STATE-CONSUMING EXPERIMENTAL DESIGN: GREEDY FAILURE AND AMORTIZED PLANNING

Anonymous authors
Paper under double-blind review

ABSTRACT

An experimental design action normally reveals information and leaves the system available for the next action. A quantum measurement does not: it consumes the state that later measurements needed. We show this is not a quantitative nuisance but a change of problem class. Under back-action the design utility stops being a set function and becomes a function of the measurement *sequence*, and we prove that greedy selection admits *no constant-factor regularized D-efficiency guarantee* over the stated family with no duration or resource restriction: on an explicit two-parameter family, for every fixed regularizer $\lambda > 0$ and every finite C some member forces regularized D-efficiency overhead above C, with the overhead tending to $1/\sin\epsilon$ as the frequency scale $\omega \to 0$. The $(1 - 1/e)$ argument loses its *premise*—the utility is not a set function—rather than a degraded ratio; unbounded overhead alone would not falsify the bound, and we locate the regularizer-coupled regime where the normalized ratio itself drops below $1 - 1/e$. Numerically, greedy has $\rho_D = 3.3$ on an instance whose fresh-copy control shows no gap at all ($5.9\times$ on a second, where fresh copies *reverse* the ranking). Exact enumeration certifies the optimum here but costs $(2d{+}1)(2|\mathcal{A}|)^k$ leaf operations, motivating—not proving the necessity of—amortization. A policy trained once per instance distribution, spending *zero* per-instance Fisher-utility evaluations at deployment, beats exact-model greedy on all four tested random-Hermitian conditions (29–40 of 30–40 paired instances, Holm $p \leq 1.3 \times 10^{-8}$). The boundary is negative on device-physics instances: the existing bare policy trails greedy, and a new fixed, post-hoc check on 90 previously evaluated held-out instances finds that learned-surrogate repair improves the bare rollout but does not significantly beat greedy ($51/90$ wins, $p = 0.105$). Separately, one preregistered hardware mechanism test on a trapped-ion processor met its frozen numerical criterion (MSE ratio 18.1, shot-resampling 90% CI $[8.9, 40.7]$); because each arm is one sequentially submitted job, this interval is conditional on those jobs and does not cover between-job drift. It is one adversarial instance, not distribution-level validation. The anonymous supplement contains the code, raw evidence, preregistration, and adverse results.

1 INTRODUCTION

Classical experimental design rests on a premise so familiar it is rarely stated: the system survives being measured. Because it does, the utility of a set of measurements is a *set function*—the premise without which the $(1-1/e)$ guarantee of greedy submodular maximization (Nemhauser et al., 1978) cannot even be stated, a lineage running from Lindley (1956) through the modern Bayesian design literature (Chaloner & Verdinelli, 1995; Ryan et al., 2016; Rainforth et al., 2024). Set-valuedness is necessary for that argument and not sufficient for the guarantee: monotonicity and submodularity are assumed as well, and some sequence objectives carry greedy guarantees of their own (Golovin & Krause, 2011; Asadpour et al., 2022; Tang & Yuan, 2024). What removes the guarantee below is therefore the explicit counterexample of §2, not the change of type on its own. A quantum measurement violates the premise directly. Projection collapses the very state a later measurement needed, so an action does not merely reveal information: it destroys some of the information that remained to be gathered.

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

The regime of interest is the one where re-preparation is not free: state preparation dominates the shot budget, the platform measures a register mid-circuit, or the sequence itself is the object of design. When fresh copies *are* free the problem collapses back to a set function—our fresh-copy control makes that collapse an instrument, not an assumption.

This paper is about the consequences of that one violation. Our position is that it changes the *class* of the design problem rather than its difficulty. When actions consume the system, the design utility becomes a function of the measurement *sequence*, and set-based reasoning is not a good approximation that degrades gracefully—it is the wrong model. We make this precise, characterize where greedy fails and where it is safe, and test an amortized alternative under a fixed planning budget:

- **The premise does not survive (§2).** We give an explicit two-parameter family on which greedy takes the same bait for every parameter value and every regularizer, and whose ρ_D exceeds any finite C at fixed $\lambda > 0$, tending to $1/\sin\epsilon$ as $\omega \to 0$: *no* constant-factor ρ_D bound exists (Proposition 1). Nor is the regularizer an escape route: any ridge large enough to bound greedy on that family is large enough to bound a rule that never reads the instance (Proposition 3). The $(1 - 1/e)$ bound loses its premise (set-function utility), not merely its constant; on the set-function relaxation the ratio drops below $1 - 1/e$ only when the surviving information reaches the ridge floor, $\sin^2\epsilon = O(\lambda)$. The loss is large and attributable: greedy pays $3.3\times$ on an instance whose fresh-copy control pays 1.001, and $5.9\times$ on a second where fresh copies *reverse* the ranking (Proposition 2); permuting the same action multiset moves utility by 1.6–$2.3\times$.
- **Amortization works on random matrices, with a device boundary (§3).** A policy trained once per instance distribution spends *zero* utility evaluations at deployment and still beats greedy on all four random-matrix conditions (29–40 of 30–40 paired instances, Holm $p \leq 1.3 \times 10^{-8}$); these are held-out instances from the same generator, so this is within-distribution amortization, not transfer. Greedy makes $k|\mathcal{A}|$ Fisher-utility queries per instance (Table 2); these are available computationally when the nominal generator, background, initial state, and expansion point are supplied, so the bare policy is not the only nominal-model-deployable planner. Its distinction is zero per-instance utility queries, not exclusive deployability or a measured speed advantage. Adding a budget-parity repair sweep—at greedy's own evaluation count—reaches the exact optimum in 17–80% of instances against greedy's 0–7%, and trains *teacher-free* at $12.6\times$ lower cost. On device-physics instances the bare policy is *worse*: null at preregistered n=30, but supported on the previously evaluated 90-instance suffix at k=3, 4 (Holm $p = 0.036$ and 6.2×10^{-6}; Appendix A.28); combined n=120 tests are descriptive because they include that prefix. Post-hoc learned repair improves the bare rollout but does not beat greedy significantly; no oracle-free device advantage is established.
- **Where greedy is safe is not where one would guess (§2).** Weak measurement does not restore submodularity: at κ=0.1 the median device-physics instance is optimal yet 29% still lose, and the random-matrix median is non-monotone in κ. The operative structure is the rank and conditioning of the Fisher information, which the proposition exploits.

Together these give a single claim: *a quantum measurement can destroy the future information an experimental design still needed; this state-consuming design problem is not a set-function problem, greedy provably loses its guarantee on it, and amortized planning at a fixed budget recovers the loss on the tested random-Hermitian family, and does not yet on the device-physics family*. Two further allocation questions raised by the same accounting—how many shots per training label, and whether to spend shots characterizing the device—are supporting cases with their own appendices (Appendices A.14 and A.16), not contributions. We close with a preregistered, deferred-measurement implementation of the order experiment whose single preregistered run on a trapped-ion processor met its frozen numerical criterion, conditionally on the four observed jobs (§4). The anonymous supplement contains the testbed, raw counts, analysis, and calibration evidence.

2 Order: back-action breaks greedy design

Setting. An experimenter learns d=2 Hamiltonian parameters $\theta = (J, \delta)$ of a two-qubit device (exchange coupling and detuning difference, with fixed ZZ crosstalk and preparation-error background) by choosing a sequence of k projective measurements from a menu $\mathcal{A}$ of $|\mathcal{A}|$=6

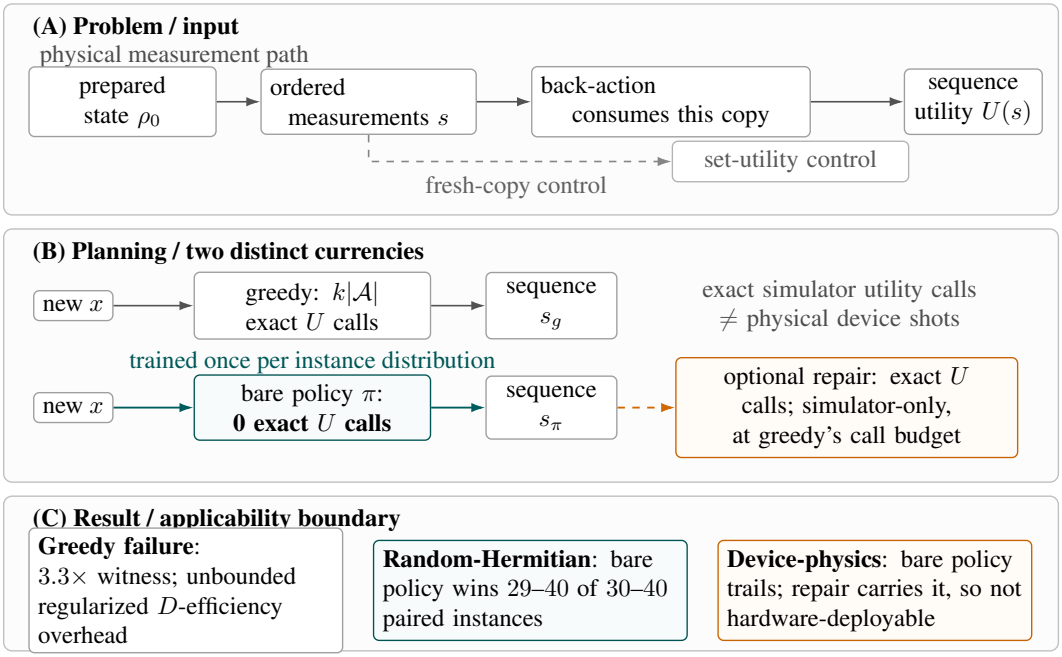


Figure 1: **State-consuming experimental design, and the amortized exit.** (A) A projective measurement consumes the probed copy, making utility a function of the measurement *sequence*; the dashed fresh-copy branch is the set-utility control. (B) Greedy spends $k|\mathcal{A}|$ true-model utility calls per benchmark instance; nominal-model greedy can instead compute utilities from the supplied descriptors. The policy is trained once per instance distribution and deploys with zero utility evaluations. The optional repair sweep also queries exact simulator utility, at greedy's call budget. (C) The $\omega \to 0$ $\rho_D = 1/\sin\epsilon$ limit is unbounded, with the k=2 witness at 3.3× (Proposition 1); the bare rollout wins 29–40 of 30–40 paired random-matrix instances (Holm $p \le 1.3 \times 10^{-8}$), while on device-physics instances it trails greedy and the repair carries it (§3).

actions (evolution time × qubit × basis). The utility is the local D-optimality criterion $U(s) = \log\det\big(F(s) + \lambda_{\text{reg}} I\big)$ of the Fisher information of the *joint* outcome distribution of the sequence s, with a fixed ridge $\lambda_{\text{reg}} = 10^{-10}$ used throughout. Because each measurement projects the register, later outcome distributions depend on what was measured earlier: U is a function of the ordered sequence. Table 3 fixes the protocol (the width-case counterpart is Table 5).

The premise collapses, measurably. On a fresh-copy control (each setting applied to a new state; utilities add), greedy is optimal to numerical precision. With back-action, greedy's median *regularized D-efficiency overhead* $\rho_D = \exp\big[(U^* - U_{\text{greedy}})/d\big] = \big(\det(F^* + \lambda_{\text{reg}} I)/\det(F_{\text{greedy}} + \lambda_{\text{reg}} I)\big)^{1/d}$—the reciprocal of the relative regularized D-efficiency of classical optimal design—grows monotonically with horizon (n=200 instances per condition; 95% bootstrap CIs, all six conditions Holm $p \le 6 \times 10^{-4}$): on device-physics instances $1.30\,[1.18, 1.38]$ (k=3), $1.43\,[1.30, 1.54]$ (k=4), $1.49\,[1.39, 1.58]$ (k=5); on random-matrix instances 1.45/1.60/1.73 (Figure 2). It is not a fixed-ridge shot factor; that interpretation holds only at $\lambda = 0$ or asymptotically (Appendix A.3). The fresh-copy control's median is exactly $1.0000\,[1.0000, 1.0000]$ (max 1.0002 over 100 instances): on these witnesses the deficit is back-action, isolated. Permuting the *same multiset* of actions changes ρ_D by 1.6–2.3× (best vs. median ordering; best vs. worst is 3.1–8.6× across all thirteen settings): U is not a set function, so the submodularity machinery does not merely degrade—its type no longer matches (a measured submodularity-ratio diagnostic $\gamma = 0.31\,[0.26, 0.37]$ in the sense of Das & Kempe (2011), a family-level diagnostic that does not rank per-instance risk, computed on a set-function relaxation of U; estimator definition in appendix, fresh-copy control exactly 1.000).

Proposition 1 (Greedy's ρ_D is unbounded under back-action)**.** *Fix the local D-optimal utility* $U(s) = \log\det(F(s) + \lambda I)$. *There is an explicit two-parameter family* $\{\mathcal{I}_{\epsilon,\omega}\}$ *of instances—one qubit,* $|\mathcal{A}|$=2, k=2, *with frequency scale* ω*—on which* (i) *greedy selects the same first action for* ev-

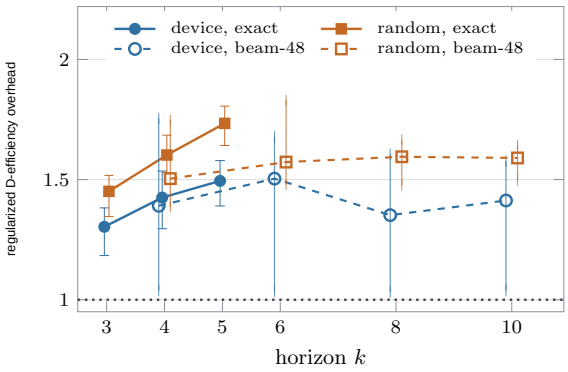


Figure 2: Greedy's median ρ_D against horizon. Filled markers and solid lines are exact-oracle medians ($k \leq 5$, n=200 per condition); open markers and dashed lines extend the trend to k=10 with a beam-48 best-found proxy, which is a lower bound on the true gap (exploratory). Bars are bootstrap 95% CIs of the median; the fresh-copy control sits exactly at 1 (measured at k=3), so on this witness the penalty is back-action. Device beam-48 lower CI endpoints sit at 1.03–1.04, above the ρ_D=1 line.

