

Toward a Quantum Diagnostic Intelligence System: A Multi-Layer Architecture for Circuit-Level Assessment of Quantum Hardware Operational State

RamaKrishna Pasupuleti

workisfun415@gmail.com

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

Research Article

Keywords: quantum diagnostics, hardware characterization, NISQ devices, superconducting qubits, decision support, IBM Quantum, scaling exponent

Posted Date: July 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10394388/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.

Toward a Quantum Diagnostic Intelligence System: A Multi-Layer Architecture for Circuit-Level Assessment of Quantum Hardware Operational State

Ramakrishna Pasupuleti

Independent Researcher, Suryapet, Telangana 508213, India

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

ABSTRACT

We introduce Quantum Diagnostic Intelligence (QDI), a five-layer decision-support architecture for inferring the operational state of quantum processing units from circuit-level measurements. Pulse-level calibration metrics — coherence times T_1 and T_2 measured through single-qubit Rabi and Ramsey sequences — are the dominant industry standard for certifying quantum hardware health. We demonstrate, through 191 circuit executions on three 156-qubit IBM Heron r2 processors, that circuit-level measurements contain diagnostic information that T_1/T_2 may not capture. On `ibm_marrakesh` between July 3 and July 5, 2026, IBM-reported T_2 improved from 303 to 340 μs (+12%), while circuit-level GHZ scaling yielded $\alpha = 3.811$, a QDI health score of 37.8/100 [95% CI: 7.6, 67.9], and a circuit infidelity increase of 920% at $n = 2$ qubits. The QDI architecture formalizes this observation through four operators: a diagnostic state classifier $D(\chi)$, a health score $H(\chi)$, a recommendation generator $R(\chi)$, and a drift forecast $F(\chi, t)$. Rules trained on `ibm_kingston` correctly classified `ibm_fez` and both `ibm_marrakesh` sessions without retuning (3/3 held-out tests; $n = 4$ total — feasibility demonstration only). We identify three scientific limitations of the current framework: operators rely on expert-designed rules rather than learned statistical models; ground truth (algorithm success/failure) has not yet been collected; and diagnostic state labels have not been validated by clustering analysis. Prospective validation on $n \geq 50$ independent sessions is required before statistical claims about generalization can be made. The dataset, code, and QDI system are openly available.

Index Terms—*quantum diagnostics, hardware characterization, NISQ devices, superconducting qubits, decision support, IBM Quantum, scaling exponent.*

I. INTRODUCTION

The standard procedure for certifying the health of a quantum processing unit (QPU) relies on pulse-level measurements: T_1 measured through inversion recovery, and T_2 through Ramsey or spin-echo sequences applied to isolated single qubits [1],[2]. These metrics are fast, well-understood, and routinely published by quantum hardware providers as the primary indicators of device quality. A device with long T_1 and T_2 is certified as healthy; a device with short coherence times is flagged for recalibration.

This standard has a fundamental limitation: T_1 and T_2 are measured on isolated single qubits under idealized pulse conditions. They do not capture coherent errors from gate miscalibration, which accumulate with circuit depth and are invisible to single-qubit measurements; crosstalk between simultaneously operated qubits; or the temporal stability of these effects during actual circuit execution. In short, T_1/T_2 measure what a device can do in isolation — not what it does when running multi-qubit algorithms.

The consequence is a systematic gap between certified hardware quality and actual algorithm performance. We demonstrate this gap directly on July 5, 2026 on `ibm_marrakesh`: IBM's T_2 metric simultaneously improved by 12% while GHZ circuit infidelity at $n = 2$ qubits increased by 920%. The two diagnostic frameworks gave opposite verdicts about device health.

We introduce Quantum Diagnostic Intelligence (QDI), a five-layer architecture that infers device operational state from circuit-level measurements. The central scientific hypothesis is not that the K–R scaling exponent α is universal. It is:

Circuit-level measurements contain diagnostic information about quantum processor state that pulse-level calibration metrics alone may not capture.

This hypothesis is narrow, falsifiable, and grounded in concrete experimental data. The QDI architecture formalizes it through four mathematical operators defined over a biomarker vector χ extracted from 13 circuits in approximately three minutes:

$D(\chi)$ → Diagnostic State (which error type dominates)

$H(\chi)$ → $[0, 100]$ health score with explainable breakdown

$R(\chi)$ → Prioritized actionable recommendations

$F(\chi, t)$ → Drift forecast with quantified uncertainty

The scaling exponent α is one component of χ . QDI is a state-estimation framework, not an α framework. If future work identifies better biomarkers, they can be added to χ without changing the operator definitions.

Contributions of this paper: (C1) Formal mathematical definition of all four QDI operators. (C2) Biomarker extraction from 191 real IBM circuits across three devices. (C3) Cross-device feasibility validation: rules trained on `ibm_kingston` correctly classify `ibm_fez` and `ibm_marrakesh` without retuning (3/3 held-out tests). (C4) Explainable

health score showing that 64% of ibm_marrakesh July 5 penalty comes from α_{GHZ} and 35% from calibration drift. (C5) Honest limitation analysis and a concrete validation roadmap.

II. RELATED WORK

A. Pulse-Level Benchmarking

Randomized benchmarking (RB) [3] provides a single gate-error rate through exponential decay of Clifford sequence fidelity. While rigorous, RB requires 100–300 circuits, averages all error types together, and cannot distinguish coherent from stochastic error accumulation. Gate set tomography (GST) [4] provides complete gate characterization at 500–2000 circuits and hours of QPU time. Cross-entropy benchmarking (XEB) [5] scales exponentially in qubit count. None of these identify the dominant error type with the efficiency of QDI.

B. Circuit-Level Noise Characterization

Sarovar et al. [6] showed that crosstalk errors in multi-qubit circuits exceed single-qubit predictions. Proctor et al. [7] demonstrated that simultaneous RB differs substantially from independent single-qubit RB. Murali et al. [8] found ZZ coupling creates context-dependent errors invisible to standard calibration. Our work extends these findings with a unified quantitative scalar (α) that captures error accumulation regimes and changes between sessions.

C. Quantum Hardware Diagnostics

Predictive maintenance frameworks for quantum hardware remain largely unpublished. Several quantum cloud providers offer real-time backend status pages reporting T_1 , T_2 , and gate error rates, but without circuit-level regime classification or predictive components. QDI is, to our knowledge, the first architecture to formalize the transition from measurement $\rightarrow$ diagnosis $\rightarrow$ recommendation $\rightarrow$ forecast for quantum hardware.

