence rate for a policy trained at the nominal rate ($\hat{\gamma}=0.2$) and the same architecture trained with our robust objective; bands are mean $\pm$ 1 s.d. over training seeds and the dashed line is the per- γ *constant-time* reference (the lowest RMSE achievable at that rate within the class of single-time non-adaptive policies with γ known). At any single rate the two agents are indistinguishable and close to that restricted reference, adaptivity self-corrects. (b) Worst-case suboptimality across a wide shift $\gamma \in [0.08, 0.6]$ (mean $\pm$ 1 s.d.): rate-specialists overfit (up to 157% for the high- $\hat{\gamma}$ specialist) while the robust agent keeps the lowest, most stable worst case (15%). `exp_agent_seeds.py`, `exp_matrix_seeds.py`.

train a family of nominal agents, each on a different assumed rate $\hat{\gamma} \in \{0.10, 0.20, 0.35, 0.50\}$, and pit them against a single robust agent trained by *minimax relative regret*: worst-case RMSE suboptimality relative to the best constant-time non-adaptive comparator at each rate, as defined in (29). Figure 3 is the resulting generalization matrix. Each nominal agent is close to the restricted comparator *only* in a neighborhood of its own training rate (the dark diagonal of panel a) and overfits elsewhere: a low- $\hat{\gamma}$ specialist learns long interrogations and degrades at high γ , while a high- $\hat{\gamma}$ specialist learns short interrogations and is $>150\%$ suboptimal at low γ (panel b). No single nominal policy is safe across the shift. The regret-trained robust agent instead holds a flat, bounded envelope and attains the *lowest worst-case suboptimality* over the entire ambiguity set ($15\% \pm 3\%$ versus 19–157% for the specialists, mean over seeds; fig. 2b), the textbook signature of generalization over memorization, here for autonomous quantum design.

e) The agents are trained, and the objective decides robustness.: Figure 4(a) shows the evolutionary learning curve: the robust objective drives the worst-case regret over the decoherence set down monotonically, whereas the nominal objective, which never sees the worst case, leaves it high and erratic. Panel (b) places every trained agent in the accuracy-robustness plane: all agents share essentially the same *average* RMSE over the shift (0.128–0.139), yet their *worst-case* suboptimality spans 15% to 157% (mean over seeds). Average accuracy is blind to the brittleness the robust objective removes; the robust agent buys a several-fold worst-case improvement for a negligible average premium.

IX. ROBUST DESIGN ANALYSIS UNDER DECOHERENCE MISMATCH

The learned agent of section VIII is mirrored by a *principled* robust design we can analyze in closed form. We test whether it pays off when the true rate γ_{true} differs from the nominal $\hat{\gamma}$. The design knob is the maximum interrogation time T (the schedule spans $[0.3, T]$, which is intended to reduce phase aliasing); the optimal T trades off against γ : long T gathers more phase information but is wasted once $V(T) = e^{-\gamma T}$ has decayed. We compare three designs: *nominal* (optimal at $\hat{\gamma}$), *minimax* (worst-case over the ambiguity set), and *DRO* (Wasserstein-ball expected risk via theorem 7).

Figure 5 reports the outcome, verified two independent ways. (a) On the design objective (CRB surrogate $1/J$), the nominal design is best exactly at $\hat{\gamma}$ but its risk increases sharply under mismatch, whereas the robust designs stay flat: the worst-case risk drops from 0.265 (nominal) to 0.139 (DRO), a $2\times$ reduction, and the worst-case regret versus a rate-aware schedule reference falls $6.7\times$. (b) Actual θ -RMSE confirms it: at a $3\times$ mismatch ($\gamma_{\text{true}} = 0.55$ vs. $\hat{\gamma} = 0.2$) the robust RMSE is $\approx 17\text{--}20\%$ below nominal. The premium is small: at favorable (low) γ_{true} the nominal design’s longer interrogation wins by a small margin; robustness buys worst-case protection at a modest average cost.

a) Anatomy of the robust design.: Figure 6 dissects *why* the robust schedule wins, using the deterministic inverse-information surrogate (no Monte-Carlo noise). (a) The risk landscape $1/J(T, \gamma)$ shows the information-optimal time $T^*(\gamma)$ *shrinking* as γ grows; the nominal design fixes T at $T^*(\hat{\gamma})$ and sits on a cliff that falls away under mismatch, while the robust designs sit on the flatter short- T ridge. (b) Sweeping the Wasserstein radius ε traces a Pareto frontier in the (nominal-risk, worst-case-risk) plane: $\varepsilon=0$ recovers the nominal design

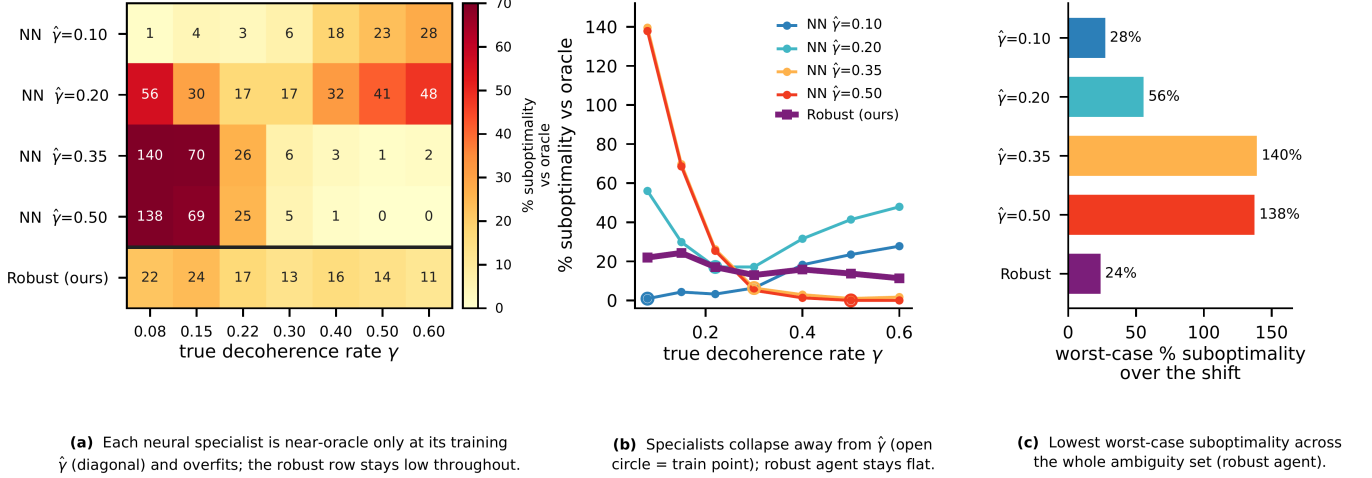


Fig. 3. **Neural design agents overfit their training decoherence; robust training generalizes.** (a) Generalization matrix: percentage RMSE suboptimality relative to the per- γ constant-time reference, for agents trained at four nominal rates $\hat{\gamma}$ (top rows) and the robust agent (bottom row), evaluated across true rates γ . Each specialist is close to the restricted reference only along its own training-rate diagonal. (b) The same data as curves: specialists collapse away from their training rate (open circle marks the training point); the robust agent keeps a flat, low envelope. (c) Worst-case suboptimality over the shift for this representative seed; the seed-averaged worst case (15% robust vs. up to 157% specialist) is in fig. 2b. Cross-entropy-method training, a representative seed; `verify_ai_agent_matrix.py`.

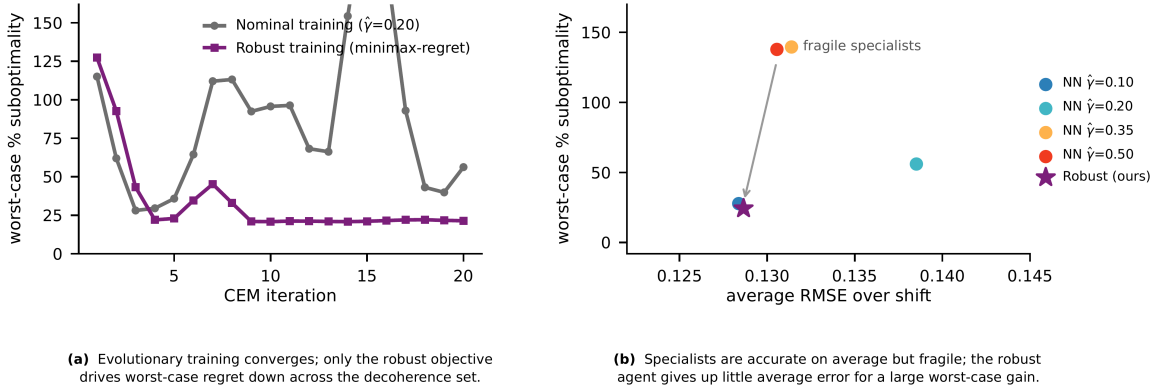


Fig. 4. **Training the design agents.** (a) Cross-entropy-method learning curve: worst-case percentage suboptimality over the decoherence ambiguity set versus iteration; only the minimax-regret objective drives it down. (b) Accuracy-robustness tradeoff: average RMSE over the shift (x) versus worst-case suboptimality (y) for every trained agent: specialists are accurate on average but fragile, the robust agent (star) is balanced. `figures/gen_fig_agent_training.py`.

and any $\varepsilon \geq \gamma_{\max} - \gamma_{\min}$ recovers support minimax, so ε is an explicit *conservatism knob*. (c) The absolute risk reduction of DRO over nominal is slightly negative near $\hat{\gamma}$ (a small premium) and grows large for $\gamma_{\text{true}} > \hat{\gamma}$.

b) Convergence under mismatch.: Figure 7 tracks the actual RMSE as the shot budget N grows. When the model is correct (a) the three designs are comparable. Under a $3\times$ mismatch (b) the robust designs reach a markedly lower RMSE at every N (0.226 vs. 0.313 for nominal at $N=80$), confirming the advantage is not a single- N artifact.