ery (ϵ, ω) *and* every $\lambda > 0$*;* (ii) ρ_D *has a closed form which tends to* $1/\sin\epsilon$ *as* $\omega \to 0$ *at any* fixed $\lambda > 0$*; every Fisher entry scales as* ω^{-2}*, so the ridge vanishes asymptotically and this limit also has the D-optimal shot-replication interpretation. Hence* (iii) *for every finite* C *and fixed* $\lambda > 0$ *some member forces overhead* $> C$*: no constant uniformly bounds this criterion. This is not a literal fixed-ridge shot count or the approximation ratio: on the normalized relaxation* $f_\lambda(S) = U(S) - d\log\lambda$ *the greedy ratio stays above* $1 - 1/e$ *(and* $\to 1$ *as* $\omega \to 0$*) except where the information surviving greedy's first action reaches the ridge floor,* $\pi^2 \sin^2\epsilon/(4\omega^2) = O(\lambda)$*. The premise is removed by the type change—*U *is not a set function—not by the overhead's size. The bait choice is structural rather than tuned: greedy selects it for every* $\omega > 0$*, for non-canonical second-action times, and under perturbation of the operating point* θ_0*; unboundedness is a property of the joint family over* (ϵ, ω)*, not of any fixed member.*

Proposition 2 (The deficit is back-action, not a worse action multiset)**.** *There exist instances of the workload family above (reproducible from stated seeds) on which greedy's* ρ_D *is* 3.30 *at* k=2 *and* 5.94 *at* k=3*, while executing each arm's own action multiset on fresh copies erases the gap entirely (ratio* 1.00 *at* k=2*)—and at* k=3 *even* reverses *it* (0.19)*: under fresh copies greedy's multiset is the better one, so on these witnesses the deficit is an ordering/back-action effect.*

Proof of Proposition 1 (closed form; verified to machine precision). Take $\psi_0 = |0\rangle$ and $H(\theta) = \theta_1 Z + \theta_2 X$ at $\theta_0 = (0, \omega)$, with actions D (evolve $\pi/2\omega$, measure in the X basis tilted by ϵ toward Z) and G (evolve $\pi/4\omega$, measure Z). Then $F_D = \mathrm{diag}(4/\omega^2, 0)$ and $F_G = \mathrm{diag}(0, \pi^2/4\omega^2)$ are both rank one, so the regularized one-step $\log\det$ comparison reduces to a trace comparison and greedy takes the bait D whenever $16 > \pi^2$—that is, always, independently of λ. Measuring D leaves the state within Bloch angle ϵ of an eigenstate of the θ_2 generator, so any follow-up carries only $O(\sin\epsilon)$ sensitivity: $\det F(D, G) = \pi^2 \sin^2\epsilon/\omega^4$ against $\det F(G, D) = \pi^2/\omega^4$, giving $\rho_D^2 = \det(F(G, D) + \lambda I)/\det(F(D, G) + \lambda I)$ in closed form. Every Fisher entry is $\Theta(\omega^{-2})$, so at any fixed $\lambda > 0$, $\omega \to 0$ sends the ratio to $1/\sin^2\epsilon$; given C, choose ϵ with $1/\sin\epsilon > C$, then ω small enough. □

Appendix A.3 checks every closed-form quantity above against numerically computed Fisher matrices and certifies the three robustness claims of Proposition 1.

Scope. Unbounded overhead alone is not a back-action effect: a fresh-copy additive submodular design already diverges as $\lambda^{-1/4}$ with ratio $\to 3/4 > 1 - 1/e$, so divergence falsifies nothing below exponent $1/e$. At any fixed member of the family and fixed λ the ridge floor caps the overhead (magnitudes at the codebase's λ in Appendix A.3), which is exactly why the proposition quantifies over the family. The result is mechanism-level, not a claim of experimental accessibility at fixed resources; restricting the family to a time budget gives a quantitative version, in which any constant-factor guarantee must degrade at least linearly in the allowed interrogation time and as $\lambda^{-1/2}$ (bounded-duration corollary, Appendix A.3). Finite-shot access to rare outcomes and deco-

herence both remain untreated. The family is deliberately minimal, and its protection lasts exactly one step: the second projective measurement resets the state off the dark manifold, and greedy escapes at $k \geq 3$ (overheads 1.03–1.26 across tilts, against unbounded at k=2). Proposition 1 is therefore an unboundedness statement at fixed k=2, not a horizon-scaling law; the horizon trend is measured, not proved (Figure 2). Both propositions are statements about the local D-optimal utility; we do not claim them for every design criterion (Appendix A.8 reports an expected-information-gain kill test). Setting $\epsilon = \arcsin(1/3.30)$ recovers the k=2 witness of Proposition 2 analytically, tying the two together.

Proof of Proposition 2 (explicit instances plus exhaustive oracle evaluation; seeds in the appendix). The anatomy is a one-step sacrifice trap: the immediately most informative measurement collapses the state component that carries the second parameter, and greedy by definition takes it. □

Nor does tuning the ridge. Proposition 1 holds λ fixed and varies the instance, which invites the counter-move of letting λ grow with the instance. It cannot rescue greedy, and the obstruction is structural: the smallest ridge that bounds greedy already bounds a rule that never looks at the instance.

Proposition 3 (Ridge dichotomy: no regularizer schedule certifies greedy)**.** *Write* $\Phi(\mathcal{I}) = \max_{s\in\mathcal{A}^k} \|F(s)\|$ *for the largest Fisher eigenvalue an instance admits over sequences, and let* $\rho_D(R;\mathcal{I},\lambda)$ *denote the* ρ_D *value of a selection rule* R *at ridge* $\lambda > 0$*.* (i) *For every instance and* every *rule—including one that ignores* $\mathcal{I}$ *and emits a fixed sequence—*$\rho_D(R;\mathcal{I},\lambda) \leq 1 + \Phi(\mathcal{I})/\lambda$*.* (ii) *On the family* $\{\mathcal{I}_{\epsilon,\omega}\}$ *of Proposition 1 there is a constant* $c_0 > 0$ *of the family such that every ridge schedule* $\lambda = \lambda(\epsilon,\omega)$ *holding greedy's overhead at or below* $C \geq 1$ *must satisfy* $\lambda \geq c_0 C^{-2}\Phi(\mathcal{I}_{\epsilon,\omega})$ *on every member with* $\sin\epsilon \leq 1/(C\sqrt{2})$*; by* (i)*, that same schedule bounds* every *rule by* $1 + C^2/c_0$ *on those members. On this family no regularizer bounds greedy without simultaneously bounding a rule that reads nothing. All quantities here are analytic.*

Proof. (i) Each $F(s) \succeq 0$, so $\det(F(s) + \lambda I) \geq \lambda^d$, while the optimum s^* has every Fisher eigenvalue at most Φ and therefore $\det(F(s^*) + \lambda I) \leq (\Phi + \lambda)^d$; the ratio and a d-th root give $\rho_D \leq (\Phi+\lambda)/\lambda$, a bound that never mentions what R returned. (ii) Here d=2, where $\det(F+\lambda I) = \det F + \lambda \operatorname{tr} F + \lambda^2$ exactly and $\operatorname{tr} F \leq 2\Phi$. Substituting the determinants of Proposition 1 into its closed form,

$$\rho_D^2 = \frac{\det(F(G,D) + \lambda I)}{\det(F(D,G) + \lambda I)} \;\geq\; \frac{\pi^2\omega^{-4}}{\pi^2 \sin^2\epsilon\, \omega^{-4} + 2\lambda\Phi + \lambda^2}\,.$$

(Either follow-up qualifies: G attains the determinant bound, and the repeated D leaves $\det F(D,D) = 0$—the post-D state is an eigenstate of the measured tilted basis, so the repeated outcome law is first-order insensitive to θ_2 at θ_0—so the display covers whichever second action the regularized comparison returns.) Imposing $\rho_D \leq C$ and $C^2 \sin^2\epsilon \leq \frac{1}{2}$ leaves $\frac{1}{2}\pi^2\omega^{-4} \leq C^2(2\lambda\Phi + \lambda^2)$. Every Fisher entry of the family scales as ω^{-2} (Proposition 1), so $\phi(\epsilon) := \omega^2\Phi$ is free of ω and, being continuous on a compact range of ϵ, has $\bar{\phi} := \sup_\epsilon \phi < \infty$; put $c_0 := \min\{2, \pi^2/(8\bar{\phi}^2)\}$. If $\lambda \leq 2\Phi$ then $2\lambda\Phi + \lambda^2 \leq 4\lambda\Phi$, so $\lambda \geq \pi^2\omega^{-4}/(8C^2\Phi) = \pi^2\Phi/(8\phi^2C^2) \geq c_0C^{-2}\Phi$; if instead $\lambda > 2\Phi$ then $\lambda > 2\Phi \geq c_0\Phi \geq c_0C^{-2}\Phi$. Either way $\lambda \geq c_0C^{-2}\Phi$, and (i) then bounds every rule by $1 + \Phi/\lambda \leq 1 + C^2/c_0$. □

The regularizer is thus not a free parameter one can tune into a guarantee: any λ that certifies greedy on this family certifies the ridge rather than the rule, and the regime it certifies is the one in which selection no longer matters.

Worst cases are not tail exotica: across 1,000 instances per condition the 90th-percentile overhead reaches 2.0–2.3 at k=3. A closed-form one-parameter family reproducing the same trap—and recovering the k=2 champion instance analytically—is given in Appendix A.4. The failure is not confined to strong measurement, and here our own data corrects a natural guess. Weakening the instrument moves the *median* device-physics instance to exactly 1.000 for $\kappa \leq 0.75$ (Figure 3a), but the median hides the distribution: at κ=0.1 the mean is still 1.072, 29.2% of instances exceed 1.01, and the worst reaches 1.905; on random-matrix instances the median is *non-monotone* in κ (1.194 at κ=0.1 against 1.129 at κ=0.25) and does not approach 1 as $\kappa \to 0$. We state this empirically: reducing κ does not uniformly drive the ratio to 1 across families. It is tempting to blame information

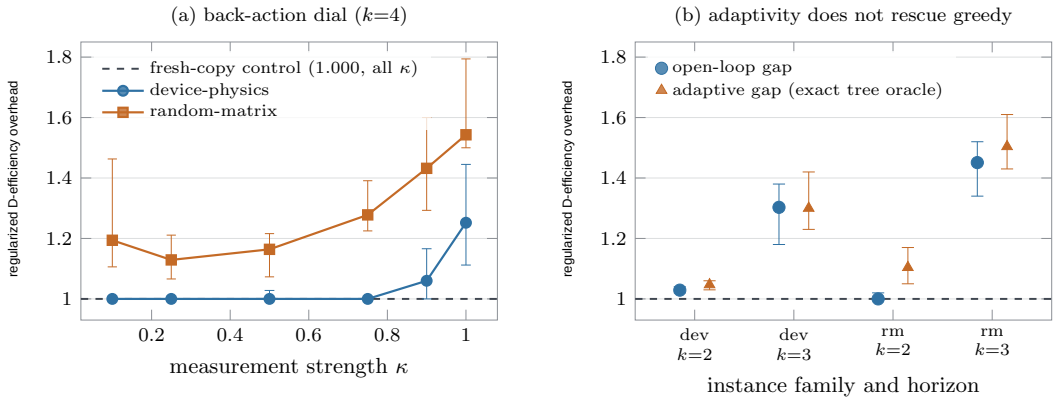


Figure 3: The order pathology, mechanistically (all exploratory; medians with bootstrap 95% CIs). **(a)** Weak-measurement dial: on device-physics instances greedy's overhead is exactly 1.000 for $\kappa \leq 0.75$ and switches on only near projective strength, while the fresh-copy control stays at 1.000 for every κ; random-matrix instances add a κ-insensitive scheduling residual (appendix). **(b)** The exact closed-loop tree oracle does not shrink the greedy gap: adaptive gaps match or exceed open-loop gaps in all four conditions.

and disturbance entering at different orders, but for these Kraus operators they do not—the non-selective channel damps coherences by $\sqrt{1-\kappa^2} = 1 - \kappa^2/2 + O(\kappa^4)$ and the Fisher information is likewise $O(\kappa^2)$ (Appendix A.23); what remains is a scheduling contribution, not a disturbance-order mismatch. Measurement strength is therefore the wrong dial for a safety condition; the rank and conditioning of F—the structure Proposition 1 exploits—is the operative one. The gap neither vanishes nor explodes with horizon (Figure 2, open markers).

Matched-budget search on the tested menu. On a harder menu ($|\mathcal{A}|$=8, n=12 instances; oracle 8^4=4,096), beam search narrows the median overhead (1.98 $\rightarrow$ 1.25 from greedy to beam-8 at k=4) without closing it, and only by raising the evaluation bill from greedy's 32 to beam-8's 200 against the oracle's 4,096—and each evaluation is itself an outcome-tree of size 2^k. Exhaustive certainty costs $|\mathcal{A}|^k$ utility evaluations, each a $(2d{+}1)\,2^k$-leaf computation—one outcome tree at θ_0 plus a central difference per parameter, so $(2d{+}1)(2|\mathcal{A}|)^k$ leaf operations in all ($2d{+}1{=}5$ here)—and local search with greedy's budget stalls at 1.52. These finite comparisons motivate amortization; they are not a lower bound and do not exclude a more structured non-amortized planner.

Neither does adaptivity (Figure 3b). Conditioning each next measurement on the observed outcomes—an exact closed-loop oracle over all decision trees, tractable at $k \leq 3$ and already costing $\approx 2.8 \times 10^5$ joint evaluations there ($6 \cdot 216^2$; Appendix A.7)—does not dissolve the gap: the adaptive greedy-to-oracle overhead matches or exceeds the open-loop one (medians 1.30 vs. 1.30 device-physics, 1.50 vs. 1.45 random-matrix at k=3; paired difference ≥ 0), adaptivity adds only 1.00–1.15$\times$ headroom over the open-loop optimum, and the k=3 maximal-gap instance of §4 retains a 4.99$\times$ adaptive gap. The order pathology is a property of back-action, not of the open-loop restriction. The effect also has a *dial* (Figure 3a): when projective readout is replaced by weak measurements of strength κ, the device-physics greedy gap is absent for $\kappa \leq 0.75$ and switches on only near projective strength (median 1.25$\times$ at κ=1), while the fresh control stays at exactly 1.000 for every κ (both exploratory, outside the declared families; appendix).

3 Amortized policies close the gap at greedy cost

We train a sequence policy on features of the instance and partial history, with an exhaustive-search teacher on *training* instances only, and evaluate on disjoint instances under strict evaluation-budget parity: the policy plus the repair pass of Algorithm 1 may use exactly greedy's number of utility evaluations.

