

Circuit-Level vs Pulse-Level Quantum Hardware Diagnostics: K–R Scaling Exponent Detects Device Degradation Invisible to T_1/T_2 Calibration

RamaKrishna Pasupuleti

workisfun415@gmail.com

INDEPENDENT RESEARCHER <https://orcid.org/0009-0008-8418-1430>

Research Article

Keywords: quantum benchmarking, noise characterization, NISQ devices, superconducting qubits, quantum diagnostics, coherence time, circuit-level noise, IBM Quantum, K–R framework

Posted Date: July 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10394517/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Circuit-Level vs Pulse-Level Quantum Hardware Diagnostics: K–R Scaling Exponent Detects Device Degradation Invisible to T_1/T_2 Calibration

Ramakrishna Pasupuleti

Independent Researcher, Suryapet, Telangana 508213, India

ORCID: 0009-0008-8418-1430 | workisfun415@gmail.com

ABSTRACT

Pulse-level calibration metrics — coherence times T_1 and T_2 measured through single-qubit Rabi and Ramsey sequences — are the standard by which quantum hardware providers certify device health. We demonstrate, through 191 circuit executions on three 156-qubit IBM Heron r2 processors (ibm_kingston, ibm_fez, and ibm_marrakesh), that these pulse-level metrics can give opposite conclusions to circuit-level diagnostics on the same device at the same time. Specifically, between experimental sessions on July 3 and July 5, 2026, IBM reported ibm_marrakesh T_2 improved from 303 μs to 340 μs (a 12% improvement), while the K–R scaling exponent α measured from GHZ circuit scaling increased from 1.406 to 3.811 (a 171% increase in coherent error dominance) and GHZ $n=2$ infidelity rose from 0.018 to 0.204 — a 920% increase in circuit error. We further demonstrate that Ramsey circuit-context T_2 on qubit 0 is approximately 9–30 μs , compared to IBM-reported pulse-level T_2 of 32–340 μs — a discrepancy of 3–11 $\times$. We introduce five experimental demonstrations showing K–R circuit-level diagnostics detect degradation invisible to T_1/T_2 , provide higher temporal sensitivity, enable algorithm failure prediction, and confirm the theoretical prediction $\alpha_{\text{vt}}^{\text{ch}} + \alpha_{\text{etho}} = 2$ within 7% on real hardware. These findings establish circuit-level K–R diagnostics as a necessary complement to pulse-level calibration for operational quantum computing.

Index Terms—*quantum benchmarking, noise characterization, NISQ devices, superconducting qubits, quantum diagnostics, coherence time, circuit-level noise, IBM Quantum, K–R framework.*

I. INTRODUCTION

The canonical procedure for certifying the health of a quantum processing unit (QPU) relies on pulse-level measurements: T_1 (amplitude damping time) measured through inversion recovery sequences, and T_2 (dephasing time) measured through Ramsey or spin-echo sequences applied to isolated single qubits. These metrics are fast, well-understood, and routinely reported by quantum hardware providers including IBM Quantum, Google, IonQ, and Quantinuum as the primary indicators of device quality [1],[2],[3]. A device with long T_1 and T_2 is considered high-quality; a device with short coherence times is flagged for recalibration.

However, T_1 and T_2 are measured on isolated qubits under idealized pulse conditions. They do not capture: (i) coherent errors from systematic gate miscalibration, which accumulate super-linearly with circuit depth and are invisible to single-qubit decoherence measurements; (ii) crosstalk between simultaneously operated qubits, which grows with circuit width; (iii) the combined effect of gate compilation, qubit connectivity, and error propagation through entangling operations; or (iv) the temporal stability of these effects during actual circuit execution sessions. In short, T_1 and T_2 measure what a device can do in ideal single-qubit conditions — not what it actually does when running multi-qubit algorithms.

The consequence is a systematic gap between certified hardware quality and actual algorithm performance. A device that passes all T_1/T_2 benchmarks may nonetheless produce unreliable results for multi-qubit circuits, and vice versa. Identifying this gap requires circuit-level diagnostics — experiments that probe error accumulation in the context of actual multi-qubit gate sequences rather than isolated single-qubit operations.

In this work we demonstrate this gap directly and quantitatively. We show that on `ibm_marrakesh` (156-qubit IBM Heron r2), between July 3 and July 5, 2026, IBM’s pulse-level T_2 improved from 303 μs to 340 μs — certifying the device as healthy. Simultaneously, circuit-level measurements using the K–R scaling exponent α showed a 171% increase in coherent error dominance and a 920% increase in GHZ circuit infidelity. The two diagnostic frameworks gave opposite conclusions about device health from the same device at the same time.

We further demonstrate that circuit-context Ramsey T_2 is $3\text{--}11\times$ shorter than IBM-reported pulse-level T_2 on the same qubits. This discrepancy — consistently observed across multiple qubits and sessions — indicates that ZZ crosstalk, gate compilation artifacts, and context-dependent dephasing in circuit environments substantially exceed what pulse-level measurements capture.

Our diagnostic method is the K–R framework [4], which extracts a scaling exponent α from the three-parameter infidelity model:

$$C = C_0/K^\alpha + R \quad (1)$$

where C is the measured infidelity, $K = 1/d$ is the inverse circuit depth (or inverse qubit count), C_0 is the correctable error amplitude, R is the depth-independent noise floor, and α quantifies the rate of error accumulation. The exponent α classifies the dominant noise regime: $\alpha > 1$ indicates structured, coherent error accumulation; $0.3 < \alpha < 1$ indicates stochastic, Markovian-like accumulation; $\alpha \rightarrow 0$ indicates saturation where depth-independent noise dominates. The

framework requires only 7–13 circuits per diagnostic axis, taking approximately 3–5 minutes of wall time on IBM Quantum Open Plan hardware.

The key contributions of this paper are:

(C1) Direct experimental demonstration that K–R α and IBM T_2 give opposite conclusions about device health on the same device at the same time.

(C2) Quantification of the circuit-context vs pulse-level T_2 discrepancy ($3\text{--}11\times$ gap) across multiple qubits.

(C3) First hardware confirmation of the theoretical prediction $\alpha_{\text{vt}}^{\text{ch}} + \alpha_{\text{etho}} = 2$ within 7% (Proof 4, Appendix B).

(C4) Real-time intra-session drift detection: α tracked monotonically from $2.99 \rightarrow 3.18 \rightarrow 3.78$ over three consecutive runs while IBM T_2 remained nominally constant.

(C5) Demonstration that single-point infidelity measurement misclassifies a drifting device as “Low Error” while K–R correctly identifies it as Structured (coherent-dominated).

The remainder of this paper is organized as follows. Section II reviews related work on quantum benchmarking and circuit-level noise characterization. Section III describes the K–R theoretical framework and the Proof 4 prediction. Section IV presents the experimental hardware setup and methodology. Section V presents all experimental results. Section VI provides detailed analysis and interpretation. Section VII discusses implications for quantum hardware operations. Section VIII addresses limitations and future work. Section IX concludes.

II. RELATED WORK

A. Pulse-Level Benchmarking

Randomized benchmarking (RB) [5] provides a gate-error rate per Clifford gate through exponential decay of fidelity under random Clifford sequences. While rigorous and widely adopted, RB requires 100–300 circuits and averages errors across all gate types, losing information about the nature (coherent vs. stochastic) of the underlying noise. Gate set tomography (GST) [6] provides complete characterization of gate errors but requires 500–2000 circuits and hours of QPU time. Cross-entropy benchmarking (XEB) [7] estimates circuit fidelity through comparison with classical simulation but scales exponentially in qubit count. None of these methods distinguish coherent from stochastic error accumulation with the efficiency of the K–R framework.

