

Ensemble Dependence of Critical Exponents at Quantum Error Correction Thresholds: Analytic Mechanisms and Statistical Consequences

Ensemble Dependence of Critical Exponents at Quantum Error Correction Thresholds: Analytic Mechanisms and Statistical Consequences

Abstract

Ensemble equivalence — the expectation that microcanonical, canonical, and grand-canonical descriptions of a system agree in the thermodynamic limit — is a cornerstone of statistical mechanics, yet it can fail for observables that probe exponentially rare events. Recent work on a simplified quantum error correction (QEC) model reports that critical exponents governing fidelity and magic at the error-correction threshold depend on the noise ensemble chosen, and that even the existence of the transition can be ensemble-dependent [1,2]. Here we develop a minimal, fully analytic treatment of this phenomenon using a random-unitary encoding model subject to depolarizing noise. We derive explicit expressions for the decoding fidelity in grand-canonical (independent per-qubit errors) and canonical (fixed error number) ensembles, and exhibit concrete numerical instances of the discrepancy: for a distance-3 code with 100 physical qubits and error rate $p = 0.01$, the grand-canonical fidelity is 0.735762 while the canonical fidelity is exactly 1. We show that the discrepancy survives the vanishing of relative fluctuations ($\sigma/\mu \rightarrow 0$) because fidelity is a large-deviation tail quantity, and we compute the associated entropy and capacity budgets. We further derive the statistical cost of resolving ensemble-dependent exponents: distinguishing two candidate exponents at distance Δp from threshold requires a sample count scaling as Δp^{-4} and system sizes scaling as Δp^{-2} , a compound burden that we quantify under stated assumptions. We relate the mechanism to the known ensemble dependence of the random transverse-field Ising chain [3], discuss saturation of the information-theoretic bound of Feldman et al. as reported in [1,2], and outline falsifiable predictions. The results clarify when ensemble choice is physically consequential rather than a matter of convenience in QEC threshold physics.

1. Introduction

Quantum error correction protects quantum information from decoherence by encoding logical states into a subspace of a larger Hilbert space designed so that common errors move the state into an orthogonal, detectable error space [10]. A central empirical fact about QEC architectures is the threshold phenomenon: below some physical error rate p_c , *increasing code size suppresses logical error rates; above it, encoding fails*. Near p_c , observables such as fidelity and magic — the non-Clifford resource content of a state, i.e., its distance from the set of states preparable by stabilizer circuits — behave as power laws, defining critical exponents.

In equilibrium thermodynamics, ensemble equivalence is a cornerstone: microcanonical, canonical, and grand-canonical descriptions coincide for generic short-range systems in the thermodynamic limit. The recent preprint [1,2] argues that QEC thresholds violate this expectation. Using a simplified model — single-step encoding and decoding by a random unitary — it reports (i) different exponents for generic channels versus quantum trajectories (unravelings of the same channel into stochastic pure-state evolutions conditioned on measurement records), (ii) saturation of a recently derived information-theoretic bound by Feldman et al. (2026), extended from grand-canonical to canonical and newly defined intermediate ensembles, and (iii) ensemble-dependent existence of the transition itself.

This paper has three goals. First, to make the mechanism of ensemble inequivalence in QEC thresholds fully explicit and quantitative, with every number derived from stated inputs, so that the phenomenon is not merely asserted but demonstrated in closed form. Second, to connect the QEC instance to the earlier condensed-matter precedent, the random transverse-field Ising chain, where microcanonical versus canonical treatments of disorder were already shown to yield different critical behavior [3]. Third, to derive the statistical consequences of the claim: what sample sizes and system sizes are required to confirm or falsify ensemble-dependent exponents, and what failure modes arise at finite size.

Our central analytic result is a clean separation of scales: relative fluctuations of the error number vanish as $1/\sqrt{np}$, yet the decoding fidelity, being a tail probability $P(k \leq t)$, remains exponentially ensemble-sensitive. Ensemble equivalence fails precisely because the order parameter is a large-deviation quantity, not a self-averaging intensive one. This observation immediately suggests which QEC observables (fidelities, magic, tail-weighted logical error rates) are ensemble-fragile and which (average syndromes, mean entropies) are not.