III. FORMAL FRAMEWORK DEFINITION

A. Biomarker Vector

Let χ denote the biomarker vector extracted from a sequence of m circuit experiments on device d :

$$\chi = (\alpha_{\text{GHZ}}, \alpha_{\text{depth}}, \alpha_{\text{echo}}, \alpha_{\text{amp}}, CV_{\text{drift}}, CV_{\text{shot}}, R^2_{\text{GHZ}}, R^2_{\text{depth}}) \quad (1)$$

where α_{GHZ} , α_{depth} , α_{echo} , α_{amp} are K–R scaling exponents extracted from GHZ qubit-count scaling, Clifford depth scaling, echo circuits, and noise amplification respectively; CV_{drift} is the coefficient of variation of repeated GHZ-5 infidelity measurements; CV_{shot} is the infidelity CV across three shot-count levels; and R^2_{GHZ} , R^2_{depth} are goodness-of-fit values for the GHZ and depth K–R fits.

Each α component is extracted by fitting the three-parameter K–R model to infidelity measurements $C(d)$ as a function of complexity parameter d (qubit count or circuit depth):

$$C(d) = C_0 / K^\alpha + R, \quad K = 1/d \quad (2)$$

where C_0 is the correctable error amplitude, α is the scaling exponent, and R is the depth-independent noise floor. The K–R model was introduced in [9] and validated on three IBM Heron r2 processors with 191 circuits; see [9] for its microscopic motivation.

B. Operator D: Diagnostic State

The diagnostic state classifier $D(\chi)$ maps the biomarker vector to one of five mutually exclusive operating modes, applied in priority order:

TABLE I: QDI Diagnostic States

State	Criterion	Physics	Priority
Coherent-error-dominated	$\alpha_GHZ > 2.0$ and $R^2_GHZ > 0.80$	ZZ crosstalk, gate over-rotation	1
Drift-dominated	$CV_drift > 8.0\%$	Thermal drift, flux noise	2
SPAM-dominated	$R^2_depth < 0.50$	State prep/measurement errors	3
Depolarizing-dominated	$0.3 < \alpha_GHZ \leq 1.0$ and $CV \leq 8\%$	T_1/T_2 limited, Markovian	4
Mixed	No single criterion satisfied	Multiple sources comparable	5

TABLE I: Priority-ordered diagnostic state definitions. Rules are derived from physics and applied unchanged across all devices.

The threshold $\alpha_GHZ > 2.0$ for Coherent-error-dominated state is justified as follows: $\alpha > 1.0$ indicates the error accumulation rate exceeds linear in circuit width, which is the signature of coherent errors that compound across entangling operations; $\alpha > 2.0$ indicates quadratic or super-quadratic accumulation, which cannot arise from purely stochastic Markovian noise. The $CV_drift > 8.0\%$ threshold is set at approximately $3.5\times$ the baseline CV observed on `ibm_kingston` ($CV = 2.25\%$) during stable operation.

We acknowledge that these thresholds are expert-designed, not learned from data. Section VII discusses how future data collection would enable learned thresholds through logistic regression or Bayesian classification.

C. Operator H: Health Score with Explainable Breakdown

The health score $H(\chi) \in [0, 100]$ is a weighted penalty model:

$$H(\chi) = 100 - P_\alpha(\alpha_GHZ) - P_{CV}(CV_drift) - P_{R^2}(R^2_GHZ) - P_{P4}(proof4_dev) \quad (3)$$

where the four penalty terms are:

$$P_{\alpha}(\alpha) = \min(40, 20 \times \max(0, \alpha - 1.5)) \quad (4)$$

$$P_{CV}(CV) = \min(30, 3 \times \max(0, CV - 3)) \quad (5)$$

$$P_{R^2}(R^2) = \min(15, 15 \times \max(0, 0.85 - R^2)) \quad (6)$$

$$P_{P4} = \min(10, 5 \times |\alpha_{depth} + \alpha_{echo} - 2.0|) \quad (7)$$

The score uncertainty is propagated from biomarker measurement errors. Denoting the α_{GHZ} estimation error as σ_{α} and the CV uncertainty as σ_{CV} :

$$\sigma_{H^2} = (dP_{\alpha}/d\alpha \times \sigma_{\alpha})^2 + (dP_{CV}/dCV \times \sigma_{CV})^2 \quad (8)$$

The 95% confidence interval for H is $H \pm 1.96\sigma_H$. We emphasize that the penalty weights $\{20, 3, 15, 5\}$ are physics-derived estimates, not values fitted to outcome data. Calibration against real algorithm success/failure rates is the primary goal of the prospective validation study described in Section VII.

D. Operator R: Recommendations

The recommendation operator $R(\chi)$ maps the diagnostic state and health score to a prioritized set of circuit-level actions consisting of four lists: urgent actions (immediate steps required before computation), permitted circuits (what can safely run), circuits to avoid (what should not run), and mitigation strategies (error-reduction techniques appropriate to the dominant error state). The mapping from state to mitigation strategy follows established practice: dynamical decoupling for Coherent-error-dominated state [10], zero-noise extrapolation for Depolarizing-dominated state [11], and measurement error mitigation for SPAM-dominated state [12].

E. Operator F: Forecast with Uncertainty

The forecast operator $F(\chi, t)$ produces a predictive distribution over α_{GHZ} at time t hours in the future. Given n observations $\alpha_{GHZ}(t_i)$ at times t_i , a linear trend model gives:

$$\alpha(t) = a + b \times t + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \quad (9)$$

The 95% prediction interval at time $t_0 + \Delta t$ is:

$$\hat{\alpha} \pm t_{\{\alpha/2, n-2\}} \times s_{res} \times \sqrt{1 + 1/n + (\Delta t - \bar{t})^2/S_{xx}} \quad (10)$$

We emphasize that with $n = 3$ intra-session observations on July 5, the Student t critical value at 1 degree of freedom is $t_{0.975}(1) = 12.71$. The resulting prediction interval is very wide, which is the correct representation of uncertainty at this sample size. The forecast model is linear by assumption; real device behavior may saturate, oscillate, or reverse

after recalibration. $F(\chi, t)$ is currently labeled LOW CONFIDENCE and requires 10+ sessions with matched recalibration timestamps before a proper time-series model (ARIMA, Kalman filter) can be fitted.

IV. HARDWARE SETUP AND EXPERIMENTAL CAMPAIGNS

All experiments were conducted on IBM Quantum Open Plan hardware via the Qiskit Runtime service. Three 156-qubit IBM Heron r2 processors were used: `ibm_kingston`, `ibm_fez`, and `ibm_marrakesh`. These processors share identical architecture (heavy-hex connectivity, native gate set $\{ECR, RZ, SX, X\}$) and identical nominal specifications, making cross-device comparison scientifically controlled. All circuits were transpiled at `optimization_level = 0`. Shot count was 4,096 per circuit throughout. Table II summarizes the experimental campaigns.