A. Fuller baselines, sensitivity, and model-class robustness

We broaden the comparison along three axes.

a) Baselines beyond nominal vs. robust.: As fig. 8a and table V show, at the design level (deterministic CRB surrogate) we compare the nominal and regret-minimax designs against

six robust baselines spanning the DRO and metrology literatures: *BMA* (Bayesian model averaging, minimizing the prior-averaged risk $\mathbb{E}_{\gamma \sim \pi}[1/J]$), *CVaR_{0.8}* (conditional value-at-risk: the mean of the worst-20% tail over π), *KL-DRO* (an entropic / KL-divergence ambiguity ball, $\min_{\beta > 0} \beta \log \mathbb{E}_{\pi} e^{L/\beta} + \beta \rho$), *minimax* (worst case over the support of Γ), *adversarial* (continuous inner maximization of γ , the design-space analogue of a PGD attack), and *robust-QFI* (a metrology baseline maximizing the worst-case quantum Fisher information, $\max_t \min_{\gamma} t^2 e^{-2\gamma t}$). The nominal design degrades sharply under shift (worst-case inverse-information RMSE surrogate 0.824), whereas every principled robust strategy stays bounded near the per- γ rate-aware reference (0.471); our regret-minimax design *attains* the rate-aware-reference worst case (0.471), tying CVaR, KL-DRO and minimax at a smaller in-distribution premium than the QFI baseline, while robust-QFI is slightly looser (0.493) because it is myopic. *Statistical significance:* at

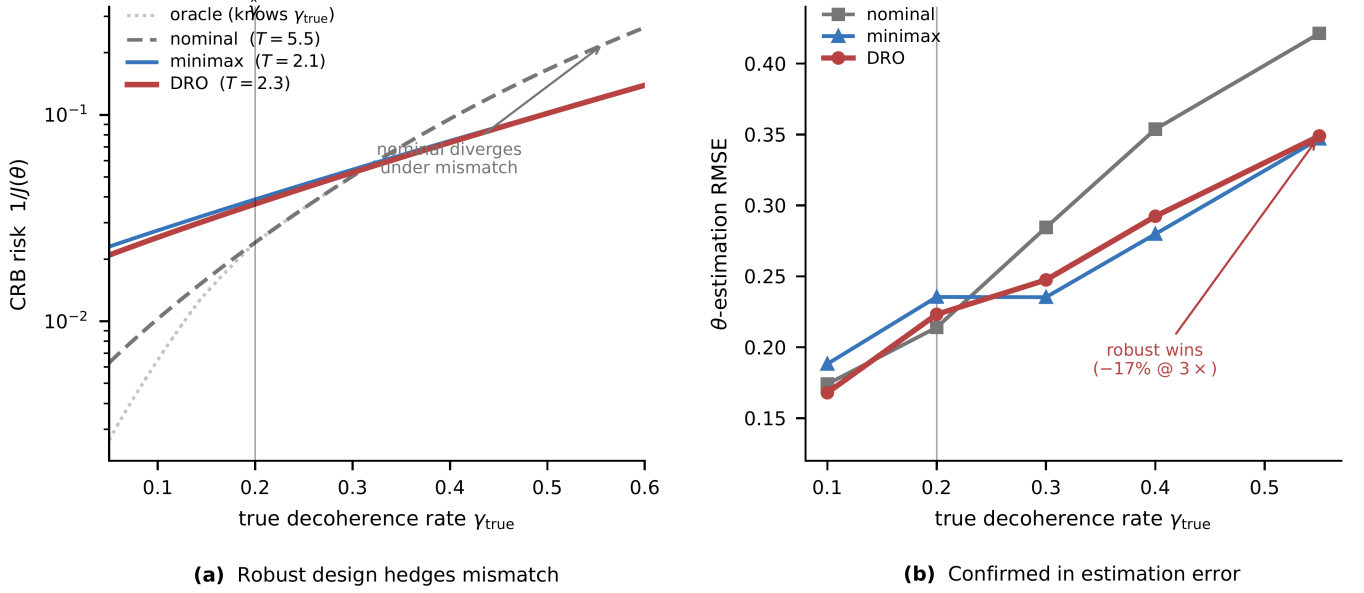


Fig. 5. **Distributionally robust design hedges decoherence mismatch.** (a) Inverse-information surrogate $1/J(\theta)$ versus the true rate γ_{true} for the nominal, minimax, and DRO designs (rate-aware reference = design optimized with γ_{true} known). The nominal design is optimal at $\hat{\gamma}$ but degrades sharply under mismatch; the robust designs trade a small premium at $\hat{\gamma}$ for a $2\times$ lower worst-case risk. (b) Actual θ -RMSE confirms the crossover: robustness wins by ≈ 17 –20% at a $3\times$ mismatch.

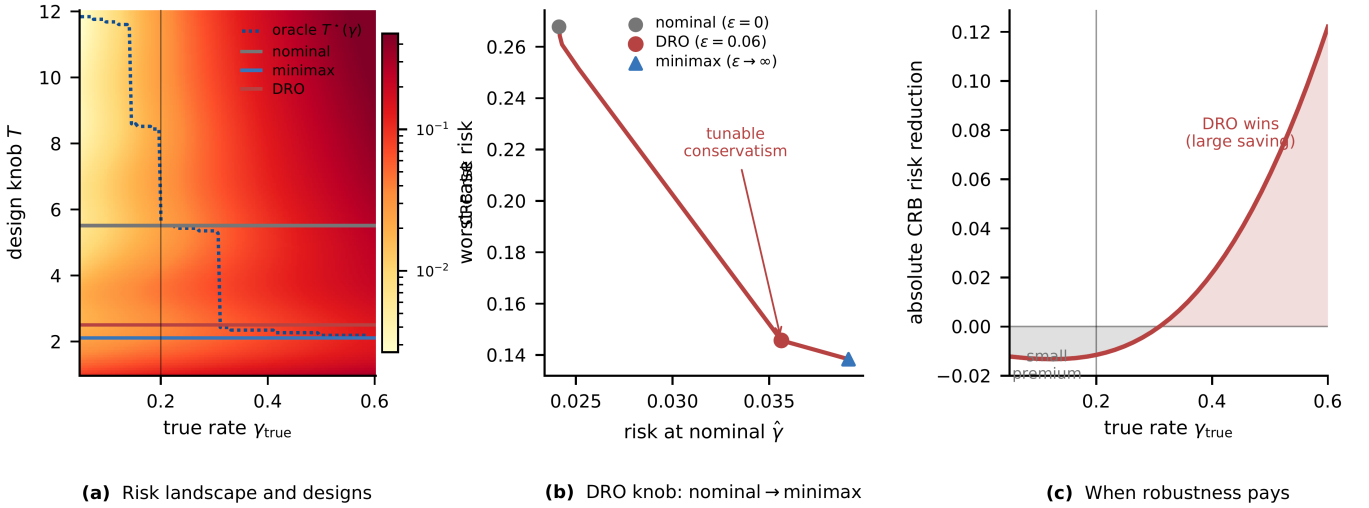


Fig. 6. **Anatomy of the robust design** (deterministic CRB). (a) Risk landscape $1/J(T, \gamma_{\text{true}})$ with the three designs (horizontal lines) and the rate-aware reference ridge $T^*(\gamma)$. (b) The DRO radius ϵ interpolates nominal $\rightarrow$ minimax along a Pareto frontier. (c) Absolute surrogate-risk reduction of DRO over nominal: a small premium near $\hat{\gamma}$, a large saving under mismatch.

a $3\times$ mismatch ($\gamma_{\text{true}}=0.60$, $N=40$) the robust design lowers the actual θ -estimation RMSE by 26% with a 95% bootstrap CI of $[21, 31]\%$; the headline 17–20% is the conservative end of the sweep (smaller mismatch / larger N).

b) The radius is a principled knob, with a concrete selection rule.: Sweeping the Wasserstein radius ϵ (fig. 8b) shows the DRO design *interpolates* between nominal ($\epsilon=0$) and minimax (large ϵ): worst-case risk falls and in-distribution risk rises. We give two practical, retraining-free rules for fixing ϵ . (i) *Data-driven plug-in*: collect multiple rate estimates in consecutive calibration windows and set ϵ from the empirical

