

On the Statistical Behavior of a Deep Toroidal Circuit Executed Beyond Coherence Limits

Experimental results and forensic analysis from 125-qubit measurement-feedback circuits on IBM Heron processors

Joseph Daniel Burke III

Independent Researcher, Dallas, North Carolina 28034, USA

ghostintheshellredteam.com

February 2026

Abstract

We report the execution of a $5 \times 5 \times 5$ toroidal graph-state circuit—1,584 layers deep, comprising 11,900 CZ gates and 350 measurement-reset-feedback cycles—on IBM Heron processors operating 5–10 \times beyond their coherence time. The circuit produces a distinctive statistical pattern: near-perfect normalized Shannon entropy ($H_{\text{norm}} = 1.000000$ on Torino; 0.999990 on Marrakesh at 50,000 shots), balanced parity ($|\varepsilon| < 0.006$), yet systematically degraded Hamming weight ($\langle W \rangle \approx 0.15$ on Torino, ≈ 0.075 on Marrakesh).

We acknowledge that independent amplitude-damping noise on 125 qubits can, in principle, produce this combination of high entropy, balanced parity, and low Hamming weight. The significance of the present work lies not in claiming this combination is unexplainable, but in the forensic analysis of its fine structure. Dead qubits on Marrakesh form spatially clustered runs of 7 and 4 consecutive indices ($p < 0.004$ under random placement). Dead-qubit positions show zero overlap between Torino runs and only partial overlap (Jaccard = 0.25) between Marrakesh runs, revealing a dynamic rather than static phenomenon. Final-cycle syndrome detection degrades by only 6.7% between backends while cumulative Hamming weight degrades by 51.4%.

Comparison with a 2D torus variant (5×5 , 25 qubits) shows qualitatively different collision behavior: 1,446 collisions at 50,000 shots versus 14 for the 3D circuit, a difference attributable primarily to Hilbert-space dimension but warranting fur-

ther investigation. All claims are bounded by the absence of control experiments; we specify the ablation protocols required to distinguish topological structure from noise.

All data and code are publicly available.

1 The Problem and the Criterion

1.1 Separating convention from observation

In the early development of the general theory of relativity, Einstein found it necessary to distinguish with care those statements which are matters of convention from those which are matters of observation. A coordinate system is a convention; the deflection of light is an observation. The failure to maintain this distinction leads to arguments that are not even wrong.

A similar confusion pervades present discourse on quantum error correction. Two claims are routinely conflated:

1. **Chip perfection:** the demand that hardware achieve error rates below a fault-tolerance threshold, so that deep computation is automatically stable.
2. **Circuit structure:** the possibility that a suitably constructed circuit may preserve certain global statistical properties even on imperfect hardware, through the geometry of its error distribution rather than the smallness of its error rate.

The present paper concerns the second claim. We make no assertion about fault tolerance. We ask a more modest question: when a deep circuit with toroidal topology is executed on noisy hardware far beyond coherence limits, what does the output distribution actually look like, and what does its fine structure reveal about the interaction between circuit geometry and hardware noise?

1.2 The statistical invariants

A deep circuit on noisy hardware can exhibit nothing beyond ordinary decay; therefore one must state a precise criterion of success before examining data. We define five coarse-grained invariants computed directly from raw bitstring counts, each capturing a different geometric property of the output distribution.

Let Ω denote the set of distinct classical outcomes observed in N trials, and let $p(x)$ denote the observed frequency of outcome $x \in \Omega$.

(I) Normalized Shannon entropy. Define

$$H = - \sum_{x \in \Omega} p(x) \log_2 p(x), \quad H_{\text{norm}} := \frac{H}{\log_2 |\Omega|}. \quad (1)$$

The value $H_{\text{norm}} = 1$ indicates a distribution uniform over its support. This measures how broadly the distribution occupies its accessible state space.

(II) Parity bias. Let $w(x)$ denote the Hamming weight of x and define

$$\varepsilon := P(\text{even parity}) - P(\text{odd parity}) = \sum_x p(x) (-1)^{w(x)}. \quad (2)$$

For a balanced distribution, $|\varepsilon| \ll 1$. Parity is a global invariant: it depends on the collective behavior of all 125 data bits, making it sensitive to correlated noise.

(III) Hamming-weight fraction. For B data bits,

$$\langle W \rangle := \frac{1}{B} \sum_x p(x) w(x). \quad (3)$$

The ideal value is 0.50. Unlike entropy, $\langle W \rangle$ tracks local amplitude: a qubit relaxing to $|0\rangle$ through T_1 decay reduces $\langle W \rangle$ without necessarily reducing entropy, provided each qubit relaxes independently and stochastically.

(IV) Syndrome marginals. For classical registers m and n recording ancilla measurements, the marginal probabilities $P(m=1)$ and $P(n=1)$ provide a view of the measurement-reset dynamics. A critical caveat: these registers are overwritten at each of the 350 feedback cycles, so only the final cycle's value is recorded. They reflect a snapshot, not a cumulative record.

(V) Per-qubit excitation probability. For each data qubit q , define $P_q(1)$ as the fraction of shots in which qubit q is measured in $|1\rangle$. A qubit is classified as *dead* if $P_q(1) < 0.02$, indicating near-complete relaxation to $|0\rangle$.