2. Background and Related Work

The source claim. The arXiv query record and abstract for 2609.21886 [1] report that in a simplified QEC model — a single encoding/decoding step by a random unitary — the critical exponents of fidelity and magic at threshold differ between generic channels and supposedly equivalent quantum trajectories, that the exponents saturate an information-theoretic bound due to Feldman et al. (2026), and that the existence of the transition itself is ensemble-dependent. The full preprint [2] carries the same content; we cite both records since they constitute the grounding material for this analysis.

Ensemble dependence in statistical physics. The phenomenon closest to the claim of [1,2] is ensemble dependence of critical exponents in disordered systems. Ref. [3] studies the random transverse-field Ising chain, contrasting a microcanonical ensemble — where disorder realizations satisfy a precise constraint on the random variables — with a canonical ensemble in which variables are drawn from an unconstrained distribution. That work asks directly whether critical exponents can differ between the two cases and answers affirmatively through a detailed study, establishing that ensemble dependence of exponents is not pathological but generic in disordered criticality. The QEC threshold, being an average over random codes and noise realizations, is precisely such a disordered critical point, so the analogy is structural, not merely rhetorical.

Foundations of QEC. Ref. [10] provides the modern definition: a code is a subspace such that common errors map the encoded state into an orthogonal error space, enabling syndrome-based recovery. Ref. [8] introduces the operational core — encoding, syndrome extraction, error operators, and code construction — and shows that general noise on two-state systems decomposes into Pauli operators, which is why Pauli-error models dominate threshold analyses; this decomposition is also what makes the "generic channel" versus "trajectory" distinction sharp, since a channel is a probabilistic mixture of Pauli actions while a trajectory is a single stochastic realization. Ref. [7] surveys the passage from classical to quantum coding, emphasizing that quantum channels behave differently from classical ones — superposition and measurement disturbance mean that naive classical error-control intuition fails, which is the historical reason ensemble questions in QEC were underexamined. Ref. [9] extends QEC beyond qubits to continuous-variable and higher-dimensional systems, noting that the discovery of QEC transformed quantum information from a curiosity into a technology; this matters here because the ensemble-dependence claim, if robust, applies to any platform whose threshold is estimated by averaging over random realizations.

Specialized frameworks. Ref. [5] introduces entanglement-assisted codes, in which pre-shared entanglement between sender and receiver relaxes the dual-containing constraint of standard stabilizer codes; the noise statistics on the ebits constitute yet another ensemble choice, and our tail-sensitivity argument transfers directly. Ref. [4] treats continuous-time QEC, in which noise and correction are simultaneous weak-measurement-and-feedback processes; this is the setting where the channel-versus-trajectory distinction is most physically acute, since continuous measurement intrinsically produces trajectories, and the subsystem principle it invokes gives an operational meaning to fidelity at threshold. Ref. [6] proposes QEC in quaternionic Hilbert spaces with quaternionic Pauli analogues; while the algebra differs, the statistical question of how errors are sampled over the generalized error algebra is identical in structure, so ensemble dependence should be expected there as well.

System-level context. Ref. [11] on spectral benchmarking of holographic quantum simulations supplies diagnostics sensitive to the same rare-event tails that govern logical fidelity; an ensemble-dependent exponent would make spectral benchmarks non-comparable across protocols. Ref. [12], on the lifecycle of a fault-tolerant quantum computer, treats threshold crossing as a lifecycle milestone; if the exponent — and per [1,2] even the existence of the transition — depends on ensemble, lifecycle cost models built on a single exponent inherit that ambiguity. Ref. [13] on thermodynamic and informational bottlenecks of scalable fault-tolerant computation argues that thermodynamic-style limits constrain scalable QEC; the claim of [1,2] sharpens this by suggesting the limits themselves are ensemble-labeled. Ref. [14] on operationalizing generalized symmetries provides mathematical language — non-invertible and higher-form symmetries — in which the ensemble constraint (fixed error weight versus fixed mean weight) can be viewed as a superselection sector structure on the noise algebra, suggesting a symmetry-based reformulation of ensemble inequivalence that we flag as an open question.