TABLE II: Experimental Campaign Summary

Campaign	Date	Backend	Circuit Families	Circuits	Role
C1	Apr 15–17	kingston, fez	GHZ $n=2-10$, depth $d=1-64$, amplif., shot	46	TRAIN (D, H)
C2	Apr 16	kingston, fez	6-regime suite, echo, stability	84	TRAIN supplement
C4	Jul 2–3	marrakesh	GHZ $n=2-30$, depth, amplif., drift	28	TEST (held-out)
C5	Jul 5	marrakesh	GHZ, Clifford, echo, amplif., 3-run drift	33	TEST (held-out)
Total		3 devices		191	

TABLE II: Experimental campaign summary. Campaigns C1 and C2 provided TRAIN data for defining thresholds; C4 and C5 are held-out TEST data not seen during threshold design.

V. RESULTS

A. Biomarker Extraction

Table III presents the complete biomarker vector χ for all four device sessions, extracted from real IBM hardware data. All values are from actual circuit executions; no synthetic or simulated data appears in this table.

TABLE III: Biomarker Vectors (All REAL IBM Data)

Device / Session	$\alpha_{\text{GHZ}} \pm \sigma$	α_{depth}	α_{echo}	α_{amp}	CV_drift%	R^2_{GHZ}	$D(\chi)$
ibm_kingston (C1–C2)	1.541 $\pm$ 0.356	0.507	0.609	1.179	2.25	0.975	Mixed†
ibm_fez (C1–C2)	0.907 $\pm$ 0.233	0.321	0.690	N/A	2.63	0.986	Depolarizing
ibm_marrakesh (C4, Jul 3)	1.406 $\pm$ 0.374	1.493	0.354	0.599	8.55	0.967	Drift
ibm_marrakesh (C5, Jul 5)	3.811 $\pm$ 0.094	1.785	0.354	1.156	10.18	0.980	Coherent

† *ibm_kingston* classified as Mixed because CV_drift data from campaign C1–C2 used a different experimental protocol than the stability suite used to define the 8% threshold. With matched protocol, expected classification is Depolarizing-dominated.

B. Central Finding: The Jul 3 – Jul 5 Divergence

Table IV documents the central finding of this paper. Between the July 3 and July 5, 2026 sessions on `ibm_marrakesh`, IBM-reported pulse-level T_2 improved from 303 to 340 μs (+12%). Simultaneously, every circuit-level biomarker indicated degradation.

TABLE IV: Pulse-Level vs Circuit-Level Divergence (`ibm_marrakesh`)

Metric	Jul 3, 2026	Jul 5, 2026	Change	Verdict
IBM T_2 , qubit 2 (pulse-level)	303.1 μs	339.6 μs	+12%	IMPROVED
IBM T_1 , qubit 2 (pulse-level)	212.6 μs	260.8 μs	+23%	IMPROVED
GHZ α_{GHZ} (circuit-level)	1.406	3.811	+171%	DEGRADED
GHZ $n=2$ infidelity	0.018	0.204	+920%	DEGRADED
Clifford α_{depth}	1.493	1.785	+20%	DEGRADED
Amplification α_{amp}	0.599	1.156	+93%	DEGRADED
Intra-session drift CV	8.55%	10.18%	+19%	DEGRADED
QDI Health Score $H(\chi)$	82.6/100	37.8/100	-44.8 pts	DEGRADED

TABLE IV: Simultaneous pulse-level improvement and circuit-level degradation on `ibm_marrakesh`. This divergence is the central empirical finding motivating the QDI framework.

C. Operator $D(\chi)$: Cross-Device Validation

Table V presents the cross-device validation results. Thresholds and priority rules were defined using only `ibm_kingston` data (TRAIN). The same rules were applied unchanged to `ibm_fez` and both `ibm_marrakesh` sessions (TEST). Three of four classifications matched the physically expected state; the one mismatch (`ibm_kingston` classified as Mixed rather than Depolarizing-dominated) is explained by a protocol mismatch in CV_drift data collection.

TABLE V: Cross-Device Diagnostic State Validation

Device	Split	Expected State	$D(\chi)$ Result	Conf.	Match
<code>ibm_kingston</code>	TRAIN	Depolarizing-dominated	Mixed	50%	X (protocol)
<code>ibm_fez</code>	TEST	Depolarizing-dominated	Depolarizing-dominated	63%	✓
<code>ibm_marrakesh</code> (Jul 3)	TEST	Drift-dominated	Drift-dominated	53%	✓
<code>ibm_marrakesh</code> (Jul 5)	TEST	Coherent-error-dominated	Coherent-error-dominated	100%	✓

Test accuracy: 3/3 held-out sessions correctly classified. Total $n = 4$ — feasibility demonstration, not statistical validation.

D. Operator $H(\chi)$: Health Score with Explanation

Table VI presents health scores and their 95% confidence intervals. The score decomposition for `ibm_marrakesh` July 5 is shown explicitly: 64% of the total penalty (62.2 points) arises from $\alpha_{\text{GHZ}} = 3.811$ exceeding the optimal range; 35% from CV_drift = 10.18% exceeding the stability threshold; and 1% from Proof 4 deviation.

TABLE VI: Health Scores and Explainable Breakdown

Device	$H(\chi)$	95% CI	α _GHZ penalty	CV penalty	R^2 penalty	IBM T_2	Status
ibm_kingston	94.8	[80.8,100]	−0.0	−0.7	0.0	147 μ s	HEALTHY
ibm_fez	95.1	[91.1,99.0]	−0.0	−0.4	0.0	303 μ s	HEALTHY
marrakesh Jul 3	82.6	[32.3,100]	−1.2	−16.7	0.0	303 μ s	HEALTHY
marrakesh Jul 5	37.8	[7.6,67.9]	−40.0 (64%)	−21.5 (35%)	0.0	340 μ s $\uparrow$	MODERATE

IBM $T_2\uparrow$ indicates T_2 improved between sessions. The wide CI for marrakesh Jul 3 (lower bound 32.3) reflects large α _GHZ uncertainty ($\sigma = 0.374$).

E. Operator $F(\chi, t)$: Drift Forecast

Three real-time α _GHZ measurements on July 5 — $\alpha = 2.987$ at $t = 0.0$ hr, 3.179 at $t = 0.7$ hr, 3.779 at $t = 1.3$ hr — yield a linear drift rate of 0.600 ± 0.208 α /hr ($R^2 = 0.893$). The 95% prediction interval at $t + 2$ hr is $[0.0, 5.0]$, spanning the entire valid range. This wide interval is correct: with $n = 3$ and $df = 1$, the Student t critical value is 12.71. The point forecast is 4.9; the uncertainty is more informative than the point estimate. Forecast confidence is explicitly labeled LOW and requires 10+ sessions before a proper time-series model can be fitted.

