Draft v0.4.6 — Manuskript

# Indhold

# Titel

**Noise-Regime Crossovers in Logical-Subspace Retention: Matched-Null Benchmarks, Manifold Probing, and Leakage-Aware Scoring (n=4)**

# Abstract

I study open-system retention of a two-dimensional logical subspace (one logical qubit) embedded in a 4-qubit Hilbert space using a matched-control design: a structured REAL construction is compared against a spectrum-matched NULL baseline under identical sampling, time grids, and Lindblad noise. To reduce probe-state (observable-regime) bias, I distinguish a coarse ZX screening set (4 canonical logical directions) from manifold probing using rand64 (≈66 probes, 64 random directions plus two basis states). Unless stated otherwise, headline results use rand64, while ZX runs are retained as diagnostic baselines. I report unconditional fidelity F_uncond(t), conditional fidelity F_cond(t) (conditioned on survival in the logical subspace), and leakage L(t), summarized by AUC over $t \in [0,5]$ and by threshold times. Under amplitude damping ($\gamma 1 > 0$) I consistently observe a coherent-but-leaky regime: REAL improves F_cond but increases leakage and degrades F_uncond. In contrast, under dephasing-only noise ($\gamma\phi=0.10$, $\gamma 1=0.00$) REAL exhibits a batch-stable net improvement in F_uncond that is robust to logical-basis choice and to single-qubit noise localization (subsets 0–3). A dephasing-strength sweep reveals a crossover: at $\gamma\phi=0.05$ no net unconditional advantage is observed, whereas at $\gamma\phi=0.20$ the unconditional advantage becomes large and leakage is reduced relative to NULL. A mixed-noise bridge run (both, $\gamma\phi=0.10$, $\gamma 1=0.02$) preserves a positive unconditional AUC gain while reintroducing early-time threshold penalties and higher leakage. Finally, I introduce a leakage-aware cost proxy to quantify when gains remain beneficial under leakage costs.

# 1. Introduction

A common failure mode in evaluating logical encodings, protected subspaces, or structured logical constructions is to over-index on a metric that conditions away failure modes. In open-system dynamics, an encoding may show improved **conditional** quality inside the logical subspace while simultaneously increasing the probability of leaving it (leakage), resulting in worse **unconditional** performance. This motivates a protocol that (i) uses a matched control, (ii) closes probe-state loopholes, (iii) checks batch stability, and (iv) explicitly tracks leakage.

This paper reports a controlled benchmark study using:

- matched-null design (REAL vs spectrum-matched NULL),
- manifold probing (rand64),
- logical basis robustness (noise_diag vs eigen),
- full single-qubit noise localization sweep (subsets 0–3),
- and a dephasing-strength crossover test.

The key empirical outcome is a regime split:

- amplitude-damping dominated noise yields a robust **tradeoff** (coherent-but-leaky),

- dephasing-only yields a **net unconditional advantage** for REAL under sufficiently strong dephasing.

# 2. Methods

## 2.1 Logical subspace and probe states

Any pure logical probe state can be parameterized on the logical Bloch sphere as |ψ_L(θ, ϕ)⟩ = cos(θ/2)|0_L⟩ + e^{iϕ} sin(θ/2)|1_L⟩, with θ ∈ [0, π] and ϕ ∈ [0, 2π). (Nielsen & Chuang, 2010; Bloch, 1946)

As (θ, ϕ) vary, |ψ_L(θ, ϕ)⟩ traces the logical Bloch sphere restricted to $\mathcal{L}$.

Probe-state regimes. I use two probe ensembles: (i) ZX screening with four canonical logical directions (fast but coarse), and (ii) manifold probing (rand64), which samples ≈66 directions (64 random plus two basis states). Unless stated otherwise, headline results use rand64; ZX runs are included as diagnostic baselines to document when coarse probing can mask direction-dependent effects. (Marsaglia, 1972)

## 2.2 Open-system dynamics (Lindblad)

Dynamics are simulated under Lindblad evolution with localized noise acting on a single-qubit subset (subset ∈ {0,1,2,3}, specified per run). Time is discretized uniformly over t ∈ [0, 5.0] using t_steps=25 points (Δt = 5.0/(25−1) ≈ 0.208).$\{q\}q \in \{0,1,2,3\}t \in [0, T]T = 5.0$ (Lindblad, 1976; Gorini, Kossakowski, & Sudarshan, 1976; Breuer & Petruccione, 2002)

Noise regimes:

- Amplitude damping (amp_damp): energy relaxation with rate $\gamma_1$; in the reported runs a small background dephasing term $\gamma\phi$ may be included as specified (e.g., $\gamma\phi$=0.01).
- Dephasing-only (dephasing): phase-randomization with rate $\gamma\phi$, with $\gamma_1$=0.00.
- Both (both): simultaneous dephasing ($\gamma\phi$) and amplitude damping ($\gamma_1$) acting together.

## 2.3 Matched-control design (REAL vs NULL)

I evaluate a structured construction (REAL) against a spectrum-matched baseline (NULL) under matched sampling. Within each run, REAL and NULL share the same probe ensemble, time grid, noise parameters (including localization to all qubits or a specified subset), and randomization schedule (3 batches; seeds_per_batch=5000; pairs_per_batch=400). An optional stable_pool filter can restrict evaluation to candidates passing a stability screen; because this can introduce selection bias, stable_pool is reported explicitly per run and is disabled for headline results unless selection and evaluation are separated (e.g., holdout seeds).

#### 2.3a Run registry and provenance

For transparency and reproducibility, each figure/table is tied to a concrete run ID (output filename) and its key settings: probe-state regime (ZX vs rand64), logical basis (eigen vs noise_diag), noise localization, and whether stable_pool was enabled.