Algorithm 1 Amortized sequence design with budget-parity repair

Require: instance features x; horizon k; action menu $\mathcal{A}$; utility oracle U; history encoding $\mathrm{hist}(\cdot)$ (one-hot actions + step index)

Ensure: sequence s, $|s| = k$, using at most $k|\mathcal{A}|$ utility evaluations at deployment (= greedy's per-instance bill)

1: **Training** (once per instance *distribution*; amortized over all future instances)
2: **for** each training instance x_i **do**
3: $s_i^\star \leftarrow$ target sequence: the exhaustive optimum, *or* (teacher-free) this policy's own repaired rollout
4: add $\big\{\big(x_i, \mathrm{hist}(s^\star_{i,<t})\big) \mapsto s^\star_{i,t}\big\}_{t=1}^{k}$ to the training set
5: **end for**
6: fit π on the training set $\{k$ examples per instance$\}$
7: **Amortized design** (zero utility evaluations)
8: **for** $t = 1, \ldots, k$ **do**
9: $a_t \leftarrow \pi\big(x, \mathrm{hist}(s)\big)$; $\quad s \leftarrow s \oplus a_t$
10: **end for**
11: **Budget-parity repair**: coordinate-ascent sweeps over (t, a), accepting any improving swap $s_t \leftarrow a$; repeat until no improvement, capped at $k|\mathcal{A}|$ evaluations (greedy's bill)
12: **return** s {teacher-free variant: use these s as targets in line 3, repeat lines 1– 6; stored labels are then polished by multi-start local search at $8\times$ the deployment cap (training-side only; deployment bill unchanged)}

Table 1: Policy + repair at *greedy's own evaluation budget* ($k|\mathcal{A}|$ utility evaluations per instance, 30–60 across rows, identical for both arms; test instances disjoint from training, n=30 per row except $(7, 6)$ with n=40; "reach" = fraction where the exact optimal sequence is found; "wins" = paired instances where the bare policy—zero utility evaluations—beats greedy, significant on all four random-matrix rows and on neither device row (§3); per-instance paired win-or-tie rates of policy+repair against greedy are 90–100% across rows). Greedy's median penalty of 1.2–2.6× drops to 1.00–1.11×; a beam-8 comparison at one setting is reported in §2, and per-condition beam/SLS tables are in the released artifacts. The k=5, $|\mathcal{A}|$=6 row is the teacher-free (self-improving) policy. All tests and intervals condition on the fitted policy; they do not estimate variation across independent training realizations.

Instances	$k, \lvert\mathcal{A}\rvert$	Greedy		Bare policy (0 evals)		Policy + repair (= greedy evals)		
		med ρ_D	reach	med ρ_D	wins	med ρ_D	reach	[Wilson 95%]
random-matrix	$5, 6$	2.03	0%	**1.000**	29/30	**1.000**	80%	[63,90]%
random-matrix	$7, 6$	2.58	0%	1.033	40/40	**1.000**	60%	[45,74]%
random-matrix	$6, 8$	1.86	0%	1.175	29/30	1.010	50%	[33,67]%
random-matrix	$5, 8$	1.97	0%	1.153	29/30	1.023	47%	[30,64]%
device	$3, 12$	1.29	7%	1.379	14/30	1.011	50%	[33,67]%
device	$5, 12$	1.22	0%	1.367	12/30	1.114	17%	[7,34]%

On the policy class. The policy is a random-forest sequence classifier, and deliberately so: the claim under test is that *amortization* escapes the order pathology, not that a particular architecture does. The regime makes a small model the honest choice—the instance description is a handful of real Hamiltonian entries, a condition supplies 100–200 training instances, and the decision at each step is a choice among $|\mathcal{A}| \leq 12$ actions. We did not tune the function class—and we ran the direct comparison: under identical teachers, seeds, inputs, and deployment budget, an MLP on the same features and a DAD-style GRU history encoder (Foster et al., 2021; Ivanova et al., 2021) tie with or lose to the forest (no significant deep win in eight paired comparisons; the three significant differences favor the forest; Appendix A.18). The gain is amortization, not architecture.

Two currencies, and what survives with only one. Table 2 makes explicit that our utility is the *local* D-optimality criterion at a known, instance-specific expansion point θ_0, evaluated with the exact Fisher information—a planning quantity, not a measurable one. Every planner in Table 1 except the bare policy spends this currency, greedy included; the shot column of Table 1 is what those plans cost *once executed*, at matched planning spend. What remains when the simulator is removed entirely? On the teacher-free headline row: most of it. The bare policy alone—zero utility

Table 2: **What each planner may look at.** Two currencies are separated: a *device shot* consumes the probed state; a *utility evaluation* consumes exact model-computation time rather than device shots. The last column separates per-instance computational queries from device shots. With the nominal model in column two, greedy can compute nominal Fisher utilities and is therefore deployable in regime (ii), but it makes $k|\mathcal{A}|$ such queries. The bare policy makes none; we measured utility quality, not wall-clock speed.

Quantity	Known at deployment (regime ii)	Used by	Simulator-only (regime i)
Action menu (time × qubit × basis)	yes	all	—
Generators A_1, A_2, background H_b, ψ_0	nominal	policy; nominal greedy	true values
Measurement history (one-hot + step)	yes	policy input	—
Estimand $\theta = (J, \delta)$	**no**	nothing	—
Expansion point θ_0 (known per instance)	yes	utility def.; all planners	—
Fisher $F(s)$ at θ_0	computable under nominal model	greedy, beam, repair	true-model benchmark
Exhaustive oracle over $\|\mathcal{A}\|^k$	**no**	teacher (removable)	✓
Device shots	yes	estimation	—

evaluations—has median $\rho_D = 1.000$ and reaches the exact optimum in 57% of instances against greedy's 2.03 and 0%, winning 29 of 30 paired instances (Wilcoxon $p = 3.7 \times 10^{-9}$); repair buys the remaining reach ($57\% \to 80\%$) and tames the tail ($2.75 \to 1.23$). We report the boundary the paired per-instance tests actually draw—a matched horizon sweep, k=3, 4, 5 in each instance family (Wilcoxon on $\log \rho_D$ with zero-differences discarded, n=30 per condition, Holm within that six-condition family; paired win counts reported alongside keep ties visible). Those tests separate the two instance families cleanly. On *all four* random-matrix rows of Table 1 the bare policy beats greedy: medians 1.00 vs. 2.03 at $(5, 6)$, 1.03 vs. 2.58 at $(7, 6)$, 1.18 vs. 1.86 at $(6, 8)$, 1.15 vs. 1.97 at $(5, 8)$, winning $29/30$, $40/40$, $29/30$ and $29/30$ paired instances (Holm $p \leq 1.3 \times 10^{-8}$; the $(5, 6)$ row is the teacher-free policy, $p = 3.7 \times 10^{-9}$). The deployable claim is thus not a median artifact. On *device-physics* instances it is: the bare policy beats greedy at no horizon—not significant at preregistered n=30 (smallest Holm $p = 0.083$), but significantly worse on the previously evaluated 90-instance suffix at k=3, 4 (Holm $p = 0.036$ and 6.2×10^{-6}; Appendix A.28). Combined n=120 tests are descriptive post-hoc because they include the inspected prefix. The Table 1 device rows give 1.38 vs. 1.29 and 1.37 vs. 1.22. Those rows' headline numbers therefore do depend on repair. We state the scope accordingly: amortization alone suffices on the tested random-Hermitian family, and the exact-utility repair is what carries the device-physics family—the clearest open problem this paper leaves is to replace that oracle with an estimate from measured statistics. Two preregistered attempts at this gap—a disturbance-aware policy feature and a learned-surrogate repair—did not clear their frozen bars. After those tests were closed, we froze a new post-hoc check before running it on the previously evaluated $i = 30, \ldots, 119$ suffix of the established held-out split (k=3, n=90). Learned-surrogate repair improved the bare rollout (median 1.150 vs. 1.325; 28 wins, 55 ties, 7 losses; paired Wilcoxon $p = 5.3 \times 10^{-6}$) but did not significantly beat greedy (median 1.243; 51 wins, 1 tie, 38 losses; $p = 0.105$). Thus it is a new adverse boundary check, not a reopened confirmatory test or evidence of oracle-free superiority over greedy (Appendices A.21 and A.19).

Two findings sharpen the claim (Table 1). (i) *Budget*: repair is not the whole story. Random initialization can also reach the optimum but needs 4–18× more repair evaluations; and giving *greedy itself* the same repair sweep on top of its own selection ($2n_g$ evaluations total, twice the policy arm's budget) still loses to policy+repair in five of six conditions (paired Wilcoxon, Holm $p \leq 0.038$; real k=4 not significant; Appendix A.17). The learned initialization is doing work the sweep cannot. (ii) *Teacher-free*: self-improving training in the expert-iteration pattern (Anthony et al., 2017) (labels from the policy's own repaired rollouts—repair still scores by the exact simulator, so training is exhaustive-teacher-free, not oracle-free) reaches exhaustive-teacher parity on all four metrics using 123,552 total training evaluations against the exhaustive teacher's 1,555,200 (12.6×)—removing the exponential teacher from the loop entirely. Training cost is amortized: it is paid once per instance *distribution*, while greedy pays its per-instance evaluations on every new instance, so the *utility-call* break-even for this $(5, 6)$ teacher-free condition is ∼4,100 deployed instances ($123{,}552/30$; training time, feature computation, and device shots are separate ledgers not included here). Throughout, an "evaluation" is one black-box call to U; its internal leaf cost is separate. In a post-prior-results addendum, ten regenerated 200-instance training sets and forest seeds for the released teacher-free $(5, 6)$ pipeline gave mean fit-level log ρ_D ratio -0.649792 (geometric ratio 0.522154; fit-bootstrap 95% CI $[-0.651812, -0.647241]$), all favorable on the fixed 120-seed bank. However, $1{,}165/1{,}200$ rollouts were $[5, 5, 5, 5, 2]$ and the other 35 were $[5, 5, 5, 2, 5]$: this is stable near-constant behavior,

not input-adaptive generalization. The inferential unit is the fit (n=10), and the interval conditions on this already-inspected bank (Appendix A.29). Separate exhaustive-supervised fits retained the random-family advantage (geometric ratio 0.5203; identical constant rollout in all fits) but were adverse on device physics (ratio 1.4938; all ten fit effects adverse). Under distribution shift the boundary remains sharp, and no finite-readout-error claim is made (Appendices A.9 and A.20).

Three information regimes. (i) *Simulation benchmark*: greedy, beam and repair call the true-model Fisher oracle—where headline scoring lives. (ii) *Nominal-model deployment*: both the policy and a model-based greedy planner may use the supplied nominal descriptors and known θ_0; greedy computes $k|\mathcal{A}|$ nominal Fisher utilities, whereas the policy computes none. We report no wall-clock comparison. (iii) *True hardware*: unknown θ, finite shots, an estimation objective, and drift. The distribution-level policy results remain regime (ii) evaluated inside the regime (i) simulator. The hardware result below is narrower: one frozen adversarial instance tests the order mechanism, not policy generalization.

4 Preregistered on-device protocol

Using exact deferred measurement for fixed open-loop sequences, the preregistration fixed seed 429, greedy and zero-evaluation policy sequences (policy equaled the oracle), and a 90% bootstrap MSE-ratio criterion. Four completed, non-debiased, sequentially submitted `qpu.forte-enterprise-1` jobs used one job per arm and 6,000 shots per arm: greedy/policy ratio 18.1 (shot-resampling 90% CI $[8.9, 40.7]$), meeting the numerical criterion. The interval conditions on each arm's single job and does not estimate between-job or temporal-drift uncertainty. The k=2 case was excluded in advance for low power; a median pair was pre-demoted to exploratory. We retain the 16–18-hour smoke reuse and record-only day-rule NO-GO deviations. This is a conditional observation consistent with the mechanism, not policy generalization (Appendix A.24).

5 Related work

Greedy guarantees. Set-function results require monotonicity and submodularity (Nemhauser et al., 1978; Krause & Golovin, 2014); weak- and sequence-submodularity quantify structured departures (Das & Kempe, 2011; Bian et al., 2017; Asadpour et al., 2022; Tang & Yuan, 2024). Adaptive submodularity likewise requires non-destructive observations (Golovin & Krause, 2011). Our contribution is the explicit state-consuming counterexample, not order-dependence alone. **Amortized design.** Original DAD uses per-round design-conditioned likelihoods (Foster et al., 2021); iDAD allows history dependence (Ivanova et al., 2021), and Step-DAD adds online refinement (Hedman et al., 2025). Dynamic-system designs distinguish fixed parameters from controlled evolving states (Iqbal et al., 2024b); see also Iqbal et al. (2024a). Fixed θ therefore does not imply fixed physical state, nor does general BED exclude back-action. RL design (Blau et al., 2022) and expert iteration (Anthony et al., 2017) precede our amortization; we claim neither history dependence nor architectural novelty. **Quantum measurement.** Fresh-copy tomography (Huszár & Houlsby, 2012), back-action-aware phase estimation (Berry & Wiseman, 2000), and learned metrology (Fiderer et al., 2021; Belliardo et al., 2024b) are established. Belliardo et al. explicitly model quantum instruments and post-measurement updates (Belliardo et al., 2024a); finite-copy strategies can beat greedy (André et al., 2026). Our narrower contribution is the explicit discrete-sequence counterexample with unbounded regularized D-efficiency overhead, under the stated measurement resources, and its amortized evaluation. Table 4 separates these modeling assumptions; mid-circuit back-action and drift are further discussed in the supplement (Chu et al., 2026; Proctor et al., 2020; Nielsen et al., 2021).

6 Limitations and conclusion

Limitations. Instances are small (2–3 qubits, $k \leq 7$) and supporting cases are narrower; hardware is one adversarial device-day with sequential one-job arms and a job-conditional CI, not distribution-level validation. **Conclusion.** State consumption makes utility sequence-valued and defeats a uniform greedy guarantee. Amortization closes the random-Hermitian gap, but no oracle-free device advantage is established.

Ethics statement

This work involves no human subjects, personal data, or crowdsourcing. The device dataset consists of publicly served machine calibration summaries. We do not foresee a direct dual-use risk: the methods allocate measurement budgets for characterizing and operating quantum devices, and the released artifacts contain no privileged device information beyond the public feed.

Reproducibility statement

The anonymous supplementary archive uploaded with this submission contains the back-action testbed and seeded experiment scripts, frozen protocols and amendments, raw simulation results, hardware counts and calibration evidence, environment lockfile, and a number-checking gate. The new post-hoc device check is separately marked and preserves the original train/test split. Figures 2–4 use manually transcribed coordinates, and three historical claims lack their generating scripts; these gaps are listed in `UNEXECUTED.md`. *Compute.* The recorded experiments used Python 3.12, NumPy, SciPy, and scikit-learn, CPU-only with BLAS threads pinned per worker; the new 90-instance check ran locally with eight workers in 478 seconds. *Dataset licensing.* The archive contains 1,498 public vendor calibration records and a collection script, but their redistribution license has not been independently established. The records are provided for anonymous review; any public repository release will retain only the collection script and checksums unless redistribution permission is confirmed.

References

A. E. Allahverdyan, R. Balian, and Th. M. Nieuwenhuizen. Maximal work extraction from finite quantum systems. *Europhysics Letters*, 67:565, 2004. arXiv:cond-mat/0401574.

Erik L. André, Jessica Bavaresco, and Mohammad Mehboudi. Strategy optimization for Bayesian quantum parameter estimation with finite copies: Adaptive greedy, parallel, sequential, and general strategies. *Quantum Science and Technology*, 11(3), 2026. doi: 10.1088/2058-9565/ae846c. arXiv:2602.09655.