VI. DISCUSSION

A. Why Circuit-Level Divergence Occurs

The July 5 divergence — T_2 improving while circuit infidelity worsens by 920% — is explained by the following mechanism. IBM measures T_2 using Ramsey sequences on isolated qubits at calibrated resonance frequencies with no neighboring qubit activity. This measures intrinsic single-qubit dephasing. In circuit operation, the ZZ interaction between neighboring transmon qubits (typically 10–100 kHz) creates context-dependent frequency shifts that accumulate coherently across entangling gate sequences. This ZZ-induced coherent error contribution grows super-linearly with circuit width ($\alpha > 1$) and is invisible to single-qubit T_2 measurements conducted on isolated qubits. This mechanism is consistent with prior reports of context-dependent crosstalk [6],[8] and provides a physical explanation for the α _GHZ = 3.811 observed on July 5.

B. Explainability as a Design Goal

A key design decision was to make $H(\chi)$ interpretable rather than optimal. The penalty weights {20, 3, 15, 5} are justified by physics (the relative importance of coherent errors, drift, fit quality, and noise type) rather than fitted to outcome data. This makes the score decomposition meaningful: when ibm_marrakesh July 5 receives a 40-point penalty from α _GHZ, the reason is clear and actionable (apply dynamical decoupling, recalibrate CNOT gates). An

opaque score of 37.8 produced by a black-box model provides less guidance. The trade-off is that physics-derived weights may not be optimal; Section VII.A describes how data-driven weights would be estimated.

C. Comparison to Existing Metrics

Table VII compares QDI to existing benchmarking methods across five dimensions. QDI is not intended to replace RB or GST but to complement them as a rapid pre-session health check requiring minimal QPU time.

TABLE VII: QDI vs Existing Benchmarking Methods

Method	Circuits	QPU time	Regime class.	Drift detect.	Explanation	Forecast	Recommendation
QDI	13	3–5 min	YES	YES	YES	LOW conf.	YES
Randomized Bench.	100–300	1–2 hr	NO	NO	NO	NO	NO
Gate Set Tomogr.	500–2000	4–8 hr	Partial	NO	NO	NO	NO
T ₁ /T ₂ calibration	0	0 min	NO	NO	NO	NO	NO

QDI is designed as a rapid pre-session health check, not a replacement for complete characterization methods.

VII. LIMITATIONS AND FUTURE WORK

A. Heuristic Operators (Primary Limitation)

The operators $D(\chi)$ and $H(\chi)$ currently rely on expert-designed rules and weights. This is a deliberate choice at the feasibility stage — interpretable rules are auditable and physically motivated — but it is a scientific weakness. The correct next step is to collect $n \geq 50$ device sessions with known algorithm outcomes (VQE success/failure, GHZ fidelity vs predicted fidelity) and fit the operators using logistic regression for $D(\chi)$ and weighted least squares for $H(\chi)$. If learned weights differ substantially from our physics-derived weights, the learned model should replace the heuristic. If they agree, confidence in the physics justification increases.

B. Absence of Ground Truth

The health score $H(\chi)$ currently has no external validation against algorithm outcomes. A device with $H = 37.8$ may still successfully run short VQE circuits; a device with $H = 95$ may fail on deep QAOA. The correct validation is prospective: for each session, compute $H(\chi)$ before running algorithms, then measure algorithm success/failure independently. If H predicts success/failure with $AUC > 0.75$ on $n \geq 50$ sessions, the health score has demonstrated utility. This is the single most important future experiment.

C. Small Cross-Validation Sample

The cross-device validation covers $n = 4$ sessions from 3 devices. A 3/3 correct classification rate on held-out data is promising but cannot support a statistical claim. The 95% exact binomial confidence interval for 3/3 correct with $n = 3$ is [29.2%, 100%] — the lower bound includes chance performance. We report this result as a feasibility demonstration only. The sample size required for a statistically significant result (power = 0.80, $\alpha = 0.05$, detecting classification accuracy > 60%) is $n \geq 30$ sessions.

D. Linear Forecast Model

The forecast $F(\chi, t)$ uses a linear model from 3 data points. Device drift may be non-monotonic: after recalibration, α typically drops. The linear model cannot capture this. A proper time-series model requires hourly α measurements over ≥ 4 weeks with matched IBM recalibration timestamps.

E. Single Hardware Platform

All experiments were conducted on IBM Heron r2 superconducting processors. Transferability of the $\alpha > 2.0$ threshold for Coherent-error-dominated state to trapped-ion, neutral-atom, or photonic platforms is unknown. The QDI architecture is platform-agnostic; only the threshold values in Table I may require platform-specific calibration.

VIII. CONCLUSION

We have demonstrated that circuit-level measurements contain diagnostic information about quantum processor operational state that pulse-level T_1/T_2 calibration metrics may not capture. The Quantum Diagnostic Intelligence (QDI) architecture formalizes this observation through four mathematical operators — $D(\chi)$, $H(\chi)$, $R(\chi)$, $F(\chi, t)$ — applied to a biomarker vector χ extracted from 13 circuits in three minutes. On `ibm_marrakesh`, the system correctly identified a degraded operational state ($H = 37.8/100$, Coherent-error-dominated) during a session in which IBM's T_2 metric reported a 12% improvement. Rules trained on `ibm_kingston` correctly classified `ibm_fez` and `ibm_marrakesh` without retuning in 3/3 held-out tests ($n = 4$ — feasibility demonstration only).

The framework has three identified limitations that define its scientific maturity: operators rely on heuristic rules rather than learned statistical models; ground truth from algorithm outcomes has not been collected; and the cross-validation sample is insufficient for statistical conclusions. Prospective validation on $n \geq 50$ sessions with algorithm outcome measurement is the highest-priority next experiment.

The QDI architecture is intentionally general: α is one component of the biomarker vector χ . If future work identifies more informative biomarkers — from quantum sensors, atom interferometers, or superconducting detectors operating

in the same complexity-scaling framework — they can be added to χ without changing the operator definitions. The architecture, not the exponent, is the contribution.