Usage map (quick audit): the following bullets link manuscript sections/tables to the corresponding output files listed in the run registry.

• **Section 3.1** (amp_damp): v23_ampd_g10_subset3_noise_diag_p400_rand64.txt; v23_ampd_g05_subset3_noise_diag_p400_rand64.txt; v23_ampd_g02_subset3_noise_diag_p400_rand64.txt.
• **Section 3.2** (dephasing $\gamma\phi$=0.10): v23_deph_gphi10_subset3_noise_diag_p400_rand64.txt; v23_deph_gphi10_subset3_eigen_p400_rand64.txt; subset sweep: v23_deph_gphi10_subset0_eigen_p400_rand64.txt; v23_deph_gphi10_subset1_eigen_p400_rand64.txt; v23_deph_gphi10_subset2_eigen_p400_rand64.txt.
• **Section 3.3** ($\gamma\phi$ sweep): v23_deph_gphi05_subset3_eigen_p400_rand64.txt; v23_deph_gphi20_subset3_eigen_p400_rand64.txt (plus $\gamma\phi$=0.10 as above).
• **Section 3.4** (both): v23_both_gphi10_g02_subset3_eigen_p400_rand64.txt (batches 1–2 in this draft).

Table 0: batch-level deltas for the above runs; Table S2b: endpoint quantiles for amp_damp + dephasing runs (mixed-noise endpoint quantiles pending).

| Run set | Run ID / output | Noise model | Rates | Probe states | Basis | Noise qubits | stable_pool | pairs/batch | t_max / steps |
|---|---|---|---|---|---|---|---|---|---|
| B (headline) | v23_ampd_g10_ subset3_ noise_diag_p400 _ rand64.txt | amp_damp | γϕ=0.01, γ1=0.10 | rand64 | noise_diag | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_ampd_g05_ subset3_ noise_diag_p400 _ rand64.txt | amp_damp | γϕ=0.01, γ1=0.05 | rand64 | noise_diag | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_ampd_g02_ subset3_ noise_diag_p400 _ rand64.txt | amp_damp | γϕ=0.01, γ1=0.02 | rand64 | noise_diag | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi1 0_subset3_ noise_diag_p400 _ rand64.txt | dephasing | γϕ=0.10, γ1=0.00 | rand64 | noise_diag | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi0 5_subset3_ eigen_p400_ rand64.txt | dephasing | γϕ=0.05, γ1=0.00 | rand64 | eigen | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi1 0_subset3_ eigen_p400_ rand64.txt | dephasing | γϕ=0.10, γ1=0.00 | rand64 | eigen | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi1 0_subset0_ eigen_p400_ rand64.txt | dephasing | γϕ=0.10, γ1=0.00 | rand64 | eigen | subset={0} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi1 0_subset1_ eigen_p400_ rand64.txt | dephasing | γϕ=0.10, γ1=0.00 | rand64 | eigen | subset={1} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi1 0_subset2_ eigen_p400_ rand64.txt | dephasing | γϕ=0.10, γ1=0.00 | rand64 | eigen | subset={2} | False | 400 | 5.0 / 25 |
| B (headline) | v23_deph_gphi2 0_subset3_ eigen_p400_ rand64.txt | dephasing | γϕ=0.20, γ1=0.00 | rand64 | eigen | subset={3} | False | 400 | 5.0 / 25 |
| B (headline) | v23_both_gphi1 0_g02_subset3_ eigen_p400_ rand64.txt | both | γϕ=0.10, γ1=0.02 | rand64 | eigen | subset={3} | False | 400 | 5.0 / 25 |
| A (diagnostic) | v23_ampd_g10_ subset3_ noise_diag_p400 .txt | amp_damp | γϕ=0.01, γ1=0.10 | ZX | noise_diag | subset={3} | False | 400 | 5.0 / 25 |

### 2.3b Evidence criteria (decision rule)

We treat each batch as an independent replicate. A claimed effect must satisfy: (i) matched advantage (REAL−NULL) under spectrum-matched controls; (ii) replicate consistency (sign matches in ≥2/3 batches, ideally 3/3); (iii) practical effect floor $|\Delta| \geq 0.01$ for the primary summary metric; and (iv) leakage accounting, interpreting unconditional and conditional fidelities jointly with leakage.

## 2.4 Metrics: leakage, conditional fidelity, unconditional fidelity

Let $P_L$ be the projector onto the logical subspace L. For an evolved state $\rho(t)$ and a target logical pure state $|\psi_L\rangle$ (the intended logical probe state), I compute three coupled quantities: leakage, conditional fidelity, and unconditional fidelity. (Jozsa, 1994; Uhlmann, 1976; Nielsen & Chuang, 2010)

Notation note: in tables and command-line outputs I use fid_uncond and fid_cond; in the prose I may also write F_uncond and F_cond for the same quantities.

Leakage: $L(t) = 1 - \mathrm{Tr}[P_L \rho(t)]$.

Conditional state given survival: $\rho_{cond}(t) = P_L \rho(t) P_L / \mathrm{Tr}[P_L \rho(t)]$ (defined when $\mathrm{Tr}[P_L \rho(t)]>0$).