3. Methods

Model. One logical qubit is encoded into n physical qubits by a Haar-random unitary U (the "random encoding" of [1,2]). Noise acts as an n -qubit depolarizing channel: each physical qubit independently suffers a Pauli error with probability p (grand-canonical ensemble), or exactly $k = pn$ qubits are corrupted, chosen uniformly (canonical ensemble). Decoding applies $U^\dagger$ and measures whether the logical state survived; for a code correcting up to t errors, decoding succeeds iff the number of actual errors k satisfies $k \leq t$.

Ensembles.

- *Grand-canonical (GC):* $k \sim \text{Binomial}(n, p)$; $P(k) = C(n, k) p^k (1-p)^{(n-k)}$.
- *Canonical (C):* $k = pn$ exactly (we take pn integer); the corrupted set is uniform among $C(n, pn)$ subsets.
- *Intermediate ensembles:* following the extension in [1,2] from grand-canonical to canonical, one can interpolate by fixing k only on a subsystem of $m < n$ qubits; we define these but compute only the two endpoint ensembles, since the endpoints suffice to demonstrate inequivalence.

Observables. (i) Decoding fidelity $F = P(k \leq t)$. (ii) Noise entropy budget $H = n \cdot h_2(p)$, where h_2 is the binary entropy. (iii) Channel capacity upper bound $C \leq 1 - h_2(p)$ bits per qubit (Holevo-type bound for the depolarizing channel). (iv) Fluctuation diagnostics: mean $\mu = np$, variance $\sigma^2 = np(1-p)$.

Statistical framework for exponent resolution. To compare ensembles near threshold we adopt the scaling ansatz $F(p, n) = F_c - A(p_c - p)^\beta G((p_c - p) n^{1/\nu})$, with $G(0) = 1$. We compute the fidelity gap ΔF between two candidate exponents at fixed $\Delta p = p_c - p$, then ask how many independent samples N are needed for the gap to exceed two standard errors of the mean, with per-sample standard deviation σF . All inputs are stated explicitly in Section 4.

Analytic program. All arithmetic below uses only closed-form expressions with stated inputs (n , p , t , and the assumed scaling parameters). No simulation data are used or reported. Where we extrapolate beyond computed points, the extrapolation is explicitly labeled a projection with stated assumptions and uncertainty bounds.

4. Analysis

Every input number below is stated with its origin; every arithmetic step is shown.

Input parameters (Part I: fidelity demonstration). $n = 100$ physical qubits (chosen as a round, computationally transparent system size); $p = 0.01$ error rate per qubit (a commonly quoted near-term physical error rate); $t = 1$ correctable errors (distance-3 code, the minimal interesting case); success iff $k \in \{0, 1\}$.

Step 1: Grand-canonical fidelity. $F_{GC} = P(k=0) + P(k=1) = (1-p)^n + n \cdot p \cdot (1-p)^{n-1}$.

Compute $(0.99)^{100}$. Using $\ln(0.99) = -0.01005034$: $100 \times (-0.01005034) = -1.005034$; $e^{(-1.005034)} = e^{(-1)} \times e^{(-0.005034)} = 0.367879 \times 0.994979 = 0.366032$. So $P(k=0) = 0.366032$.

$(0.99)^{99} = 0.366032 / 0.99 = 0.369729$. $P(k=1) = 100 \times 0.01 \times 0.369729 = 1 \times 0.369729 = 0.369729$.

$F_{GC} = 0.366032 + 0.369729 = \mathbf{0.735762}$.

Step 2: Canonical fidelity. In the canonical ensemble $k = pn = 100 \times 0.01 = 1$ exactly. Since $t = 1$ and $k = 1 \leq t$, decoding always succeeds: $F_C = \mathbf{1.000000}$.

The gap is $\Delta F = 1.000000 - 0.735762 = \mathbf{0.264238}$ at identical mean error number. This is the concrete demonstration of ensemble inequivalence of the order parameter.

Step 3: Why relative fluctuations do not rescue equivalence. $\mu = np = 100 \times 0.01 = 1$. $\sigma^2 = np(1-p) = 100 \times 0.01 \times 0.99 = 0.99$; $\sigma = \sqrt{0.99} = 0.994987$. Relative fluctuation $\sigma/\mu = 0.994987$ at $n = 100$; in general $\sigma/\mu = \sqrt{(1-p)/(np)} \propto n^{-(1/2)} \rightarrow 0$. Standard self-averaging arguments would therefore predict ensemble equivalence. But fidelity is not a function of k near its mean — it is the tail indicator $1[k \leq t]$. The tail mass $P(k \geq 2) = 1 - F_{GC} = 0.264238$ is generated by fluctuations of size $O(1)$ around the mean $\bar{k} = 1$, which are $O(1)$ in absolute terms for all n when $p = t/n$ scaling is held. Equivalence of means does not imply equivalence of tail functionals; this is the analytic core of the phenomenon reported in [1,2].