Thomas Anthony, Zheng Tian, and David Barber. Thinking fast and slow with deep learning and tree search. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. arXiv:1705.08439.

Andrew Arrasmith, Lukasz Cincio, Rolando D. Somma, and Patrick J. Coles. Operator sampling for shot-frugal optimization in variational algorithms. *arXiv preprint arXiv:2004.06252*, 2020.

Arash Asadpour, Rad Niazadeh, Amin Saberi, and Ali Shameli. Sequential submodular maximization and applications to ranking an assortment of products. *Operations Research*, 2022. Conference version appeared in ACM EC 2022.

Federico Belliardo, Fabio Zoratti, and Vittorio Giovannetti. Applications of model-aware reinforcement learning in Bayesian quantum metrology. *Physical Review A*, 109:062609, 2024a. arXiv:2403.05706.

Federico Belliardo, Fabio Zoratti, Florian Marquardt, and Vittorio Giovannetti. Model-aware reinforcement learning for high-performance Bayesian experimental design in quantum metrology. *Quantum*, 8:1555, 2024b. doi: 10.22331/q-2024-12-10-1555.

D. W. Berry and H. M. Wiseman. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Physical Review Letters*, 85:5098–5101, 2000. arXiv:quant-ph/0009117.

Andrew An Bian, Joachim M. Buhmann, Andreas Krause, and Sebastian Tschiatschek. Guarantees for greedy maximization of non-submodular functions with applications. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 498–507, 2017.

Tom Blau, Edwin V. Bonilla, Iadine Chades, and Amir Dezfouli. Optimizing sequential experimental design with deep reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2022. arXiv:2202.00821.

Francesco Campaioli, Stefano Gherardini, James Q. Quach, Marco Polini, and Gian Marcello Andolina. Colloquium: Quantum batteries. *Reviews of Modern Physics*, 96:031001, 2024. arXiv:2308.02277.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3), 1995. doi: 10.1214/ss/1177009939.

Chia-Tung Chu, Su-un Lee, Han Zheng, Senrui Chen, Bibek Pokharel, Alireza Seif, and Liang Jiang. Learning mid-circuit measurement backaction from three repeated measurements. *arXiv preprint arXiv:2606.00433*, 2026.

Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. *arXiv preprint arXiv:1102.3975*, 2011.

Chandler Davis and William M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):20–28, 1979. doi: 10.2307/2346806.

Dario Ferraro, Michele Campisi, Gian Marcello Andolina, Vittorio Pellegrini, and Marco Polini. High-power collective charging of a solid-state quantum battery. *Physical Review Letters*, 120: 117702, 2018.

Lukas J. Fiderer, Jonas Schuff, and Daniel Braun. Neural-network heuristics for adaptive Bayesian quantum estimation. *PRX Quantum*, 2(2):020303, 2021. doi: 10.1103/PRXQuantum.2.020303.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. In *International Conference on Machine Learning (ICML)*, 2021. arXiv:2103.02438.

Benoît Frénay and Michel Verleysen. Classification in the presence of label noise: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5):845–869, 2014. doi: 10.1109/TNNLS.2013.2292894.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011. arXiv:1003.3967.

Andi Gu, Angus Lowe, Pavel A. Dub, Patrick J. Coles, and Andrew Arrasmith. Adaptive shot allocation for fast convergence in variational quantum algorithms. *arXiv preprint arXiv:2108.10434*, 2021.

Ju-Yeon Gyhm, Dominik Šafránek, and Dario Rosa. Quantum charging advantage cannot be extensive without global operations. *Physical Review Letters*, 128:140501, 2022. arXiv:2108.02491.

Marcel Hedman, Desi R. Ivanova, Cong Guan, and Tom Rainforth. Step-DAD: Semi-amortized policy-based bayesian experimental design. In *International Conference on Machine Learning (ICML)*, volume 267 of *PMLR*, pp. 22904–22923, 2025. arXiv:2507.14057.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16:1050–1057, 2020. arXiv:2002.08953.

Ferenc Huszár and Neil M. T. Houlsby. Adaptive bayesian quantum tomography. *Physical Review A*, 85:052120, 2012. arXiv:1107.0895.

Panagiotis G. Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10)*, pp. 64–67. ACM, 2010. doi: 10.1145/1837885.1837906.

Sahel Iqbal, Hany Abdulsamad, Sara Pérez-Vieites, Simo Särkkä, and Adrien Corenflos. Recursive nested filtering for efficient amortized bayesian experimental design. *arXiv preprint arXiv:2409.05354*, 2024a. NeurIPS Bayesian Decision-making and Uncertainty workshop.

Sahel Iqbal, Adrien Corenflos, Simo Särkkä, and Hany Abdulsamad. Nesting particle filters for experimental design in dynamical systems. In *International Conference on Machine Learning (ICML)*, 2024b. arXiv:2402.07868.

Desi R. Ivanova, Adam Foster, Steven Kleinegesse, Michael U. Gutmann, and Tom Rainforth. Implicit deep adaptive design: Policy-based experimental design without likelihoods. *arXiv preprint arXiv:2111.02329*, 2021.

Andreas Krause and Daniel Golovin. Submodular function maximization. In Lucas Bordeaux, Youssef Hamadi, and Pushmeet Kohli (eds.), *Tractability: Practical Approaches to Hard Problems*, pp. 71–104. Cambridge University Press, 2014. doi: 10.1017/CBO9781139177801.004.

Jonas M. Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum*, 4:263, 2020.

D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956. doi: 10.1214/aoms/1177728069.

Klaus Mølmer and Anders Sørensen. Multiparticle entanglement of hot trapped ions. *Physical Review Letters*, 82:1835–1838, 1999. arXiv:quant-ph/9810040.

Charles Moussa, Max Hunter Gordon, Michal Baczyk, M. Cerezo, Lukasz Cincio, and Patrick J. Coles. Resource frugal optimizer for quantum machine learning. *Quantum Science and Technology*, 2023.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14(1):265–294, 1978. doi: 10.1007/BF01588971.

Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography. *Quantum*, 5:557, 2021. doi: 10.22331/q-2021-10-05-557.

Timothy Proctor, Melissa Revelle, Erik Nielsen, Kenneth Rudinger, Daniel Lobser, Peter Maunz, Robin Blume-Kohout, and Kevin Young. Detecting and tracking drift in quantum information processors. *Nature Communications*, 11:5396, 2020. doi: 10.1038/s41467-020-19074-4.

Tom Rainforth, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1), 2024. arXiv:2302.14545.

Elizabeth G. Ryan, Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. A review of modern computational algorithms for bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016. doi: 10.1111/insr.12107.

Abhay Shastry, Abhijith Jayakumar, Apoorva Patel, and Chiranjib Bhattacharyya. Shot-frugal and robust quantum kernel classifiers. *arXiv preprint arXiv:2210.06971*, 2022.

Victor S. Sheng, Foster Provost, and Panagiotis G. Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 614–622, 2008. doi: 10.1145/1401890.1401965.

Matthew J. Streeter and Daniel Golovin. An online algorithm for maximizing submodular functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2008.

Shaojie Tang and Jing Yuan. Non-monotone sequential submodular maximization. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012. arXiv:1004.4389.

Yudian Zheng, Guoliang Li, Yuanbing Li, Caihua Shan, and Reynold Cheng. Truth inference in crowdsourcing: Is the problem solved? *Proceedings of the VLDB Endowment*, 10(5):541–552, 2017. doi: 10.14778/3055540.3055547.

A APPENDIX

Table 3: Experimental protocol for the sequencing problem. All quantities are exact simulations; the analysis plan with hypothesis families was fixed before the confirmatory computations.

Instances	device-physics family: exchange + detuning, ZZ-crosstalk background, prep. error; random-Hermitian family as a distribution control
Actions	first 6 of time $\{0.6, 1.8, 4.0\}$ × basis $\{X, Z\}$ × qubit; $\lvert\mathcal{A}\rvert$=6 (8, 12: Table 1)
Utility	$\log\det(F + \lambda_{\text{reg}} I)$ of the joint sequence-outcome distribution (d=2)
Oracle	exhaustive over all $\lvert\mathcal{A}\rvert^k$ ordered sequences (exact)
Splits	policy: 100–200 train / 30–40 test per condition (§3), disjoint; CIs: seeds 5000–5199
Statistics	percentile bootstrap (10^4), Holm within declared families (released plan)

Table 4: **Modeling ledger.** Fixed inferential parameters do not imply an unchanged physical state. Expressive scope alone neither establishes an identical objective nor proves that a prior result subsumes our counterexample.

Line of work	Observation/state model	Distinction relevant here
Submodular design (Nemhauser et al., 1978; Krause & Golovin, 2014)	Utility is a set function	Greedy guarantee requires the stated premises
Adaptive tomography (Huszár & Houlsby, 2012)	Fresh i.i.d. copies	Next-basis selection, not within-copy sequencing
Original DAD (Foster et al., 2021)	Per-round design-conditioned likelihood	Its factorization is not a restriction on all BED
iDAD / Step-DAD (Ivanova et al., 2021; Hedman et al., 2025)	History-dependent likelihood	Step-DAD includes online refinement
Dynamic-system BED (Iqbal et al., 2024b); Iqbal et al. (2024a)	Fixed parameters; controlled Markov state	Requires the specified transition/observation model
Model-aware quantum RL (Belliardo et al., 2024b); Belliardo et al. (2024a)	Quantum instruments and state updates	Back-action already modeled; objectives/resources differ
Finite-copy strategies (André et al., 2026)	Finite quantum-copy resources	Per-instance strategy optimization
This work	Sequential, state-updating measurements	Explicit regularized D-efficiency counterexample; learned sequence evaluation

A.1 PER-CONDITION GREEDY PENALTIES

Median greedy regularized D-efficiency overheads with bootstrap 95% CIs (n=200 instances per condition, seeds 5000–5199; family-wise Holm $p \leq 6 \times 10^{-4}$ for all six): device-physics $1.303\,[1.184, 1.383]$ (k=3), $1.425\,[1.296, 1.536]$ (k=4), $1.495\,[1.387, 1.579]$ (k=5); random-matrix $1.451\,[1.345, 1.517]$, $1.602\,[1.529, 1.685]$, $1.734\,[1.644, 1.806]$. Fresh-copy control (k=3, n=50 per family): median $1.0000\,[1.0000, 1.0000]$, maximum 1.0002. *Horizon trend (exploratory)*: replacing the exact oracle by a beam-48 best-found proxy (which recovers the exact optimum at k=4 to within 1.00–1.03×), the median greedy overhead *lower bound* stays at a stable level through k=10: device-physics $1.39/1.50/1.35/1.41$ and random-matrix $1.50/1.57/1.60/1.59$ at k=4/6/8/10 (n=24 per point)—the pathology neither vanishes nor explodes with horizon; true gaps may exceed these bounds if proxy slack grows with k.

A.2 PROPOSITION 1 INSTANCES

The maximal-gap instances of Proposition 1 are generated by the released device-physics generator with seeds 159 (k=2, regularized D-efficiency overhead 3.30; fresh ratio 1.001) and 429 (k=3, 5.94; fresh ratio 0.19), action menu = the first six of the ordered product times $\{0.6, 1.8, 4.0\}$ × bases $\{X, Z\}$ × qubits (so t=4.0 is absent and t=1.8 acts on qubit 0 only), $\lvert\mathcal{A}\rvert$=6. Here A_1 =

$(XX+YY)/2$, $A_2 = (ZI-IZ)/2$, and $H_b = c_{ZZ}ZZ + c_X(XI+IX)/2 + c_0 ZI + c_1 IZ$. The artifact-derived certificate is:

seed	θ_0	(c_{ZZ}, c_X, c_0, c_1)	greedy / oracle sequence	F_g	F_*
159	$(1.399865, 0.927413)$	$(.252485, .194101, .046512, -.016475)$	$(3,5)/(5,1)$	$\begin{bmatrix} 1.285740 & -.097493 \\ -.097493 & .187599 \end{bmatrix}$	$\begin{bmatrix} 3.710434 & 1.735220 \\ 1.735220 & 1.491833 \end{bmatrix}$
429	$(1.358464, 0.205565)$	$(.188011, .060872, -.116542, .016690)$	$(5,1,2)/(5,0,2)$	$\begin{bmatrix} 14.212593 & .873636 \\ .873636 & .082189 \end{bmatrix}$	$\begin{bmatrix} 12.816963 & .818483 \\ .818483 & 1.166341 \end{bmatrix}$

Action indices map to (t, q, b) in the released menu; the exact ψ_0, matrices, actions, utilities, and full-precision values are in `proposition2_witnesses_20260920.json`, generated by `export_prop2_witnesses.py`. Thus Proposition 2 is an explicit computational certificate rather than a seed-only witness.

A.3 The analytic trap: numerical certificate

Metric scope. At fixed ridge, ρ_D is regularized D-efficiency overhead, not a literal replication count: $\det(rF + \lambda I) \neq r^d \det(F + \lambda I)$ in general. For example, with $F_* = \mathrm{diag}(100, 1)$, $F_g = \mathrm{diag}(10, 10)$, and $\lambda = 1$, the determinant-root ratio is 1.2921, whereas solving $\det(rF_g + I) = \det(F_* + I)$ gives $r = 1.3213$. The two interpretations coincide at $\lambda = 0$ and asymptotically when the relevant Fisher eigenvalues dominate the ridge, as in Proposition 1's $\omega \to 0$ limit.

Corollary (bounded duration). Among members of the family with total interrogation time $t_D + t_G = 3\pi/4\omega \leq T$, the closed form of Proposition 1 saturates, as $\epsilon \to 0$ at $\omega = 3\pi/4T$, at overhead $4T/(3\sqrt{5}\,\lambda^{1/2})$ (machine-verified to $< 1\%$ across $T \in [0.5, 10]$): any constant-factor guarantee must degrade at least linearly in the allowed duration and as $\lambda^{-1/2}$.

Proposition 1 is proved in the main text; `darkstate_trap_verify.py` checks every closed-form quantity against numerically computed Fisher matrices (agreement $\leq 4 \times 10^{-7}$ in all entries), verifies the uniform quantifier of Proposition 1 and its bounded-duration corollary directly (at fixed $\lambda = 10^{-10}$ and $\epsilon = 10^{-5}$, the ω=1 member saturates at 0.81 of $1/\sin\epsilon$ while $\omega = 10^{-2}$ attains it to four decimal places), and confirms the three robustness claims: the conclusion is unchanged for other ω, for a non-canonical second-action time (t_G at half the canonical value: greedy still takes the bait, $\rho_D = 3.97/11.89/39.61$ at $\epsilon = 0.30/0.10/0.03$), and under $\theta_0 = (\delta, 1)$ with $\delta \neq 0$, where the overhead *grows* before the mechanism finally disappears near $\delta \approx 0.5$ ($\rho_D = 10.02/11.12/2941$ at $\delta = 0/0.01/0.1$, ϵ=0.1). At the codebase's fixed $\lambda = 10^{-10}$ the closed form saturates at $\sqrt{\pi^2/5\lambda} \approx 1.4 \times 10^5$, which is where the $\lambda \to 0$ statement stops being observable numerically; at the same λ the bounded-duration corollary saturates at $\approx 5.96 \times 10^4\, T$.