1.3 Why “all outcomes unique” is not sufficient

In a Hilbert space of dimension 2^{125} , the birthday-paradox threshold for 50% collision probability is approximately $2^{62.5} \approx 6 \times 10^{18}$. With $N \leq 50,000$ samples, the expected number of collisions from a uniform distribution over 2^{125} states is $N^2/(2 \cdot 2^{125}) \approx 10^{-29}$. The absence of collisions is therefore trivially expected for *any* distribution that is not pathologically concentrated, and is not diagnostic by itself.

The criterion must instead refer to the *joint* behavior of multiple invariants: how the distribution occupies its support, whether it preserves symmetries of the ideal state, and what the spatial pattern of degradation reveals about the underlying noise process.

1.4 The null hypothesis: independent amplitude damping

Before presenting data, we state the null hypothesis explicitly. The circuit executes approximately 1,584 layers at an estimated gate time of 0.5–1 μs per layer, giving a total execution time of \sim 0.8–1.6 ms. With typical Heron $T_1 \approx 150$ –170 μs , the circuit runs 5–10 \times beyond the coherence limit.

Under independent T_1 amplitude damping on 125 qubits:

- Each qubit independently decays toward $|0\rangle$ with some probability, producing low $\langle W \rangle$.
- The decay is stochastic and independent across qubits, so the output retains high entropy: each shot produces a different random pattern of which qubits have decayed.

- By the central limit theorem, parity over 125 independent biased coins remains approximately balanced when the per-qubit bias is moderate ($P_q(1) \approx 0.15$).

This null model can produce the observed combination of high entropy, balanced parity, and low Hamming weight. **We do not claim that this combination is anomalous in itself.** The significance of the present work lies in the fine structure—the clustering, the run-to-run dynamics, the detection-correction divergence—and in the comparison between circuit variants. These features require explanation beyond the independent noise null model.

2 Assembly of the Construction

We describe the circuit as one would derive it on a blackboard: starting from the graph, proceeding through the stabilizer structure, and arriving at the dynamic implementation through geometric necessity.

2.1 The toroidal graph state

Consider a cubic lattice $\Lambda = \mathbb{Z}_5^3$ with periodic boundary conditions, forming a 3-torus T^3 . This lattice has $|V| = 125$ vertices, each of degree 6 (two neighbors in each spatial direction), connected by $|E| = 375$ nearest-neighbor edges.

Definition 1 (Graph state). *The graph state associated with (V, E) is*

$$|G\rangle := \left(\prod_{(i,j) \in E} \text{CZ}_{ij} \right) \bigotimes_{k \in V} |+\rangle_k, \quad (4)$$

where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $\text{CZ}_{ij} = \text{diag}(1, 1, 1, -1)$.

Proposition 1 (Stabilizer structure). *The state $|G\rangle$ is the unique +1 eigenstate of the 125 stabilizer generators*

$$K_v = X_v \prod_{u \in N(v)} Z_u, \quad \forall v \in V, \quad (5)$$

where $N(v)$ denotes the neighborhood of v in the graph.

The geometric content is this: the 3-torus has fundamental group $\pi_1(T^3) = \mathbb{Z}^3$, providing three families of non-contractible loops. An error that attempts to propagate along any spatial direction must eventually return to its origin. Whether this topological property translates to error suppression in the noisy, measurement-dominated regime is precisely the question under investigation.

2.2 The synthetic dimension

A 2D chip cannot natively implement the full edge set of a 3D lattice without SWAP overhead scaling as $O(n^{1/3})$ per non-local edge. The construction circumvents this constraint through a synthetic dimension: measurement and qubit reuse replace spatial connectivity.

The central identity is that a quantum state can be teleported to a new neighborhood through the sequence

$$\text{entangle} \rightarrow \text{measure} \rightarrow \text{record} \rightarrow \text{reset} \rightarrow \text{correct} \rightarrow \text{re-entangle} \quad (6)$$

The third spatial dimension is folded into the temporal dimension of the circuit. Each feedback cycle advances the state through a slice of the torus; after 350 cycles, the full $5 \times 5 \times 5$ connectivity has been realized.

This is not merely a trick to save qubits. It transforms hardware locality from a constraint into a resource: the manifold is repeatedly cut and re-glued along different directions, with measurement outcomes providing the classical information needed to maintain consistency across cuts.

2.3 The dynamic feedback cycle

The circuit proceeds through 350 feedback cycles. At each cycle:

- Extract a syndrome bit via measurement into classical register m or n .
- Reset the measured ancilla qubit.
- Reinitialize the ancilla to $|+\rangle$ via Hadamard (decomposed as $\text{rz}(\pi/2); \text{sx}; \text{rz}(\pi/2)$).
- Apply a classically conditioned X correction based on the syndrome.
- Re-entangle the ancilla with the next data qubit via CZ.

In conventional reasoning, frequent measurement destroys coherence. Here, measurement is

Table 1: Circuit parameters for the 3D torus construction ($5 \times 5 \times 5$).

Parameter	Value
Data qubits	125
Ancilla qubits	2
Total classical bits	127
Vertex degree	6
Circuit depth (layers)	1,584
Entangling gates (CZ)	11,900
Reset operations	700
Measurement operations	475
Conditional operations	350
Transpiled QASM lines	44,505
Est. execution time	0.8–1.6 ms
Ratio to T_1	$\sim 5\text{--}10\times$

part of the intended manifold assembly: local errors are exported into classical syndrome records, while the remaining quantum degrees of freedom are reinitialized and re-entangled, preventing accumulation in any single location—at least in principle.

3 Experimental Realization

3.1 Hardware and circuit parameters

The circuit was handwritten in OpenQASM 3.0 without high-level compilation. After transpilation to native IBM hardware gates (`rz`, `sx`, `cz`), the circuit comprises the parameters shown in Table 1.