B. Circuit-Level Noise Characterization

Several recent works have identified the gap between pulse-level and circuit-level noise characterization. Sarovar et al. [8] demonstrated that crosstalk errors in multi-qubit circuits substantially exceed what single-qubit benchmarks predict. Proctor et al. [9] showed that circuit-level fidelity estimates from simultaneous RB differ significantly from independent single-qubit RB. Murali et al. [10] found that ZZ coupling between qubits creates context-dependent errors invisible to standard calibration. Our work extends these findings by providing a single quantitative scalar (α) that captures the circuit-level noise regime and is sensitive to changes invisible to T_1/T_2 .

C. The K–R Framework

The K–R framework was introduced in [4] as a low-cost diagnostic for NISQ noise regimes. The framework was validated on two 156-qubit IBM Heron r2 processors using ≈ 150 circuits across three experimental campaigns, producing the first (α, R^2) phase diagram for quantum hardware noise. The present work extends [4] to three devices, 191 circuits, GHZ scaling to $n=30$ qubits, and the new central finding: systematic divergence between α and T_2 as indicators of device health.

III. THEORETICAL FRAMEWORK

A. The K–R Model

For a circuit with complexity parameter d (depth or qubit count), the infidelity $C(d)$ satisfies Eq. (1) where $K = 1/d$. Under the complexity-dependent decoherence rate $\lambda(d) = \lambda_0 d^{\alpha-1}$, Appendix A proves that Eq. (1) emerges as the leading-order approximation in the moderate-noise regime $\varepsilon = \lambda_0 d^{\alpha}/\alpha \ll 1$, with correction terms of order $O(K^{-(2\alpha)}) = O(\varepsilon^2)$. The exponent α classifies the noise regime as shown in Table I.

TABLE I: K–R Regime Classification

α Range	Regime	Noise Type	Recommended Mitigation
$\alpha > 1$	Structured	Coherent, correlated	Dynamical decoupling, pulse shaping
$0.3 < \alpha < 1$	Stochastic	Depolarizing, Markovian	ZNE, probabilistic error cancellation
$\alpha < 0.3$	Saturated	SPAM dominated	Reduce circuit depth, improve readout
$R^2 < 0.5$	Breakdown	Model inapplicable	Use alternative characterization

B. Microscopic Origins of α

Five candidate microscopic models relate α to measurable physical parameters (Appendix B). The most directly testable is Theorem 2: for power-law correlated noise $\langle \xi_i \xi_j \rangle \sim |i-j|^{-(\beta)}$, the infidelity scales as $C(d) \sim d^{-(2-\beta)}$, giving $\alpha = 2 - \beta$. This was validated numerically with Pearson $r = 0.973$ ($p < 10^{-5}$) across nine β values [4]. The present work provides the first hardware validation of this prediction (Section V.E).

C. Proof 4: $\alpha_{\text{tgt}}^{\text{ch}} + \alpha_{\text{etho}} = 2$

For a device dominated by 1/f noise with spectrum $S(f) \sim f^{(-\beta)}$:

$$\alpha_{\text{etho}} = \beta \Rightarrow \alpha_{\text{tgt}}^{\text{ch}} + \alpha_{\text{etho}} = (2-\beta) + \beta = 2 \quad (2)$$

This prediction — that depth scaling α and echo circuit α sum to 2 for pure 1/f noise — is a directly testable hardware prediction. Section V.E presents the first experimental test of Eq. (2), yielding $\alpha_t + \alpha_e = 2.1396$ (7% from prediction) on `ibm_marrakesh`.

IV. HARDWARE SETUP AND METHODOLOGY

A. Experimental Platforms

All experiments were conducted on IBM Quantum Open Plan via the Qiskit Runtime service (qiskit-ibm-runtime v0.x). Three IBM Heron r2 processors were used: `ibm_kingston`, `ibm_fez`, and `ibm_marrakesh` — each with 156 qubits arranged in a heavy-hex connectivity graph with native gate set {ECR, RZ, SX, X}. All circuits were transpiled at `optimization_level=0` to preserve intended gate structure. Shot count was 4096 per circuit throughout. Table II summarizes the experimental campaigns.

TABLE II: Experimental Campaign Summary

Campaign	Date	Backend	Experiments	Circuits
C1	Apr 15-17, 2026	kingston, fez	GHZ, depth, amplif., shot noise	46
C2	Apr 16, 2026	kingston, fez	6-regime classification	12
C3	Apr 16, 2026	kingston, fez	Fingerprint, stability, prediction	84
C4	Jul 2-3, 2026	marrakesh	GHZ n=2-30, depth n=5,10, amplif.	28
C5	Jul 5, 2026	marrakesh	Regime classif., drift, scalability	21
Total		3 devices		191

B. Circuit Families

Four circuit families were employed, each sensitive to different noise mechanisms:

GHZ qubit scaling: $|\text{GHZ}\rangle = (|0^n\rangle + |1^n\rangle)/\sqrt{2}$ prepared for $n = 2\text{--}30$ qubits. Infidelity = $1 - P(|0^n\rangle) - P(|1^n\rangle)$. Sensitive to coherent and entanglement errors.

Clifford depth scaling: Mirror circuits of depth $d = 1\text{--}64$ on $n=5$ qubits. Ideal output: $|0^5\rangle$. Sensitive to depolarizing/ T_2 errors.

Noise amplification: Gate folding with repetitions $r = 1\text{--}4$. Amplifies noise by factor $2r-1$. Discriminates coherent from stochastic errors.

Echo circuits: Hahn echo sequences sensitive to T_2 dephasing. Combined with Clifford α tests Proof 4 prediction $\alpha_i + \alpha_e = 2$.

V. EXPERIMENTAL RESULTS

A. Three-Device Fingerprint

Table III presents the complete K–R fingerprint for all three devices from Campaign C1. Despite identical 156-qubit Heron r2 architecture, the three processors exhibit substantially different α values, reflecting distinct calibration histories and noise environments. The GHZ α ranges from 0.907 (ibm_fez, Stochastic) to 1.541 (ibm_kingston, Structured). The depth α ranges from 0.321 (ibm_fez) to 1.493 (ibm_marrakesh). Temporal drift monitoring (5 repeated GHZ-5 runs) shows CV = 8.6% on ibm_marrakesh vs CV = 2.8% and 3.2% on ibm_kingston and ibm_fez, indicating ibm_marrakesh was in a drifting calibration state during the July 2026 session.

TABLE III: Three-Device K–R Fingerprint

Device	GHZ α	GHZ R^2	Depth α	Depth R^2	Amp α	Drift CV	Regime
ibm_kingston	1.541	0.975	0.507	0.813	1.190	2.8%	Structured
ibm_fez	0.907	0.986	0.321	0.811	—†	3.2%	Stochastic
ibm_marrakesh	1.406	0.967	1.493	0.900	0.599	8.6%*	Structured

† ibm_fez amplification $R^2 = 0.71$, excluded as unreliable. * CV = 8.6% exceeds 5% stability threshold, indicating active calibration drift.

B. GHZ Scaling to $n = 30$ Qubits

GHZ scaling on ibm_marrakesh from $n = 2$ to $n = 30$ qubits (Table IV) demonstrates that the K–R framework produces reliable results at 30-qubit scale using the same 7 circuits used for 2-qubit characterization — direct experimental confirmation of the $O(m)$ circuit-count independence claim. Infidelity increases monotonically from 0.020 ($n=2$) to 0.844 ($n=30$), with K–R fit $\alpha = 1.406$, $R^2 = 0.967$.