Step 4: Scaling of the gap (projection, labeled). If we scale the code distance so that t grows with n while p is fixed, the GC tail $P(k > t)$ is governed by the large-deviation rate function of the binomial. For the computed point, the excess tail probability 0.264238 is the quantity whose n -scaling defines the GC critical exponent; in the canonical ensemble the same observable is identically zero for $k = pn \leq t$ and identically one for $pn > t$ — a step function with no critical scaling at all. This reproduces, in closed form, the report of [1,2] that the transition's very existence is ensemble-dependent: the canonical ensemble has no fidelity transition in this model, only a jump at $pn = t$.

Step 5: Entropy and capacity budgets. $h_2(0.01) = -0.01 \cdot \log_2(0.01) - 0.99 \cdot \log_2(0.99)$. $\log_2(0.01) = -6.643856$; $\log_2(0.99) = \ln(0.99)/\ln 2 = -0.01005034/0.693147 = -0.0144996$. $h_2(0.01) = 0.01 \times 6.643856 + 0.99 \times 0.0144996 = 0.0664386 + 0.0143545 = 0.0807931$ bits/qubit.

Noise entropy at $n = 100$: $H = 100 \times 0.0807931 = \mathbf{8.07931}$ bits. Holevo-type capacity bound per qubit: $1 - 0.0807931 = 0.9192069$ bits; over 100 qubits: $\mathbf{91.92069}$ bits.

The canonical ensemble carries zero entropy of the error number (k is fixed), but full combinatorial entropy of the error location: $\log_2 C(100,1) = \log_2 100 = 6.643856$ bits. The GC ensemble carries both: 8.07931 bits total. The difference, $8.07931 - 6.643856 = 1.435454$ bits, is exactly the entropy of the binomial fluctuation of k around its

mean at these parameters — the same fluctuations that generate the fidelity gap of Step 2. This quantitative bookkeeping makes precise how the "supposedly equivalent" ensembles differ in information content, the channel-side counterpart of the information-theoretic bound saturation reported in [1,2].

Part II: statistical resolution of exponent differences (all inputs are assumptions, labeled as such).

Input parameters (Part II): candidate exponents $\beta_1 = 1$, $\beta_2 = 2$ (hypothesis chosen to make the saturation claim of [1,2] concrete; the actual exponents of [1,2] may differ, and every number below rescales accordingly); amplitude $A = 0.1$ (assumed, dimensionless); threshold fidelity $F_c = 0.5$ (assumed; plausible by symmetry between correct and incorrect decoding under Haar averaging); per-sample fidelity standard deviation $\sigma F = 0.01$ (assumed, typical for single-shot fidelity estimates on $n \sim 10^2$ – 10^3 qubits); shift exponent $v_s = 1$ (assumed).

Step 6: Fidelity gap at $\Delta p = 10^{-2}$. $F_1 = F_c - A \cdot (\Delta p)^{\beta_1} = 0.5 - 0.1 \times 0.01 = 0.499$. $F_2 = 0.5 - 0.1 \times (0.01)^2 = 0.5 - 0.00001 = 0.49999$. $\Delta F = 0.49999 - 0.499 = 0.00099 = 9.9 \times 10^{-4}$.

Step 7: Fidelity gap at $\Delta p = 10^{-3}$. $F_1 = 0.5 - 0.1 \times 0.001 = 0.4999$. $F_2 = 0.5 - 0.1 \times 10^{-6} = 0.4999999$. $\Delta F = 0.0000999 = 9.99 \times 10^{-5}$. The gap scales as $\Delta p - \Delta p^2 \approx \Delta p$: linear, because $\beta_1 = 1$ dominates.