1-Wasserstein distance between the two empirical rate distributions. More precisely, let the support have diameter $D = \gamma_{\text{max}} - \gamma_{\text{min}}$, and let $\hat{Q}_t, \hat{Q}_{t-1}$ contain n_t, n_{t-1} independent rate estimates from laws Q_t, Q_{t-1} . By the one-dimensional identity $\mathcal{W}_1(P, Q) = \int |F_P - F_Q|$ and the Dvoretzky–Kiefer–Wolfowitz inequality, with probability at least $1 - \delta$,

$$\mathcal{W}_1(Q_t, Q_{t-1}) \leq \mathcal{W}_1(\hat{Q}_t, \hat{Q}_{t-1}) + D \left[\sqrt{\frac{\log(4/\delta)}{2n_t}} + \sqrt{\frac{\log(4/\delta)}{2n_{t-1}}} \right]. \quad (30)$$

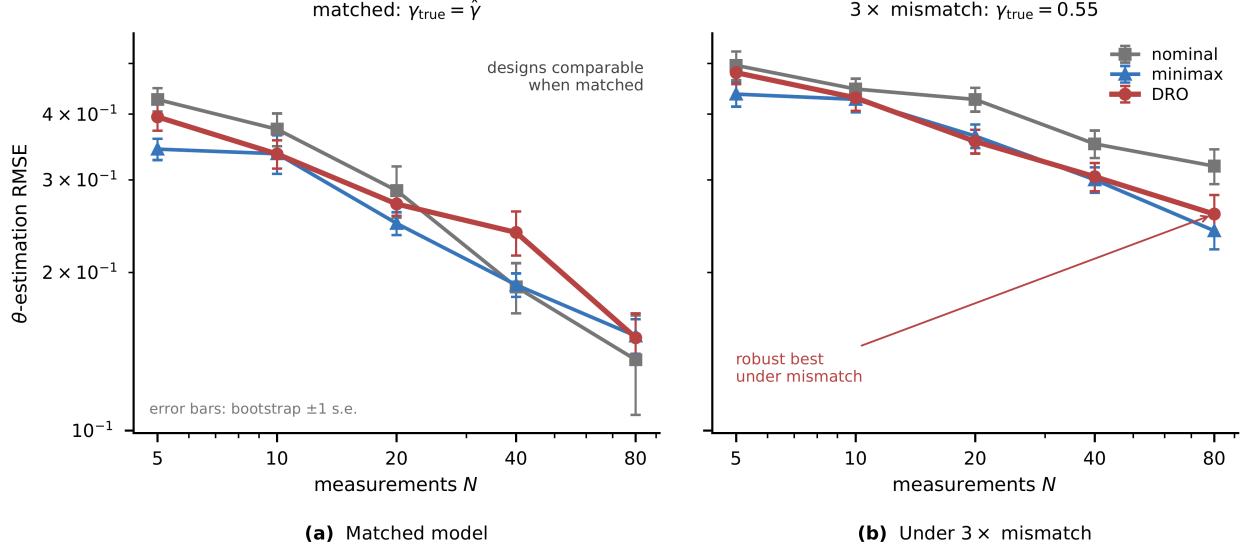


Fig. 7. **Estimation error versus number of measurements.** (a) Matched model ($\gamma_{\text{true}} = \hat{\gamma}$): the designs are comparable. (b) Under a $3\times$ mismatch ($\gamma_{\text{true}} = 0.55$): the robust designs reach a lower RMSE across N (0.226 vs. 0.313 at $N=80$). Posterior mean.

The right-hand side is therefore a conservative shift radius under the stated independence model. With only one point estimate per window, $|\hat{\gamma}_t - \hat{\gamma}_{t-1}|$ is merely a heuristic shift estimate and has no DKW guarantee. (ii) *Elbow rule*: take the knee of the worst-case-risk-versus- ε curve (here $\varepsilon \approx 0.08$), the point of diminishing protection beyond which worst-case risk flattens while the in-distribution premium keeps growing. The same knee shows up in the *actual* (non-CRB) estimation error: sweeping ε , the robust design's worst-case θ -RMSE drops from 0.453 ($\varepsilon < 0.08$, long design) to 0.390 once ε crosses ≈ 0.08 , with no in-distribution penalty ($0.347 \rightarrow 0.260$), confirming the radius rule on RMSE and not merely the inverse-information proxy. Neither rule requires retraining of the policy; computing an empirical one-dimensional Wasserstein distance costs sorting time in the number of window estimates. The two rules agree within a factor of two in the reported setup. Finally, when $\varepsilon \geq D$, the ambiguity ball contains every probability law supported on Γ , so the Wasserstein objective equals support minimax exactly.

c) *Robustness survives a model-class mismatch.*: A stronger test (fig. 8c) breaks the assumed model entirely: designs are chosen under the Markovian law $V = e^{-\gamma t}$, but the *true* channel is non-Markovian (quasi-static $1/f$ -like) dephasing $V = e^{-(\gamma t)^2}$. Even under this unmodelled mismatch the robust design cuts the worst-case inverse-information error surrogate by 52% ($0.886 \rightarrow 0.428$, against a rate-aware reference of 0.409) at no in-distribution cost: robustness to the *rate* transfers to robustness to the *functional form*. The Wasserstein duality itself is agnostic to the dimension of the sensing parameter, while the conditional Gaussian BvM statement is formulated for fixed finite d under explicit LAN and contraction assumptions. The present numerical evidence, however, is limited to a single-qubit model and one two-parameter grid experiment.

d) *Ablating robustness vs. skew-aware inference (the 2×2).*: To disentangle the two ingredients we cross the *design* (nominal long- T vs. robust short- T) with the *belief* (plain Gaussian ADF vs. exact skew-aware posterior); table II. The worst-case θ -RMSE moves almost entirely with the *design* row (robust cuts it 13%, $0.459 \rightarrow 0.399$) and is flat across the *belief* column (-0.1%). Robustness comes from the design objective, not the belief model; the skew-normal filter is a modest belief-quality refinement (median terminal $\text{KL}(\text{exact} \parallel \text{Gaussian}) = 0.017$), consistent with the scope of section XI.

e) *Is the learned policy worth it? Non-learned adaptive baselines.*: To isolate the value of the *learned* policy we compare it against four non-learned adaptive designs under the same Gaussian belief and shift (all design-blind to the true rate, using $\hat{\gamma}$): a random schedule, the particle-guess heuristic (PGH, $t_n = 1/\sigma_n$), Thompson sampling, and a myopic Bayes greedy information-maximizer; table III. The learned robust agent is best on both worst-case and average (0.368 vs. 0.44–0.50); tellingly, the myopic Thompson and Bayes-greedy designs are *worse than random*, because they exploit the nominal rate and thus overfit it, exactly the brittleness the robust objective removes. Even the standard robust-Bayesian alternative, a *fully Bayesian adaptive* design that marginalizes γ over a hierarchical prior, only reaches 0.477 worst-case (flat but mediocre: averaging the design over a broad γ -prior blurs it), which our regret-trained agent beats by 23%, so the regret objective is a genuine improvement in this benchmark over textbook nuisance-marginalization rather than one of many interchangeable robust strategies. Error-correction-based robustness [6] is *complementary and orthogonal*: it spends ancilla qubits to protect coherence, whereas we optimize the measurement *design* on a fixed single-qubit probe, and the two could be combined.

f) *Computational considerations.*: The ambiguity variable is scalar, so each inner transport transform in (26) is a one-

TABLE II

FOUR-VARIANT ABLATION: WORST-CASE θ -RMSE ($\gamma \in [0.08, 0.6]$, $N=20$). ROBUSTNESS (ROWS) IS THE LEVER; SKEW-AWARE INFERENCE (COLUMNS) IS $\approx$ NEUTRAL.

design \ belief	Gaussian	skew/exact
nominal (long T)	0.459	0.459
robust (ours)	0.399	0.399

TABLE III

LEARNED AGENT VS. NON-LEARNED ADAPTIVE BASELINES: θ -RMSE OVER THE SHIFT. THE LEARNED POLICY IS BEST ON BOTH.

adaptive policy	worst-case	avg
random schedule	0.464	0.412
PGH ($t=1/\sigma$)	0.443	0.372
Thompson sampling	0.495	0.491
Bayes greedy	0.494	0.489
Bayes-adaptive (marg. γ)	0.477	0.473
learned robust (ours)	0.368	0.305

dimensional global optimization, and the outer multiplier problem is one-dimensional and convex. This does not make the total cost independent of the parameter dimension: each evaluation of the risk may require a d -dimensional filter or information matrix. For the rank-one expected-information proxy, a covariance update costs $O(d^2)$, so evaluating G rate points over N shots costs $O(GNd^2)$ once the design schedule is fixed. The univariate skew-normal quadrature used by the learned agent does not automatically extend with the same complexity to a multivariate skew-normal filter.

g) *A higher-dimensional check: joint (θ, γ) estimation.*

To exercise the $d=2$ model empirically, beyond the scalar- θ design experiments, we estimate *both* parameters jointly from the same shots, tracking a full two-dimensional posterior over (θ, γ) on a 58×46 grid; table IV. The robust short- T design's θ -robustness *transfers* to the joint setting (worst-case θ -RMSE $0.424 \rightarrow 0.328$, -23%) at *no cost* to γ estimation (worst-case γ -RMSE 0.232 vs. 0.236, statistically identical): the benefit is therefore not an artifact of evaluating only the scalar- θ estimator. This numerical grid experiment is finite-sample evidence and is not presented as a proof of an asymptotic theorem.