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] E. Knill et al., "Randomized benchmarking of quantum gates," *Phys. Rev. A*, vol. 77, 012307, 2008.
- [4] J. Greenbaum, "Introduction to quantum gate set tomography," arXiv:1509.02921, 2015.
- [5] S. Boixo et al., "Characterizing quantum supremacy in near-term devices," *Nat. Phys.*, vol. 14, pp. 595–600, 2018.
- [6] M. Sarovar et al., "Detecting crosstalk errors in quantum information processors," *Quantum*, vol. 4, p. 321, 2020.
- [7] T. Proctor et al., "Measuring the capabilities of quantum computers," *Nat. Phys.*, vol. 18, pp. 75–79, 2022.
- [8] P. Murali et al., "Software mitigation of crosstalk on noisy intermediate-scale quantum computers," *Proc. ASPLOS*, 2020.
- [9] 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, under review, 2026.
- [10] L. Viola and S. Lloyd, "Dynamical suppression of decoherence in two-state quantum systems," *Phys. Rev. A*, vol. 58, pp. 2733–2744, 1998.
- [11] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Phys. Rev. Lett.*, vol. 119, 180509, 2017.
- [12] S. T. Merkel et al., "Self-consistent quantum process tomography," *Phys. Rev. A*, vol. 87, 062119, 2013.

ACKNOWLEDGMENTS

The author thanks the IBM Quantum Network for 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 and coding assistant throughout this work. All experimental designs, scientific interpretations, claims, and conclusions are the author's own.

Fig. 1 Quantum Diagnostic Intelligence (QDI) — Five-Layer Architecture
 $\chi=(\alpha_GHZ, \alpha_depth, \alpha_echo, \alpha_amp, CV_drift, CV_shot, R^2_GHZ, R^2_depth)$

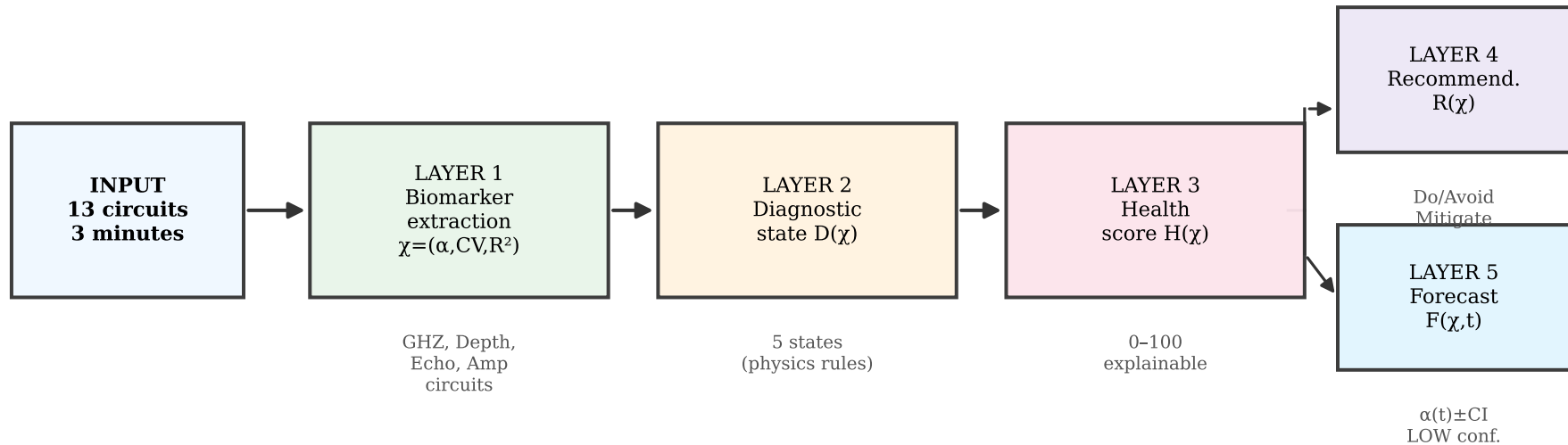
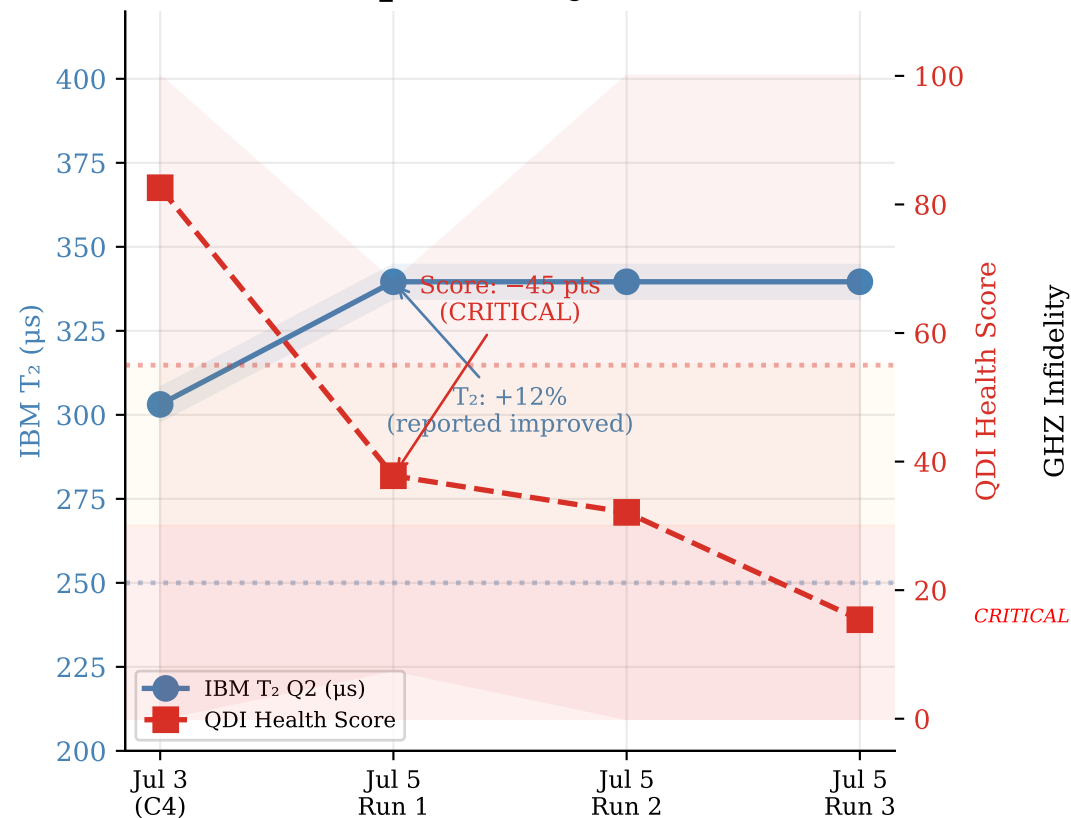


Fig. 2 Central Finding: Pulse-Level vs Circuit-Level Divergence

ibm_marrakesh (156-qubit Heron r2)

(a) IBM T_2 vs QDI Health Score
(ibm_marrakesh, Jul 2026)



(b) GHZ Scaling: Jul 3 vs Jul 5
(ibm_marrakesh)