Conditional fidelity: $F_{cond}(t) = \langle\psi_L| \rho_{cond}(t) |\psi_L\rangle$. (Jozsa, 1994; Uhlmann, 1976; Nielsen & Chuang, 2010)

Unconditional fidelity (survival-weighted): $F_{uncond}(t) = \mathrm{Tr}[P_L \rho(t)] \cdot F_{cond}(t)$. (Jozsa, 1994; Uhlmann, 1976; Nielsen & Chuang, 2010)

Summaries: I report AUC over $t\in[0, t_{max}]$ for each metric, along with endpoint quantiles at $t=t_{max}$ and median threshold times $t_{F90}$ (time to $F_{uncond}<0.90$) and $t_{L10}$ (time to $L>0.10$), computed per batch. (Bradley, 1997)

Sign convention: unless stated otherwise, deltas are Δ = REAL − NULL. For fidelity metrics, positive ΔAUC is favorable; for leakage, negative ΔAUC(L) is favorable (REAL leaks less). (Jozsa, 1994; Uhlmann, 1976; Nielsen & Chuang, 2010) (Bradley, 1997)

Evidence threshold: unless stated otherwise, the practical effect floor $|\Delta| \geq 0.01$ is applied to the primary KPI ΔAUC(F_uncond); secondary metrics (F_cond, leakage, thresholds) are reported to explain mechanism and tradeoffs. (Bradley, 1997)

## 2.5 Leakage-aware cost proxy

To translate tradeoffs into a simple system-level score, define a leakage-aware score $S(\alpha) = \mathrm{AUC}(F_{uncond}) - \alpha\cdot\mathrm{AUC}(L)$, where $\alpha \geq 0$ encodes the application's relative cost of leakage. I report $\Delta S(\alpha) = S_{REAL}(\alpha) - S_{NULL}(\alpha) = \Delta\mathrm{AUC}(F_{uncond}) - \alpha\cdot\Delta\mathrm{AUC}(L)$.

Interpretation of α: α sets how costly leakage is relative to unconditional retention. For example, α≈1 treats a +0.01 increase in leakage AUC as canceling a +0.01 gain in AUC(F_uncond) in the combined score.

When $\Delta\mathrm{AUC}(L) > 0$ (REAL leaks more), the crossover $\alpha^* = \Delta\mathrm{AUC}(F_{uncond}) / \Delta\mathrm{AUC}(L)$ gives the leakage-cost level above which ΔS becomes negative. When $\Delta\mathrm{AUC}(L) < 0$ and $\Delta\mathrm{AUC}(F_{uncond}) > 0$, $\Delta S(\alpha)$ is favorable for any $\alpha \geq 0$.$\Delta S(\lambda) = S_{\mathrm{REAL}}(\lambda) - S_{\mathrm{NULL}}(\lambda)$

Choosing α in practice depends on the application: if leakage events are effectively fatal (discard, reset, or costly recovery), use a larger α; if small leakage is tolerable, use a smaller α. Reporting α* lets readers map conclusions to their own leakage-cost assumptions.

$$\lambda^{\star} \approx \frac{\Delta\mathrm{AUC}(F_{\mathrm{uncond}})}{\Delta\mathrm{AUC}(L)} (\Delta\mathrm{AUC}(L) > 0).$$

# 3. Results

## 3.1 Amplitude damping: robust “coherent-but-leaky” tradeoff

Under amp_damp with $\gamma_1$=0.10 ($\gamma\phi$=0.01), subset=3, basis=noise_diag, rand64, REAL shows per-batch deltas (REAL−NULL):

- $\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) \in [-0.021, -0.016]$ΔAUC(F_uncond) = [−0.021, −0.016, −0.017] (batches 1–3).
- $\Delta\mathrm{AUC}(F_{\mathrm{cond}}) \in [+0.027, +0.030]$ΔAUC(F_cond) = [+0.027, +0.030, +0.027].
- $\Delta\mathrm{AUC}(L) \in [+0.032, +0.047]$ΔAUC(L) = [+0.047, +0.041, +0.032].

v23_ampd_g10_subset3_noise_diag_p400_rand64.txt

Reducing damping strength scales the effect down without changing its sign structure:

- $\gamma_1$=0.05: ΔAUC(F_uncond) = [−0.011, −0.014, −0.012], while ΔAUC(F_cond) = [+0.015, +0.014, +0.014], and ΔAUC(L) = [+0.026, +0.026, +0.020].$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) \approx -0.011 - 0.014\Delta\mathrm{AUC}(F_{\mathrm{cond}}) \approx +0.014 + 0.015\Delta\mathrm{AUC}(L) \approx +0.020 + 0.026$

v23_ampd_g05_subset3_noise_diag_p400_rand64.txt

- $\gamma_1$=0.02: effects drop below the practical floor: ΔAUC(F_uncond) = [−0.005, −0.006, −0.005], ΔAUC(F_cond) = [+0.006, +0.006, +0.006], and ΔAUC(L) = [+0.011, +0.012, +0.009].$\Delta\mathrm{AUC}(F_{\mathrm{cond}}) \approx +0.006\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) \approx -0.005$

v23_ampd_g02_subset3_noise_diag_p400_rand64.txt

**Interpretation:** amp_damp runs consistently support a regime where REAL improves “within-subspace quality given survival” but pays with leakage and worse unconditional retention.

## 3.2 Dephasing-only at γϕ=0.10: net unconditional advantage

Switching to dephasing-only with $\gamma\phi$=0.10, $\gamma_1$=0.00, subset=3, rand64, basis=noise_diag yields: Note that for leakage, negative ΔAUC(L) is favorable.

This is a net gain despite higher leakage (positive ΔAUC(L)); the improvement is therefore driven primarily by in-subspace quality (F_cond) rather than reduced leakage.

- $\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.015, +0.019, +0.028$ΔAUC(F_uncond) = [+0.015, +0.019, +0.028] across batches,
- $\Delta\mathrm{AUC}(F_{\mathrm{cond}}) \approx +0.067$ΔAUC(F_cond) = [+0.071, +0.069, +0.067],$+0.071$
- $\Delta\mathrm{AUC}(L) \approx +0.022$ΔAUC(L) = [+0.035, +0.037, +0.022].$+0.037$

v23_deph_gphi10_subset3_noise_diag_p400_rand64.txt

This net unconditional advantage is **basis-robust** (also present with basis=eigen).

v23_deph_gphi10_subset3_eigen_p400_rand64.txt

It is also **noise-localization robust**: with basis=eigen and rand64 at $\gamma\phi$=0.10, positive $\Delta\mathrm{AUC}(F_{\mathrm{uncond}})$is observed for subsets 0, 1, 2, and 3. For example:

For a complete batch-by-batch listing across all subsets, see Table 1 and Supplementary Note S1; the examples below illustrate the pattern.