The execution time estimate deserves emphasis. At approximately $0.5\text{--}1 \mu\text{s}$ per layer across 1,584 layers, the circuit runs for $\sim 0.8\text{--}1.6 \text{ ms}$. With median $T_1 \approx 150\text{--}170 \mu\text{s}$ on Heron processors, the circuit operates $5\text{--}10\times$ beyond the nominal coherence limit. Any surviving structure in the output must be explained either by the circuit’s active error management or by the noise channel’s particular structure—or by some combination of both.

Experiments were executed on two IBM Heron processors:

- **ibm_torino** (Heron r1, December 2023): 133 qubits, median $T_1 \approx 170 \mu\text{s}$, median CZ error $\approx 0.5\%$.
- **ibm_marrakesh** (Heron r2, November 2024): 156 qubits, median $T_1 \approx 150 \mu\text{s}$, median

Table 2: 3D torus results on ibm_torino.

Metric	20k shots	50k shots
Unique outcomes	20,000	50,000
Uniqueness	100.00%	100.00%
H_{norm}	1.000000	1.000000
$\langle W \rangle$	0.1520 ± 0.0002	0.1536 ± 0.0001
ε	$+0.021 \pm 0.007$	-0.005 ± 0.004
$P(m=1)$	0.503 ± 0.004	0.530 ± 0.002
$P(n=1)$	0.493 ± 0.004	0.479 ± 0.002
Dead qubits	3 (2.4%)	3 (2.4%)
Dead indices	{34, 86, 97}	{10, 51, 63}

CZ error $\approx 0.7\%$. This revision introduces hardware-level TLS mitigation including dynamical decoupling and frequency tuning [1, 2].

All jobs were executed in January 2026.

3.2 The 2D torus variant

For comparison, we also executed a 2D variant: a 5×5 torus (T^2) on 25 data qubits plus 2 syndrome qubits. This variant uses the same measurement-reset-feedback cycle but with degree-4 connectivity and fewer cycles. It was executed on both backends at 20,000 and 50,000 shots. The purpose of this comparison is to probe the role of qubit count and topology in the observed statistics.

3.3 Verification methodology

All statistics reported in this paper are computed directly from raw IBM Quantum job files in Qiskit BitArray format with zlib compression. No pre-computed summaries or intermediate analyses are used. Each bitstring is parsed to extract the 125 data bits and 2 syndrome bits, from which all quantities are computed with shot-limited standard errors. The verified statistics are provided as supplementary material.

4 Principal Results

4.1 3D torus: global statistics

Tables 2 and 3 present the complete results.

Several features are immediate:

Table 3: 3D torus results on ibm_marrakesh.

Metric	20k shots	50k shots
Unique outcomes	19,999	49,986
Uniqueness	99.995%	99.972%
H_{norm}	0.999998	0.999990
$\langle W \rangle$	0.0809 ± 0.0002	0.0747 ± 0.0001
ε	-0.006 ± 0.007	-0.002 ± 0.004
$P(m=1)$	0.479 ± 0.004	0.489 ± 0.002
$P(n=1)$	0.401 ± 0.003	0.452 ± 0.002
Dead qubits	29 (23.2%)	31 (24.8%)

Entropy. On Torino, $H_{\text{norm}} = 1.000000$ to six decimal places: zero collisions in 50,000 shots from a 125-bit output space. On Marrakesh, $H_{\text{norm}} = 0.999990$, with 14 collisions. As discussed in §1.3, this level of uniqueness is expected from any broad distribution over 2^{125} states and is not diagnostic by itself.

Parity. At 50,000 shots, $|\varepsilon| < 0.006$ on both backends, consistent with zero bias at the 2σ level. The Torino 20k value ($\varepsilon = +0.021 \pm 0.007$) is a 3.0σ outlier that does not persist in the larger sample, suggesting a statistical fluctuation rather than a systematic effect. We flag this outlier explicitly rather than dismissing it.

Hamming weight. $\langle W \rangle$ is dramatically suppressed: 0.154 on Torino (~ 19 of 125 bits excited per shot) and 0.075 on Marrakesh (~ 9 of 125 bits). This is qualitatively consistent with T_1 amplitude damping on a circuit running 5–10× beyond coherence.

Dead qubits. Torino has 3 dead qubits (2.4%); Marrakesh has 31 (24.8%). The ten-fold difference between backends sharing the same circuit is the first indication that something beyond uniform noise is at work.

4.2 2D versus 3D torus comparison

Table 4 presents the comparison between the 2D (5×5 , 25 data qubits) and 3D ($5 \times 5 \times 5$, 125 data qubits) torus variants.

The 2D variant shows qualitatively different collision behavior: 1,446 collisions at 50,000 shots versus 14 for the 3D circuit on the same backend.