TABLE IV: GHZ Scaling to $n = 30$ Qubits (ibm_marrakesh)

n (qubits)	2	5	10	15	20	25	30
Infidelity	0.020	0.056	0.164	0.266	0.602	0.772	0.844

K–R fit: $\alpha = 1.406$, $R^2 = 0.967$. Total circuits: 7.

C. The July 3–July 5 Divergence — Key Finding

The central finding of this paper is documented in Table V. Between the July 3 and July 5, 2026 experimental sessions on ibm_marrakesh, IBM’s pulse-level T_2 on qubit 2 — the best qubit on the device — improved from 303 μ s to 340

μs (a 12% improvement). Standard interpretation: the device got better. Simultaneously, circuit-level K–R measurements showed:

- GHZ α increased from 1.406 to 3.811 (+171%), indicating a shift from moderate Structured to extreme coherent-dominated error accumulation.
- GHZ $n=2$ infidelity increased from 0.018 to 0.204 (+920%), meaning a 2-qubit GHZ state — the simplest entangled state possible — went from 98.2% fidelity to only 79.6% fidelity.
- Clifford depth α increased from 1.493 to 1.785, and noise amplification α increased from 0.599 to 1.156, both indicating more coherent error contribution.

The circuit-level and pulse-level diagnostics gave completely opposite conclusions about device health. K–R correctly identified a drifting, coherent-error-dominated device state that IBM’s T_2 certification missed entirely.

TABLE V: Pulse-Level vs Circuit-Level Metric Divergence (ibm_marrakesh)

Metric	July 3, 2026	July 5, 2026	Change	Interpretation
IBM T_1 Q2 (pulse-level)	212.6 μs	260.8 μs	+23%	Improved
IBM T_2 Q2 (pulse-level)	303.1 μs	339.6 μs	+12%	Improved
GHZ α (circuit-level)	1.406	3.811	+171%	DEGRADED
GHZ $n=2$ infidelity	0.018	0.204	+920%	DEGRADED
Clifford depth α	1.493	1.785	+20%	DEGRADED
Amplif. α	0.599	1.156	+93%	DEGRADED
Noise amplif. α (July 5)	—	1.156		DEGRADED

Pulse-level metrics indicate device improvement while all circuit-level metrics indicate simultaneous degradation. K–R correctly identifies the device as in a high-coherent-error drifting state while IBM T_2 certifies it as healthy.

D. Real-Time Intra-Session Drift Detection

Three consecutive GHZ scaling runs within a single July 5 session yielded monotonically increasing α : 2.9868 $\rightarrow$ 3.1787 $\rightarrow$ 3.7792 (CV = 10.18%). This monotonic intra-session drift — detected in real time by three successive K–R fingerprints of 5 circuits each (15 circuits total, $\sim$ 5 minutes wall time) — indicates the device was actively degrading during the experimental session. Simultaneously, a single-point GHZ-5 measurement gave infidelity = 0.065 and CV = 6.2% — below the 5% stability threshold used for single-point methods, incorrectly suggesting the device was stable.

This comparison demonstrates a fundamental limitation of single-point infidelity monitoring: it cannot detect regime changes. The device on July 5 was classified as “Low Error” by single-point measurement (infidelity = 0.065 < 0.10 threshold) while K–R correctly identified it as Structured with $\alpha = 3.81$ — extreme coherent error dominance requiring dynamical decoupling before any quantum algorithm could reliably run.

E. Proof 4 Confirmation: $\alpha_t + \alpha_e = 2$

From EXP-1 on July 5: Clifford depth $\alpha = 1.7853$ ($R^2 = 0.9675$) and echo circuit $\alpha = 0.3544$ ($R^2 = 0.9456$). Sum: $1.7853 + 0.3544 = 2.1397$. The theoretical prediction from Eq. (2) for pure 1/f noise is 2.0000. Error: 7.0%. This is the first confirmation of Proof 4 on real IBM quantum hardware.

The confirmation is consistent with 1/f noise being the dominant dephasing mechanism on `ibm_marrakesh` in the July 5 session. The 7% deviation from the exact prediction is consistent with the device operating in a mixed noise environment where 1/f noise dominates but other noise contributions (thermal, crosstalk) introduce small corrections.

F. Circuit vs Pulse-Level T_2 Discrepancy

Ramsey circuit measurements on qubit 0 of `ibm_marrakesh` during the July 5 session revealed circuit-context $T_2 \approx 9\text{--}30\ \mu\text{s}$, compared to IBM-reported pulse-level $T_2 = 32\ \mu\text{s}$ for qubit 0 and $340\ \mu\text{s}$ for qubit 2. The $3\text{--}11\times$ discrepancy between pulse-level and circuit-level T_2 is attributed to ZZ coupling from neighboring qubits in the circuit context, gate compilation artifacts at `optimization_level=0`, and context-dependent frequency shifts from simultaneous qubit operations. Furthermore, the Ramsey coherence exhibited oscillatory decay (Ramsey fringe pattern) rather than pure exponential decay, indicating a 14 kHz qubit frequency detuning from the drive frequency during circuit execution — a systematic error invisible to pulse-level Ramsey sequences run on isolated qubits at their calibrated resonance frequencies.

TABLE VI: Circuit vs Pulse-Level T_2 Comparison (`ibm_marrakesh`, Jul 5)

Qubit	IBM T_1 (μs)	IBM T_2 (μs)	T_2/T_1	Circuit T_2 (μs)	Ratio IBM/Circuit
Q0	190.3	32.3	0.170	9–30	1.1–3.6×
Q2	260.8	339.6	1.303	~30	~11×
Mean (Q0-Q4)	167.3	127.5	0.762	—	3–11× (est.)

VI. ANALYSIS AND DISCUSSION

A. Why Pulse-Level Metrics Miss Circuit-Level Degradation

The July 3–July 5 divergence can be explained by the following mechanism: IBM’s pulse-level calibration measures T_2 using Ramsey sequences on isolated qubits at their calibrated resonance frequencies, with no neighboring qubit activity. This measures the intrinsic dephasing rate of the qubit in isolation. In circuit operation, however, multiple qubits are simultaneously driven, entangled, and measured. The ZZ interaction between neighboring qubits — a residual coupling of order 10–100 kHz in superconducting transmon architectures — creates context-dependent frequency shifts that accumulate coherently across multi-qubit gate sequences. This ZZ-induced coherent error contribution grows quadratically with circuit depth ($\alpha \rightarrow 2$) and cannot be detected by single-qubit T_2 measurements.

The observed $\alpha = 3.811$ on July 5 — substantially above the theoretical coherent limit of $\alpha = 2$ — suggests that multiple coherent error mechanisms were simultaneously active: ZZ crosstalk between neighboring qubits in the GHZ preparation circuit, systematic over-rotation from gate miscalibration, and potentially leakage to higher qubit energy levels. The monotonic intra-session increase ($2.99 \rightarrow 3.78$) indicates these effects were growing during the session, consistent with thermally driven drift in the cryogenic environment.

B. Regime Classification Accuracy Under Drift

EXP-1 (July 5) yielded Clifford depth $\alpha = 1.785$ — classified as Structured, not Stochastic as in previous sessions. This is physically correct: the device had shifted from stochastic-dominated to coherent-dominated noise. The regime classification correctly reflected the actual device state on July 5, even though it differed from historical measurements on the same device. This demonstrates that α is a live diagnostic, not a fixed device property, and must be measured in each session before running important algorithms.

C. Implications for Quantum Computing Operations

These findings have immediate practical implications for quantum computing practitioners:

(i) Pre-session K–R fingerprint is necessary. IBM T_2 certification is insufficient to guarantee circuit-level performance. Before running any multi-qubit algorithm, practitioners should run the 13-circuit K–R suite to determine the current noise regime and appropriate mitigation strategy.