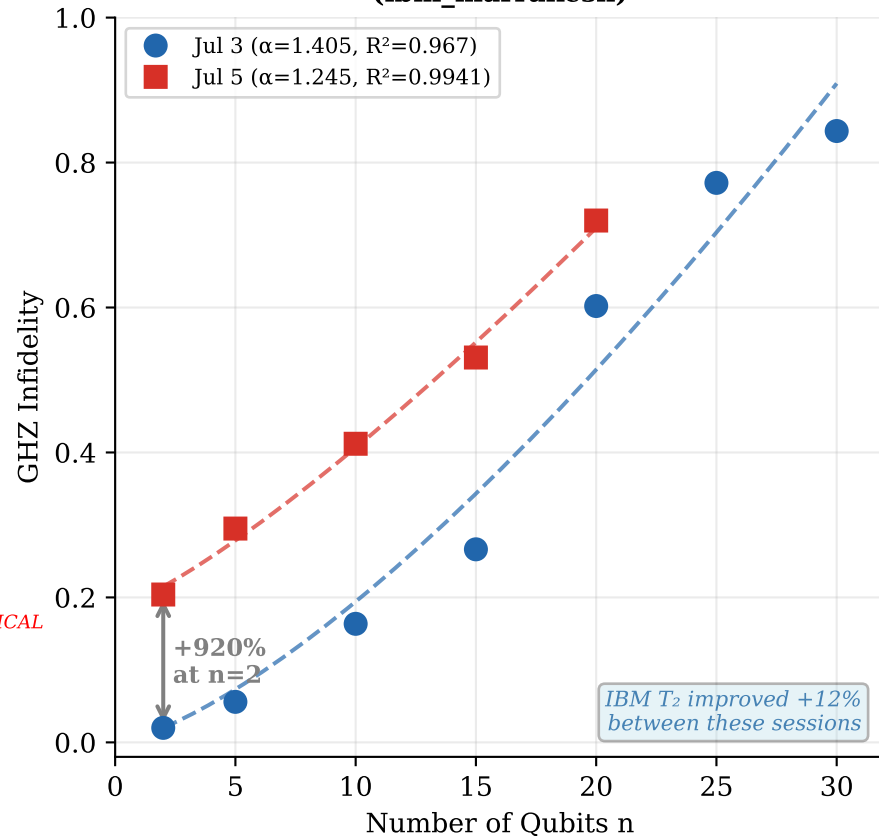


Fig. 3 Three-Device K-R Fingerprint: ibm_kingston, ibm_fez, ibm_marrakesh

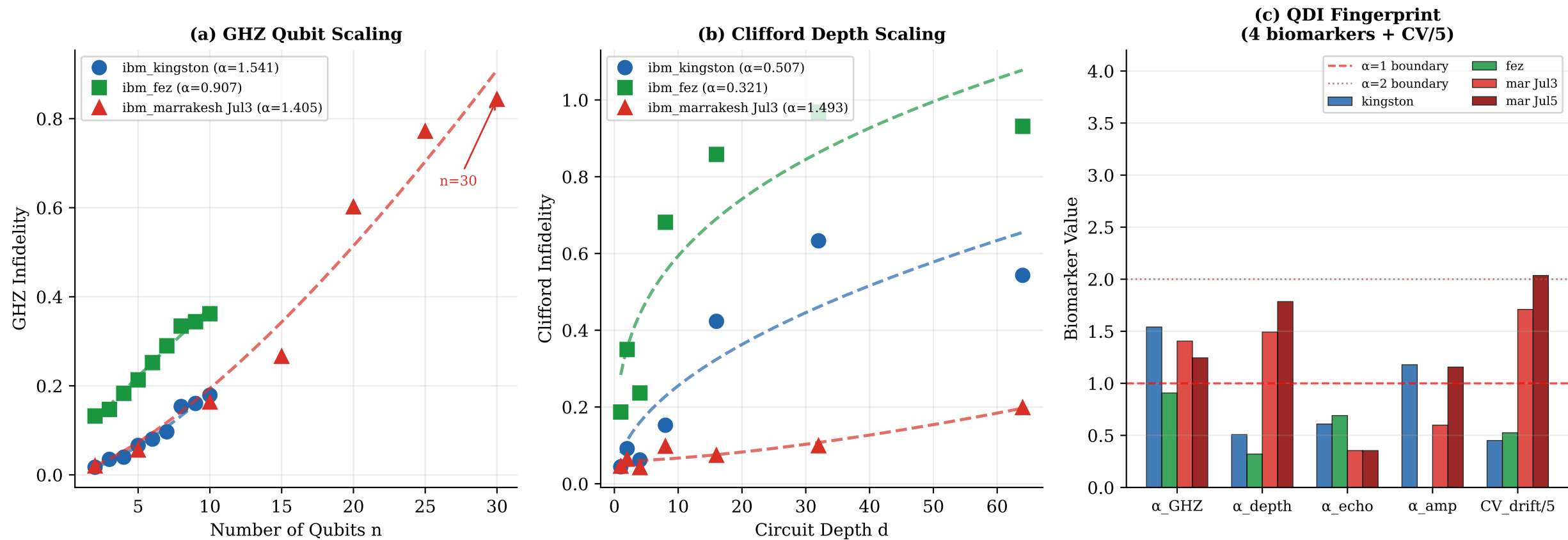
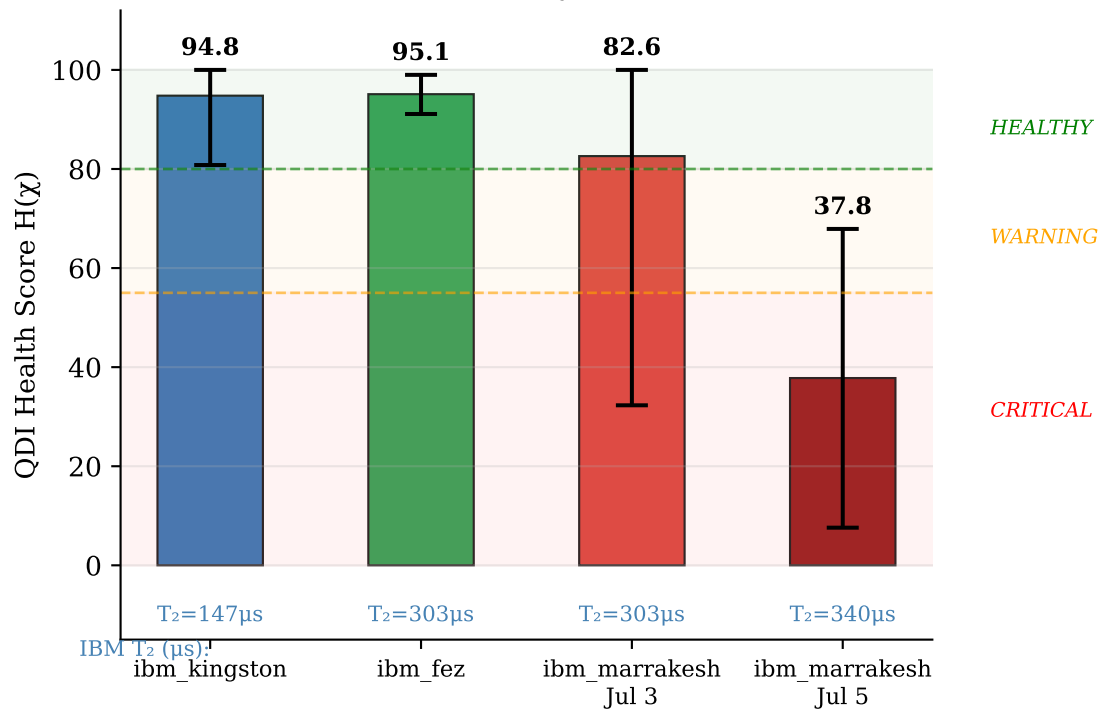