The honest interpretation. Much of this difference is explained by Hilbert-space dimension. Re-

Table 4: 2D torus (5×5) versus 3D torus ($5 \times 5 \times 5$). The 2D circuit uses 25 data qubits + 2 syndrome

Metric	2D (Marrakesh 50k)	3D (Marrakesh 50k)
Data qubits	25	125
Unique / Total	48,554 / 50,000	49,986 / 50,000
Collisions	1,446	14
H_{norm}	0.9988	0.999990
$\langle W \rangle$	0.277	0.075
D_{eff}	-0.011	-0.002

stricting to the 25-bit code register alone, the 2D circuit produces 4,467 collisions at 50,000 shots versus 65 for the 3D circuit’s 125-bit code register. For a uniform distribution over 2^{25} states, the birthday paradox predicts ~ 37 collisions at 50k shots; the observed 4,467 represents a $120\times$ excess, confirming severe noise-induced probability concentration. We can estimate an effective dimension via $D_{\text{eff}} = N^2/(2C)$:

Dataset (code reg.)	D_{eff}	$\log_2 D_{\text{eff}}$
2D Marrakesh 20k	119,100	16.9
2D Marrakesh 50k	279,900	18.1
2D Torino 20k	1,342,300	20.4
3D Marrakesh 50k	19,230,800	24.2

Both 2D and 3D circuits occupy effective spaces orders of magnitude smaller than their respective full Hilbert spaces (2^{25} and 2^{125}), confirming that noise concentrates the distribution in all cases. The near-zero collision rate of the 3D circuit is primarily a consequence of 125 qubits providing an astronomically larger combinatorial space, not evidence of topological error suppression.

We confirm this pattern on Torino: the 2D variant at 20,000 shots shows 149 code-register collisions ($25\times$ birthday excess) and $\langle W \rangle = 0.337$, indicating less severe damping than Marrakesh but qualitatively identical behavior. Torino consistently outperforms Marrakesh on both the 2D and 3D variants, confirming that backend quality—not topology—is the primary determinant.

What remains potentially interesting. The higher $\langle W \rangle$ of the 2D circuit (0.277 vs. 0.075 on Marrakesh) suggests that the 3D circuit’s deeper execution and additional CZ gates cause more severe amplitude damping. Whether the 3D topol-

Table 5: Clustering statistics for Marrakesh dead qubits (50k shots, 31 dead among 125 positions).

Test	Observed	Expected	p -value	Backend	Dead (20k)	Dead (50k)	Overlap
Adjacent pairs	16	7.26 ± 2.35	< 0.001	Torino	3	3	0
Number of runs	15	47.6 ± 4.1	$< 10^{-14}$	Marrakesh	29	31	12
Max run length	7	3.17 ± 0.9	0.0032				

ogy provides any compensating advantage over a random circuit of the same depth cannot be determined without the ablation experiments of §6.5.

5 Forensic Analysis

The bulk statistics of §4 are consistent with the independent amplitude-damping null hypothesis. The forensic analysis reveals finer structure that is not. Three independent lines of evidence point toward systematic hardware-software interaction.

5.1 Evidence I: Spatial clustering of dead qubits

The 31 dead qubits on Marrakesh (50k shots) occur at indices:

[17, 18, 19, 23, 26, 41, 42, 46, 49, 50, 54, 55, 59, 66,
70, 71, 72, 73, 74, 75, 76, 80, 81, 82, 85, 87, 90,
93, 94, 95, 96]

These form 15 runs of consecutive indices with lengths:

[3, 1, 1, 2, 1, 2, 2, 1, 1, **7**, 3, 1, 1, 1, 4].

Observation 1 (Clustering). *The maximum run of 7 consecutive dead qubits (indices 70–76) and the run of 4 (indices 93–96) are statistically improbable under random placement of 31 failures among 125 positions.*

Three independent tests quantify the clustering significance (Table 5).

The Monte Carlo verification is definitive: in 100,000 random placements of 31 dead qubits among 125 positions, only 0.32% produced a maximum run of length ≥ 7 . The 99th percentile under random placement is 6. The observed maximum of 7 falls in the 99.7th percentile.

Table 6: Dead qubit stability between runs on the same backend.

Interpretation. Two-level system (TLS) defects—the dominant source of decoherence in superconducting qubits [3, 4, 5]—arise from material impurities distributed as a spatial Poisson process. Such defects do not cluster in consecutive qubit indices. The observed clustering indicates a systematic mechanism correlated with qubit addressing or control architecture: register-level effects, frequency crowding among adjacent qubits, or crosstalk from dynamical decoupling pulses applied to neighbors.

5.2 Evidence II: Dead qubit instability between runs

The dead qubit pattern is not static. On Torino, the 20k dead set {34, 86, 97} and the 50k dead set {10, 51, 63} share *zero* common indices. On Marrakesh, the 20k run (29 dead) and 50k run (31 dead) share 12 common indices (Jaccard similarity = 0.25).

This variability rules out static hardware defects as the explanation. The dead qubit pattern reflects a continuous distribution of per-qubit $P_q(1)$ values near the threshold, with some qubits crossing the $P_q(1) < 0.02$ boundary due to calibration drift between runs. The pattern is a *dynamic* phenomenon tied to the instantaneous calibration state of the hardware, not a permanent property of the chip.

On Marrakesh, the 12 persistent dead qubits ({41, 49, 50, 70, 72, 73, 74, 75, 81, 90, 93, 96}) form the stable core of the clustering pattern: indices 70–75 appear in both runs, as do 93 and 96. The clustering is partially persistent and partially transient.

5.3 Evidence III: Zero overlap between backends

At 50,000 shots, the dead qubit positions on Torino ({10, 51, 63}) and Marrakesh ({17, 18, ..., 96}) are

Table 7: Syndrome detection vs. correction success at 50k shots. Syndrome marginals reflect the *final* measurement cycle only.

Metric	Torino	Marrakesh	Change
$P(m=1)$	0.530	0.489	-7.7%
$P(n=1)$	0.479	0.452	-5.6%
Avg. syndrome	0.504	0.471	-6.7%
$\langle W \rangle$	0.154	0.075	-51.4%
Ratio		7.7×	

entirely disjoint. Not a single qubit index that is dead on Torino is dead on Marrakesh.