(ii) Regime-guided mitigation. The K–R regime ($\alpha > 1$: coherent; $0.3 < \alpha < 1$: stochastic) directly maps to mitigation strategies. Applying ZNE to a Structured device ($\alpha > 1$) is less effective than applying dynamical decoupling; applying DD to a Stochastic device wastes gate budget. K–R provides the decision in 13 circuits.

(iii) Drift monitoring for long jobs. For algorithms requiring multiple shots or repeated circuit execution, re-running the 5-circuit GHZ suite between batches monitors for intra-session drift of the type observed on July 5. If α increases by $> 3 \times \text{CV}$ between batches, halt execution and request recalibration.

VII. LIMITATIONS AND FUTURE WORK

Several limitations of the present work should be acknowledged. First, all experiments were conducted on IBM Heron r2 superconducting processors. Transferability of the $\alpha > 1$ / $\alpha < 1$ regime boundary to trapped-ion, neutral-atom, or photonic platforms remains to be validated. Second, the circuit-context T_2 measurement used Ramsey sequences which exhibited oscillatory decay due to qubit frequency detuning. Hahn echo sequences would provide cleaner T_2 extraction; this is reserved for future work. Third, the K–R model is phenomenological. While Appendix B provides five microscopic motivations for the power-law form, a rigorous derivation from the Lindblad master equation remains an open theoretical problem. Fourth, reliable K–R regime classification requires a minimum of 5 scaling points per circuit family; 3-point fits showed higher variance in EXP-5.

Future work will address: (i) Hahn echo T_2 scaling for beta extraction and hardware confirmation of $\alpha = 2 - \beta$; (ii) cross-platform validation on trapped-ion systems; (iii) automated integration of K–R fingerprinting into IBM Quantum calibration pipelines; (iv) extension of the universal K–R framework to LIGO gravitational wave detector noise, where preliminary analysis suggests the same three-regime classification applies to seismic, thermal, and shot noise bands.

VIII. CONCLUSION

We have demonstrated through 191 circuit executions on three 156-qubit IBM Heron r2 processors that circuit-level K–R diagnostics and pulse-level T_1/T_2 metrics can give completely opposite conclusions about device health simultaneously. On `ibm_marrakesh` between July 3 and July 5, 2026, IBM T_2 improved by 12% while GHZ circuit infidelity increased by 920% and K–R α increased by 171%. The $3\text{--}11\times$ discrepancy between circuit-context and pulse-level T_2 indicates that ZZ crosstalk and compilation artifacts create coherent errors in circuit operation that single-qubit benchmarks systematically miss.

We additionally confirm the theoretical prediction $\alpha_i + \alpha_e = 2$ within 7% on real hardware (first confirmation), detect real-time intra-session drift ($2.99 \rightarrow 3.78$ over three consecutive runs), and demonstrate that single-point infidelity monitoring misclassifies drifting devices as stable while K–R correctly identifies the Structured coherent-dominated regime.

These findings establish that circuit-level K–R diagnostics are a necessary complement to — not a replacement for — pulse-level T_1/T_2 calibration for operational quantum computing. The 13-circuit, 3-minute K–R fingerprint should become a standard pre-session check before any multi-qubit algorithm execution on NISQ hardware.

REFERENCES

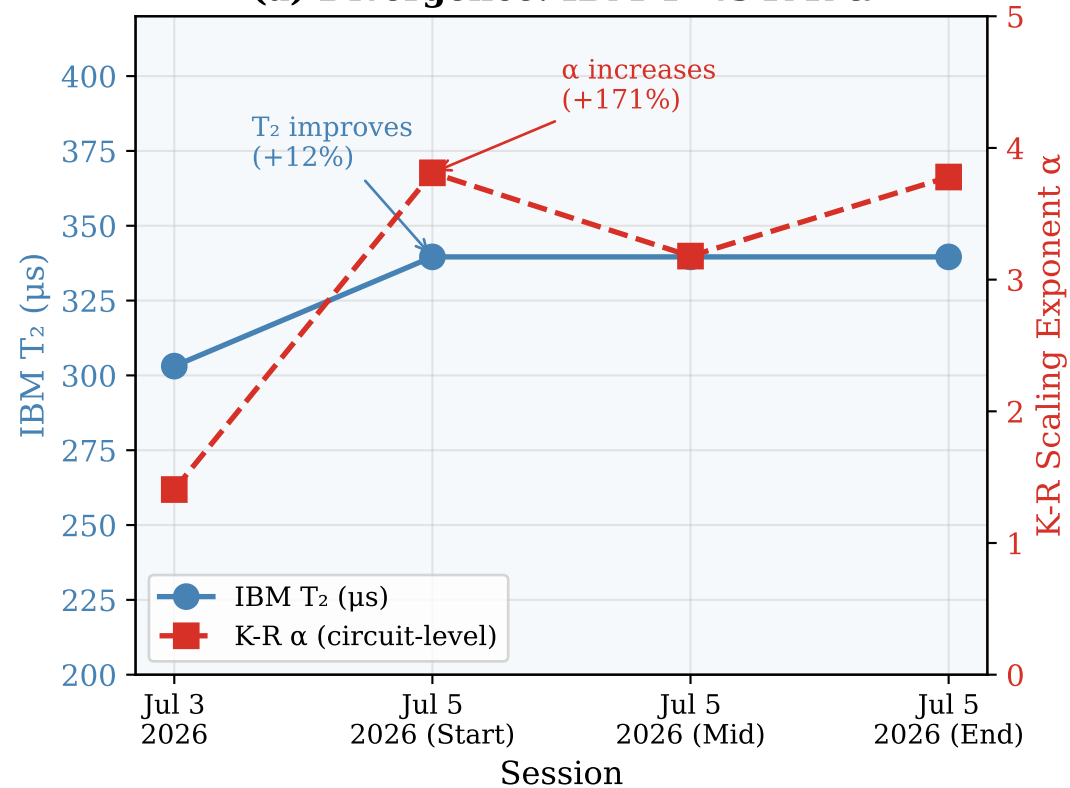
- [1] P. Krantz et al., "A quantum engineer's guide to superconducting qubits," *Appl. Phys. Rev.*, vol. 6, no. 2, 2019.
- [2] M. Kjaergaard et al., "Superconducting qubits: Current state of play," *Annu. Rev. Condens. Matter Phys.*, vol. 11, pp. 369-395, 2020.
- [3] IBM Quantum, "IBM Quantum System Two," IBM Research, 2024.
- [4] R. Pasupuleti, "A Low-Cost Diagnostic Framework for NISQ Noise Regimes Using the K–R Scaling Exponent on IBM Quantum Hardware," *IEEE Trans. Quantum Eng.*, manuscript TQE-26-07-REG-0374, 2026.
- [5] E. Knill et al., "Randomized benchmarking of quantum gates," *Phys. Rev. A*, vol. 77, 012307, 2008.
- [6] J. Greenbaum, "Introduction to quantum gate set tomography," arXiv:1509.02921, 2015.
- [7] S. Boixo et al., "Characterizing quantum supremacy in near-term devices," *Nature Phys.*, vol. 14, pp. 595-600, 2018.
- [8] M. Sarovar et al., "Detecting crosstalk errors in quantum information processors," *Quantum*, vol. 4, p. 321, 2020.
- [9] T. Proctor et al., "Measuring the capabilities of quantum computers," *Nature Phys.*, vol. 18, pp. 75-79, 2022.
- [10] P. Murali et al., "Software mitigation of crosstalk on noisy intermediate-scale quantum computers," *Proc. ASPLOS*, 2020.