- subset=0 (batch 1): ΔAUC(F_uncond)=+0.014, ΔAUC(F_cond)=+0.070, ΔAUC(L)=+0.031.$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.014\Delta\mathrm{AUC}(F_{\mathrm{cond}}) = +0.070\Delta\mathrm{AUC}(L) = +0.031$

v23_deph_gphi10_subset0_eigen_p400_rand64.txt

- subset=2 (all batches): ΔAUC(F_uncond) = [+0.018, +0.023, +0.021], ΔAUC(F_cond) = [+0.069, +0.071, +0.074], and ΔAUC(L) = [+0.032, +0.031, +0.032].$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.018, +0.023, +0.021\Delta\mathrm{AUC}(F_{\mathrm{cond}})$

v23_deph_gphi10_subset2_eigen_p400_rand64.txt

- subset=1 (all batches): ΔAUC(F_uncond) = [+0.016, +0.015, +0.021], ΔAUC(F_cond) = [+0.068, +0.072, +0.065], and ΔAUC(L) = [+0.023, +0.035, +0.035].$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.016, +0.015, +0.021$

v23_deph_gphi10_subset1_eigen_p400_rand64.txt

**Interpretation:** under sufficiently strong dephasing, REAL becomes net-beneficial even under unconditional metrics, and this finding survives key robustness checks (probe ensemble, basis, and localization).

## 3.3 Dephasing strength sweep: crossover behavior

At subset=3, basis=eigen, rand64:

- γϕ=0.05 yields no net unconditional advantage: ΔAUC(F_uncond) = [−0.006, −0.004, −0.002] across batches, despite ΔAUC(F_cond) = [+0.036, +0.034, +0.033] and ΔAUC(L) = [+0.039, +0.039, +0.034].$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = -0.006, -0.002, -0.002$

v23_deph_gphi05_subset3_eigen_p400_rand64.txt

- γϕ=0.20 yields a strong net unconditional advantage, and notably reduced leakage relative to NULL: ΔAUC(F_uncond) = [+0.084, +0.105, +0.116], ΔAUC(F_cond) = [+0.153, +0.147, +0.144], and ΔAUC(L) = [−0.017, −0.028, −0.048].$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.084, +0.105, +0.116\Delta\mathrm{AUC}(F_{\mathrm{cond}}) = +0.153, +0.147, +0.144\Delta\mathrm{AUC}(L) = -0.017, -0.028, -0.048$

v23_deph_gphi20_subset3_eigen_p400_rand64.txt

**Interpretation:** the sign of the unconditional advantage depends on dephasing strength; the sweep indicates a crossover between ~0.05 and 0.10, and a markedly stronger regime at 0.20.

## 3.4 Mixed noise bridge ("both"): advantage survives, tradeoff reappears in thresholds

For combined noise (both) with γϕ=0.10 and $\gamma_1$=0.02 at subset=3, basis=eigen, rand64:

- $\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) = +0.014, +0.020, +0.027$ΔAUC(F_uncond) = [+0.014, +0.020] (batches 1–2; batch 3 omitted),

Scope note: mixed-noise conclusions are preliminary in this draft because batch 3 and endpoint quantiles for "both" are not reported here (formatting artifact); the reported pattern is based on batches 1–2 only.

- $\Delta\mathrm{AUC}(F_{\mathrm{cond}}) = +0.070, +0.073, +0.059$ΔAUC(F_cond) = [+0.070, +0.073],
- $\Delta\mathrm{AUC}(L) = +0.039, +0.034, +0.034$ΔAUC(L) = [+0.039, +0.034], with Δt_F90_med = −0.21 and Δt_L10_med = −0.21 (REAL crosses earlier).

v23_both_gphi10_g02_subset3_eigen_p400_rand64.txt

However, median threshold deltas are negative (Δt_F90_med = −0.21 and Δt_L10_med = −0.21), indicating that REAL crosses the failure thresholds earlier despite positive AUC gains.$t_{F90} t_{L10}$

v23_both_gphi10_g02_subset3_eigen_p400_rand64.txt

# 4. Leakage-aware interpretation via cost proxy

Using ΔS(α)=ΔAUC(F_uncond) − α·ΔAUC(L):$S(\lambda) = \mathrm{AUC}(F_{\mathrm{uncond}}) - \lambda\, \mathrm{AUC}(L)$

- **γϕ=0.05: since ΔAUC(F_uncond)<0 and ΔAUC(L)>0, ΔS(α)<0 for all α ≥ 0 (REAL is unfavorable under this cost model).**$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) < 0 \Delta\mathrm{AUC}(L) > 0 \Delta S(\lambda) < 0 \lambda \geq 0$

v23_deph_gphi05_subset3_eigen_p400_rand64.txt

- **γϕ=0.10: with ΔAUC(F_uncond)>0 and ΔAUC(L)>0, the crossover lies at α* ≈ 0.43–1.27 across batches (median ≈ 0.66); REAL is favorable when leakage is not "too expensive" (α < α*).**$\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) \approx +0.02 \Delta\mathrm{AUC}(L) \approx +0.03 \lambda^* \sim 0.5 - 0.7$
- **γϕ=0.20: since ΔAUC(F_uncond)>0 and ΔAUC(L)<0, ΔS(α)>0 for any α ≥ 0; here REAL improves unconditional retention while also reducing leakage.**$\Delta\mathrm{AUC}(L) < 0 \Delta S(\lambda) \lambda > 0$