Step 8: Sample size to resolve the gap. Require $\Delta F \geq 2\sigma F/\sqrt{N}$, i.e., $N \geq (2\sigma F/\Delta F)^2$. At $\Delta p = 10^{-2}$: $N \geq (0.02 / 0.00099)^2 = (20.202)^2 = 408.12$, hence $N = 409$ samples. At $\Delta p = 10^{-3}$: $N \geq (0.02 / 0.0000999)^2 = (200.2002)^2 = 40080.12$, hence $N = 40081 \approx 4.0 \times 10^4$ samples.

Step 9: Sample size to resolve the exponent itself. Distinguishing $\beta_2 = 2$ from $\beta_1 = 1$ requires detecting the quadratic term $A \cdot \Delta p^2$ against the linear term, i.e., $2\sigma F/\sqrt{N} \leq A \cdot \Delta p^2$, giving $N \geq (2\sigma F/(A \cdot \Delta p^2))^2 = (0.2/\Delta p^2)^2 = 0.04/\Delta p^4$. At $\Delta p = 10^{-2}$: $N \geq 0.04/10^{-8} = 4 \times 10^6$ samples. At $\Delta p = 10^{-3}$: $N \geq 0.04/10^{-12} = 4 \times 10^{10}$ samples. This Δp^{-4} divergence is a central quantitative result: exponent separation is cheaply resolvable only far from threshold, but far from threshold the asymptotic power law may not hold, creating a squeeze.

Step 10: Minimum system size. With finite-size rounding $\delta p \sim 1/n$ ($v_s = 1$), distinguishability requires $\Delta p^2 \gg 1/n$, i.e., $n \gg 1/\Delta p^2$: $n \gg 10^4$ qubits at $\Delta p = 10^{-2}$ and $n \gg 10^6$ at $\Delta p = 10^{-3}$. Combined with Step 9, resolving $\beta_2 = 2$ at $\Delta p = 10^{-3}$ needs both $n > 10^6$ and $N > 4 \times 10^{10}$ — a compound cost of $10^6 \times 4 \times 10^{10} = 4 \times 10^{16}$ qubit-samples (projection: product of the two lower bounds, assuming independence of system-size and sample-count costs; uncertainty: the bound could be relaxed by up to $\sim 10^3$ if σF is reduced by variance-reduction techniques, giving a projected range 4×10^{13} – 4×10^{16}).

Step 11: Magic (projection). If magic density $m = M/n$ obeys $m(p) = m_c - B \cdot \Delta p^{\beta_2} M$ with $B = 0.1$ (assumed) and the same exponent split, identical arithmetic applies: the magic gap at $\Delta p = 10^{-2}$ is 9.9×10^{-3} in magic density, resolvable with $N = 409$ samples at $\sigma m = 0.01$ (assumed equal to σF). This is a projection, not a computation from measured magic data.

5. Results

All numbers below are outputs of the arithmetic in Section 4 from the stated inputs; none are empirical measurements.

1. **Grand-canonical fidelity:** $F_{GC} = 0.735762$ (from $P(k=0) = 0.366032$ and $P(k=1) = 0.369729$).
2. **Canonical fidelity:** $F_C = 1.000000$ ($k = 1 = t$ exactly).
3. **Ensemble gap:** $\Delta F = 0.264238$ at identical mean error number $\mu = 1$.
4. **Fluctuation diagnostics:** $\sigma^2 = 0.99$, $\sigma = 0.994987$, $\sigma/\mu = 0.994987$, scaling as $n^{(-1/2)}$ — yet the gap persists, demonstrating failure of tail-observable equivalence despite self-averaging of the mean.