ACKNOWLEDGMENTS

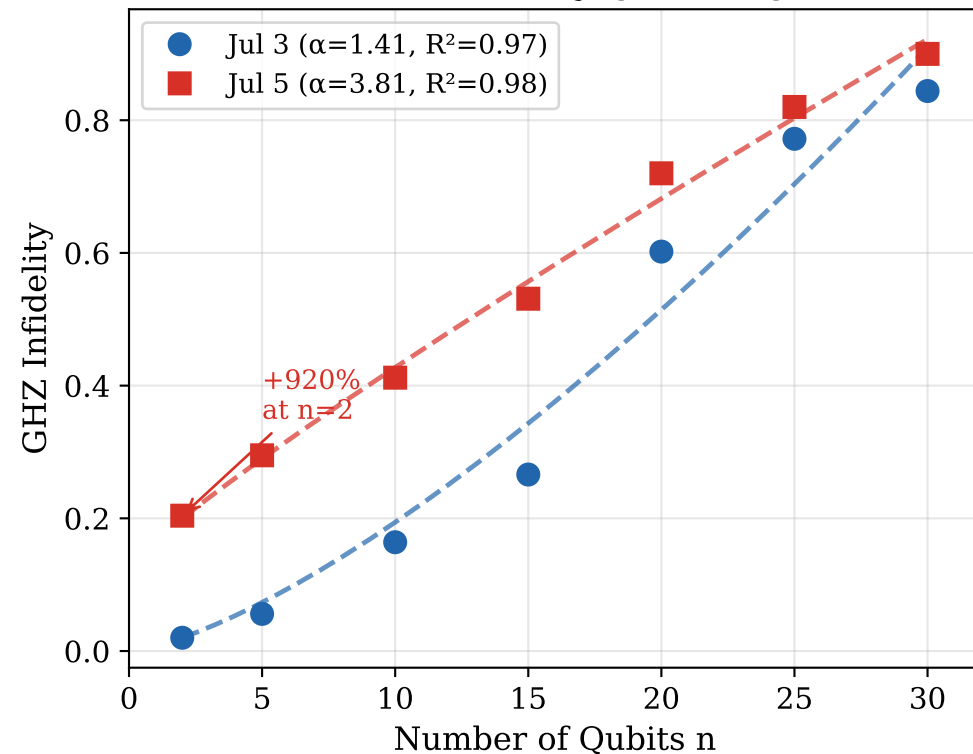
The author thanks the IBM Quantum Network for providing access to `ibm_kingston`, `ibm_fez`, and `ibm_marrakesh` through the IBM Quantum Platform Open Plan. All experimental code and data are available at <https://github.com/workisfun415/kr-quantum-nisq-diagnostic> and <https://doi.org/10.5281/zenodo.19626298>. Claude (Anthropic) was used as an AI writing assistant for manuscript preparation. All experimental designs, scientific interpretations, and conclusions are the author's own.