v23_deph_gphi20_subset3_eigen_p400_rand64.txt

- **both (γϕ=0.10, $\gamma_1$=0.02): ΔAUC(F_uncond)>0 and ΔAUC(L)>0 in the reported batches, giving α* ≈ 0.36–0.59; early-time threshold penalties (negative Δt) indicate caution when applications emphasize time-to-failure rather than integrated AUC.**$\lambda^*$

v23_both_gphi10_g02_subset3_eigen_p400_rand64.txt

# 5. Discussion

The experiments isolate a clean phenomenon: whether REAL is "net beneficial" depends strongly on the noise channel. Under amplitude damping, REAL repeatedly exhibits a coherent-but-leaky pattern, consistent across probe coverage and damping strength scaling. Under dephasing-only noise, the pattern changes: at

$\gamma\phi$=0.10 REAL becomes net beneficial in unconditional AUC, and at $\gamma\phi$=0.20 the advantage becomes large and simultaneously reduces leakage relative to NULL.

The combination of (i) matched NULL control, (ii) manifold probing, (iii) basis robustness, (iv) full single-qubit localization sweeps, and (v) a dephasing-strength crossover provides a compact but stringent evidentiary basis. The cost proxy $S(\lambda)$clarifies how "goodness" depends on the application's leakage penalty.

## 1) Box: How to use this in practice

**Practical use of this framework (engineering workflow).**
The central contribution of this work is a leakage-aware, matched-control evaluation protocol for open-system retention studies. In practical benchmarking or design selection (e.g., comparing two encodings, control schedules, or Hamiltonian families), I recommend the following workflow:

1. **Define a matched control (NULL).** Construct a spectrum-matched or otherwise constraint-matched NULL baseline to eliminate trivial advantages unrelated to the hypothesized structural effect.
2. **Evaluate three metrics jointly: unconditional retention F_uncond(t) (primary KPI), conditional retention F_cond(t) (mechanistic diagnostic), and leakage L(t) (critical failure channel).** $F_{\text{uncond}} F_{\text{cond}} L$
3. **Use manifold coverage (rand64) for headline claims.** Treat small canonical probe sets (e.g., ZX) as diagnostics only, since direction-dependent effects can be missed by sparse probing.
4. **Require replication across batches. Treat each batch as an independent replicate. Enforce a pre-registered decision rule (e.g., consistent sign in ≥2/3 batches and a practical floor |ΔAUC(F_uncond)| ≥ 0.01).** $|\ \Delta\mathrm{AUC}(F_{\text{uncond}})\ | \geq 0.01$
5. **Interpret tradeoffs explicitly. If ΔAUC(F_cond) > 0 while ΔAUC(F_uncond) ≤ 0 and ΔAUC(L) > 0, the effect is best described as coherent-but-leaky rather than a practical advantage.** $\Delta\mathrm{AUC}(F_{\text{cond}}) > 0 \Delta\mathrm{AUC}(L) > 0 \Delta\mathrm{AUC}(F_{\text{uncond}}) \leq 0$
6. **If needed, map results to application cost via α (and α*). Use the leakage-aware cost proxy to determine whether fidelity gains remain favorable once leakage penalties reflect the target application.** $\alpha$

## 2) Decision tree: Go/No-Go for "REAL advantage"

**Decision rule for practical adoption (go/no-go).**
Use the following checklist when deciding whether REAL shows a practical advantage over NULL in an open-system regime.

**Step 1 (probe regime):**

- Are headline results computed with **rand64** manifold coverage?
  - If **No**, treat findings as diagnostic only.

**Step 2 (primary KPI):**

- Is ΔAUC(F_uncond) > 0?$\Delta\mathrm{AUC}(F_{\text{uncond}}) = \mathrm{AUC}_{REAL}(F_{\text{uncond}}) - \mathrm{AUC}_{NULL}(F_{\text{uncond}})$
  - If **No**, conclude **no net retention advantage** (unless the application explicitly uses conditional/post-selected operation).

**Step 3 (replication):**

- Is the sign of ΔAUC(F_uncond) consistent in ≥2/3 batches and above the practical floor |ΔAUC(F_uncond)| ≥ 0.01?$\Delta\mathrm{AUC}(F_{\text{uncond}}) \mid \Delta \mid \geq 0.01$
  - If **No**, label the effect **inconclusive** and increase sampling or refine the regime.

**Step 4 (leakage check):**

- Is leakage favorable, i.e. ΔAUC(L) < 0 (negative is favorable)?$\Delta\mathrm{AUC}(L) < 0$
  - If **Yes**, classify the result as **strong practical evidence**.
  - If **No** (leakage worse), proceed to Step 5.

**Step 5 (mechanistic diagnosis):**

- Is ΔAUC(F_cond) > 0 while ΔAUC(L) > 0?$\Delta\mathrm{AUC}(F_{\text{cond}}) > 0\Delta\mathrm{AUC}(L) > 0$
  - If Yes, classify as coherent-but-leaky; use cost proxy α* to determine whether fidelity gains can justify leakage.$\alpha$
  - If **No**, the regime likely reflects mixed mechanisms; report both AUC and threshold metrics and treat the conclusion as **application-dependent**.

**Step 6 (application mapping):**

- If an application has strong leakage intolerance (e.g., error-correction compatibility), require ΔAUC(L) ≤ 0 and/or α < α*.$\Delta\mathrm{AUC}(L) \leq 0\alpha < \alpha^{\backslash *}$
- If an application supports erasure/post-selection, conditional improvements may remain valuable even when unconditional gains are modest.