A.4 A champion-derived analytic family (exploratory)

Reverse-engineering the k=2 maximal-gap instance yields a one-parameter analytic family on the *full* symmetric eight-action menu, in which a fringe measurement baits the greedy design while a dressed node hides the information: as $\varepsilon \to 0$ the measured greedy regularized D-efficiency overhead (actual greedy vs. exhaustive oracle) collapses onto a closed-form curve $S_1(s)$, within 2% at ε/s=0.01 across $s \in [0.02, 0.4]$ and within 19% over the whole validity region, with $s\,S_1(s) \to 1/(2\sqrt{3})$. The family is valid for $\varepsilon/s \lesssim 0.1$ (deviation $\lesssim 7\%$, growing at up to $\approx 0.65\,\varepsilon/s$, breaking down near $\varepsilon/s \approx 1$); an apparent second branch on a truncated six-action menu is a truncation artifact—the full menu contains the symmetric partner sequence and the branch disappears (utility gap $< 10^{-9}$). All quantities use finite differences of width $\varepsilon/100$ with the legacy probability cut-off disabled, since the information is carried by a rare outcome of probability $p_{\mathrm{rare}} \approx \varepsilon^2 c_x^2 s^2 q_f$ (measured-to- predicted ratio 0.84–0.98). The $1/s$ divergence of the gap is therefore not free: resolving it empirically requires $n \gtrsim 1/p_{\mathrm{rare}} = 3/(\varepsilon^2 s^2)$ shots (within 1.02–1.19× of measurement across the validity region), e.g. $\sim 10^9$ shots for an overhead of 5.9 at s=0.05—an in-principle rather than practical separation.

A.5 Submodularity-ratio diagnostic

Since U depends on the ordered sequence, the Das & Kempe (2011) ratio is computed on its permutation-maximal set-function relaxation $f(S) = \max_\pi U(\pi(S))$: γ is the largest constant satisfying $\sum_{a \in T \setminus S} [f(S \cup \{a\}) - f(S)] \geq \gamma\, [f(S \cup T) - f(S)]$ over enumerated subset pairs with

$|S|, |T| \le 2$ and $|S \cup T| \le 4$; $\gamma = 1$ recovers the weakly-submodular set-function case and back-action drives $\gamma < 1$. Measured on the $|\mathcal{A}|$=6 family (n=48 instances): $\gamma = 0.312\,[0.260, 0.366]$, while the fresh-copy control gives exactly $\gamma = 1.000$ on every instance. γ is a *family-level* diagnostic: per-instance γ does not rank per-instance greedy risk (Spearman with the regularized D-efficiency overhead -0.12, p=0.40 on this family; $+0.59$ on device-physics instances—the wrong direction; exploratory), so we do not propose γ-thresholded routing. The 1.6–2.3× spread quoted in the main text is best-vs-median over orderings of the same multiset; best-vs-worst is 3.1–8.6×. Both are regularized D-efficiency ratios $\exp(\Delta U/d)$, averaged over ten 3-subsets per instance and over instances, and both ranges span *all thirteen* settings—including the d=4, λ=10^{-2} corner that gives the minimum. An earlier draft quoted 4.4–8.6×, which silently dropped that corner. The best-vs-median statistic was not archived by the original sweep and was recomputed on 2026-08-10 with the settings and seeds recorded in the released script; it is therefore an independent recomputation rather than a replay.

A.6 Full statement and proof of Proposition 4

Noise model. The plug-in estimate is an average of m i.i.d. snapshots, $\hat{\rho}_m = \frac{1}{m}\sum_{s=1}^{m} X_s$, with X_s Hermitian, $\mathbb{E}X_s = \rho$, $\|X_s\| \le C_X$, $\operatorname{tr} X_s = 1$; classical shadows satisfy all three (Huang et al., 2020). Write $E_m = \hat{\rho}_m - \rho$ and $W(\rho) = \operatorname{tr}(\rho H_B) - \sum_i \lambda_i^{\downarrow}(\rho)\,\varepsilon_i^{\uparrow}$.

Assumptions. *(A1)* ρ has an eigenvalue r of multiplicity $\kappa \ge 2$ with eigenprojector P; all other eigenvalues are simple; the minimum isolation gap is $g > 0$; and $\operatorname{spec}(\rho) \subset [\delta, 1-\delta]$, $\delta > 0$. *(A2)* the passive pairing is non-constant on the block: with I the sorted positions occupied by the block, $\Delta\varepsilon := \max_{i\in I}\varepsilon_i - \min_{i\in I}\varepsilon_i > 0$. *(A3)* the traceless compressed snapshot noise $\tilde{B}_1 := P(X_1-\rho)P - \kappa^{-1}\operatorname{tr}[P(X_1-\rho)P]\,P$ is nondegenerate: $v := \mathbb{E}\|\tilde{B}_1\|_F^2 > 0$.

Proposition (Proposition 4, full form)**.** *Under (A1)–(A3) there are $c, C, m_0 > 0$, depending only on $(C_X, d, g, \delta, \kappa, v, \varepsilon_{\max})$, such that for all $m \ge m_0$*

$$\mathbb{E}\,W(\hat{\rho}_m) - W(\rho) \;\ge\; \frac{\Delta\varepsilon}{2\kappa^{3/2}}\sqrt{\frac{v}{4m}} \;-\; \frac{C}{m} \;\ge\; c\,m^{-1/2},$$

and the same bound holds for the PSD-projected plug-in.

Proof. Let $B_m := PE_mP$ (a $\kappa\times\kappa$ Hermitian on $\operatorname{ran} P$) and $\tilde{B}_m := B_m - \kappa^{-1}(\operatorname{tr} B_m)P$, and let $G := \{\|E_m\| \le g/4\}$; matrix Bernstein (Tropp, 2012) gives $\Pr(G^c) \le 2d\,e^{-c_1 m}$.

Step 1 (block reduction). On G, Weyl's inequality keeps the κ perturbed block eigenvalues separated from the rest, and the Davis–Kahan $\sin\Theta$ theorem (Davis & Kahan, 1970) bounds the perturbed block projector by $\|\hat{P} - P\| \le 2\|E_m\|/g$; comparing the compressions $\hat{P}\hat{\rho}_m\hat{P}$ and $rP + B_m$ and applying Weyl on $\operatorname{ran} P$ yields, for the sorted positions $i_j \in I$, $\big|\lambda_{i_j}^{\downarrow}(\hat{\rho}_m) - r - \lambda_j^{\downarrow}(B_m)\big| \le 8\|E_m\|^2/g$, and for each simple outside eigenvalue $\big|\lambda_i^{\downarrow}(\hat{\rho}_m) - \lambda_i^{\downarrow}(\rho) - u_i^{\dagger}E_m u_i\big| \le 8\|E_m\|^2/g$.

Step 2 (only the block survives in expectation). $W(\hat{\rho}_m) - W(\rho) = \operatorname{tr}(E_m H_B) - \sum_i\big[\lambda_i^{\downarrow}(\hat{\rho}_m) - \lambda_i^{\downarrow}(\rho)\big]\varepsilon_i^{\uparrow}$. The linear term is exactly unbiased. By Step 1 the outside first-order terms $u_i^{\dagger}E_m u_i$ are centered, the trace part of the block contributes $\kappa^{-1}(\operatorname{tr} B_m)\sum_{i\in I}\varepsilon_i$ (also centered), and all second-order remainders obey $\mathbb{E}\|E_m\|^2 \le \mathbb{E}\|E_m\|_F^2 \le d^2C_X^2/m$. Hence

$$\mathbb{E}\,W(\hat{\rho}_m) - W(\rho) \;=\; -\,\mathbb{E}\Big[\textstyle\sum_{j=1}^{\kappa}\lambda_j^{\downarrow}(\tilde{B}_m)\,\varepsilon_{i_j}\Big] \;+\; O\big(m^{-1}g^{-1}\big) + O\big(e^{-c_1 m}\big).$$

Step 3 (pairing bound, per realization). For reals $a_1 \ge \cdots \ge a_\kappa$ with $\sum_j a_j = 0$ and $b_1 \le \cdots \le b_\kappa$, the Chebyshev sum identity $\kappa\sum_j a_j b_j = -\sum_{i<j}(a_i - a_j)(b_j - b_i)$ gives $\sum_j a_j b_j \le -\kappa^{-1}(a_1 - a_\kappa)(b_\kappa - b_1)$. Applied to $a = \lambda^{\downarrow}(\tilde{B}_m)$ (traceless) and $b = \varepsilon_{i\cdot}$: $\sum_j \lambda_j^{\downarrow}(\tilde{B}_m)\varepsilon_{i_j} \le -\kappa^{-1}\big[\lambda_1^{\downarrow}(\tilde{B}_m) - \lambda_\kappa^{\downarrow}(\tilde{B}_m)\big]\Delta\varepsilon$.

Step 4 (the spread does not vanish in expectation). For traceless Hermitian T, $\lambda_1^{\downarrow}(T) - \lambda_\kappa^{\downarrow}(T) \ge \max(\lambda_1, -\lambda_\kappa) \ge \|T\|_F/\sqrt{\kappa}$. For $Y = \|\tilde{B}_m\|_F$, Hölder's inequality $\mathbb{E}Y^2 \le (\mathbb{E}Y)^{2/3}(\mathbb{E}Y^4)^{1/3}$ gives $\mathbb{E}Y \ge (\mathbb{E}Y^2)^{3/2}(\mathbb{E}Y^4)^{-1/2}$, and the i.i.d. fourth-moment expansion gives $\mathbb{E}Y^4 \le (3 +$

$C_0/m)(\mathbb{E}Y^2)^2$ with $\mathbb{E}Y^2 = v/m$. Hence $\mathbb{E}Y \geq \sqrt{v/(4m)}$ for $m \geq C_0$. Combining Steps 2–4 (the indicator corrections on G^c are exponentially small since $|\lambda_1 - \lambda_\kappa| \leq 4C_X$) proves the display; the second inequality holds for $m \geq m_0$.

PSD projection. On $G' := \{\|E_m\| \leq \delta/2\}$, $\mathrm{spec}(\hat{\rho}_m) \subset [\delta/2, 1-\delta/2]$ and $\mathrm{tr}\,\hat{\rho}_m = 1$, so eigenvalue clipping and trace renormalization both act as the identity: $\Pi(\hat{\rho}_m) = \hat{\rho}_m$. Since W is bounded on the relevant set and $\Pr(G'^c) \leq 2d\, e^{-c_2 m\delta^2}$, the bias changes by an exponentially small amount. □

Near-degenerate corollary. If the block eigenvalues are $r + s_i$ with spread $s := \max_i |s_i|$ rather than exactly equal, Weyl applied to the compressed matrix changes each sorted block eigenvalue by at most s, hence the main term by at most $\kappa s\, \varepsilon_{\max}$: for $s \leq c_3\, m^{-1/2}$ with $c_3 = \Delta\varepsilon\sqrt{v}\,/\,(8\kappa^{5/2}\varepsilon_{\max})$ the bound survives with halved constant. This is the "near-degenerate set" of Proposition 4; a direct simulation of this exact setting (degenerate 4×4 spectrum, trace-preserving Gaussian snapshots) confirms the sign, the $m^{-1/2}$ rate, and the leading constant ($\mathrm{bias} \cdot \sqrt{m} \to 1.20$ at $m = 2.6 \times 10^5$ against the predicted block value $\mathbb{E}\lambda_1^{\downarrow}(\tilde{B}) = 1.13$, the residual matching the $O(m^{-1}g^{-1})$ term).

A.7 Adaptivity control (exploratory)

The closed-loop oracle maximizes the log-determinant of the Fisher information of the *joint* outcome distribution over all depth-k decision trees (root action, then outcome-conditioned actions); restricted to constant trees this reduces exactly to the open-loop utility, so the comparison is apples-to-apples. Exhaustive tree search is tractable at $k \leq 3$ ($6 \cdot 216^2 \approx 2.8 \times 10^5$ joint evaluations via a branch-decomposition). Medians over n=200 instances per condition (bootstrap 95% CIs): device-physics k=3 — open-loop greedy overhead 1.30 [1.18, 1.38], adaptive greedy-to-oracle overhead 1.30 [1.23, 1.42], adaptivity headroom over the open-loop optimum 1.10 [1.10, 1.12]; random-matrix k=3 — 1.45 [1.34, 1.52], 1.50 [1.43, 1.61], 1.15 [1.13, 1.17], with paired (adaptive−open) gap difference +0.08 [+0.03, +0.14]. At k=2 the headroom medians are 1.00 and 1.10. The preregistered k=3 maximal-gap instance keeps a 4.99× gap under the adaptive oracle (open-loop 5.94×; headroom 1.05); the k=2 instance is unchanged at 3.30× with headroom exactly 1.00. These runs postdate the frozen analysis plan and are labeled exploratory.

A.8 Utility robustness: expected information gain (exploratory)

A natural objection is that local D-optimality is a fragile objective and that a global Bayesian utility would spare greedy. We repeated the greedy-vs-exhaustive comparison with the expected information gain $\mathrm{EIG} = I(Y;\Theta)$ under a uniform prior on the ±0.35 operating box (9 × 9 grid; k=3, n=50 per family). On random-matrix instances the pathology persists: greedy attains a median 0.87 (worst 0.53) of the optimal EIG and misses the EIG-optimal sequence on 84% of instances, while the fresh-copy control is exactly optimal throughout—so the failure is back-action, not the choice of utility. On device-physics instances at k=3, global-EIG greedy usually escapes (median efficiency 1.00, miss rate 6%), and the preregistered k=3 maximal-gap instance's trap is specific to the local objective: its EIG-greedy finds an EIG-optimal sequence (the k=2 instance retains a 4.5% EIG deficit). The practical reading: the *local-calibration* regime—precisely the fine-tuning setting our on-device protocol optimizes and scores by MSE—is where greedy is most exposed, and global-EIG design does not transfer a guarantee to it.

A.9 Out-of-distribution policy evaluation (exploratory)

We expose a policy trained once under a fixed convention to three distribution shifts with no retraining (n=30 per condition; oracle and greedy recomputed per condition; ranges over 10 training configurations—two seed blocks × five forest seeds—reported alongside medians; absolute values here use a different action convention from Table 1 and are not cross-citable). Shifting the operating-point prior by ±50% leaves policy+repair at median overheads 1.000 and 1.11 (range 1.07–1.14), but the oracle-reach rate drops from 70–80% in-distribution to 33–40% and the pre-repair policy median degrades to 1.46. Swapping the action menu produces the same grade of degradation (raw 1.46, policy+repair 1.10, reach 30%). Swapping the instance family degrades the median to 1.07–1.33, repair does not recover it, and the difference from greedy is not significant under a paired sign test (p=0.078). Zero-shot robustness therefore holds only in the sense that policy+repair stays at

or below that condition's greedy; the policy itself degrades significantly under operating-point and menu shifts, and a family change requires retraining.