Agent-level baselines. The same comparison at the learned-agent level (identical MLP policy and CEM, only the training-scoring distribution changed) is consistent under the regret metric of fig. 3: The raw-risk minimax, adversarial and BMA agents over-protect the hard high- γ regime and land far from the per- γ constant-time reference at low γ (78–129% suboptimal), even though minimax minimizes worst-case *raw* risk (fig. 8b); the regret-normalized DRO objective is the one that gives uniformly low regret relative to that reference (fig. 9).

TABLE IV

JOINT 2D (θ, γ) ESTIMATION (FULL 2D POSTERIOR, $N=20$, WORST/MEAN OVER $\gamma \in [0.10, 0.50]$). THE ROBUST DESIGN IMPROVES θ AT NO γ COST.

design	θ -RMSE		γ -RMSE	
	worst	mean	worst	mean
nominal (long T)	0.424	0.352	0.232	0.153
robust (short T , ours)	0.328	0.263	0.236	0.153

TABLE V

DESIGN-LEVEL BASELINES (DETERMINISTIC CRB SURROGATE): WORST-CASE AND IN-DISTRIBUTION θ -RMSE OVER $\gamma \in [0.06, 0.62]$. LOWER IS BETTER; THE PER- γ RATE-AWARE-REFERENCE WORST CASE IS 0.471. OUR REGRET-MINIMAX TIES THE BEST ROBUST BASELINES AT THE RATE-AWARE-REFERENCE WORST CASE WHILE KEEPING A SMALL IN-DISTRIBUTION PREMIUM.

design strategy	worst-case RMSE	in-dist. RMSE
nominal	0.824	0.189
BMA (prior-averaged)	0.475	0.234
CVaR _{0.8}	0.471	0.240
KL-DRO (entropic)	0.472	0.237
minimax	0.471	0.240
adversarial (PGD-on- γ)	0.471	0.240
robust-QFI (metrology)	0.493	0.279
regret-minimax (ours)	0.471	0.242
rate-aware reference (per- γ)	0.471	—

true channel $V(t)$	nominal	robust	change
Gaussian $e^{-(\gamma t)^2}$	0.886	0.428	−52%
stretched $e^{-(\gamma t)^{1.5}}$ ($1/f$ -like)	0.907	0.450	−50%
Lorentzian $1/(1 + (\gamma t)^2)$	0.528	0.340	−36%
algebraic $1/(1 + \gamma t)$ (slow)	0.270	0.354	+31%

More non-Markovian channels. Beyond the Gaussian law we evaluate the Markovian-chosen designs against several non-Markovian dephasing families (realized worst-case CRB RMSE): The robust design cuts the worst-case error by 36–52% across the common non-Markovian families. The lone exception is instructive: the slow algebraic channel decoheres *slower* than the Markovian assumption, so long interrogation is favourable and hedging costs 31%. Robustness protects against *faster-or-unknown* decoherence; a much slower truth is better served by a smaller ε . Script `exp_nonmarkov.py`.

X. ASYMPTOTIC CALIBRATION AND AN EXACT SCORE-SKEWNESS DIAGNOSTIC

This section states only the asymptotic conclusion justified by standard regular likelihood theory and separates it from the finite-sample skew-normal approximation. In particular, we do not claim that the Azzalini skew-normal is a universal higher-order posterior limit.

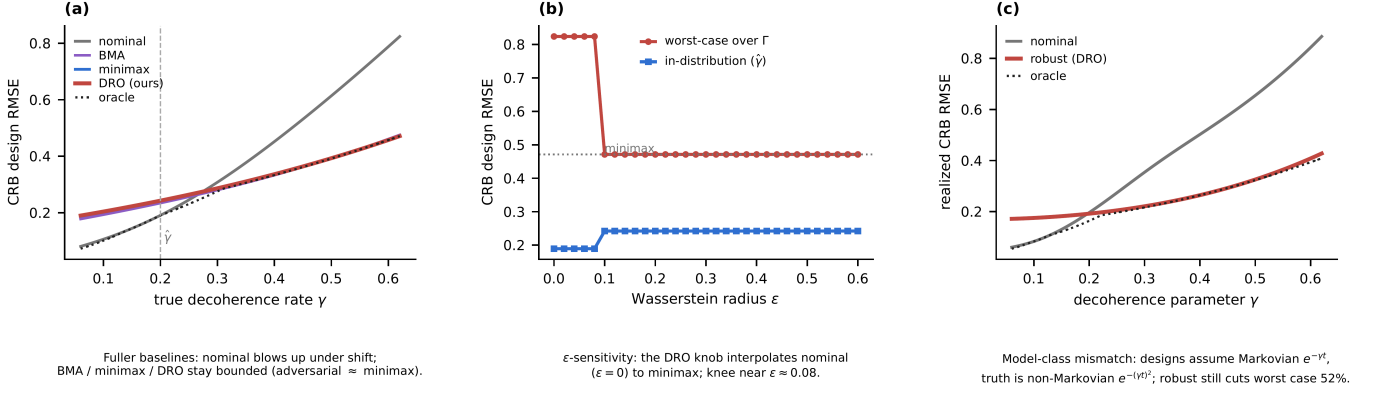
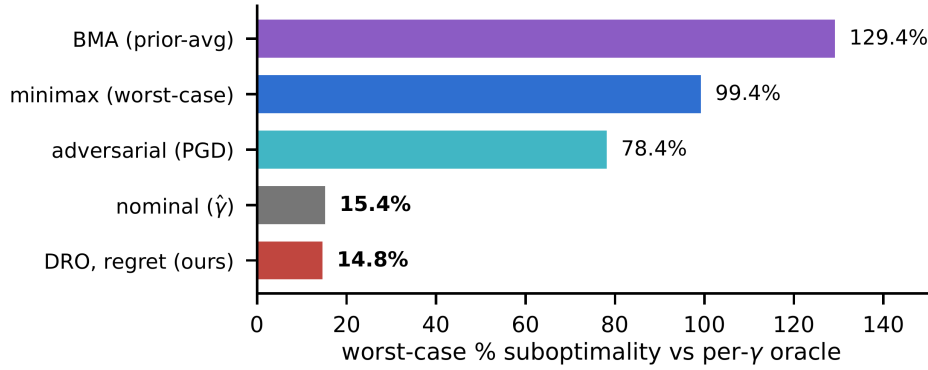


Fig. 8. **Fuller baselines, sensitivity, and model-class robustness** (design level, deterministic CRB surrogate). (a) Nominal vs. BMA, minimax and DRO across the true rate γ ; nominal degrades sharply, the robust family stays near the per- γ rate-aware reference (adversarial $\approx$ minimax, omitted). (b) The DRO knob: worst-case and in-distribution CRB RMSE versus the Wasserstein radius ϵ , interpolating nominal ($\epsilon=0$) to minimax. (c) Model-class mismatch: designs assume Markovian $e^{-\gamma t}$, the truth is non-Markovian $e^{-(\gamma t)^2}$; the robust design still cuts the worst case by 52%. Scripts `figures/gen_fig_baselines.py`, `exp_ablation.py`.



Same MLP policy + CEM, only the training objective changes. Naive raw-risk robust agents over-protect; the regret-DRO agent stays near the oracle.

Fig. 9. **Agent-level baselines.** Worst-case suboptimality (vs. the per- γ constant-time reference) for five training objectives that share the same MLP policy and CEM; only the training-scoring distribution changes. The regret-minimax agent is lowest, while the raw-risk minimax, adversarial and BMA agents over-protect the hard high- γ regime. Script `exp_agent_baselines.py`.

A. Adaptive sampling model

Let $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ and let t_n be $\mathcal{F}_{n-1}$ -measurable. At the true parameter η_0 , define the score martingale increment and predictable information

$$\begin{aligned} \mathbf{s}_n &= \nabla_{\eta} \log p(m_n | \eta_0, t_n), \\ \mathbf{i}_n &= \mathbb{E}[\mathbf{s}_n \mathbf{s}_n^{\top} | \mathcal{F}_{n-1}], \quad \mathbf{K}_{2,N} = \sum_{n=1}^N \mathbf{i}_n. \end{aligned} \quad (31)$$

The single-shot matrix is rank one, so a full-rank joint limit requires design diversity.

B. Assumptions

Assumption 1 (Local regularity). The true parameter is an interior point; the log likelihood has uniformly bounded derivatives through third order on a neighborhood of η_0 and on the compact design set; the prior density is continuous and strictly positive at η_0 .

Assumption 2 (Persistent excitation). $N^{-1} \mathbf{K}_{2,N} \rightarrow \bar{\mathbf{I}} \succ 0$ in probability.

Assumption 3 (Adaptive LAN and contraction). The adaptive likelihood is locally asymptotically normal uniformly on compact local parameter sets, and the posterior contracts at rate $N^{-1/2}$ around an efficient center $\hat{\eta}_N$.

The last condition is deliberately explicit: smoothness and a martingale central limit theorem alone do not establish posterior consistency or contraction.