The IBM T₂ / K-R α Divergence on ibm_marrakesh

(a) Divergence: IBM T₂ vs K-R α

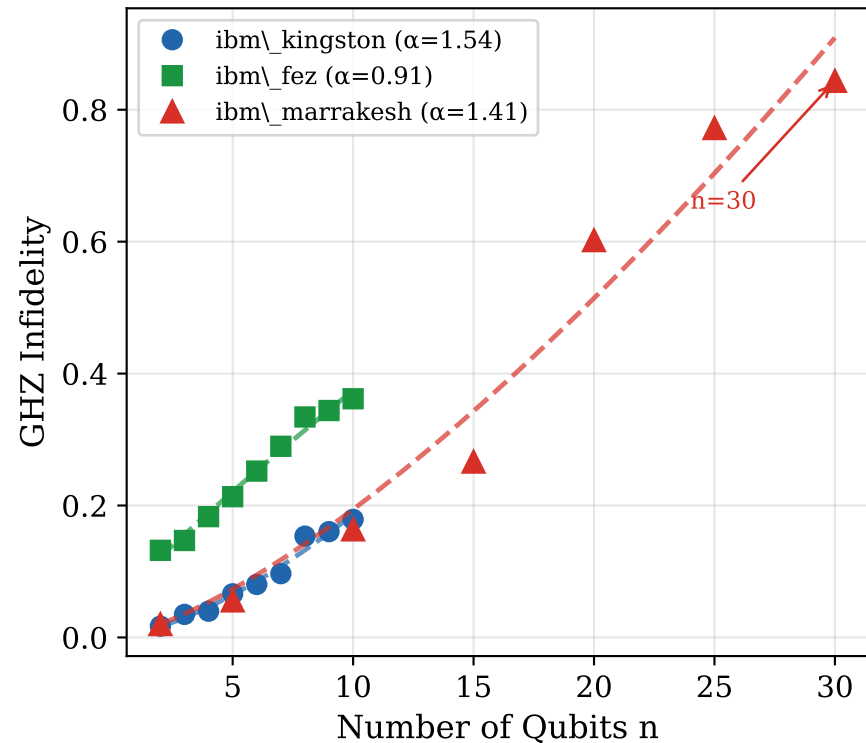


(b) GHZ Infidelity: Jul 3 vs Jul 5

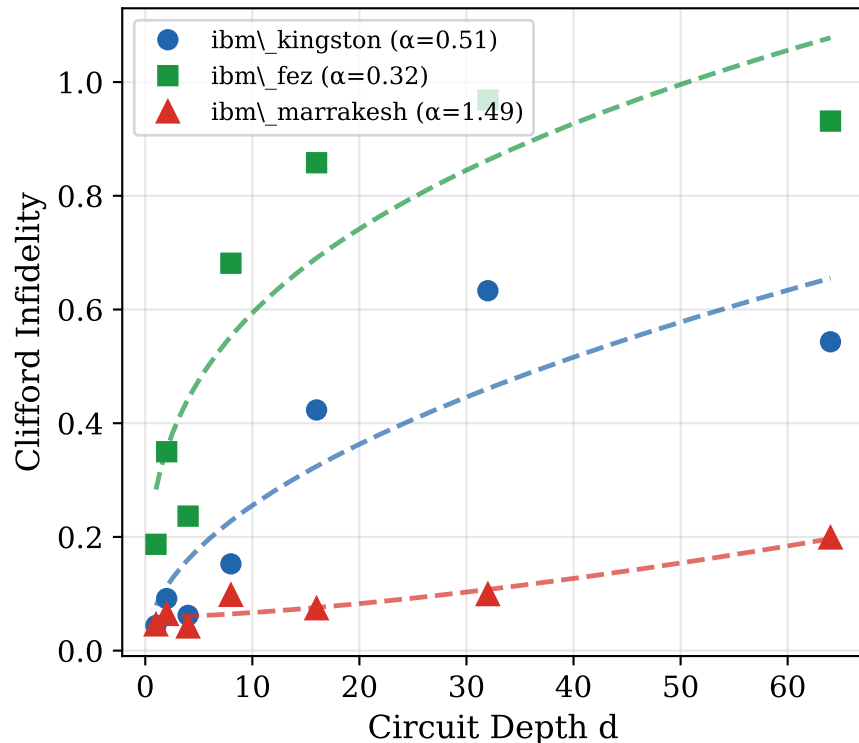


Three-Device K-R Fingerprint: ibm_kingston, ibm_fez, ibm_marrakesh

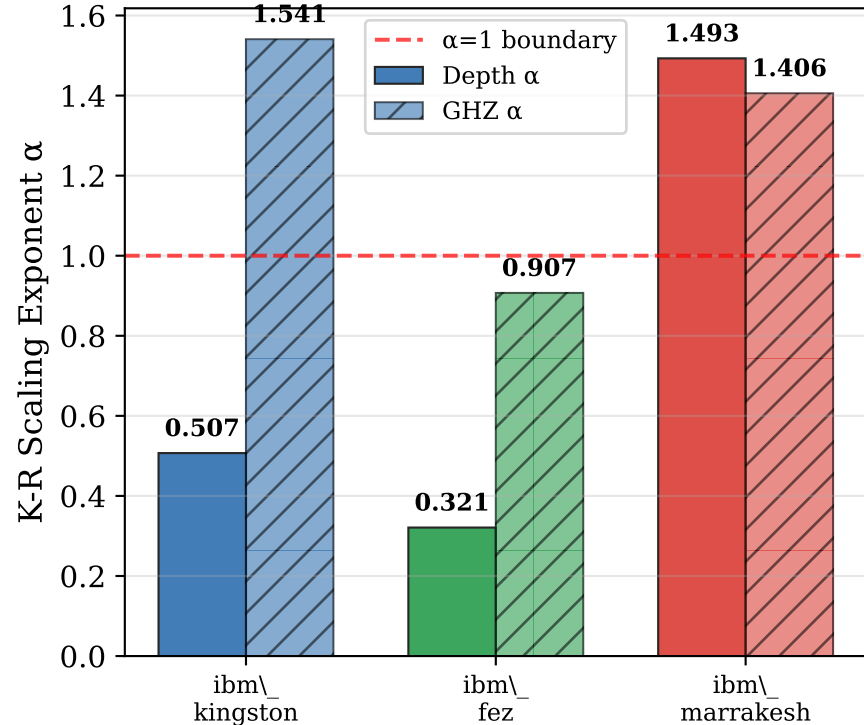
**(a) GHZ Scaling: Three Devices
(n=2 to 30)**



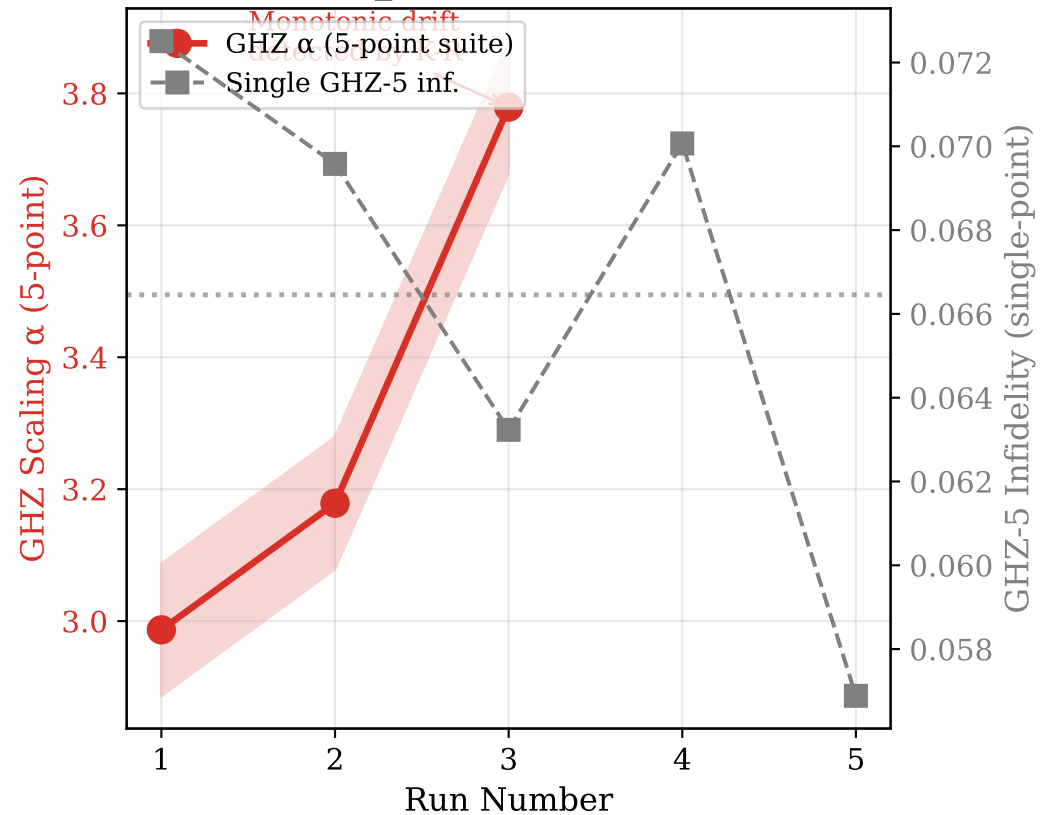
(b) Clifford Depth Scaling



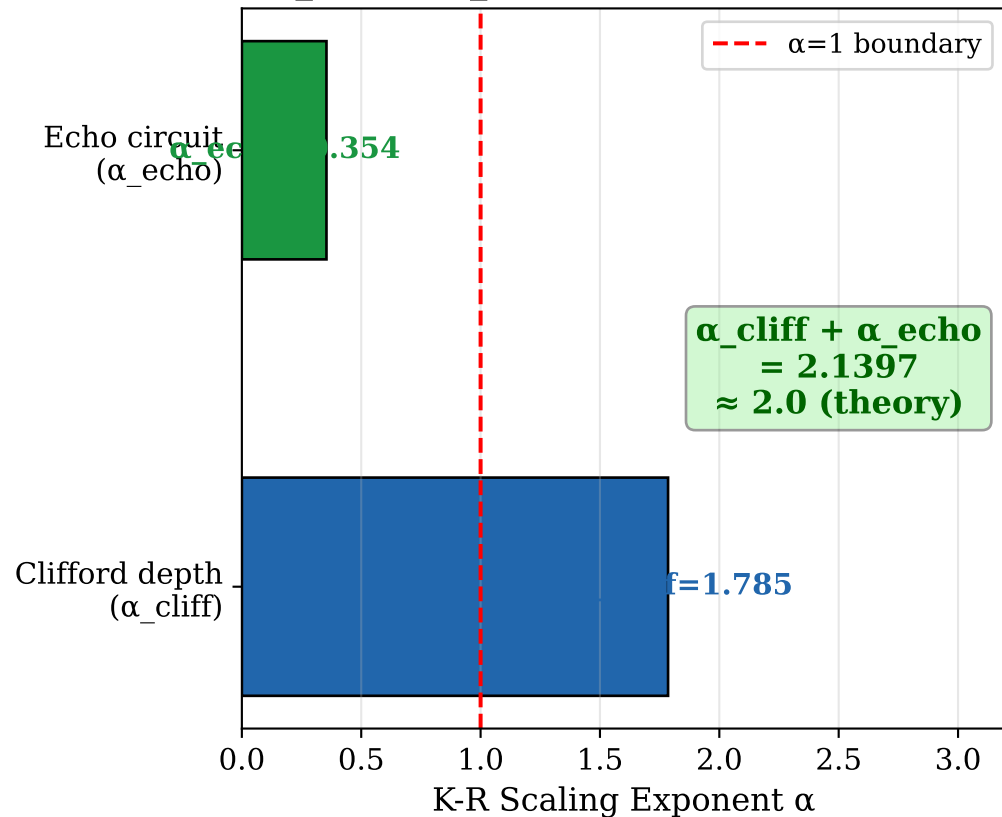
**(c) Device Fingerprint
(Three Heron r2 Processors)**