A.10 BACK-ACTION DIAL (EXPLORATORY)

Each projective readout is replaced by a weak measurement with Kraus operators $K_{\pm}(\kappa) = \sqrt{(1 \pm \kappa)/2}\, P_{\pm} + \sqrt{(1 \mp \kappa)/2}\, P_{\mp}$, so κ=1 is the projective case (verified to reproduce it exactly) and $\kappa \to 0$ removes both disturbance and information. Median greedy overheads at k=4, n=48 per point: device-physics $1.000/1.000/1.000/1.000/1.06/1.25$ for $\kappa = 0.1/0.25/0.5/0.75/0.9/1.0$—the gap switches on only near projective strength—while the fresh control is 1.000 at every κ. Random-matrix instances retain a κ-insensitive residual (1.13–1.19) rising to 1.54 at κ=1: because evolution times accumulate along a sequence, ordering matters there even at weak coupling (a scheduling effect distinct from back-action), and the back-action-driven rise sits on top of it.

A.11 TARGET SUPPORTING CASE: PRIMARY-CELL TESTS

Primary cell (n_{MS}=12, R/T=25%), paired over 20 MC seeds, Holm $p \leq 3 \times 10^{-4}$ in both families. Stale-info gain (threshold−uniform): forte-1 $+0.0117\,[+0.0096, +0.0140]$, aria-1 $+0.0189\,[+0.0161, +0.0216]$, forte-ent. $+0.0132\,[+0.0120, +0.0144]$; against oracle-minus-uniform gains of $+0.0367$, $+0.0644$, $+0.0257$ these recover 32.0%, 29.3%, 51.4% of the oracle gap. Probe loss (probe$_{50}$−threshold): $-0.0542\,[-0.0550, -0.0534]$, $-0.0571\,[-0.0583, -0.0559]$, $-0.0523\,[-0.0532, -0.0515]$.

A.12 TEMPORAL ROBUSTNESS OF THE TARGET CASE (EXPLORATORY)

The device-day series are temporally correlated and non-stationary (lag-1 autocorrelation of e_2: 0.16–0.38; forte-1 lag-3 reaches 0.51), so we complement the declared seed-resampling CIs in two ways that respect time order. As a diagnostic, a naive i.i.d. day-level bootstrap is *invalid* for this estimand: it destroys the day-to-day persistence the policy exploits and collapses the gain to zero ($P(g{>}0) \approx 0.52$). Our primary temporal uncertainty is therefore (i) time-ordered segment t-CIs—forte-1 $+0.0127\,[+0.0045, +0.0208]$, aria-1 $+0.0197\,[-0.0072, +0.0466]$, forte-ent. $+0.0134\,[+0.0037, +0.0231]$—and (ii) a strict $60/40$ temporal holdout with the threshold frozen on the first 60% of each history: the gain stays positive on all backends ($+0.0099\,[+0.0034, +0.0187]$, $+0.0150\,[-0.0001, +0.0297]$, $+0.0110\,[+0.0051, +0.0153]$; pair-preserving block bootstrap), with aria-1 borderline at the 95% level ($P(g{>}0) = 0.974$). The gain is not an early-budget-exhaustion artifact (it also beats a first-R-days policy by $+0.008$–$+0.034$ and a uniform policy restricted to the same used window by $+0.011$–$+0.021$). Moving-block bootstrap percentile CIs (weekly and monthly blocks) also exclude zero on all backends but carry a replicate-mean bias of 36–58% on forte-1 and aria-1 (about 3% on forte-enterprise-1) of the point estimate under non-stationarity, so—as a declared deviation from the percentile-bootstrap convention of the analysis plan—we do not use them as the headline uncertainty. The frozen threshold's test-period execution rate rises to 0.38–0.51 against the 0.25 design (two-qubit error improves over the record): the recorded histories are non-stationary, and all conclusions are conditional on them.

A.13 A BREAK-EVEN MAP FOR PAID RE-CHARACTERIZATION (EXPLORATORY)

To upgrade the budget-dependence caveat of Appendix A.16 into an explicit applicability boundary, we sweep the daily budget $D \in \{2{\times}10^2, \ldots, 1.25{\times}10^5\}$ shots, the probe price ratio $S_c/D \in \{1\%, 5\%, 20\%\}$, and the workload depth $n_{\mathrm{MS}} \in \{6, \ldots, 96\}$ over all three recorded histories (675 cells, 20 seeds each, Holm-corrected across the full grid). A break-even boundary exists but is narrow: paying 20% of the budget never wins ($0/225$ cells); a 5% price wins only at $D \geq 2.5 \times 10^4$ with deep workloads ($n_{\mathrm{MS}} \geq 48$; 13 cells); at a 1% price the log-interpolated crossing sits at $D^* \approx 3.6 \times 10^4$ (median over deep-workload cells). On the lowest-drift backend (forte-enterprise-1) paying never wins anywhere in the grid, and at the paper's operating budget (D=10^3) $133/135$ cells lose significantly. The boundary is priced, not informational: with the price removed, the decision value of a fresh estimate is significantly positive in $306/675$ cells—paying fails because the price and estimation noise dominate, not because the information is worthless.

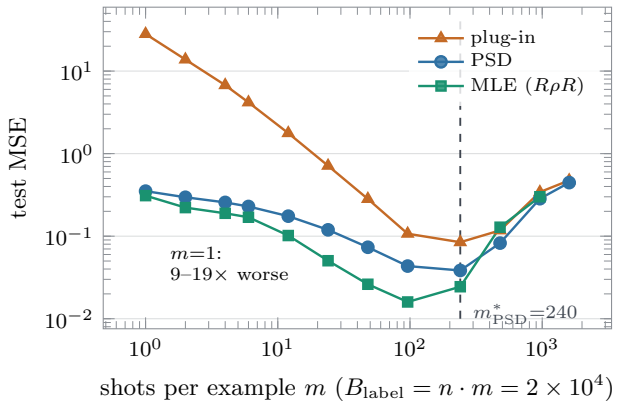


Figure 4: Width supporting case: test MSE vs. shots-per-example m at fixed budget B_{label}=2×10^4. Every physicality-constrained estimator (PSD, MLE) preserves the interior optimum and the two-sided U; corrections change constants, not structure.

Table 5: Width supporting case: experimental protocol, in the same form as Table 3. The empirical plug-in label is the only nonlinear ingredient; everything else is held fixed across the m sweep so that the n vs. m trade-off is the sole moving part.

Instances	2-qubit battery driven by collective excitation exchange from a 2-qubit charger; $\theta = (t, \phi) \in [0.2, 4.0] \times [0, \pi]$; battery spectrum $\{0, 1, 1, 2\}$ (degenerate)
Label	ergotropy $W(\rho)$ from classical shadows — a nonlinear, sorted-eigenvalue functional
Estimators	plug-in, PSD projection, Richardson, PSD+Richardson, and full MLE ($R\rho R$)
Learner	Nyström random features (300 components) + ridge (α=10^{-3})
Splits	4×10^4-instance pool; 400 held-out test instances, fixed across all m
Budgets	$B_{\mathrm{label}} \in \{2 \times 10^3,\ 2 \times 10^4,\ 2 \times 10^5\}$; m grids in Appendix A.15; 10 seeds
Statistics	percentile bootstrap, Holm within the declared six-test family (left/right of m^* per budget)

A.14 Supporting case A — width: the $n \times m$ split

The policies of §3 are trained against exactly computed utilities—an exhaustive teacher, or the policy's own repaired rollouts scored by the exact oracle. Once a surrogate in this pipeline is trained on the measured empirical-mean or MLE labels studied here instead, a second allocation question appears. Splitting the label budget $B_{\mathrm{label}} = n \cdot m$ (examples × shots-per-example), in the repeated-labeling literature the choice is a labeler-quality/data-cost trade-off (Sheng et al., 2008); with cheap examples and noise that averages out it favours m=1. For *linear* labels ($\langle H \rangle$ from classical shadows (Huang et al., 2020)) we confirm the prescription: test MSE is flat in m. For the *nonlinear* label we care about—ergotropy $W(\rho) = \mathrm{tr}(\rho H) - \min_U \mathrm{tr}(U\rho U^\dagger H)$, the maximal unitarily extractable work (Allahverdyan et al., 2004), a sorted-eigenvalue functional—it fails by two orders of magnitude: at B_{label}=2×10^4 the m=1 split is 354× worse than the interior optimum m^*=240 (paired geometric mean 352, bootstrap 95% CI $[285, 432]$; an uncorrected point estimate outside the declared six-test family). All estimates are paired over seeds, and every interval below is a 95% bootstrap CI, Holm-corrected within the declared six-test family.

The strongest standard fix does not restore $m = 1$. Projecting the shadow estimate onto the PSD cone before evaluating the functional removes the *global* bias almost entirely (mean bias ≈ 0 even at m=1), yet m=1 still loses: the paired geometric-mean ratio is 9.3 $[8.2, 10.3]$ (Figure 4) at B_{label}=2×10^4 (1.9 $[1.7, 2.2]$ at 2×10^3; 34 $[30, 39]$ from the smallest grid point at 2×10^5)—and the interior optimum is two-sided: the largest-m grid point loses by 2.1–68× (all six tests Holm $p \leq 6 \times 10^{-4}$). The optimum and its scaling are seed-stable ($m^* = 24/240/960$; bootstrap re-selection concentrates on the same m^*). Richardson extrapolation and median-of-means behave the same or worse, and full maximum-likelihood ($R\rho R$) tomography does so empirically: the MLE labels still lose 19× at m=1 (interior m^*=96 at B_{label}=2×10^4), with the better estimator shifting m^* *down*, exactly as the bias–variance account predicts. The mechanism:

Proposition 4 (θ-conditional bias)**.** *For the empirical-mean estimator $\hat{\rho}_m = m^{-1}\sum_{s=1}^{m} X_s$ of i.i.d. bounded, unbiased, trace-one Hermitian snapshots satisfying (A1)–(A3), the sorted-eigenvalue plug-in (and its PSD projection) has a positive local θ-conditional ergotropy gap $cm^{-1/2} - O(m^{-1})$ at exact degeneracy or in the stated local neighborhood with spectral spread $s \le c_3 m^{-1/2}$; this is not a fixed nonzero-gap claim. The sorted eigenvalues of a centered random perturbation have nonzero mean at first order, the passive-energy pairing transmits it, and projection acts as the identity with high probability on interior spectra. (Full statement—snapshot noise model, isolation and nondegeneracy assumptions—and proof in Appendix A.6.)*

Empirically the mechanism is confirmed quantitatively: the measured squared conditional bias at $m{=}1$ (0.336) is consistent with the learner's MSE floor ($0.352{\pm}0.002$ across seeds); conditional bias correlates with the local spectral gap (Spearman -0.53 to -0.62); and in the deep-degenerate regime $|\mathrm{bias}| \cdot \sqrt{m}$ is an $O(1)$ constant. A learner fits $\mathbb{E}[y|\theta]$, so what matters is exactly this conditional bias—global unbiasedness is the wrong target. The resulting allocation law, $m^* \propto \sqrt{cB_{\mathrm{label}}/a}$ from $\mathrm{MSE}(m) \approx c/m + am/B_{\mathrm{label}}$, matches the measured scaling $m^* \propto B_{\mathrm{label}}^{0.4\text{–}0.6}$.

Related literature for the width case. **Shot allocation in variational algorithms.** Parallel work allocates shots *within* one optimization trajectory—across gradient components or Hamiltonian terms (Kübler et al., 2020; Arrasmith et al., 2020; Gu et al., 2021; Moussa et al., 2023). Quantum-kernel work instead reduces repeated measurements needed for reliable classification (Shastry et al., 2022). Our width case allocates orthogonally—*how many* training examples versus *how precisely* each is labeled—the failure driven by the label's nonlinearity, not gradient noise. **Repeated labeling.** The n versus m question is classical in crowdsourcing (Sheng et al., 2008; Ipeirotis et al., 2010). Label-aggregation methods include worker-bias and truth-conditional error models (Dawid & Skene, 1979; Zheng et al., 2017), and the label-noise literature surveys broader corruption mechanisms (Frénay & Verleysen, 2014). Our setting instead studies θ-conditional bias generated by nonlinear plug-in labels from repeated quantum measurements. **Estimation of nonlinear functionals; workload physics.** Classical shadows (Huang et al., 2020) give unbiased linear estimates; our width case concerns what their nonlinear plug-ins do to *learning*; workload and instance physics follow the collective charging literature (Ferraro et al., 2018; Campaioli et al., 2024; Gyhm et al., 2022; Mølmer & Sørensen, 1999).

A.15 Width supporting case: full estimator grids

The PSD estimator eigendecomposes the linear-inversion (shadow) estimate, clips negative eigenvalues at zero, and renormalizes to unit trace before evaluating the label functional. Six-test family (left/right of m^* at each budget, PSD estimator, paired over 10 seeds), geometric-mean ratios with 95% CIs: $B_{\mathrm{label}}{=}2\times10^3$: $1.92\,[1.65, 2.22]$ and $2.11\,[1.79, 2.43]$; $B_{\mathrm{label}}{=}2\times10^4$: $9.31\,[8.23, 10.33]$ and $11.28\,[8.83, 14.20]$; $B_{\mathrm{label}}{=}2\times10^5$: $34.2\,[29.9, 39.0]$ and $67.6\,[55.9, 82.4]$ (all Holm $p \le 6\times10^{-4}$). MLE ($R\rho R$, 50 iterations) at $B_{\mathrm{label}}{=}2\times10^4$: $m^*{=}96$, $m{=}1$ ratio $19\times$.

A.16 Supporting case B — target: the price of device information

Any workload that will actually run on hardware—including the on-device experiment of §4, whose queue day is chosen by the rule derived here—faces the last allocation: spend today's shots *characterizing* the device or *executing* on it? We price this decision on 1,498 device-days of public IonQ calibration records (three backends—forte-1, aria-1, forte-enterprise-1—covering 2021–2026; 1,463 enter the analysis, the 35 records lacking a two-qubit median being dropped; 6 spike days are flagged but retained, with the list released). The recorded two-qubit gate error p_t on day t varies by $81.6\times$ over the record, with median day-over-day relative change 18% on forte-1 (14–16% on the others).