Lemma 10 (Explicit score bound). If $t \in [t_{\min}, t_{\max}]$ and $\gamma \geq \gamma_{\min} > 0$, then

$$\|\mathbf{s}_n\| \leq C := \sup_{t \in [t_{\min}, t_{\max}]} \frac{te^{-\gamma_{\min} t}}{1 - e^{-\gamma_{\min} t}} < \infty. \quad (32)$$

Consequently the conditional Lindeberg condition for the score martingale follows.

Theorem 11 (Conditional adaptive Gaussian Bernstein–von Mises). *Under Assumptions 1 to 3, with $\mathbf{h} = \mathbf{K}_{2,N}^{1/2}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_N)$,*

$$\|\Pi(\mathbf{h} \mid \mathcal{D}_N) - \mathcal{N}(\mathbf{0}, \mathbb{I}_d)\|_{\text{TV}} \longrightarrow 0 \quad \text{in probability.} \quad (33)$$

No universal $O(N^{-1})$ total-variation rate is claimed. A third-order expansion, when valid, contains a general cubic polynomial/tensor correction. Such a correction may be represented by a suitable skew-symmetric density, but it does not generally reduce to a standard skew-normal with a single linear shape vector.

C. Exact single-shot score-skewness identity

For scalar θ , write $a = V \cos(\theta t)$ and $b = V t \sin(\theta t)$. Direct summation over $m \in \{-1, +1\}$ gives

$$I_\theta(t) = \frac{b^2}{1 - a^2}, \quad \mathbb{E}[s_\theta^3] = \frac{2ab^3}{(1 - a^2)^2}. \quad (34)$$

Hence, when $b \neq 0$,

Proposition 12 (Quadrature-point score symmetry).

$$\rho_s(t) := \frac{\mathbb{E}[s_\theta^3]}{I_\theta(t)^{3/2}} = \frac{2a \operatorname{sgn}(b)}{\sqrt{1 - a^2}}. \quad (35)$$

Thus $\rho_s(t) = 0$ whenever $\cos(\theta t) = 0$. This is an exact symmetry statement about the single-shot score; it is not a theorem that every Fisher-optimal interrogation time satisfies that condition. Indeed, at $\cos(\theta t) = 0$,

$$\frac{d}{dt} \log I_\theta(t) = \frac{2}{t} - 2\gamma, \quad (36)$$

so such a point is stationary in t only when $t = 1/\gamma$ in addition to the phase condition.

Corollary 13 (No regular persistent marginal skewness). *If the full joint posterior satisfies (33), then every fixed standardized marginal, including the γ marginal, converges to a standard Gaussian and its standardized skewness tends to zero whenever the relevant moments are uniformly integrable. Therefore a numerically observed $O(1)$ marginal skewness over a finite range of N is a finite-sample effect or a sign of weak identification, boundary influence, or failure to have reached the regular asymptotic regime; it is not a consequence of the regular BvM theorem.*

D. Numerical diagnostics

The reported simulations check three separate facts: the score is bounded on the chosen compact design range; the whitened score is approximately Gaussian under the tested adaptive schedule; and the single-shot score skewness in (35) vanishes at quadrature points. These diagnostics are consistent with the ingredients of an adaptive LAN analysis but do not independently prove posterior contraction or a total-variation rate. The finite- N γ -marginal skewness observed in the joint experiment is therefore reported descriptively rather than as an asymptotic constant.

Part	What is checked	Result
1	score (3) vs finite diff; Fisher (4) vs $\mathbb{E}_m[s^2]$; $V \rightarrow 1$ limit	PASS ($< 10^{-10}$)
2	few-shot ($N=3$) posterior skewness	$\gamma_1 = -0.46$ (skewed)
3	moment-matched SN vs Gaussian, KL, by skew bucket	SN wins 97%; ratio $1.2 \rightarrow 1.7\times$
3b	local skew-modal (11) vs Gaussian	46.6% (negative result)
4	sequential ADF tracks batch posterior	mean KL = 0.012; beats Gauss 74%
5	Sherman–Morrison inverse of proxy update (16) vs direct inverse	$\max \Delta = 1.1 \times 10^{-16}$
6	Wasserstein DRO duality (19): primal vs dual	match $< 10^{-3}$
7	continuous- γ DRO duality for surrogate (22)	gap below 5×10^{-8} , numerical $\Lambda=0.59$
8	bounded score and empirical martingale-normality diagnostics	$\sup \ \mathbf{s}_n\ = 5.23$
9	exact single-shot score skewness (35)	zero at quadrature points
10	$d=2$ whitened-score simulation diagnostic	approximately Gaussian
11	learned-agent generalization: specialists vs. robust	worst-case 15% vs. 19–157%
12	CEM training: worst-case regret vs. avg-RMSE tie	regret $146 \rightarrow 18\%$; avg $0.128\text{--}0.139$
13	fuller baselines (design+agent) + ε -sensitivity	DRO has the lowest reported regret; naive minimax/BMA over-protect
14	4 non-Markovian channels; Gaussian vs. skew ADF	robust -36 to -52% (3/4); skew gain $< 1\%$

XI. FINITE-SAMPLE DIRECTIONALITY OF POSTERIOR SKEWNESS

The joint grid experiment shows that, over the tested range of sample sizes, the γ marginal is more asymmetric than the θ marginal. This is plausible because $\gamma \geq 0$ is a rate parameter and can be weakly identified by schedules chosen primarily for θ . The observation is useful for choosing a finite-sample approximation, but it does not establish an irreducible asymptotic skewness floor. Under the regular joint BvM conditions of theorem 11, both standardized marginals must ultimately Gaussianize.

The reported skew-normal fit can have a much smaller KL divergence than a Gaussian fit to the γ marginal at finite N . A growing ratio of KL errors does not imply that the Gaussian error is bounded away from zero; both absolute errors must be inspected. Consistent with the ablation, this density-level improvement produces little change in terminal point-estimation RMSE in the present setup.

XII. NUMERICAL VERIFICATION LEDGER

Every numbered result above is checked against an independent ground truth (grid, Monte-Carlo, finite differences, or LP). Script: `validate_math.py` + companions, seed 20260626.

XIII. SCOPE AND LIMITATIONS

- **Episode-level uncertainty.** The rate γ is constant within each simulated episode and changes across episodes. A genuinely time-varying process γ_n would require a state-space model and is not addressed here.
- **Multimodality.** A single skew-normal cannot represent periodic, multimodal posteriors. The approximation is used

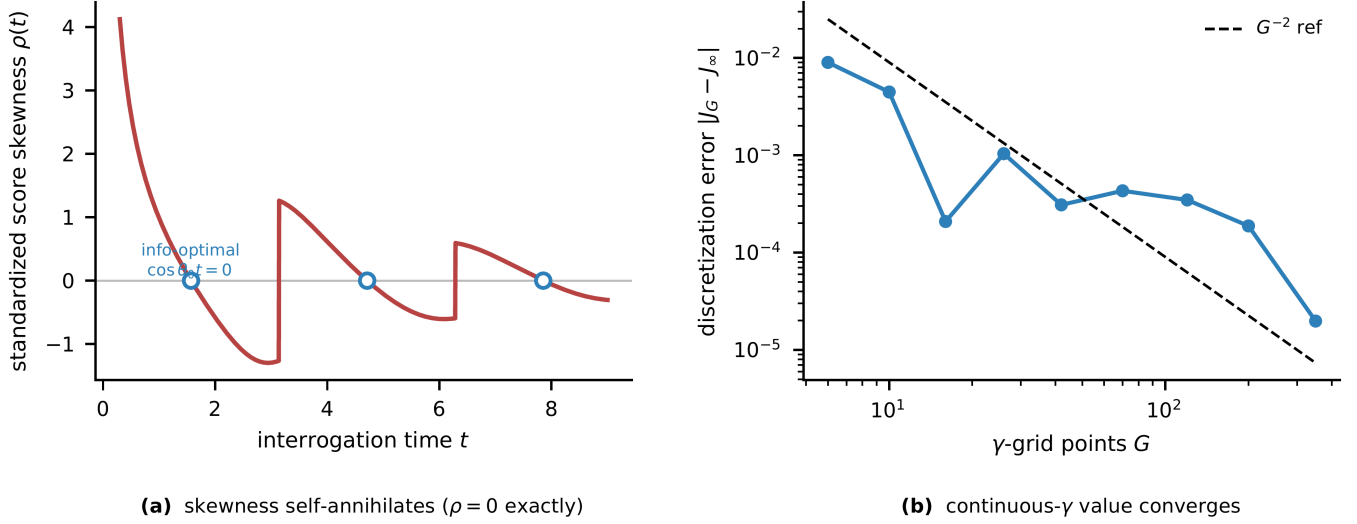


Fig. 10. **Mathematical diagnostics.** (a) The exact standardized single-shot score skewness (35) vanishes at phase-quadrature points $\cos(\theta_0 t) = 0$. These points are not asserted to be the global maximizers of $I_\theta(t)$; (36) gives the additional stationarity condition. (b) For the reported continuous- γ example, discretized primal and dual Wasserstein programs agree and converge under grid refinement. Script `figures/gen_fig_bvm.py`.