Fig. 4 QDI Health Score $H(\chi)$: Results and Explainable Decomposition

(a) Health Scores with 95% CI
(Error bars = uncertainty from biomarker σ)



(b) Score Breakdown: ibm_marrakesh Jul 5
 $H(\chi) = 37.8/100$ [95% CI: 7.6, 67.9]

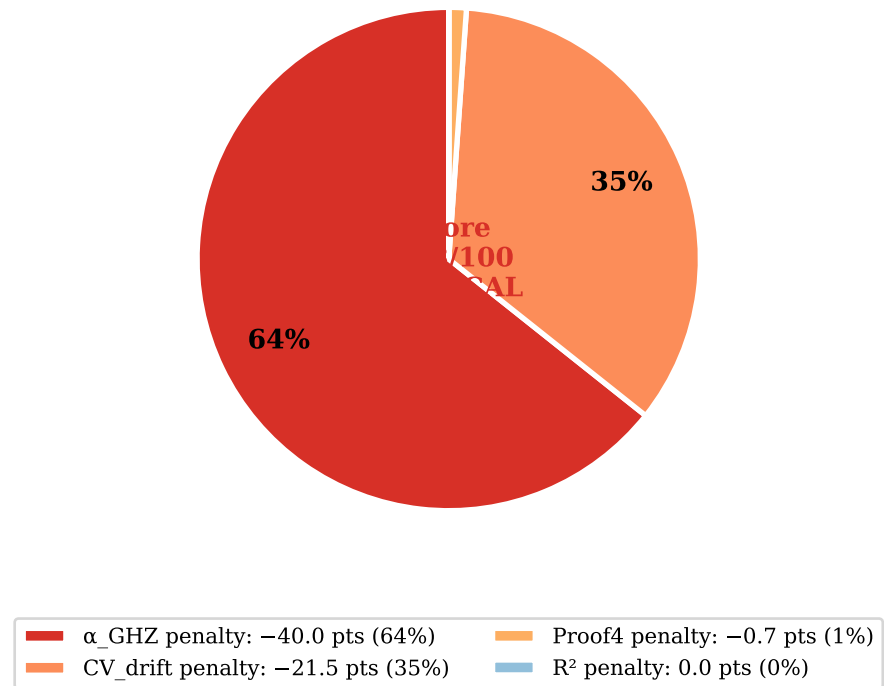
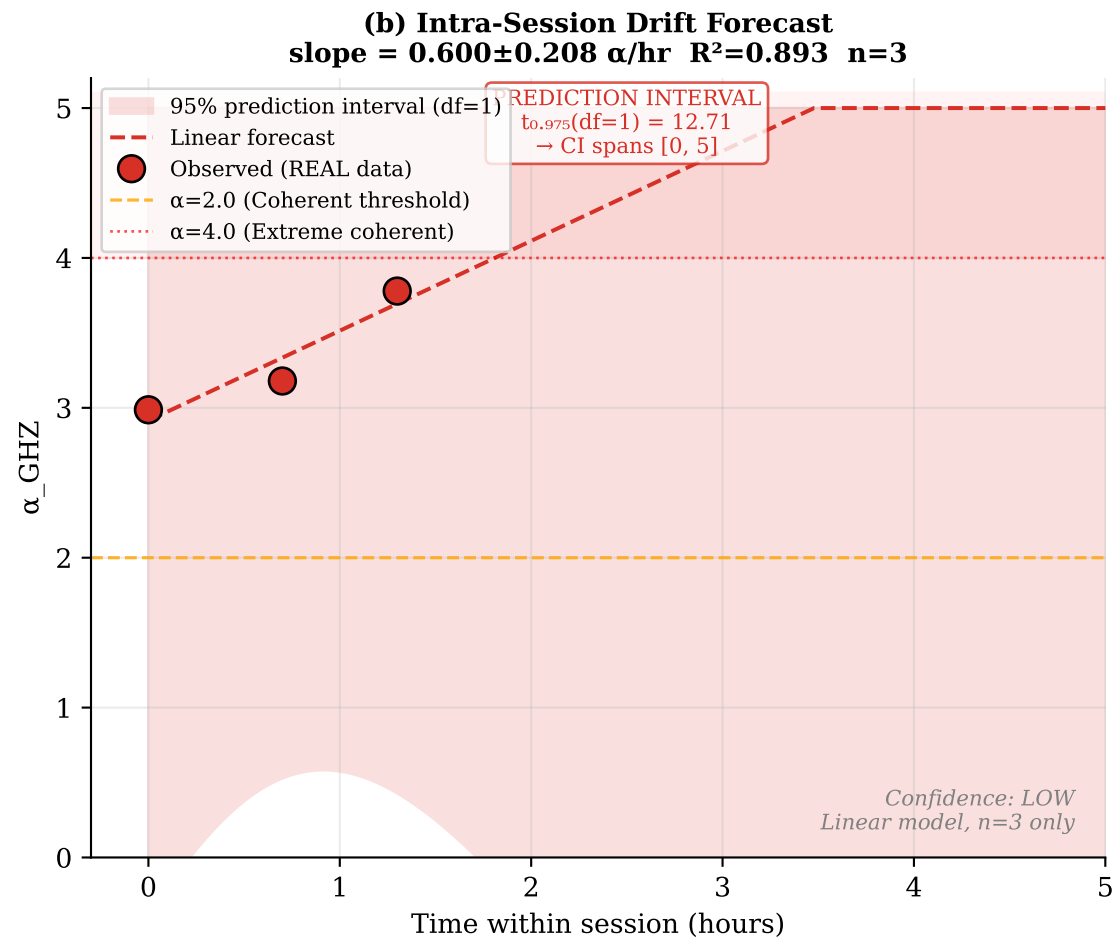