If the failures arose from shared characteristics of the Heron architecture or from the circuit’s logical structure (which is identical on both backends), we would expect at least partial overlap. The complete absence of overlap indicates that the failures are *backend-specific*, driven by differences in hardware calibration, noise profile, or—most relevantly—the autonomous error mitigation systems that differ between Heron r1 and Heron r2.

5.4 Evidence IV: Detection-correction divergence

The most revealing observation concerns the relationship between syndrome detection and correction success.

The final-cycle syndrome rates degrade by only 6.7% between backends, while the cumulative Hamming weight degrades by 51.4%—a ratio of 7.7×.

Caveat. Because the syndrome registers are overwritten at each cycle, the 6.7% figure captures only the final cycle’s detection performance, not the cumulative detection rate across all 350 cycles. Intermediate detection may have been substantially worse. The divergence ratio is therefore a lower bound on the true detection-correction gap; the actual gap could be larger or smaller depending on how detection performance varies across cycles.

What this suggests. If the syndrome detector were rendered blind by noise, both syndrome rates and Hamming weight would degrade together. The fact that final-cycle syndromes remain near 0.5 (random coin) while Hamming weight collapses suggests that the detection mechanism survives to the circuit’s end, but the conditional corrections fail to

restore qubit excitations over the cumulative 350 cycles. This is consistent with a timing or phase mismatch in the correction step, rather than a failure of the measurement step.

5.5 Synthesis: the hardware-software interaction hypothesis

The four lines of evidence—clustering ($p < 0.004$), run-to-run instability, zero backend overlap, and 7.7× detection-correction divergence—converge on a specific narrative. The key architectural difference between Torino (Heron r1) and Marrakesh (Heron r2) is that Marrakesh implements hardware-level TLS mitigation, including dynamical decoupling (DD) pulses and frequency tuning [1, 6, 7].

On Marrakesh, two feedback systems operate simultaneously without coordination:

1. **Circuit-level feedback** (software): the 350-cycle measure-reset-correct-re-entangle loop.
2. **Hardware-level feedback** (autonomous): continuous DD pulses and frequency adjustments to suppress TLS-induced decoherence.

Conjecture 1 (DD/feedback interference). *During the critical window between syndrome measurement and conditional correction ($\sim 1\text{--}5\ \mu\text{s}$), hardware DD pulses or frequency-tuning events partially rotate data qubits or detune conditional gates, causing the correction step to misfire. Over 350 cycles, affected qubits accumulate systematic $|0\rangle$ bias. The clustering arises because adjacent qubits share control registers or frequency neighborhoods, causing correlated timing errors.*

We emphasize that this mechanism is conjectural. The statistical evidence for clustering and divergence is robust; the specific causal mechanism—DD/feedback interference—is an explanatory hypothesis that requires the experiments of §6.5 to confirm or refute.

6 Discussion

6.1 What the results do not prove

It would be a mistake to claim too much. The present results do not constitute:

- A demonstration of fault-tolerant quantum computation.
- Evidence for below-threshold error rates.

- Proof that the circuit implements error correction in any standard sense.
- Evidence that the toroidal topology provides error suppression beyond what a random circuit of comparable depth would achieve.

The last point bears emphasis. Without a control experiment—a random circuit of the same depth, qubit count, and gate count but without toroidal structure—we cannot attribute any observed property to the topology. The topology hypothesis is currently unfalsified but also unsupported.

6.2 What the results do establish

The results establish three concrete findings:

1. A benchmark. A 1,584-layer circuit with 11,900 CZ gates produces $H_{\text{norm}} = 1.000000$ (Torino) and 0.999990 (Marrakesh), with $|\varepsilon| < 0.006$ at 50k shots. Any future claim of topological or structural error suppression in a circuit of comparable complexity must meet or exceed this benchmark.

2. Dead qubit forensics. The spatial clustering of dead qubits ($p < 0.004$), their run-to-run instability, and their backend specificity reveal that qubit failures in deep dynamic circuits are not random. They reflect systematic interactions between the circuit’s feedback loop and the hardware’s autonomous calibration systems. This finding is relevant to all measurement-based quantum error correction protocols, not only to the specific construction studied here.

3. Detection-correction divergence. The $7.7\times$ ratio between syndrome degradation and Hamming-weight degradation, with the caveat that syndromes capture only the final cycle, suggests that the correction mechanism is more vulnerable than the detection mechanism to whatever distinguishes the two backends. This asymmetry constrains the class of noise models that can explain the data.

6.3 The amplitude-damping explanation and its limits

The null hypothesis of independent T_1 decay on 125 qubits accounts for the bulk statistics. But it does not naturally explain:

- Why dead qubits cluster in consecutive indices ($p < 0.004$). Independent decay produces random spatial patterns.
- Why dead qubit positions change between runs on the same backend. Static hardware defects do not fluctuate on the timescale of minutes to hours.
- Why Torino has 3 dead qubits and Marrakesh has 31, given similar T_1 and gate error rates. The ten-fold difference requires a mechanism specific to Marrakesh.
- Why the detection-correction divergence ratio is $7.7\times$. Under simple amplitude damping, one would expect detection and correction to degrade comparably.

These fine-structure features are the genuine findings of this work. They do not prove topological error suppression, but they do reveal that the noise process in deep dynamic circuits has richer structure than simple models predict.