For a workload of n_{MS} two-qubit (Mølmer–Sørensen) gates, hence fidelity $(1-p_t)^{n_{\mathrm{MS}}}$, and a queue budget of R run-days out of T: uniform day selection incurs a 2–15% fidelity regret against the *oracle* day choice (larger for heavier workloads and noisier backends); a threshold rule using only *yesterday's free public* calibration recovers 29–51% of that gap (paired gain over uniform $+0.012$ to $+0.019$ fidelity across the three backends, against oracle gains of $+0.026$ to $+0.064$ on the same primary cell, all bootstrap CIs excluding zero, Holm $p \le 3\times10^{-4}$); and paying $S_c \in \{50, 200\}$ of 1,000 daily shots for a fresh estimate *loses at this operating budget* (probe minus stale-rule: -0.052

to -0.057, CIs excluding zero, Holm $p \leq 3 \times 10^{-4}$), because day-over-day persistence caps the value of freshness ($\sim 1.5\%$ fidelity) below its shot price at these budgets (the trade-off point scales with the daily budget). We also report an honest null: across the grid, the noise-optimal circuit size never changes: writing $\Gamma(N, p)$ for the simulated collective-charging advantage of an N-qubit workload at two-qubit error p, Γ is monotone in N over the tested grid ($N \leq 5$, p spanning the recorded range), so its maximizer never leaves the grid edge and $\arg\max_N \Gamma(N, p)$ is constant in p: "characterize to choose N" has no decision value in this family, and we report the null rather than a tuned N. A 675-cell break-even map (Appendix A.13) bounds the narrow larger-budget regime where paying could win; a temporal-holdout analysis (Appendix A.12) confirms the stale-information gain out of sample.

A.17 Attribution: policy initialization vs. the repair sweep

Because the headline arm is policy+repair, one can ask whether the repair sweep alone is responsible. On the same test instances (identical seeds) we ran the sweep from two other starts: greedy's own sequence, and a uniformly random sequence, each with the same first-improvement budget n_g. Medians (greedy / greedy+repair / random+repair / policy+repair): real k=3: $1.29/1.12/1.14/1.01$; real k=4: $1.24/1.14/1.22/1.13$; real k=5: $1.22/1.16/1.19/1.11$; abstract k=3: $1.58/1.27/1.43/1.08$; abstract k=4: $1.66/1.23/1.34/1.09$; abstract k=5: $1.70/1.28/1.37/1.04$. Note the accounting: greedy+repair spends $2n_g$ evaluations (its own n_g selection plus the sweep)—twice the policy arm's bill—and still loses to policy+repair in five of six conditions (paired Wilcoxon on the log regularized D-efficiency overhead, Holm $p = 0.001/0.002/0.033/0.033/0.038$; real k=4: $p = 0.31$), winning 19–24 of 30 paired instances. A random start with the same sweep is worse than greedy's start everywhere. The sweep is a local operator; where it lands depends on where it starts, and the policy provides the start.

A.18 Architecture ablation: forest vs. deep amortized policies

A fair question is whether a deep history-conditioned policy in the DAD family would change the picture. We preregistered the comparison (design frozen before execution): identical behavioral-cloning teachers, train/test seeds, inputs, and zero-evaluation deployment for three policy classes—the random forest of the main text, an MLP (256×128) on the same flat features, and a GRU history encoder consuming the action sequence (all recipes fixed in advance; *no* architecture was tuned, including the forest). Bare-rollout medians (greedy / forest / MLP / GRU): real k=3: $1.29/1.38/1.13/1.44$; real k=4: $1.24/1.64/1.32/7.04$; abstract k=3: $1.58/1.51/1.69/1.33$; abstract k=4: $1.66/1.33/1.60/2.33$. In eight paired Wilcoxon comparisons (deep vs. forest, Holm-corrected), no deep variant wins significantly; the three significant differences (GRU at both k=4 conditions, MLP at abstract k=4) all favor the forest. The GRU's k=4 failure is plausibly a training-budget artifact rather than an architectural verdict—we report it as observed under the frozen recipe. One exploratory direction (not a claim): the MLP's real k=3 median 1.13 sits below both forest and greedy (raw $p = 0.025$, Holm 0.125). The conclusion we do draw is narrow: on these problem sizes, the amortization advantage of §3 is not produced by policy-class capacity.

A.19 Learned-surrogate repair: preregistered attempt, negative

The exact-utility repair of §3 is the one deployment-time oracle left in the pipeline. We preregistered its most direct replacement: a random-forest surrogate $\hat{g}(\text{instance}, \text{sequence})$ trained only on training instances (where the oracle is legitimately available), used at deployment to adjudicate the same repair sweep with zero utility evaluations. The first pilot failed its bar (the sweep wandered: surrogate pairwise ranking accuracy on one-edit neighbours of policy rollouts was 67.9%), and a single preregistered amendment—training samples targeted at the query distribution (all one-edit neighbours of optimal, greedy, and random anchor sequences) plus an out-of-bag acceptance margin—improved every number (device k=3 median $1.398 \to 1.136$ vs. greedy's 1.287, now winning $19/30$ paired instances; k=4 $1.473 \to 1.393$ vs. 1.242) but cleared the frozen significance bar in neither condition (Holm $p = 0.45$ and 0.26). Per the frozen protocol, sequential testing closed with a negative result. A later check, explicitly frozen as post-hoc before execution, kept the model and 140 training seeds fixed and evaluated k=3 on the previously evaluated 90-instance suffix of the prior held-out extension. Learned repair improved the bare policy (1.150 vs. 1.325; $28/55/7$

wins/ties/losses, $p = 5.3 \times 10^{-6}$) but did not significantly beat greedy (1.243; $51/1/38$, $p = 0.105$). Exact utility was used only for offline scoring; deployment used the learned surrogate. This new check is not confirmatory and leaves the oracle-free device advantage unestablished.

A.20 Feature-noise robustness: nominal inputs suffice

Table 2 lists the policy's inputs as nominal at deployment while training uses exact values; a fair objection is that the exact values might be doing hidden work. We test the deployment-mismatch scenario directly: the policy receives a perturbed nominal model—each generator A_j, background H_b, and ψ_0 shifted by relative Hermitian noise of size σ—while dynamics, oracle, and evaluation use the true instance. No retraining: all simulation tests in this subsection condition on the single frozen trained forest and do not estimate variation over training replicates. At $\sigma \in \{1\%, 5\%, 10\%\}$, on real k=3, real k=4, and abstract k=4 (n=30 each), the paired median shift of the bare policy's regularized D-efficiency overhead relative to σ=0 is 0.000 with bootstrap 95% CI exactly $[0, 0]$ in seven of nine cells—the perturbed policy usually emits the *same sequence*—and the worst cell (real k=4, σ=10%) has CI $[0.000, +0.045]$ (per-cell 95% CIs; a Bonferroni-simultaneous bound over the nine cells reaches $+0.33$), and the remaining non-degenerate cell (real k=4, σ=5%) has CI $[-0.024, 0.000]$; no comparison is significant (Wilcoxon with zero-differences discarded, effective n=1–20; exact-test recomputation, all $p \geq 0.15$; normal-approximation $p \geq 0.13$). On the abstract family the medians are $1.33/1.33/1.29/1.33$ across $\sigma = 0/0.01/0.05/0.10$ against greedy's 1.66. Two scope caveats: absence of a detected effect is not formal equivalence (no preregistered margin was set), and the perturbations are independent Hermitian noise, not correlated calibration drift. Within those limits, the abstract-family advantage of §3 survives 10% input error, and the device-physics gap neither opens nor closes under perturbation.

We also reran greedy on the *same perturbed nominal model* supplied to the policy, then scored both sequences in the unchanged true simulator. Nominal greedy makes $k|\mathcal{A}|$ model-utility calls per instance. Its median overheads at $\sigma = 0/.01/.05/.10$ are $1.287/1.287/1.263/1.401$ (real k=3), $1.242/1.255/1.319/1.387$ (real k=4), and $1.661/1.602/1.738/1.870$ (abstract k=4), versus frozen-policy medians $1.379/1.515/1.495/1.366$, $1.641/1.816/1.489/1.653$, and $1.328/1.328/1.290/1.333$, respectively (n=30 each). These bounded comparisons establish neither a universal winner nor a speed benefit. They retract the exclusive nominal-deployability reading: the policy's measured distinction is zero per-instance utility queries, while nominal greedy is a computational-query alternative.

A.21 A disturbance-aware policy term: preregistered ablation, negative

A natural next step is a policy that scores each candidate action not only by its immediate information gain but by the information channel it would destroy. We preregistered exactly this (design frozen before execution, released with the artifacts): per-step, per-action features on the nonselectively propagated average state ρ_j—the one-step Fisher diagonal g (myopic gain), the sensitivity-survival ratio $r_{aj'} = \|[A_{j'}, \Phi_a(\rho_j)]\|_F / \|[A_{j'}, \rho_j]\|_F$, and a measurement–generator commutator norm—all computable from nominal quantities, so the policy's deployment status (Table 2) is unchanged. The ablation had three arms: baseline features (A), $+[g, r, c]$ (B), and $+[g]$ only (C), the last isolating "disturbance-awareness" from "action descriptors help." On the device-physics family at k=3 a preregistered confirmatory replication (fresh n=100 test instances, deterministic retraining) gave: B improves on the baseline (median 1.09 vs. 1.28, Holm $p = 0.016$)—but B vs. C did not clear the frozen bar (Holm $p = 0.093$), and at k=4 B and C were indistinguishable ($p = 0.82$). Per the frozen decision rule we therefore claim no disturbance-aware mechanism: the improvement is attributable to per-action descriptors generally, not to the destruction-prediction term specifically. We report this because the negative is informative—the bare-policy gap on the device family (§3) is real but is not closed by the most physically motivated feature we could construct for it.

A.22 On-device probe, smoke gate, and preregistered run

Before the preregistered run, the frozen protocol requires a 500-shot billing probe (the compiled champion-greedy circuit) to measure the actual per-unit price. We executed this probe on `qpu.forte-enterprise-1` (day calibration: two-qubit error 0.84%, inside the frozen 1.6% envelope). Beyond its billing purpose the probe is a hardware smoke test of the compiled measure-

ment and retrieval pipeline: the observed 500-shot outcome distribution matches the frozen target distribution at total-variation distance 0.053—below the preregistered smoke-abort threshold (0.15) and within the pure 500-shot sampling band around the target (median 0.046, 95th percentile 0.070), so device noise is not yet resolvable above shot noise at this sample size. Agreement with the pessimistic noise-envelope simulation is comparable (0.053). This certifies the deferred-measurement compilation, bit conventions, and retrieval path end-to-end on hardware; it is exploratory and makes no confirmatory claim about the order effect.

Preregistered hardware run. The separate 2026-09-02 main run comprised four completed, sequentially submitted 6,000-shot jobs, one per arm on `qpu.forte-enterprise-1`, with debiasing off and frozen five-qubit circuits (12 two-qubit gates each), costing $505.88 in total. The reused smoke TVDs were 0.045, 0.045, 0.068, and 0.048 (threshold 0.15). Reanalysis of the archived counts with the frozen R=12 batch analyzer gives, for seed 429, MSE 0.0356 (greedy) versus 0.00197 (policy/oracle), ratio 18.1, 90% CI $[8.9, 40.7]$. The exploratory seed 5098 pair gives 0.0164 versus 0.00819, ratio 2.0, CI $[1.07, 4.37]$. Earlier failed smokes and simulator jobs are excluded. The passing four-arm smoke gate was reused after a runner interruption, preceding the main jobs by 16–18 hours; the noise-envelope gate passed, while the non-enforced day-selection rule was NO-GO. Raw counts, job metadata, calibration snapshot, costs, and all amendments are in the supplement. Submission order was champion greedy, champion policy, median greedy, median oracle; UTC request times were 07:34:01, 08:26:42, 08:56:01, and 09:07:35, and the corresponding responses were 08:26:34, 08:55:53, 09:07:27, and 09:23:14. Because the R=12 batches are random partitions of each arm's one aggregate job, the bootstrap CI quantifies conditional shot noise only. It cannot estimate job-to-job variation, order effects, or device drift.

A.23 Order counting for the weak-measurement dial

The dial uses Kraus operators $K_- = c_- P_+ + c_+ P_-$ and $K_+ = c_+ P_+ + c_- P_-$ with $c_\pm = \sqrt{(1 \pm \kappa)/2}$ and $\{P_-, P_+\}$ the projectors of the chosen setting, so $K_-^\dagger K_- + K_+^\dagger K_+ = I$; κ=1 recovers the projective case and κ=0 gives $K_\pm = I/\sqrt{2}$. Three quantities that are easy to conflate scale differently. Writing $q = \mathrm{tr}(P_+\rho)$, the outcome law is $p_+ = \frac{1}{2} + \kappa(q - \frac{1}{2})$, so the *outcome-probability shift* is $O(\kappa)$, and the conditional post-measurement state likewise reweights the two blocks at $O(\kappa)$. The Fisher information of that two-outcome law is $\kappa^2(\partial_\theta q)^2/\big(\frac{1}{4} - \kappa^2(q - \frac{1}{2})^2\big) \to 4\kappa^2(\partial_\theta q)^2$, i.e. $O(\kappa^2)$. The *non-selective* channel is $\sum_\pm K_\pm \rho K_\pm^\dagger = P_+\rho P_+ + P_-\rho P_- + \sqrt{1-\kappa^2}\,(P_+\rho P_- + P_-\rho P_+)$, so it preserves the diagonal blocks and damps coherences by $\sqrt{1-\kappa^2} = 1 - \kappa^2/2 + O(\kappa^4)$: the averaged disturbance is $O(\kappa^2)$, the *same* order as the information. A numerical check reproduces both constants (coherence damping 0.0050126 against $\kappa^2/2 = 0.005$ at κ=0.1; Fisher information over κ^2 equal to 4.0016, 4.0004, 4.0001 at $\kappa = 0.1, 0.05, 0.025$). Consequently the non-monotone κ behaviour of §2 cannot be explained by information and disturbance entering at different orders. We attribute the residual at small κ to the scheduling contribution that survives when back-action is switched off, which is why we limit the back-action attribution of Proposition 2 to the designated workload witnesses.