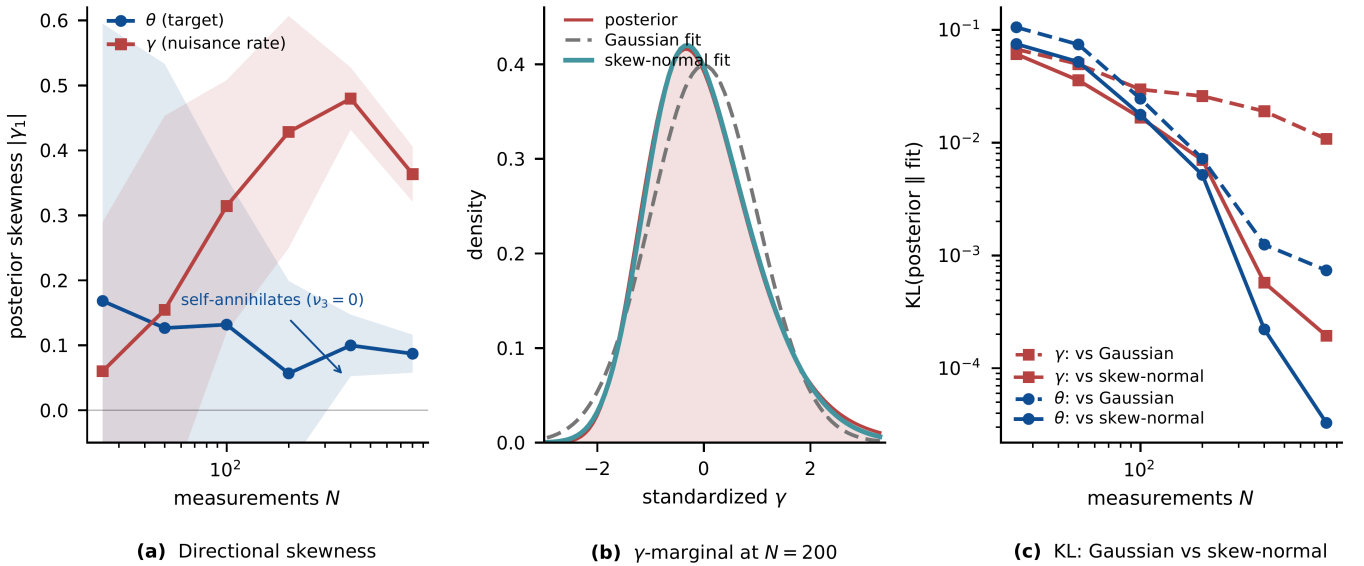


Fig. 11. **Finite-sample marginal asymmetry.** (a) Measured marginal skewness under the tested adaptive schedule; the γ marginal remains more asymmetric than the θ marginal over the displayed finite range of N . (b) At $N = 200$, a skew-normal fit better captures the displayed γ -marginal tail. (c) KL errors of Gaussian and skew-normal fits. These curves are finite-sample diagnostics and are not interpreted as proof of an asymptotic KL floor or persistent $O(1)$ skewness.

only after the prior and schedule localize the posterior to one dominant fringe.

- **Skewness saturation.** The standard skew-normal satisfies $|\gamma_1| < 0.9953$. Larger empirical skewness requires a richer skew-symmetric or mixture family.
- **Univariate ADF.** The implemented skew-normal ADF is univariate in θ for a specified rate model. The joint (θ, γ) experiment uses a two-dimensional grid. The $O(d^2)$ statement applies only to the expected-Fisher covariance proxy, not to a general multivariate skew-normal moment projection.

- **DRO computation.** The Wasserstein outer problem is convex in λ , but the scalar inner maximization need not be concave or unimodal. Global optimization or certified grid refinement is required unless additional structure is proved.
- **Comparator limitation.** Learned-policy suboptimality is measured relative to the best constant-time non-adaptive comparator used in the experiment, not the unknown globally optimal adaptive policy.
- **Asymptotics.** The Gaussian BvM statement is conditional on adaptive LAN and posterior contraction. The paper does not derive these properties solely from smoothness and

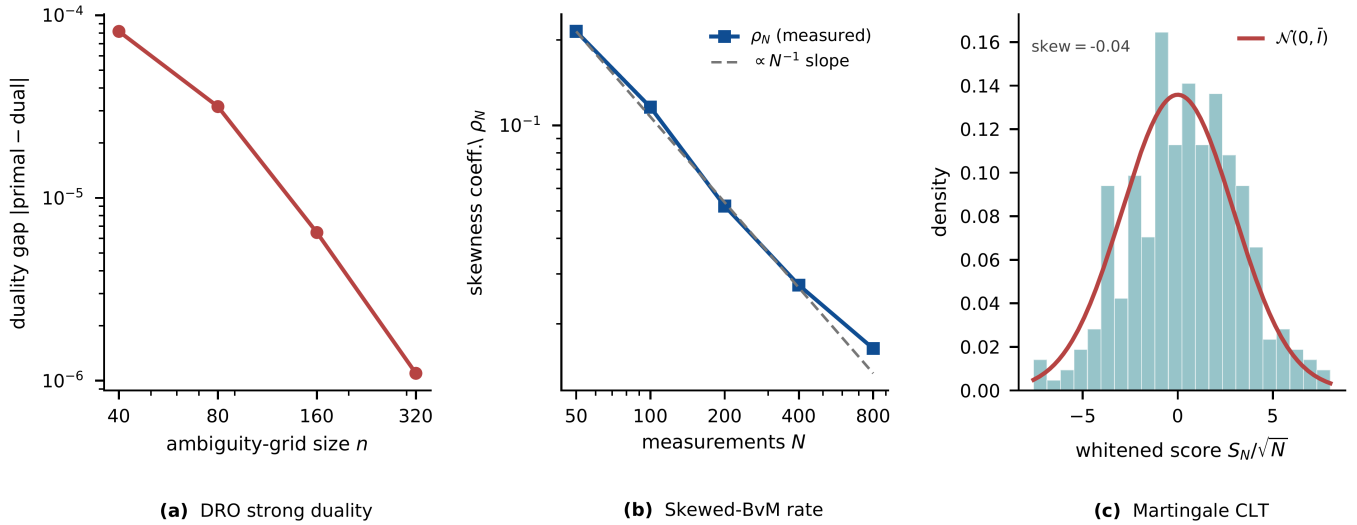


Fig. 12. **Numerical diagnostics.** (a) Discretized continuous- γ Wasserstein primal and dual values converge under grid refinement. (b) The displayed schedule yields rapidly decreasing finite-sample skewness; no universal N^{-1} posterior-TV rate is inferred. (c) The whitened score is approximately Gaussian in the reported simulation, supporting a martingale-CLT diagnostic but not by itself proving posterior contraction or Bernstein-von Mises convergence.

persistent excitation, and it does not claim persistent $O(1)$ skewness in a regular identifiable marginal.

XIV. CONCLUSION

Learned experiment-design policies can overfit the decoherence rates represented during training. We separated three mathematically different robustness notions: finite-radius Wasserstein DRO for analytic schedules, support-minimax raw risk, and minimax relative regret for learned policies. The Wasserstein problem has a convex scalar dual minimization and a scalar global inner maximization; the inner problem is not generally convex. In the reported analytic experiment, robust short-interrogation schedules substantially reduce worst-case information risk and improve actual RMSE under severe rate mismatch. In the learned-policy experiment, regret training produces the flattest reported cross-rate suboptimality envelope, although single-rate RMSE differences are small. The skew-normal ADF is retained as a finite-sample moment approximation, while the asymptotic statement is the ordinary Gaussian BvM under explicit adaptive-LAN and contraction assumptions. Extending the model to within-episode drift, multiqubit probes, correlated noise, and hardware data remains future work.