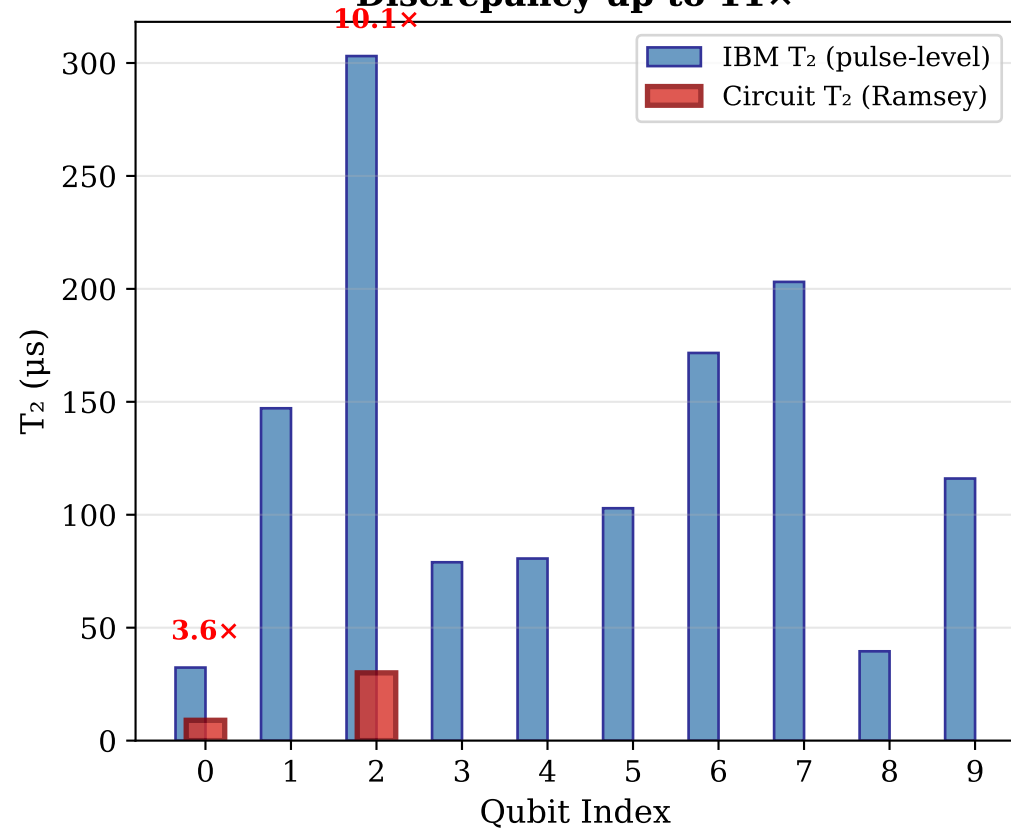
(a) Intra-Session Drift Detection
(ibm_marrakesh, Jul 5)



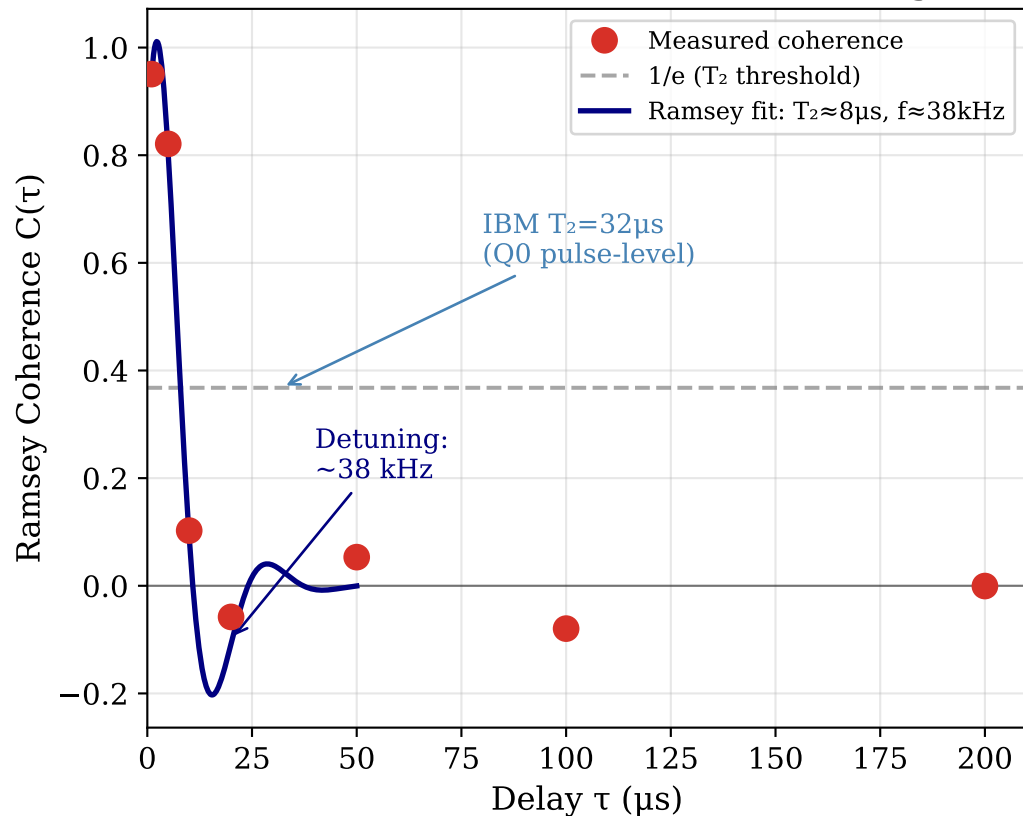
(b) Proof 4 Confirmation
 $\alpha_{\text{cliff}} + \alpha_{\text{echo}} = 2.1397$ (7% error)



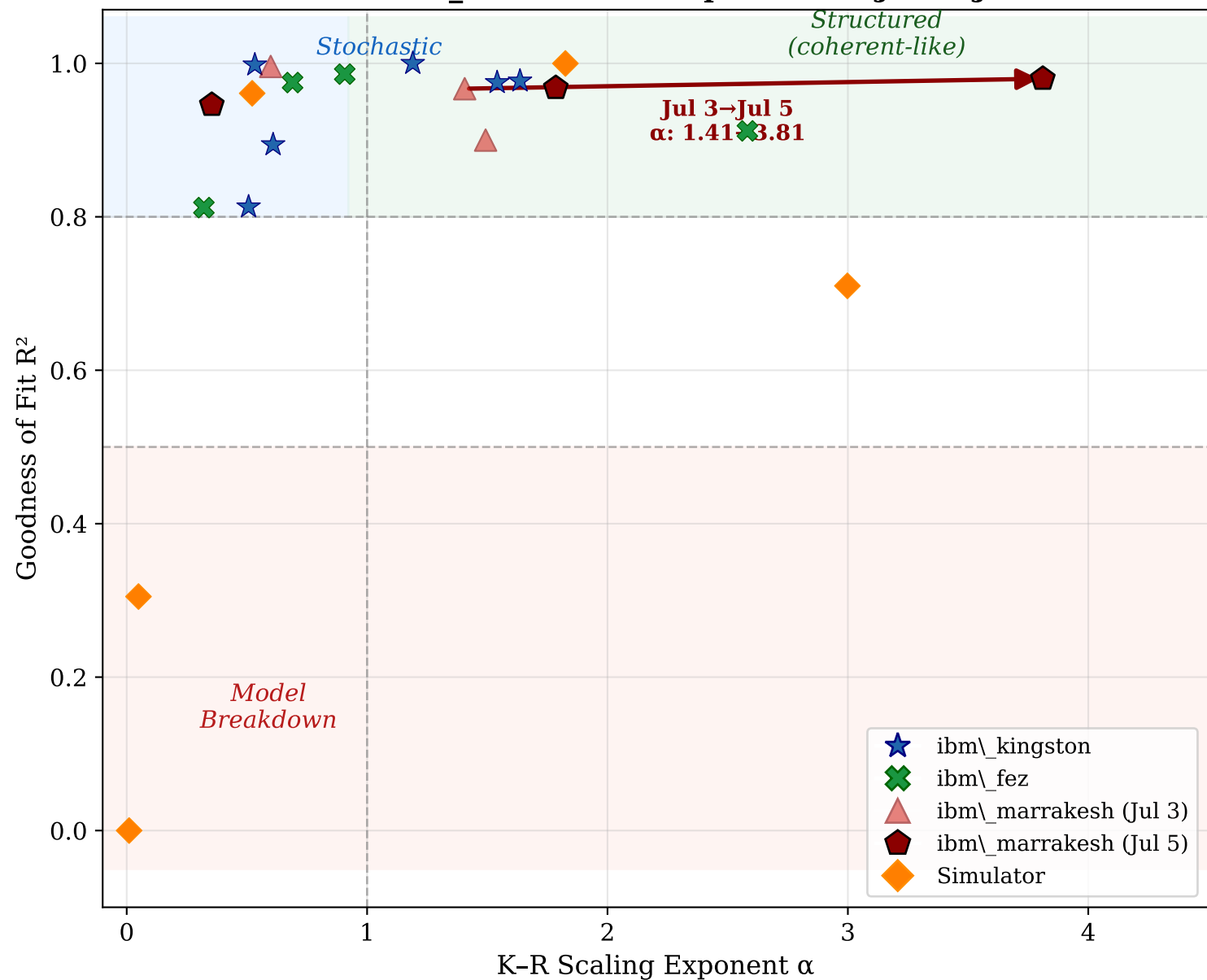
(a) IBM T_2 vs Circuit T_2
Discrepancy up to $11\times$