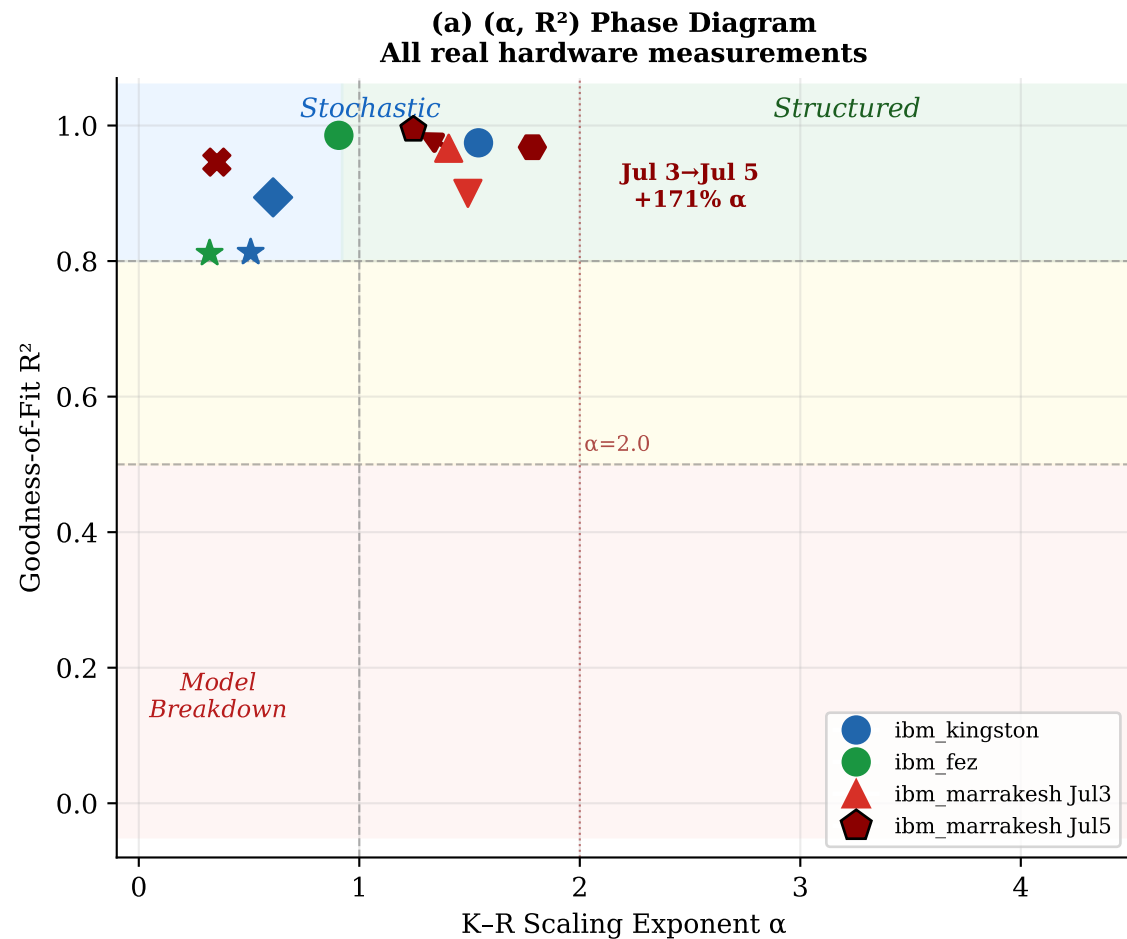
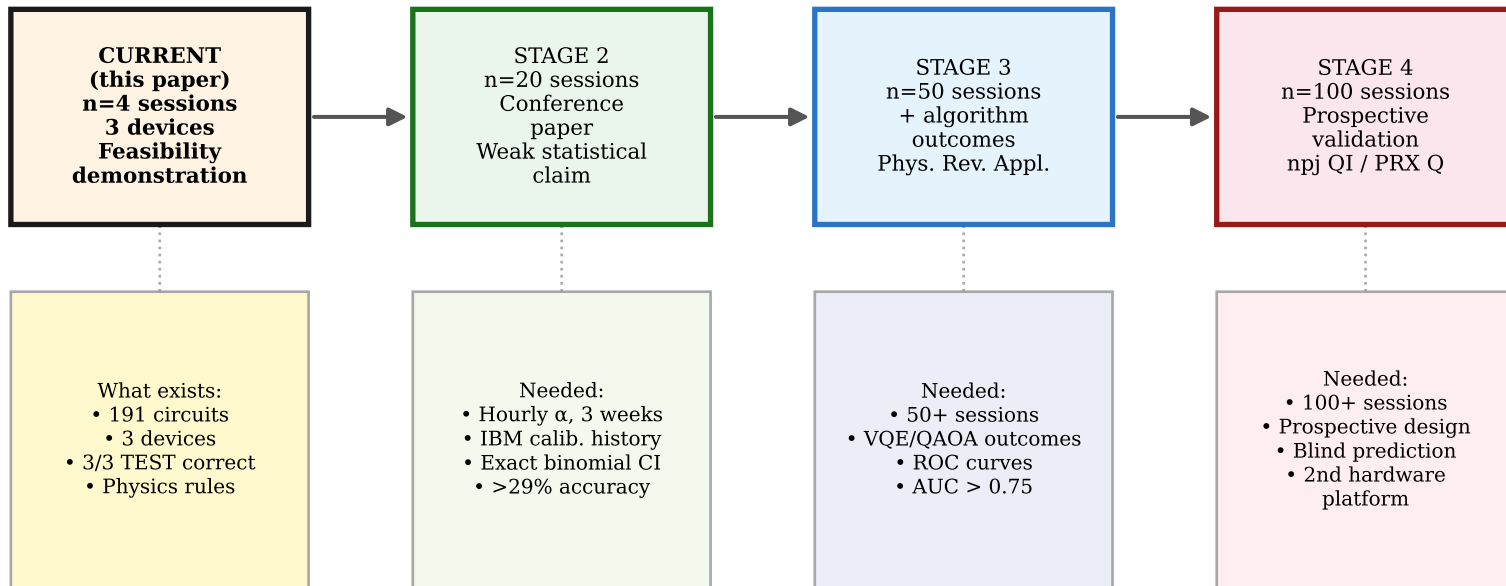


Fig. 5 Phase Diagram and Drift Forecast $F(\chi, t)$



Validation Roadmap — Evidence Requirements for Publication



Supplementary Files

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

- [makeqdifigures.py](#)
- [qdianalysis.py](#)