## 3) Table: Common failure modes (and how this paper addresses them)

Table X. **Common evaluation failure modes and mitigations used in this work.**

| Failure mode | What it looks like in metrics | Why it misleads | Mitigation in this paper |
|---|---|---|---|
| **Coherent-but-leaky** | $\Delta\mathrm{AUC}(F_{\mathrm{cond}}) >$ 0ΔAUC(F_cond) > 0 but ΔAUC(F_uncond) ≤ 0 and ΔAUC(L) > 0 $\Delta\mathrm{AUC}(L) >$ $0\Delta\mathrm{AUC}(F_{\mathrm{uncond}}) \leq 0$ | Conditional quality improves only when the state remains in-subspace; practical retention does not improve | Joint reporting of F_uncond, F_cond, L; leakage-aware interpretation and cost proxy α, α* $F_{\mathrm{uncond}}, F_{\mathrm{cond}}, L\alpha$ |
| **Probe blind spot (sparse probes)** | Effect appears under ZX but disappears under rand64 (or vice versa) | Direction-dependent effects can be missed with a small probe set | Treat ZX as diagnostic; use rand64 manifold coverage for headline claims |
| **Cherry-picking / selection bias** | Strong effect only after filtering candidates; weak or inconsistent otherwise | Selection on the same population used for evaluation inflates effects | Explicit run registry and batch replication; disclose selection procedures (or disable selection for headline runs) |
| **Basis artefact** | Advantage appears only in one basis (eigen vs noise-diagonal) | Coordinate choice can hide/induce apparent structure | Basis robustness checks (eigen vs noise_diag) as part of evaluation |
| **Single-run overinterpretation** | Large Δ in one run; sign flips across batches | Random sampling fluctuations dominate | Treat batches as independent replicates; require ≥2/3 sign-consistency and effect floor |
| **Metric monoculture** | Large AUC gain but thresholds worsen (or vice versa) | AUC averages across time; threshold captures first critical failure | Report both AUC and threshold metrics; interpret tradeoffs explicitly |
| **Endpoint vs trajectory confusion** | Endpoint quantiles suggest improvement while AUC suggests degradation (or vice versa) | Curves may cross; endpoint misses mid-trajectory behavior | Use AUC for trajectory summary; annotate cases where quantiles and AUC diverge |
| **Unmatched baselines** | REAL beats a “random” NULL by a lot | Trivial differences (e.g., spectrum) can explain the win | Spectrum-matched NULL controls to isolate structural effects |

# 6. Limitations

This work is a controlled pilot in an n=4 testbed with Markovian Lindblad noise. Internal validity is strengthened by matched NULL controls, batch replication, explicit leakage tracking, basis checks, and single-qubit noise localization sweeps. However, external validity depends on scaling to larger encodings and on testing richer noise models (correlated/non-Markovian and hardware-specific channels). (Rivas, Huelga, & Plenio, 2014) A further limitation is observable-regime dependence: coarse ZX probing can under-sample the logical manifold and may mask direction-dependent effects; for this reason, headline claims are restricted to the manifold-probe regime (rand64), while ZX results are treated as diagnostics. (Marsaglia, 1972) Finally, when stable_pool selection is used, selection bias is a potential loophole unless selection and evaluation are separated; I therefore report stable_pool explicitly per run and do not rely on stable_pool-restricted runs as primary evidence unless accompanied by holdout controls.

Not tested / not addressed in this draft:

• Scaling beyond n=4 (no asymptotic claims).
• Non-Markovian or correlated noise (only Markovian Lindblad channels are modeled). (Rivas, Huelga, & Plenio, 2014)
• Hardware-calibrated noise parameters or device-specific validation (simulation-only).
• Mixed-noise completeness (batch 3 and endpoint quantiles for “both” are pending due to a formatting artifact).

# 7. Conclusion

I provide a matched-null benchmark protocol for logical-subspace retention under open-system noise with explicit leakage tracking and manifold probing. The results show (i) a robust coherent-but-leaky tradeoff under amplitude damping, (ii) a dephasing-dependent crossover where REAL becomes net beneficial under unconditional AUC at $\gamma\phi=0.10$ and strongly so at $\gamma\phi=0.20$, and (iii) persistence of the unconditional AUC advantage under mixed noise, albeit with reintroduced threshold penalties. A leakage-aware score $S(\lambda)$offers a compact system-level interpretation.

***"The results demonstrate that evaluating logical constructions without explicit leakage accounting risks overestimating practical performance, particularly in energy-relaxation-dominated environments."***

***(Draft) Figure/Table plan (minimal)***

- Table 1: Summary of deltas for each regime (amp_damp g1=0.10/0.05/0.02; dephasing gphi=0.05/0.10/0.20; both).
- Fig 2: Crossover plot: ΔAUC(F_uncond) vs $\gamma\phi$ (0.05, 0.10, 0.20).

# 8. References


Nielsen, M. A., & Chuang, I. L. (2010). Quantum computation and quantum information (10th anniversary ed.). Cambridge University Press.

Bloch, F. (1946). Nuclear induction. Physical Review, 70(7-8), 460-474. https://doi.org/10.1103/PhysRev.70.460

Breuer, H.-P., & Petruccione, F. (2002). The theory of open quantum systems. Oxford University Press.

Gorini, V., Kossakowski, A., & Sudarshan, E. C. G. (1976). Completely positive dynamical semigroups of N-level systems. Journal of Mathematical Physics, 17(5), 821-825. https://doi.org/10.1063/1.522979

Lindblad, G. (1976). On the generators of quantum dynamical semigroups. Communications in Mathematical Physics, 48(2), 119-130. https://doi.org/10.1007/BF01608499

Uhlmann, A. (1976). The "transition probability" in the state space of a *-algebra. Reports on Mathematical Physics, 9(2), 273-279.