A.24 On-device protocol: the full preregistered plan

The order experiment of §2 is compiled for a trapped-ion processor *without* mid-circuit measurement: for open-loop sequences the deferred-measurement principle is exact, so each intermediate measurement becomes a basis rotation, a CNOT onto a fresh ancilla, and the inverse rotation, with one terminal readout (2 system qubits + 3 ancillas at k=3). A frozen preregistration fixes the maximal-gap instance of Proposition 1 at k=3; the k=2 instance was excluded in advance because power analysis shows device noise erases its signal. It is a single adversarial instance, distinct from the distribution-level conditions of Table 1: the on-device test is a conditional observation consistent with the mechanism, not a certification of the mechanism or the median. The preregistration also fixes the two arms (greedy against the policy sequence, which our frozen policy emits with zero utility evaluations and which coincides with the oracle), the decision rule (90% bootstrap CI of the batch-MLE MSE ratio; predicted ratio $\approx$ 14, power 1.00 at n=500 shots/batch under measured noise, an upper bound on the frozen unpaired rule, Appendix A.25), the day-selection rule of Appendix A.16, and kill conditions. The arms were intended for one run-day but were submitted as

sequential jobs, so drift or submission order can tilt the comparison and is not represented by the conditional CI; the only *statistical* day gate is the frozen noise envelope ($e_2 \approx 1.6\%$), the operational gates (availability, calibration freshness $\leq$ 48h, cost cap) are released, and the protocol permits exactly one confirmatory run, negative results included. A preregistered *secondary* run targets typicality rather than the mechanism: a median-gap instance, selected by a deterministic preregistered rule, is compiled identically and scored by the same rule (presimulated power 0.96–0.98 at 50,000 shots; archive 0.93–1.00 across noise models and partitions); a frozen amendment demoted it to *exploratory* before execution after a noiseless-simulator check returned a CI covering 1; it runs after the primary if queue time remains, results reported as observed. The completed primary run met the frozen numerical criterion with ratio 18.1 and conditional shot-resampling 90% CI $[8.9, 40.7]$; the secondary exploratory result and deviations are reported in Appendix A.22.

A.25 Bootstrap procedure: presimulation versus the run-day analyzer

The frozen power figure quoted in §4 comes from a presimulation that resampled the two arms on *shared* batch indices (400 paired replicates), whereas the analyzer that scored the preregistered run resamples each arm independently (2,000 replicates per arm), matching the unpaired batch-bootstrap decision rule the preregistration declares. Pairing removes the shared run-day component and therefore reports a narrower interval than the rule that will actually be applied, so the presimulated power is an upper bound on the power of the frozen rule. We keep the unpaired rule as preregistered and flag the mismatch rather than restating the power figure, because recomputing it would change a frozen quantity after the fact.

A.26 Anticipated questions

We record the objections we expect and our answers, with the evidence each rests on.

Are the reported p-values measured or floored? Floored, and we say so here. The percentile bootstrap uses $B{=}10^4$ resamples, so the smallest attainable raw p is 10^{-4} and the smallest Holm-adjusted value in a six-test family is 6×10^{-4}; a three-test family floors at 3×10^{-4}. Every such value in this paper is written with $\leq$ and should be read as "at the resolution floor," not as a measured quantity. The same applies to the paired Wilcoxon figures: at $n{=}30$ the two-sided minimum is 1.9×10^{-9}, so 1.3×10^{-8} and 3.7×10^{-9} carry no information beyond "nearly every pair favours the policy," which the win counts already state. We report the win counts as the primary evidence.

Which robustness claims apply to the deployable arm? The bare policy and the new learned-surrogate repair run without a utility oracle. The former carries the paired results of Table 1, Appendix A.28, and the feature-noise check; the latter has only the post-hoc $k{=}3$, $n{=}90$ check above and does not significantly beat greedy. The exact-optimum reach rates use exact-utility repair and do not carry over. This paper reports no statistics on short-horizon-to-long transfer, on a closed-loop adaptive policy variant, or on finite readout error, and no claim here rests on any of them. Teacher-free training concerns the training pipeline, not deployment.

On device-physics instances, is "not distinguishable" the whole story? No. Every device point estimate favours greedy (1.38 against 1.29 at $k{=}3$; 1.64 against 1.24 at $k{=}4$; 1.37 against 1.22 at $(5, 12)$) and both win counts sit below half ($14/30$, $12/30$). The correct reading is a consistent directional shortfall. The ordinary combined $n{=}120$ tests are not confirmatory because the inspected first 30 instances are their prefix. On the previously evaluated 90-instance suffix, however, the bare policy is significantly worse than greedy on device-physics instances at $k{=}3$ and $k{=}4$ after Holm adjustment over all six conditions ($p = 0.036$ and 6.2×10^{-6}; Appendix A.28). We state the boundary as a post-hoc suffix-supported deficit, not a resolved preregistered claim.

Is a greedy failure without a constant-factor guarantee new? That greedy's guarantees on sequence objectives rest on string submodularity is classical (Streeter & Golovin, 2008; Golovin & Krause, 2011; Asadpour et al., 2022; Tang & Yuan, 2024), and we claim no novelty for greedy degrading without that premise. Our claim is narrower: the utility here is not a set function at all, so the $(1-1/e)$ argument loses its premise rather than its ratio, and we exhibit an explicit family where the regularized D-efficiency overhead is unbounded uniformly in the regularizer.

Can greedy use the same nominal information? Yes. Given the supplied nominal generators, background, initial state, and known expansion point θ_0, greedy can compute nominal Fisher utilities without device shots. The policy has no exclusive information advantage; it replaces per-instance utility queries with a trained mapping. The same-input comparison in Appendix A.20 has mixed results. At 2–3 qubits the oracle is cheap, and we claim no measured wall-clock advantage or deployment-count break-even.

Does the bounded-duration corollary leave only decoherence untreated? No: finite-shot rare-outcome access is untreated too. The corollary's saturation is approached as $\epsilon \to 0$, and in that limit the separating information sits on a rare outcome, so observing it needs a shot count that grows like $1/(\epsilon^2 s^2)$ (Appendix A.3). A finite shot budget is untreated as well, which is why we present the proposition as a mechanism-level impossibility rather than an experimentally accessible loss.

What does a clean-room reproduction actually verify? The released bundle carries an automated numerical check over every headline quantity in this paper; the packaged gate reports its current pass count with no failures. It verifies numbers against result files, not the scientific claims built on them.

Does the anonymous repository contain the raw calibration records? Yes. The bundle ships the raw per-day JSON for all three backends together with the collection script, checksums and a datasheet, so the drift analysis can be rerun without contacting the vendor API. The conditional wording in §6 covers redistribution terms, not availability.

A.27 Implementation accounting for the evaluation budget

The budget-parity claim is stated in black-box utility calls, and a reviewer may ask what the implementation actually pays. We instrumented the released code with counting wrappers on a device-physics instance at k=3, $|\mathcal{A}|$=6 (d=2, so each call builds $2d+1 = 5$ outcome trees):

Planner	Reported bill	Actual U calls	Distinct sequences	Outcome trees	Leaves
Greedy	18 ($= k\|\mathcal{A}\|$)	19	18	95	760
Repair (cap 18)	18	18	18	90	720
Bare policy	0	1 (scoring only)	1	5	40
Exhaustive	216	216	216	1,080	8,640

Three conventions matter. *(i)* Greedy's implementation pays $k|\mathcal{A}|+1$ calls: it re-evaluates its own final sequence once without caching. The repair arm is capped at the *reported* $k|\mathcal{A}|$, so parity is enforced in greedy's favour. *(ii)* Repair accepts an improving swap immediately and continues the same sweep, so while improvements keep arriving it consumes the full budget; sequence re-evaluations are full outcome-tree recomputations with no prefix reuse, identically for every arm. *(iii)* The bare policy's single call exists only to score the rollout for our tables and is not needed at deployment. Sequence-level memoization would remove greedy's duplicate call and let repair see more distinct candidates within the same budget, so no caching convention can flip the headline in our favour.

A.28 Expanded held-out evaluation: combined description and previously evaluated suffix

The extension uses the *same* training set (140 instances, identical seeds, the same forest—the policy is not retrained or reselected) and extends each held-out condition from n=30 to n=120 with seed rule $880000 + 977k + i$. The original 30 instances are the prefix, so an ordinary test on all 120 observations is not an independent confirmation after the prefix has been inspected. We therefore retain its adverse and favorable numerical outcomes but label all combined n=120 p-values descriptive post-hoc. Summarizing only the prefix reproduces every previously reported median exactly (18/18 quantities).

For a post-hoc exploratory analysis excluding the original prefix, we analyze the $i = 30, \ldots, 119$ suffix (n=90 per condition), using the same two-sided paired Wilcoxon test on log regularized D-efficiency overhead with zero differences discarded and Holm adjustment over the six conditions.

On random-matrix instances, the bare policy has $38/1/51$, $61/0/29$, and $74/0/16$ wins/ties/losses at k=3, 4, 5, with Holm $p = 0.365$, 1.82×10^{-3}, and 2.58×10^{-9}. On device-physics instances the corresponding counts are $41/1/48$, $26/0/64$, and $41/0/49$; policy-minus-greedy is significantly positive (worse) at k=3, 4 (Holm $p = 0.0362$ and 6.16×10^{-6}) and unresolved at k=5 ($p = 0.262$). Thus the previously evaluated suffix supports the device deficit at k=3, 4 without treating the inspected prefix as new evidence.

For completeness, the combined n=120 descriptive summaries are unchanged. On the random-matrix family the bare-policy wins are $85/120$ at k=4 and $99/120$ at k=5 (combined Holm $p = 1.7 \times 10^{-5}$ and 2.7×10^{-13}), while k=3 remains unresolved. On device-physics instances the combined medians are 1.34 vs. 1.25 at k=3 and 1.77 vs. 1.32 at k=4 (combined Holm $p = 8.2 \times 10^{-3}$ and 3.1×10^{-7}); k=5 remains unresolved ($p = 0.066$). These combined values are reported descriptively, not as confirmatory or "resolved" evidence. The repair-arm combined medians remain 1.04–1.13 with win rates 0.79–0.98.

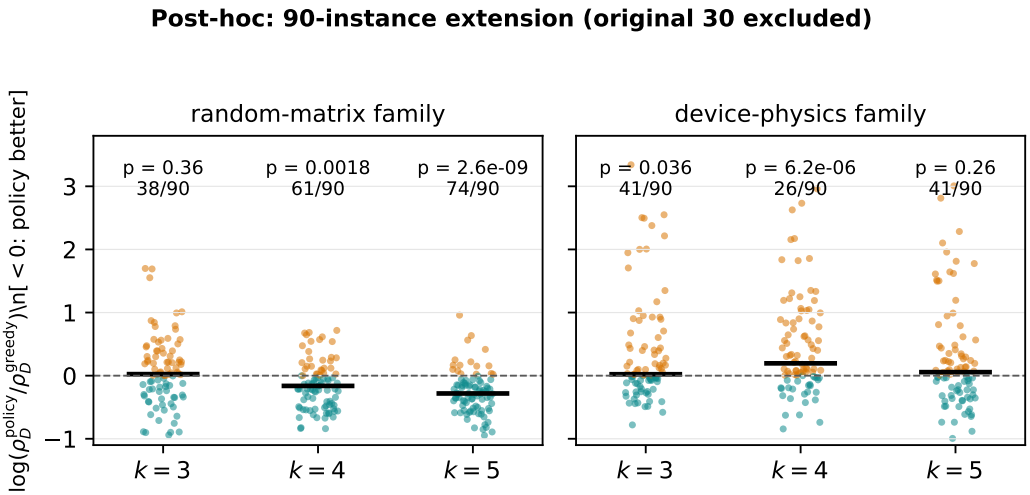


Figure 5: **Post-hoc boundary analysis on the previously evaluated 90-instance suffix.** Each dot is one suffix instance's paired log regularized D-efficiency-overhead ratio $\log(\rho_D^{\text{policy}}/\rho_D^{\text{greedy}})$; negative means the bare policy is better on that instance. Bars mark medians; counts give policy wins; p-values are Holm-adjusted Wilcoxon tests over the six-condition family. This suffix excludes the previously inspected 30-instance prefix.

These six conditions are the $|\mathcal{A}|$=12 evaluation family of §3, not the four headline conditions of Table 1. The n=30 numbers in the main text remain the preregistered originals; the suffix analysis is explicitly post-hoc and uses no new simulation.

A.29 Training-realization experiments

The first study independently regenerated ten training sets and forest seeds for each of two exhaustive-supervised policies. At random-Hermitian (k=5, $|\mathcal{A}|$=6), all fits emitted $[5, 5, 5, 5, 2]$ on every input and had geometric $\rho_D(\text{policy})/\rho_D(\text{greedy}) = 0.5203$; this is not evidence of input adaptivity. At device (k=3, $|\mathcal{A}|$=12) the ratio was 1.4938, and every fit-level effect was adverse. The random arm is not a replication of the released teacher-free row: it used 100 exhaustive-labeled training instances, while the released pipeline used 200 instances, four self-improvement rounds, and two polish rounds.

The separate teacher-free addendum was frozen after those outcomes had been inspected but before its own test outcomes. It regenerated ten disjoint 200-instance training sets and ten forest seeds. Exhaustive labels were used only for a post-training audit. Its builder function and representation differ from the supervised path, but for all 120 seed IDs the primitive arrays, oracle utilities, and greedy outputs agree exactly: this is the same physical test bank reached through an equivalent builder path, with all values freshly recomputed. Table 6 gives every unselected fit effect.

Table 6: All fit-level effects, defined as the mean over the fixed bank of $\log[\rho_D(\text{policy})/\rho_D(\text{greedy})]$; negative favors the policy. TF is the post-prior-results teacher-free addendum; ES is exhaustive-supervised.

Fit	TF random $(5, 6)$	ES random $(5, 6)$	ES device $(3, 12)$
1	-0.646142887	-0.653420918	+0.422468014
2	-0.651626433	-0.653420918	+0.291471951
3	-0.649719468	-0.653420918	+0.424719384
4	-0.653420918	-0.653420918	+0.468463637
5	-0.640818792	-0.653420918	+0.356612808
6	-0.653420918	-0.653420918	+0.356888158
7	-0.651705761	-0.653420918	+0.295545803
8	-0.650372294	-0.653420918	+0.573387942
9	-0.648299206	-0.653420918	+0.488492469
10	-0.652394407	-0.653420918	+0.335324955
Mean	-0.649792108	-0.653420918	$+0.401337512$
Geometric ratio	0.522154	0.520263	1.493821
Fit bootstrap 95% CI	$[-0.651812, -0.647241]$	$[-0.653421, -0.653421]$	$[0.349963, 0.456190]$

Effects are fit-level means over a shared 120-input bank, so n=10 for inference; the 1,200 fit/input rows are descriptive. Bootstrap intervals capture training-fit variation but no test-bank uncertainty. The enumerated sign-flip $p = 0.0009766$ for each favorable random arm is exact only under sign symmetry/exchangeable signs, not from a mean-zero null alone. These are exact simulations with a regularized Fisher objective: they establish neither new-bank, other-condition, hardware, finite-shot, nor literal shot-ratio performance. Protocols, executed code, source-kernel hashes, raw training and test traces, and a reanalysis gate are in the anonymous supplement; reproducible model and utility caches are omitted because they exceed 2 GiB and can be regenerated. The successful teacher-free runner was bound before execution and all its successful-run caches missed. An earlier serialization failure produced no test outputs, but its runner was not retained or hashed, so the claim that its edit changed only serialization and provenance is not independently diff-verifiable.

A.30 Remaining appendix material

Beam/SLS per-condition tables and the frozen preregistration text are included with the released artifacts.