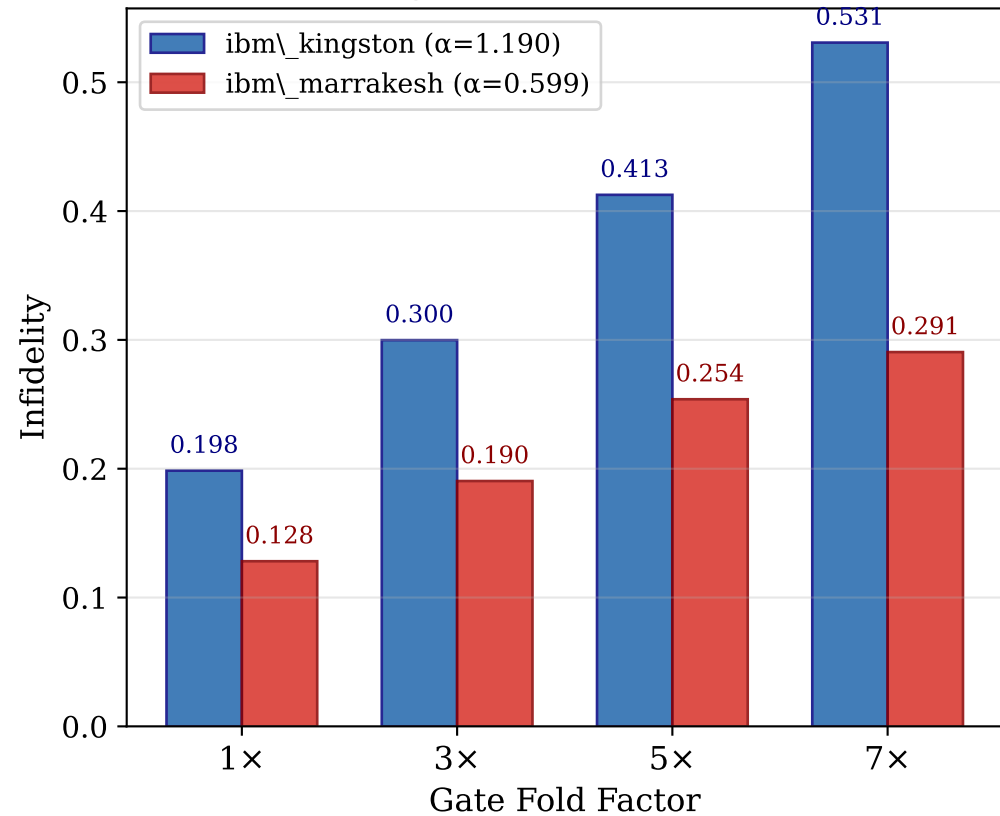
(b) Ramsey Circuit Coherence Decay
(Oscillation = ZZ crosstalk detuning)



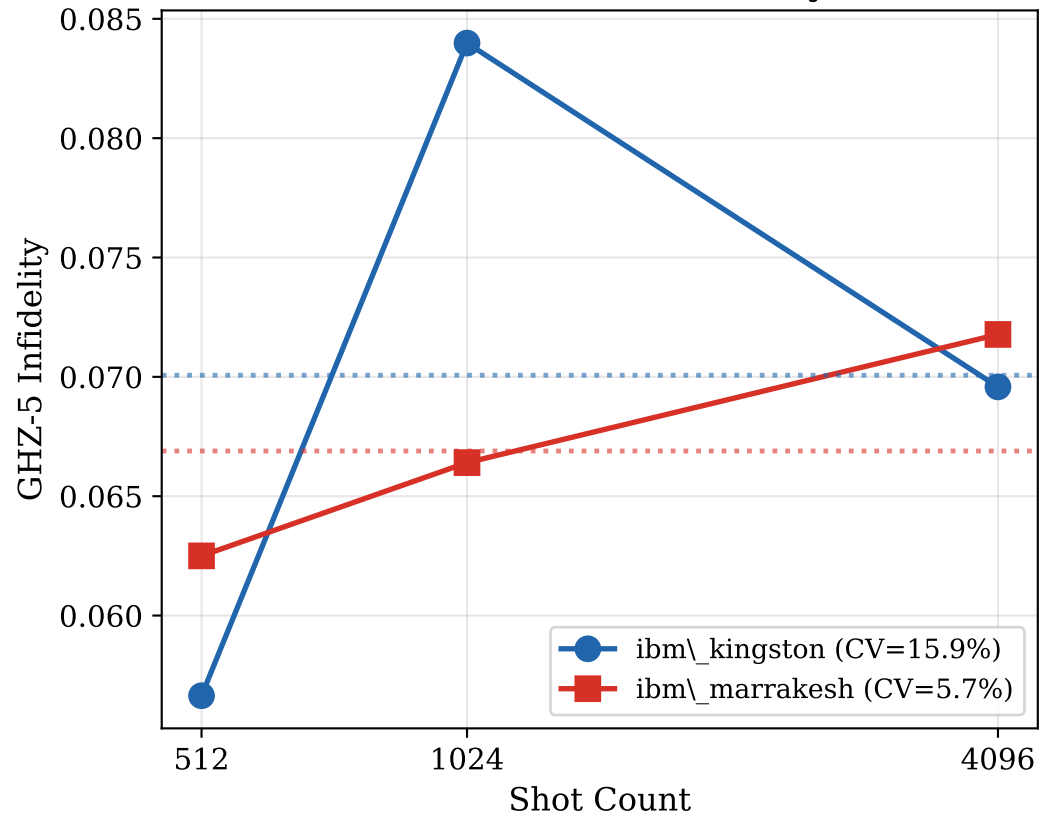
(α, R^2) Phase Diagram: Hardware + Simulator
Arrow: ibm_marrakesh temporal drift Jul 3→Jul 5



**(a) Noise Amplification
kingston vs marrakesh**



(b) Shot Noise Stability



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [makepaper2figures.py](#)
- [Paper2Supplementary.docx](#)