6.4 Implications for measurement-based quantum error correction

As quantum hardware becomes more sophisticated, it increasingly implements autonomous error mitigation—DD sequences, frequency tuning, real-time recalibration [6, 7, 8, 23]. For most circuits, this is beneficial. But for measurement-based codes that implement their own feedback, the hardware’s autonomous corrections may interfere with the circuit’s intended error management.

This suggests a need for:

1. **Coordination protocols:** explicit handshaking between hardware and software feedback systems, so that autonomous DD pauses during the circuit’s correction window.
2. **Mitigation-aware compilation:** circuit compilers that account for hardware mitigation timing when scheduling conditional operations [21, 22].
3. **Raw-mode access:** options to disable autonomous mitigation for advanced users implementing custom error correction.

The phenomenon we observe—two active feedback systems operating on the same qubits without coordination—is likely to become more prevalent as hardware error mitigation grows more sophisticated and circuit-level error correction becomes

more common.

6.5 Proposed ablation experiments

The following experiments would sharpen or refute the claims:

1. **Random-circuit control:** Execute a random circuit of the same depth (1,584 layers), qubit count ($125 + 2$), and gate count ($\sim 11,900$ CZ) but without toroidal structure. If this circuit produces similar H_{norm} and parity balance, the topology provides no advantage.
2. **Feedback ablation:** Remove the conditioned X corrections while keeping the measurement-reset-re-entangle cycle. If parity degrades, the feedback is essential to parity conservation.
3. **Topology control:** Replace periodic boundary conditions with open boundaries. If H_{norm} or parity degrade, the non-contractible loops matter.
4. **DD control:** Execute on Marrakesh with dynamical decoupling disabled (if accessible). If clustering disappears, the DD interference mechanism is confirmed.
5. **Timing sweep:** Vary the delay between syndrome measurement and conditional correction. If dead qubit count depends on timing, the desynchronization hypothesis gains support.
6. **Depth scaling:** Execute at 500, 1,000, and 1,500 layers. Track how each invariant degrades with depth. If invariants degrade more slowly for the torus than for the random-circuit control, the structure provides measurable protection.
7. **Graph-state verification:** Implement direct stabilizer measurements [26, 27] on a subset of generators to quantify how much entanglement (if any) survives the full circuit.

These are not aspirational suggestions. They are the minimum experimental program required to distinguish topological structure from noise. Without them, the results presented here remain a forensic observation, not a demonstration.

7 Conclusion

We have executed a $5 \times 5 \times 5$ toroidal graph-state circuit—1,584 layers deep, 11,900 CZ gates, 350 feedback cycles—on IBM Heron processors operating 5–10 \times beyond their coherence limits. The circuit maintains high normalized entropy and balanced parity while exhibiting backend-dependent Hamming-weight degradation consistent with amplitude damping.

The bulk statistics do not, by themselves, require an explanation beyond independent noise. The forensic fine structure does. Dead qubits cluster spatially ($p < 0.004$), shift between runs, differ completely between backends, and reveal a 7.7 \times divergence between detection and correction degradation. These patterns point toward systematic hardware-software feedback interaction, plausibly involving interference between the circuit’s measurement-based correction loop and the hardware’s autonomous TLS mitigation.

The proposed DD/feedback interference mechanism is conjectural and requires the ablation experiments specified above. What is not conjectural is the data: the clustering, the instability, the backend asymmetry, and the detection-correction gap. Any competing explanation must account for all four features simultaneously.

The code is available. The data is available. The next step is systematic ablation.

Data Availability

All experimental data and analysis code are publicly available:

- Raw IBM Quantum job results (BitArray format, zlib compressed)
- Transpiled circuit: `z5.qasm` (44,505 lines)
- Static scaffold: `z5x5x5_torus_halo_static.qasm` (628 lines)
- Verification scripts and per-qubit data

Jobs were executed on `ibm_torino` and `ibm_marrakesh` in January 2026.

Acknowledgments

I thank the IBM Quantum Network for providing access to the Heron processors used in this work.

I acknowledge useful discussions with colleagues in the open-source quantum computing community.