Jozsa, R. (1994). Fidelity for mixed quantum states. Journal of Modern Optics, 41(12), 2315-2323. https://doi.org/10.1080/09500349414552171

Marsaglia, G. (1972). Choosing a point from the surface of a sphere. The Annals of Mathematical Statistics, 43(2), 645-646. https://doi.org/10.1214/aoms/1177692644

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Rivas, Á., Huelga, S. F., & Plenio, M. B. (2014). Quantum non-Markovianity: Characterization, quantification and detection. Reports on Progress in Physics, 77(9), 094001. https://doi.org/10.1088/0034-4885/77/9/094001

Watrous, J. (2018). The theory of quantum information. Cambridge University Press.

# 9. Tables

## Table 1

### Table 1 — Caption

Table 1. Batch-level AUC deltas (Δ = REAL − NULL) for open-system retention: unconditional fidelity AUC(F_uncond), conditional fidelity AUC(F_cond), and leakage AUC(L), reported for each batch (3 batches) together with median threshold-time deltas Δt_F90 and Δt_L10 where available. Positive ΔAUC for fidelity is favorable; negative ΔAUC(L) is favorable (reduced leakage). All runs use n=4, t_max=5.0, t_steps=25, pairs_per_batch=400, and the probe-state regime specified in the run registry.

**amp_damp ($\gamma\phi$=0.01, $\gamma_1$=0.02), subset=3, basis=noise_diag**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | -0.005 | +0.006 | +0.011 | +0.00 | +0.00 |
| 2 | -0.006 | +0.006 | +0.012 | +0.00 | +0.00 |
| 3 | -0.005 | +0.006 | +0.009 | +0.00 | +0.00 |

**amp_damp ($\gamma\phi$=0.01, $\gamma_1$=0.05), subset=3, basis=noise_diag**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | -0.011 | +0.015 | +0.026 | -0.20 | -0.42 |
| 2 | -0.014 | +0.014 | +0.026 | -0.20 | -0.42 |
| 3 | -0.012 | +0.014 | +0.020 | -0.20 | -0.42 |

**amp_damp ($\gamma\phi$=0.01, $\gamma_1$=0.10), subset=3, basis=noise_diag**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | -0.021 | +0.027 | +0.047 | -0.42 | -0.42 |
| 2 | -0.016 | +0.030 | +0.041 | -0.42 | -0.42 |
| 3 | -0.017 | +0.027 | +0.032 | -0.42 | -0.42 |

**dephasing ($\gamma\phi$=0.05, $\gamma_1$=0.00), subset=3, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | -0.006 | +0.036 | +0.039 | -0.21 | -0.21 |
| 2 | -0.004 | +0.034 | +0.039 | -0.21 | -0.21 |
| 3 | -0.002 | +0.033 | +0.034 | -0.21 | -0.21 |

**dephasing (γϕ=0.10, $\gamma_1$=0.00), subset=3, basis=noise_diag**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.015 | +0.071 | +0.035 | +0.00 | -0.10 |
| 2 | +0.019 | +0.069 | +0.037 | +0.00 | +0.00 |
| 3 | +0.028 | +0.067 | +0.022 | +0.00 | +0.00 |

**dephasing (γϕ=0.10, $\gamma_1$=0.00), subset=3, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.017 | +0.068 | +0.037 | +0.00 | +0.00 |
| 2 | +0.020 | +0.073 | +0.037 | +0.00 | +0.00 |
| 3 | +0.029 | +0.070 | +0.023 | +0.00 | +0.00 |

**dephasing (γϕ=0.10, $\gamma_1$=0.00), subset=0, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.014 | +0.070 | +0.031 | +0.00 | +0.00 |
| 2 | +0.029 | +0.063 | +0.040 | +0.00 | +0.00 |
| 3 | +0.023 | +0.073 | +0.026 | +0.00 | +0.00 |

**dephasing (γϕ=0.10, $\gamma_1$=0.00), subset=1, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.016 | +0.068 | +0.023 | +0.00 | +0.00 |
| 2 | +0.015 | +0.072 | +0.035 | +0.00 | +0.00 |
| 3 | +0.021 | +0.065 | +0.035 | +0.00 | +0.00 |

**dephasing (γϕ=0.10, $\gamma_1$=0.00), subset=2, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.018 | +0.069 | +0.032 | +0.00 | +0.00 |
| 2 | +0.023 | +0.071 | +0.031 | +0.00 | +0.00 |
| 3 | +0.021 | +0.074 | +0.032 | +0.00 | +0.00 |

**dephasing ($\gamma\phi$=0.20, $\gamma_1$=0.00), subset=3, basis=eigen**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.084 | +0.153 | -0.017 | +0.00 | +0.00 |
| 2 | +0.105 | +0.147 | -0.028 | +0.00 | +0.00 |
| 3 | +0.116 | +0.144 | -0.048 | +0.00 | +0.00 |

**both ($\gamma\phi$=0.10, $\gamma_1$=0.02), subset=3, basis=eigen (Note: Batch 3 omitted here due to a formatting artifact in the extracted log; batches 1–2 are reported.)**

| Batch | ΔAUC(fid_uncond) | ΔAUC(fid_cond) | ΔAUC(leak) | Δt_F90_med | Δt_L10_med |
|---|---|---|---|---|---|
| 1 | +0.014 | +0.070 | +0.039 | -0.21 | -0.21 |
| 2 | +0.020 | +0.073 | +0.034 | -0.21 | -0.21 |

## Supplementary Note S1 (Batch-Level AUC Deltas and Robustness Criteria)

Table 1 reports batch-level deltas (Δ = REAL − NULL) in AUC summary metrics for unconditional fidelity F_uncond(t), conditional fidelity F_cond(t), and leakage L(t), with AUC computed over $t\in[0,5]$. Positive ΔAUC for fidelity indicates net improvement; for leakage, negative ΔAUC(L) is favorable (REAL leaks less). Because F_cond conditions on survival in the logical subspace, it can improve even when unconditional performance worsens; Table 1 therefore makes it explicit when regimes are coherent-but-leaky (ΔAUC(F_cond)>0 while ΔAUC(F_uncond)<0 and ΔAUC(L)>0). Batch reporting supports the evidence criteria: effects must be consistent in sign in ≥2/3 batches and exceed a practical floor.