5. **Entropy budgets:** $h_2(0.01) = 0.0807931$ bits/qubit; GC noise entropy $H = 8.07931$ bits; canonical location entropy 6.643856 bits; difference 1.435454 bits attributable to number fluctuations.
6. **Capacity bound:** 91.92069 bits over 100 qubits at $p = 0.01$ (Holevo-type upper bound 0.9192069 bits/qubit).
7. **Fidelity gap between candidate exponents (projection):** $\Delta F = 9.9 \times 10^{-4}$ at $\Delta p = 10^{-2}$; 9.99×10^{-5} at $\Delta p = 10^{-3}$; linear in Δp to leading order.
8. **Samples to resolve the gap at 2σ (projection):** $N = 409$ at $\Delta p = 10^{-2}$; $N = 40,081$ at $\Delta p = 10^{-3}$.
9. **Samples to resolve the exponent (projection):** $N \geq 4 \times 10^6$ at $\Delta p = 10^{-2}$ and $N \geq 4 \times 10^{10}$ at $\Delta p = 10^{-3}$, following $N = 0.04/\Delta p^4$.
10. **Minimum system size (projection):** $n \gg 1/\Delta p^2$, i.e., $n \gg 10^4$ at $\Delta p = 10^{-2}$ and $n \gg 10^6$ at $\Delta p = 10^{-3}$; compound cost $\geq 4 \times 10^{16}$ qubit-samples at $\Delta p = 10^{-3}$ (projected range 4×10^{13} – 4×10^{16}).
11. **Structural result:** the canonical ensemble in this model exhibits a jump rather than a critical transition, reproducing in closed form the ensemble dependence of the transition's existence reported in [1,2]. We do not compute the GC exponent v_{GC} here; extracting it requires finite-size data we do not possess.

6. Discussion

Limitations. Our analysis uses a distance-3 code at a single parameter point for Part I, and assumed scaling parameters for Part II; the full exponent extraction of [1,2] requires finite-size scaling across n , which we have not performed. The random-unitary encoding is a simplification: real codes have structured syndromes, and the identification "success iff $k \leq t$ " ignores degenerate codes and correlated error structures. The canonical ensemble with pn integer is an idealization; for non-integer pn the canonical model requires a rounding convention that can itself affect near-threshold behavior. We have not computed magic values, only argued structurally. The Feldman et al. (2026) bound is cited in [1,2] but is not in our bibliography; we cannot verify its value, form, or existence beyond the source abstract.

Failure modes and self-criticism. A skeptic could argue that the fidelity gap of 0.264238 is a finite-size artifact: at fixed p and growing n with t fixed, both ensembles eventually give $F \rightarrow 0$, and the gap closes. This is correct — which is precisely why the meaningful statement concerns exponents and scaling windows near threshold, not single-point fidelities. A second objection is that trajectory and channel ensembles are mathematically equivalent for any linear observable of the density matrix: the trajectory average is the channel. Any reported exponent difference must therefore arise from (a) nonlinear functionals (magic measures are nonlinear in p , so this is plausible for magic but less so for fidelity), (b) non-commutation of the order of limits — averaging over trajectories before or after taking $n \rightarrow \infty$ need not commute at criticality, echoing the constrained-versus-unconstrained averaging of [3], or (c) finite-size artifacts misread as asymptotic exponents. Our Step 9 result sharpens (c): at $\Delta p = 10^{-2}$ with $n = 10^3$, the quadratic term is $10^{-4} \times A$, below the finite-size rounding $A/n = 10^{-3}$, so a simulation at that scale *cannot* see $\beta_2 = 2$ and might report spurious effective exponents. Conversely, if exponent differences were observed at modest sizes, that itself would be evidence against the asymptotic interpretation.

What would falsify the claims. (i) A demonstration that fidelity exponents agree between channel and trajectory ensembles in the joint limit ($n \rightarrow \infty$, then $\Delta p \rightarrow 0$, samples $\rightarrow \infty$) at fixed averaging order. (ii) An analytic solution of the random-unitary model showing $F(p)$ is exactly linear in Δp for both ensembles. (iii) Failure of the predicted Δp^{-4} sample-cost law in well-controlled simulations. (iv) An explicit construction in which the QEC order parameter is a self-averaging mean (not a tail functional) yet still shows ensemble-dependent exponents.

Open questions. (i) Do the exponents of [1,2] saturate the Feldman et al. bound for *all* intermediate ensembles, or only at the endpoints? Our entropy bookkeeping (Step 5) suggests the 1.435454-bit fluctuation entropy is the relevant currency, but this is a conjecture. (ii) Is the intermediate-ensemble family of [1,2] a one-parameter family

connecting the exponents continuously, and if so, is $\beta(\lambda)$ monotone? (iii) Can ensemble inequivalence be reformulated as superselection of noise sectors in the generalized-symmetry language of [14]? (iv) Do architecture-level scaling analyses [12,13] and spectral diagnostics [11] inherit a systematic ensemble bias, and can it be bounded? (v) Does the quaternionic generalization [6] or the entanglement-assisted setting [5] admit ensembles with *larger* inequivalence than the depolarizing case studied here?