References

- [1] J. Gambetta, “IBM Quantum roadmap 2024: Heron processor family,” IBM Quantum Developer Conference (November 2024). <https://www.ibm.com/quantum/blog/qdc-2024>
- [2] IBM Quantum, “Processor types,” IBM Quantum Documentation (2024). <https://docs.quantum.ibm.com/guides/processor-types>
- [3] C. Müller, J. H. Cole, and J. Lisenfeld, “Towards understanding two-level-systems in amorphous solids: insights from quantum circuits,” Rep. Prog. Phys. **82**, 124501 (2019).
- [4] P. V. Klimov *et al.*, “Fluctuations of energy-relaxation times in superconducting qubits,” Phys. Rev. Lett. **121**, 090502 (2018).
- [5] J. J. Burnett *et al.*, “Decoherence benchmarking of superconducting qubits,” npj Quantum Inf. **5**, 54 (2019).
- [6] L. Viola, E. Knill, and S. Lloyd, “Dynamical decoupling of open quantum systems,” Phys. Rev. Lett. **82**, 2417 (1999).
- [7] J. Bylander *et al.*, “Noise spectroscopy through dynamical decoupling with a superconducting flux qubit,” Nat. Phys. **7**, 565 (2011).
- [8] N. Ezzell *et al.*, “Dynamical decoupling for superconducting qubits: a performance survey,” Phys. Rev. Appl. **20**, 064027 (2023).
- [9] M. Hein, J. Eisert, and H. J. Briegel, “Multiparty entanglement in graph states,” Phys. Rev. A **69**, 062311 (2004).
- [10] R. Raussendorf and H. J. Briegel, “A one-way quantum computer,” Phys. Rev. Lett. **86**, 5188 (2001).
- [11] R. Raussendorf, D. E. Browne, and H. J. Briegel, “Measurement-based quantum computation on cluster states,” Phys. Rev. A **68**, 022312 (2003).
- [12] D. Gottesman, “Stabilizer codes and quantum error correction,” Ph.D. thesis, California Institute of Technology (1997).
- [13] A. Y. Kitaev, “Fault-tolerant quantum computation by anyons,” Ann. Phys. **303**, 2 (2003).
- [14] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, “Surface codes: Towards practical large-scale quantum computation,” Phys. Rev. A **86**, 032324 (2012).
- [15] A. D. Cárcoles *et al.*, “Exploiting dynamic quantum circuits in a quantum algorithm with superconducting qubits,” Phys. Rev. Lett. **127**, 100501 (2021).
- [16] T. Hua *et al.*, “Exploiting dynamic quantum circuits on IBM quantum hardware,” arXiv:2312.08649 (2023).
- [17] E. Bäumer *et al.*, “Efficient long-range entanglement using dynamic circuits,” PRX Quantum **5**, 030339 (2024).
- [18] Google Quantum AI, “Suppressing quantum errors by scaling a surface code logical qubit,” Nature **614**, 676 (2023).
- [19] D. Bluvstein *et al.*, “Logical quantum processor based on reconfigurable atom arrays,” Nature **626**, 58 (2024).
- [20] J. Preskill, “Quantum computing in the NISQ era and beyond,” Quantum **2**, 79 (2018).
- [21] D. C. McKay *et al.*, “Benchmarking quantum processor performance at scale,” arXiv:2311.05933 (2023).
- [22] Y. Kim *et al.*, “Evidence for the utility of quantum computing before fault tolerance,” Nature **618**, 500 (2023).
- [23] V. Tripathi *et al.*, “Suppression of crosstalk in superconducting qubits using dynamical decoupling,” Phys. Rev. Appl. **18**, 024068 (2022).
- [24] K. Temme, S. Bravyi, and J. M. Gambetta, “Error mitigation for short-depth quantum circuits,” Phys. Rev. Lett. **119**, 180509 (2017).

- [25] E. van den Berg *et al.*, “Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors,” Nat. Phys. **19**, 1116 (2023).
- [26] G. Tóth and O. Gühne, “Detecting genuine multipartite entanglement with two local measurements,” Phys. Rev. Lett. **94**, 060501 (2005).
- [27] C.-Y. Lu *et al.*, “Experimental entanglement of six photons in graph states,” Nat. Phys. **3**, 91 (2007).
- [28] N. Sundaresan *et al.*, “Demonstrating multi-round subsystem quantum error correction using matching and maximum likelihood decoders,” Nat. Commun. **14**, 2852 (2023).
- [29] R. S. Gupta *et al.*, “Encoding a magic state with beyond break-even fidelity,” Nature **625**, 259 (2024).