## Table S2b — Endpoint quantiles ved t=5.0 (q10 / q50 / q90)

### Table S2b — Caption

Table S2b. Endpoint quantiles and threshold medians for open-system logical retention. Reported are q10/q50/q90 at t=5.0 for F_uncond and leakage L for REAL and NULL, plus (where applicable) median threshold times t_F90 (time to F_uncond<0.90) and t_L10 (time to L>0.10). Threshold deltas are defined as Δt = t_REAL − t_NULL; negative Δt means REAL crosses earlier (worse on that threshold).

| **Run** | **F_uncond(t=5) REAL** | **F_uncond(t=5) NULL** | **Leak(t=5) REAL** | **Leak(t=5) NULL** |
|---|---|---|---|---|
| amp_damp (γϕ=0.01, $\gamma_1$=0.10), subset={3}, basis=noise_diag | 0.777/0.792/0.837 | 0.751/0.797/0.841 | 0.117/0.202/0.219 | 0.145/0.189/0.235 |
| amp_damp (γϕ=0.01, $\gamma_1$=0.05), subset={3}, basis=noise_diag | 0.879/0.887/0.913 | 0.865/0.891/0.916 | 0.061/0.110/0.119 | 0.077/0.101/0.128 |
| amp_damp (γϕ=0.01, $\gamma_1$=0.02), subset={3}, basis=noise_diag | 0.948/0.952/0.964 | 0.943/0.955/0.965 | 0.026/0.047/0.050 | 0.032/0.042/0.053 |
| dephasing (γϕ=0.10, $\gamma_1$=0.00), subset={3}, basis=noise_diag | 0.645/0.671/0.752 | 0.638/0.651/0.674 | 0.152/0.322/0.354 | 0.290/0.327/0.351 |
| dephasing (γϕ=0.10, $\gamma_1$=0.00), subset={0}, basis=eigen | 0.645/0.672/0.761 | 0.638/0.651/0.674 | 0.131/0.323/0.354 | 0.289/0.327/0.350 |
| dephasing (γϕ=0.10, $\gamma_1$=0.00), subset={1}, basis=eigen | 0.645/0.672/0.758 | 0.638/0.651/0.674 | 0.150/0.323/0.355 | 0.289/0.327/0.350 |
| dephasing (γϕ=0.10, $\gamma_1$=0.00), subset={2}, basis=eigen | 0.645/0.671/0.758 | 0.638/0.651/0.674 | 0.130/0.323/0.354 | 0.289/0.327/0.350 |
| dephasing (γϕ=0.10, $\gamma_1$=0.00), subset={3}, basis=eigen | 0.646/0.673/0.768 | 0.638/0.651/0.674 | 0.114/0.322/0.354 | 0.290/0.327/0.351 |

| Run | F_uncond(t=5) REAL | F_uncond(t=5) NULL | Leak(t=5) REAL | Leak(t=5) NULL |
|---|---|---|---|---|
| dephasing ($\gamma\phi$=0.05, $\gamma_1$=0.00), subset={3}, basis=eigen | 0.792/0.818/0.869 | 0.792/0.803/0.836 | 0.075/0.188/0.204 | 0.142/0.189/0.205 |
| dephasing ($\gamma\phi$=0.20, $\gamma_1$=0.00), subset={3}, basis=eigen | 0.467/0.520/0.609 | 0.440/0.468/0.499 | 0.451/0.461/0.532 | 0.447/0.501/0.538 |

### Supplementary Note S2 (Thresholds and Endpoint Quantiles)

Table S2b reports median threshold times t_F90 (time until F_uncond<0.90) and t_L10 (time until L>0.10) for REAL and the spectrum-matched NULL. Threshold deltas are defined as $\Delta t$ = t_REAL − t_NULL, so negative $\Delta t$ means REAL crosses earlier (worse on that specific threshold metric). Threshold metrics capture first-crossing behavior, whereas AUC integrates over the full interval $t \in [0,5]$; therefore, AUC and thresholds can disagree in sign when trajectories differ in shape (e.g., early penalty but later improvement). Table S2b complements this by reporting endpoint quantiles (q10/q50/q90) at t=5.0, indicating whether effects are broad shifts or tail-dominated. Endpoint quantiles for the mixed-noise ("both") run are omitted in this revision due to the same formatting artifact noted in Table 1.

Example ($\gamma\phi$=0.10, subset=3): Table 1 shows $\Delta$AUC(L)>0 (REAL leaks more on average over $t \in [0,5]$), while Table S2b may still show smaller early-time leakage quantiles for REAL (e.g., lower q10). This is consistent if REAL starts with lower leakage but rises faster later, yielding a larger time-integrated leakage even when early endpoint quantiles look favourable.

# 10. Code and Data Availability

The simulation framework, run registry, and all raw data files supporting the findings of this study are available on GitHub at https://github.com/tomstevns/qubits. This includes the source code for Lindblad evolution, the rand64 probe generation, and the leakage-aware cost-proxy analysis scripts. For reproducibility, a command registry (command_registry_v23_recent.docx) listing the most commonly used PowerShell one-liner scripts for executing the Python runs is included in the repository (https://github.com/tomstevns/qubits/tree/main/documentation).