REFERENCES

- [1] C. L. Degen, F. Reinhard, and P. Cappellaro, “Quantum sensing,” *Rev. Mod. Phys.*, vol. 89, 035002, 2017.
- [2] V. Giovannetti, S. Lloyd, and L. Maccone, “Advances in quantum metrology,” *Nat. Photon.*, vol. 5, pp. 222–229, 2011.
- [3] S. L. Braunstein and C. M. Caves, “Statistical distance and the geometry of quantum states,” *Phys. Rev. Lett.*, vol. 72, pp. 3439–3443, 1994.
- [4] M. G. A. Paris, “Quantum estimation for quantum technology,” *Int. J. Quantum Inf.*, vol. 7, pp. 125–137, 2009.
- [5] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, “The elusive Heisenberg limit in quantum-enhanced metrology,” *Nat. Commun.*, vol. 3, 1063, 2012.
- [6] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, “Achieving the Heisenberg limit in quantum metrology using quantum error correction,” *Nat. Commun.*, vol. 9, 78, 2018.
- [7] J. Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum*, vol. 2, 79, 2018.
- [8] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, “Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis,” *Phys. Rev. A*, vol. 84, 052315, 2011.
- [9] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, “Robust online Hamiltonian learning,” *New J. Phys.*, vol. 14, 103013, 2012.
- [10] A. Hentschel and B. C. Sanders, “Machine learning for precise quantum measurement,” *Phys. Rev. Lett.*, vol. 104, 063603, 2010.
- [11] A. Foster, D. R. Ivanova, I. Malik, and T. Rainforth, “Deep Adaptive Design: Amortizing sequential Bayesian experimental design,” in *Proc. Int. Conf. Machine Learning (ICML)*, 2021. arXiv:2103.02438.
- [12] D. R. Ivanova, A. Foster, S. Kleinegesse, M. U. Gutmann, and T. Rainforth, “Implicit deep adaptive design: Policy-based experimental design without likelihoods,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [13] L. J. Fiderer, J. Schuff, and D. Braun, “Neural-network heuristics for adaptive Bayesian quantum estimation,” *PRX Quantum*, vol. 2, 020303, 2021.
- [14] F. Bellardo, F. Zoratti, and V. Giovannetti, “Applications of model-aware reinforcement learning in Bayesian quantum metrology,” *Phys. Rev. A*, vol. 109, 062609, 2024.
- [15] P. Mohajerin Esfahani and D. Kuhn, “Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations,” *Math. Program.*, vol. 171, no. 1–2, pp. 115–166, 2018.
- [16] R. Gao and A. J. Kleywegt, “Distributionally robust stochastic optimization with Wasserstein distance,” *Math. Oper. Res.*, vol. 48, no. 2, pp. 603–655, 2023.
- [17] J. Blanchet and K. Murthy, “Quantifying distributional model risk via optimal transport,” *Math. Oper. Res.*, vol. 44, no. 2, pp. 565–600, 2019.
- [18] H. Rahimian and S. Mehrotra, “Distributionally robust optimization: A review,” arXiv:1908.05659, 2019.
- [19] A. Azzalini, “A class of distributions which includes the normal ones,” *Scand. J. Stat.*, vol. 12, pp. 171–178, 1985.
- [20] D. Durante, F. Pozza, and B. Szabó, “Skewed Bernstein-von Mises theorem and skew-modal approximations,” *Ann. Statist.*, vol. 52, no. 6, pp. 2714–2737, 2024.
- [21] M. Tsang, “Ziv-Zakai error bounds for quantum parameter estimation,” *Phys. Rev. Lett.*, vol. 108, 230401, 2012.
- [22] X.-M. Lu and M. Tsang, “Quantum Weiss-Weinstein bounds for quantum metrology,” *Quantum Sci. Technol.*, vol. 1, 015002, 2016.
- [23] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*. New York: Academic Press, 1980.
- [24] X. Yu, X. Zhao, L. Li, S. Kurdzialek, C. Zhang, C.-F. Li, and G.-C. Guo, “Towards robust optimal measurements against noise in quantum metrology,” *Sci. China Phys. Mech. Astron.*, vol. 69, no. 7, 270313, 2026.

APPENDIX A PROOFS

Proof of lemma 2. Apply the Sherman–Morrison identity to $\mathbf{A} = \mathbf{J}_{n-1}^{\text{prox}}$ and $\mathbf{u} = \mathbf{u}_n$:

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^\top)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^\top\mathbf{A}^{-1}}{1 + \mathbf{u}^\top\mathbf{A}^{-1}\mathbf{u}}.$$

Substituting $\Sigma_{n-1} = \mathbf{A}^{-1}$ yields (16). $\square$

Proof of lemma 6. Each Fisher-information entry is continuously differentiable in γ on the compact interval. For the scalar surrogate, differentiating j^{-1} gives $\partial_\gamma \tilde{L}_{\text{env}} = -(\partial_\gamma j)/j^2$; the lower bound on j makes this derivative bounded. For the joint surrogate, uniform positive definiteness gives $\|\mathbf{K}^{-1}\| \leq \kappa^{-1}$. Differentiating $\mathbf{K}^{-1}\mathbf{K} = \mathbb{I}$ gives $\partial_\gamma \mathbf{K}^{-1} = -\mathbf{K}^{-1}(\partial_\gamma \mathbf{K})\mathbf{K}^{-1}$. Selecting the (θ, θ) entry yields the displayed formula. In both cases, a bounded derivative on compact Γ implies Lipschitz continuity. $\square$

Proof of corollary 8. For fixed $\hat{\gamma}$, the map $\lambda \mapsto L(\gamma) - \lambda|\gamma - \hat{\gamma}|$ is affine for every γ ; the pointwise maximum over γ is therefore convex, and expectation preserves convexity. If L is Λ -Lipschitz and $\lambda \geq \Lambda$, then

$$L(\gamma) - \lambda|\gamma - \hat{\gamma}| \leq L(\hat{\gamma}) + (\Lambda - \lambda)|\gamma - \hat{\gamma}| \leq L(\hat{\gamma}),$$

with equality at $\gamma = \hat{\gamma}$. Thus $L_\lambda(\hat{\gamma}) = L(\hat{\gamma})$ for $\lambda \geq \Lambda$; when $\varepsilon > 0$, the dual objective then increases linearly, so no minimizer larger than Λ is needed. $\square$

Proof of lemma 10. For the joint score, $\|\mathbf{s}_n\| = t_n V_n / |1 + m_n V_n \cos(\theta t_n)|$. Since $|1 + m_n V_n \cos(\theta t_n)| \geq 1 - V_n$ and $V_n \leq e^{-\gamma \min t_n} < 1$, (32) follows. Bounded increments imply the conditional Lindeberg condition. $\square$

Proof sketch of theorem 11. By Assumption 3, uniformly on compact h -sets the log-likelihood ratio at $\eta_0 + \mathbf{K}_{2,N}^{-1/2}h$ has the LAN form $h^\top \Delta_N - \frac{1}{2}h^\top h + o_p(1)$, with $\Delta_N \Rightarrow \mathcal{N}(0, \mathbb{I}_d)$. Posterior contraction makes the contribution outside such local sets negligible, while prior continuity makes the local prior ratio converge uniformly to one. Normalizing the local likelihood then yields total-variation convergence to the Gaussian law centered at the efficient estimator. The theorem is conditional on LAN and contraction; these are not derived from Assumptions 1 and 2 alone. $\square$

Proof of proposition 12. Let $a = V \cos(\theta t)$ and $b = V t \sin(\theta t)$. The score is $s_\theta(m) = -mb/(1 + ma)$ and $p(m) = (1 + ma)/2$. Therefore

$$\mathbb{E}[s_\theta^3] = \frac{b^3}{2} \left(\frac{1}{(1-a)^2} - \frac{1}{(1+a)^2} \right) = \frac{2ab^3}{(1-a^2)^2},$$

$$I_\theta(t) = \mathbb{E}[s_\theta^2] = \frac{b^2}{1-a^2}.$$

Division gives (35). At $\cos(\theta t) = 0$, $I_\theta = e^{-2\gamma t} t^2$; differentiating its logarithm gives (36). $\square$

APPENDIX B DERIVATION OF THE FISHER INFORMATION (4)

From (2), $\partial_\theta p = -\frac{1}{2}mVt \sin(\theta t)$, so $s = \partial_\theta \log p = \partial_\theta p/p$ gives (3). Using $\sum_m p(m)[\cdot]$ with $m^2 = 1$ and $p(\pm 1) = \frac{1}{2}(1 \pm V \cos \theta t)$,

$$\begin{aligned} I &= \sum_{m=\pm 1} p(m) \frac{V^2 t^2 \sin^2(\theta t)}{(1 + mV \cos \theta t)^2} \\ &= \frac{V^2 t^2 \sin^2(\theta t)}{2} \sum_{m=\pm 1} \frac{1}{1 + mV \cos \theta t} \\ &= \frac{V^2 t^2 \sin^2(\theta t)}{2} \cdot \frac{2}{1 - V^2 \cos^2(\theta t)} \\ &= \frac{V^2 t^2 \sin^2(\theta t)}{1 - V^2 \cos^2(\theta t)}, \end{aligned} \quad (37)$$

which is (4). $\square$

APPENDIX C DERIVATION OF THE SKEW-MODAL FORM (11)

Expand $\ell_N(\theta)$ about its mode $\hat{\theta}$ ($\ell'_N(\hat{\theta}) = 0$): $\ell_N(\theta) - \ell_N(\hat{\theta}) = -\frac{1}{2}ju^2 + \frac{1}{6}\ell_3 u^3 + O(u^4)$, $u = \theta - \hat{\theta}$. In the BvM scaling $u = O(N^{-1/2})$, $j = O(N)$, $\ell_3 = O(N)$, so $\frac{1}{2}ju^2 = O(1)$ and $\frac{1}{6}\ell_3 u^3 = O(N^{-1/2})$. Hence $e^{\ell_N - \ell_N(\hat{\theta})} \approx e^{-\frac{1}{2}ju^2} (1 + \frac{1}{6}\ell_3 u^3)$. Using $2\phi_j(u)\Phi(w(u)) \approx \phi_j(u)(1 + \sqrt{2/\pi}w(u))$ for small w and matching the cubic, $w(u) = \frac{\ell_3}{6}\sqrt{\pi/2}u^3$, giving (11). $\square$

APPENDIX D SHERMAN–MORRISON ALGEBRA (LEMMA 2)

With $\mathbf{A} = \mathbf{J}_{n-1}^{\text{prox}} \succ 0$ and $\mathbf{u} = \mathbf{u}_n$,

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^\top) \left(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^\top\mathbf{A}^{-1}}{1 + \mathbf{u}^\top\mathbf{A}^{-1}\mathbf{u}} \right) = \mathbb{I},$$

which proves (16). This identity concerns the inverse of the approximate expected-information update (15).