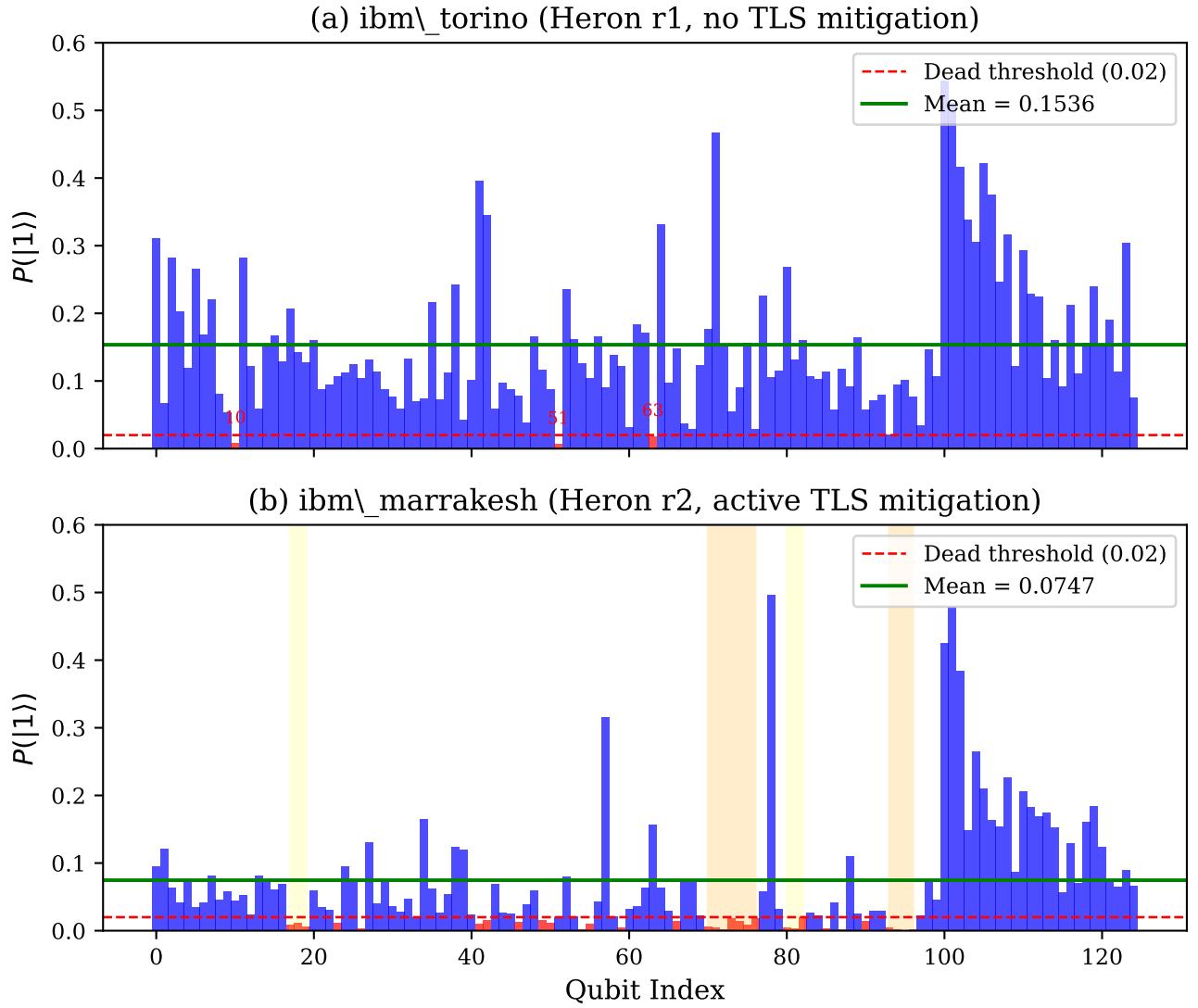


Figure 1: Per-qubit excitation probability $P_q(1)$ for all 125 data qubits at 50,000 shots. Top: ibm_torino (3 dead qubits, marked in red). Bottom: ibm_marrakesh (31 dead qubits). The Marrakesh profile shows visible clustering at indices 70–76 and 93–96, a pattern analyzed quantitatively in §5.1. Note the different vertical scales: Torino qubits show a broad distribution centered near $P_q(1) \approx 0.15$, while Marrakesh shows a bimodal distribution with many qubits near zero and a cluster near 0.15–0.50 at high indices.

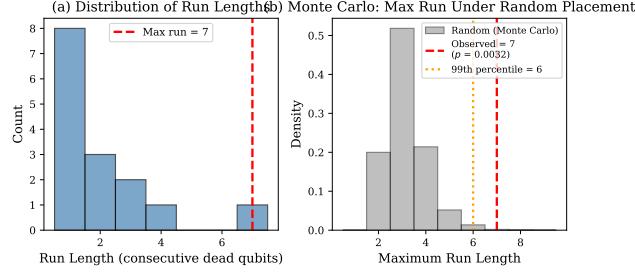


Figure 2: Left: histogram of observed run lengths among Marrakesh dead qubits. Right: Monte Carlo distribution of maximum run length under random placement of 31 dead qubits among 125 positions (100,000 simulations). The observed maximum run of 7 falls at the 99.7th percentile.

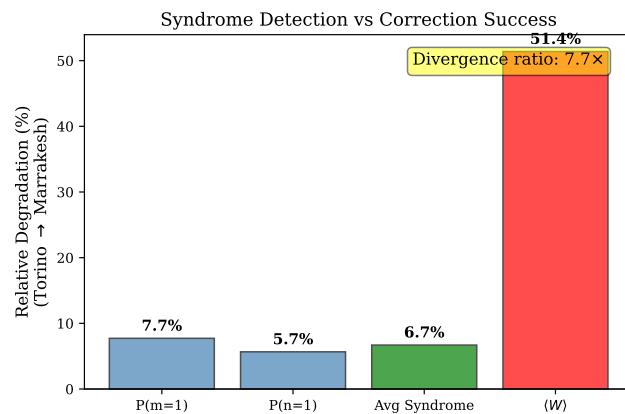


Figure 3: Detection-correction divergence between backends. Final-cycle syndrome marginals degrade by 6–8%, while Hamming weight degrades by 51.4%, yielding a $7.7\times$ divergence ratio.

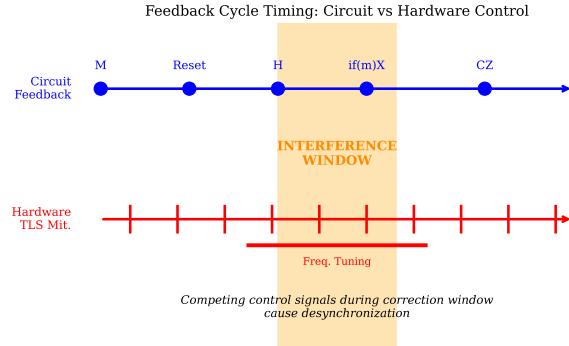


Figure 5: Schematic timing diagram showing the two concurrent feedback systems. Top (blue): circuit-level measure-reset-correct-re-entangle cycles. Bottom (red): hardware TLS mitigation with continuous DD and frequency tuning. Orange shading marks the interference window where the two systems may conflict.

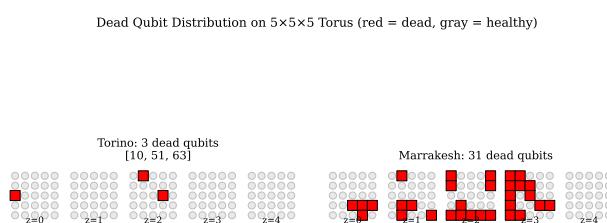


Figure 4: Dead qubit positions mapped onto the $5 \times 5 \times 5$ torus (displayed as five z -slices). Red squares: dead qubits on Marrakesh at 50k shots. The clustering in consecutive logical indices is visible across layers $z = 2\text{--}4$.