7. Conclusion

Using a fully analytic random-unitary QEC model, we have shown with explicit arithmetic that the decoding fidelity at fixed mean error number differs by 0.264238 between grand-canonical and canonical noise ensembles (0.735762 versus 1.000000 at $n = 100$, $p = 0.01$, $t = 1$), even though relative fluctuations vanish as $n^{-(1/2)}$. The mechanism is that fidelity is a large-deviation tail functional, immune to the self-averaging that enforces ensemble equivalence for means. We quantified the information-theoretic asymmetry between ensembles (8.07931 versus 6.643856 bits of noise entropy) and showed that the canonical ensemble in this model exhibits a jump rather than a critical transition, reproducing in closed form the ensemble dependence of the transition's existence reported in [1,2]. We further derived the statistical cost of confirming ensemble-dependent exponents: a sample count scaling as Δp^{-4} and system sizes scaling as Δp^{-2} , with a compound burden of $\sim 4 \times 10^{16}$ qubit-samples at $\Delta p = 10^{-3}$ under stated assumptions. This squeeze means ensemble dependence is simultaneously easy to miss (finite-size experiments cannot see it near threshold) and hard to confirm (far from threshold the asymptotic regime fails). The practical lesson — from foundational treatments [7,8,9,10] to specialized frameworks [4,5,6] and system-level scaling studies [11,12,13,14] — is that threshold exponents are not ensemble-invariant observables, and any reported exponent must be accompanied by a statement of the noise ensemble. The precedent of the random transverse-field Ising chain [3] indicates this is a general feature of disordered critical phenomena, not a peculiarity of quantum information.

References

- [1] TITLE: arXiv Query: `searchquery=&idlist=2609.21886&start=0&max_results=1` [2] Ensemble Dependence of the Critical Exponent at a Quantum Error Correction Threshold. arXiv:2609.21886v1. <https://arxiv.org/abs/2609.21886v1> [3] Ensemble dependence in the Random transverse-field Ising chain. arXiv:cond-mat/0305664v1. <https://arxiv.org/abs/cond-mat/0305664v1> [4] Continuous-time quantum error correction. arXiv:1311.2485v2. <https://arxiv.org/abs/1311.2485v2> [5] Entanglement-Assisted Quantum Error-Correcting Codes. arXiv:1610.04013v1. <https://arxiv.org/abs/1610.04013v1> [6] Quantum Error Correction in Quaternionic Hilbert Spaces. arXiv:2504.19833v1. <https://arxiv.org/abs/2504.19833v1> [7] An Introduction to Error-Correcting Codes: From Classical to Quantum. arXiv:quant-ph/0602157v1. <https://arxiv.org/abs/quant-ph/0602157v1> [8] Quantum Computing and Error Correction. arXiv:quant-ph/0304016v2. <https://arxiv.org/abs/quant-ph/0304016v2> [9] Quantum error correction beyond qubits. arXiv:0811.3734v1. <https://arxiv.org/abs/0811.3734v1> [10] Quantum Error Correction. arXiv:1910.03672v1. <https://arxiv.org/abs/1910.03672v1> [11] DOI 10.5281/zenodo.18327721. QNFO: Spectral Benchmarking of Holographic Quantum Simulations. [12] DOI 10.5281/zenodo.18000790. QNFO: Lifecycle of a Fault-Tolerant Quantum Computer. [13] DOI 10.5281/zenodo.17955898. QNFO: Thermodynamic and Informational Bottlenecks of Scalable Fault-Tolerant Quantum Computation. [14] DOI 10.5281/zenodo.18199396. QNFO: Operationalizing Generalized Symmetries.

Appendix A. Divergence report

D1. Nature of the quantitative claims (DIVERGENT; convention chosen: analytic derivation with stated assumptions).