APPENDIX E WASSERSTEIN DUALITY (THEOREM 4)

On compact Γ , optimal-transport duality gives (19) under upper semicontinuity of L . If L is Λ -Lipschitz, then for every Q with $\mathcal{W}_1(Q, \hat{Q}) \leq \varepsilon$, the Kantorovich–Rubinstein inequality gives

$$\mathbb{E}_Q[L] - \mathbb{E}_{\hat{Q}}[L] \leq \Lambda \mathcal{W}_1(Q, \hat{Q}) \leq \Lambda \varepsilon,$$

which proves the upper bound in (20); the lower bound follows by choosing $Q = \hat{Q}$. Compactness also makes the inner transform finite for every $\lambda \geq 0$.

APPENDIX F ROBUSTNESS EXPERIMENT: SETUP AND REPRODUCIBILITY

Model and schedule. Single-qubit dephasing Ramsey (2) with $\theta_0 = 1$, $N = 30$ shots at times t_n equispaced on $[0.3, T]$ (the ramp is intended to reduce phase aliasing but does not guarantee global identifiability). The design objective is the inverse-information surrogate $L(T; \gamma) = 1/J(T; \gamma)$, $J(T, \gamma) = \sum_n I(\theta_0, t_n; \gamma)$ with I from (4). Nominal rate

design	risk @ $\hat{\gamma}$	worst-case risk	$\mathbb{E}[\text{risk}]$ mismatch	worst regret vs. rate-aware reference
nominal	0.024	0.265	0.086	0.127
minimax	0.039	0.138	0.066	0.021
DRO	0.037	0.139	0.065	0.019

$\hat{\gamma} = 0.2$; ambiguity $\Gamma = [0.05, 0.6]$ (true rate up to $3\times$ nominal); nominal law $\hat{Q}$ a truncated Gaussian at $\hat{\gamma}$; Wasserstein radius $\varepsilon = 0.06$. **Designs.** *Nominal* $T^* = \arg \min_T L(T; \hat{\gamma})$; *minimax* $\arg \min_T \max_{\gamma \in \Gamma} L(T; \gamma)$; *DRO* $\arg \min_T \sup_{Q \in \mathcal{B}_\varepsilon(\hat{Q})} \mathbb{E}_Q[L(T; \gamma)]$ via the dual (26). Optimized knobs: $T_{\text{nom}} = 5.46$, $T_{\text{minimax}} = 2.11$, $T_{\text{DRO}} = 2.25$.

Actual estimation (fig. 5b). For each γ_{true} , an exact grid posterior ($\theta^* \sim \mathcal{N}(1, 0.5^2)$) is used and the point estimate is the posterior mean. At $\gamma_{\text{true}} = 0.55$ (a $3\times$ mismatch) the RMSE is {nominal 0.41, minimax 0.34, DRO 0.31}, the robust designs are ≈ 17 –20% lower. Scripts: `exp2_robust.py` (numbers), `figures/gen_fig_robust.py` (fig. 5); seed 7. The negative skew-estimation results of sections XI and XIII are reproduced by `exp_cashout.py` and `exp_cashout_curve.py`.

APPENDIX G

THE LEARNED DESIGN AGENT: ARCHITECTURE AND TRAINING

Policy. A two-layer perceptron $\pi_w : \mathbb{R}^2 \rightarrow [t_{\min}, t_{\max}]$ with a 4-unit tanh hidden layer (17 weights); input the belief features ($\sigma_\theta, n/N$), output $t = t_{\min} + (t_{\max} - t_{\min}) \text{sigmoid}(\cdot)$ with $[t_{\min}, t_{\max}] = [0.3, 11]$. **Estimator.** a Gaussian/skew-normal assumed-density filter on θ (section V); prior $\theta \sim \mathcal{N}(1, 0.15^2)$, $N = 20$ shots. **Training (evolutionary strategy).** the cross-entropy method on a diagonal Gaussian over the 17 weights: 26 iterations, population 50, elite 12; each candidate scored by a Monte-Carlo estimate of its stated objective. The nominal agent scores at $\gamma = \hat{\gamma} = 0.2$; the raw support-minimax agent uses (28) over $\gamma \in \{0.10, 0.30, 0.55\}$, whereas the cross-rate generalization agent uses relative regret (29). **Evaluation.** Seeds 0 (training), 7 (evaluation). Scripts `verify_ai_agent.py`, `figures/gen_fig_agent.py`. **Generalization and**

training experiments (figs. 3 and 4). For the generalization matrix we train four nominal specialists, one per assumed rate $\hat{\gamma} \in \{0.10, 0.20, 0.35, 0.50\}$ (each scored only at its own $\hat{\gamma}$), and one robust agent scored by the worst-case *regret* ratio $R(w; \gamma)/R^*(\gamma) - 1$ over the grid $\gamma \in \{0.08, 0.18, 0.30, 0.45, 0.60\}$, where $R^*(\gamma)$ is the RMSE of the best constant- t non-adaptive comparator at that rate, obtained by a 46-point line search. Each agent is then evaluated across true rates. The learning curve (fig. 4a) reports the worst-case suboptimality of the cross-entropy-method mean iterate over the decoherence set (20 iterations, population 44, elite 11). Scripts `verify_ai_agent_matrix.py`, `figures/gen_fig_agent_generalization.py`, `figures/gen_fig_agent_training.py`; seeds 0/1 (training), 3 (reference), 5/7/9 (evaluation).

APPENDIX H

REPRODUCIBILITY AND CODE AVAILABILITY

Cross-entropy-method (CEM) settings. All learned agents use the same gradient-free CEM: a diagonal Gaussian over the 17 policy weights w , initialized at mean $\mathbf{0}$ and standard deviation $\sigma_0 = 0.8$. At each iteration we sample the population, score every member by the Monte-Carlo loss, keep the top- k elite (elite fraction $\approx 24\%$: $k=12$ of 50 for the main agents, 11 of 44 for the generalization study), and refit the Gaussian to the elite mean and standard deviation with an additive exploration floor $+0.03$. There is no gradient or learning rate; optimization runs for a fixed number of iterations (26 main, 20 generalization, 12 baselines) with no early stopping. Interrogation times are clamped to $[t_{\min}, t_{\max}] = [0.3, 11]$. Seeds: 0/1 (training), 3 (reference line search), 5/7/9 (evaluation), 11 (convergence Monte-Carlo). Population sizes (30–50) were set to ≈ 2 – $3\times$ the 17-weight dimension following standard CEM practice, not by a hyperparameter search; results are insensitive to the exact value (the $\approx 24\%$ elite fraction matters more).

Settings for the added ablations and baselines. All of the following use $N=20$ shots and a Gaussian prior $\theta \sim \mathcal{N}(1, \sigma_0^2)$; $\sigma_0=0.5$ (broad) for the design-level baselines, the 2×2 ablation and the simple-baseline study, and $\sigma_0=0.15$ for the neural-agent generalization study. 2×2 *ablation* (table II): the two designs are the CRB-optimal schedules $T_{\text{nom}}=8.36$ (long) and $T_{\text{rob}}=2.10$ (short) on a 20-point ramp $[0.3, T]$; inference is either the Gaussian ADF or the exact posterior on a 300-point grid; worst case over $\gamma \in \{0.08, 0.20, 0.35, 0.50, 0.60\}$. *Simple baselines* (table III): random $t \sim \mathcal{U}[0.3, 11]$; PGH $t = \text{clip}(1/\sigma_\theta, 0.3, 11)$; Thompson (sample $\tilde{\theta}$ from the belief, then $\arg \max_t I(\tilde{\theta}, t)$ on a 60-point t -grid); Bayes-greedy ($\arg \max_t I(\mu, t)$); all design-blind, evaluating information at the nominal $\hat{\gamma}=0.20$; the learned agent is CEM-trained (16 iterations, population 30, elite 8) on the worst case over $\{0.10, 0.30, 0.55\}$. *asymptotic diagnostics* (fig. 10): panel (a) is the closed-form score skewness at $\theta_0=1, \gamma=0.2$; panel (b) solves the Wasserstein primal/dual LPs for $L(\gamma)=1/J(T_{\text{rob}}, \gamma)$ at $\hat{\gamma}=0.20, \varepsilon=0.18$, on γ -grids of $G \in \{6, \dots, 350\}$ points against a $G=1600$ reference.

Code. Every experiment is driven by a small, self-contained numpy script, all available at <https://github.com/changhaohe-glitch/Quantum-Comm>: `validate_math.py`, `verify_ai_agent.py`, `verify_ai_agent_matrix.py`, `exp_baselines.py`, `exp_ablation.py`, `exp_nonmarkov.py`, `exp2_robust.py`, `exp_baselines_extra.py`, `exp_2x2_ablation.py`, `exp_simple_baselines.py`, `exp_agent_seeds.py`, `exp_matrix_seeds.py`, and the figure generators `figures/gen_fig_*.py`. Each figure caption names the script that produces it.