- Draft A reports specific simulated numerical exponents and thresholds: canonical $v = 2.5$, $\beta F = 0.40$, $\beta M = 0.52$, $pc = 0.152$; *grand-canonical* $v = 2.0$, $\beta F = 0.50$, $\beta M = 0.61$, $pc = 0.148$; fidelity ratio $FC/FG = 1.34$ at $p = 0.10$; all attributed to a 10^5 -instance simulation of a $[[5,1,3]]$ code that no other draft performed or could verify.
- Drafts B and C both explicitly refuse to report simulation data: B treats exponent values as hypotheses ($\beta_1 = 1$, $\beta_2 = 2$) and derives statistical consequences; C derives closed-form fidelities with no simulation.
- *Convention adopted*: the main text presents only analytically derivable quantities (Part I, from Draft C) and clearly labeled projections with stated assumptions (Part II, from Draft B). Draft A's simulated exponents are SINGLE-source claims (see Appendix B) and are not incorporated into the main results, because they cannot be cross-validated within this corpus and their provenance (a simulation the other drafts did not perform) is unverifiable. Should Draft A's simulation be independently reproduced, its values would replace the hypothesis-driven inputs of Part II.

D2. Model definition (DIVERGENT; convention chosen: Draft C's success-criterion model).

- Draft A defines a $[[5,1,3]]$ stabilizer code with fixed Clifford encoding, depolarizing noise on each qubit, and exact state-vector simulation; it also defines a micro-canonical ensemble with fixed k .
- Drafts B and C both use the Haar-random unitary encoding of $[1,2]$ with the success criterion "decoding succeeds iff $k \leq t$ "; B frames the channel/trajectory distinction, C frames the canonical/grand-canonical distinction.
- *Convention adopted*: the Haar-random model with the $k \leq t$ success criterion (2 of 3 drafts; directly traceable to $[1,2]$). Draft A's stabilizer-code variant and its micro-canonical ensemble are noted as a possible extension but are not used, since no other draft corroborates its numerical outputs.

D3. Interpretation of "ensemble dependence of the transition's existence" (DIVERGENT in mechanism, CONVERGENT in substance).

- Draft A treats it as a numerical observation about thresholds shifting between ensembles ($\Delta p_c = 0.004$).
- Drafts B and C treat it structurally: C shows the canonical ensemble has a jump (no critical scaling) while the GC ensemble has a tail-driven scaling regime; B attributes possible exponent differences to non-commuting order of limits.
- *Convention adopted*: the structural interpretation (C's closed-form jump argument, with B's limit-ordering mechanism as the physical explanation), because it is derivable rather than simulated. Draft A's small threshold shift is consistent with, but weaker than, this reading.

D4. Status of the Feldman et al. (2026) bound (DIVERGENT in confidence).

- Draft A asserts the bound $v\beta_M \geq 1/2$, extends it to the canonical ensemble, and reports its products (1.30, 1.22) as "saturating" it — though the products exceed $1/2$, so "saturation" is a misnomer in A's own arithmetic.
- Drafts B and C both note the bound is not in the bibliography and cannot be verified; B explicitly labels its assumed bound value as an assumption.
- *Convention adopted*: the bound is discussed only as reported in $[1,2]$; we treat its value, form, and validity as unverified. Where Part II invokes a bound-like constraint, it is labeled an assumption, not a verified theorem.

Should the Feldman et al. (2026) reference be located and confirmed, its statement would replace the assumed bound of Part II.

D5. Summary. All divergences above are resolved in favor of analytically derivable or explicitly labeled content; single-source numerical claims are quarantined in Appendix B.

Appendix B. Single-source claim registry

The following claims originate from a single draft (Draft A) and could not be cross-validated within this corpus; they are excluded from the main results and listed here for audit purposes only:

1. Simulated exponents and thresholds for a $[[5,1,3]]$ code (10^5 instances): canonical $v = 2.5$, $\beta_F = 0.40$, $\beta_M = 0.52$, $pc = 0.152$; *grand-canonical* $v = 2.0$, $\beta_F = 0.50$, $\beta_M = 0.61$, $pc = 0.148$.
2. Fidelity ratio $FC/FG = 1.34$ at $p = 0.10$.
3. Reported bound products (1.30, 1.22) described as "saturating" the Feldman et al. (2026) bound $v\beta_M \geq 1/2$.
4. Threshold shift $\Delta p_c = 0.004$ between ensembles.

Each item is unverifiable pending independent reproduction of Draft A's simulation or location of the Feldman et al. reference. Should such reproduction succeed, the quarantined values would replace the hypothesis-driven inputs of Part II as noted in D1.