

Toward Live Noise Fingerprinting for Discrepancy Analysis in Quantum Software Engineering

Avner Bensoussan
King’s College London, UK

Elena Chachkarova
King’s College London, UK

Karine Even-Mendoza
King’s College London, UK

Sophie Fortz
Inria, Univ. Rennes, CNRS, IRISA, France

Vasileios Klimis
Queen Mary University of London, UK

Mohammad Reza Mousavi
King’s College London, UK

Abstract—Contemporary quantum computers are inherently noisy, posing significant challenges for the development and testing of quantum software. Simplified or outdated noise assumptions can lead to incorrect assessments of program correctness, obscure debugging, and hinder cross-platform portability, creating a critical quantum software development gap. Providing accurate, practical noise characterisation is challenging as traditional reconstruction methods scale exponentially and rapidly become outdated.

In this vision paper, we address this gap via a novel *classical shadow tomography*-based pipeline, SIMSHADOW, enabling efficient, continuously updatable *noise fingerprinting* from empirical observations, suitable for integration into quantum software development workflows, including testing and validation. We prototyped the pipeline to investigate fingerprints’ ability to capture structured, interpretable noise and cross-platform discrepancies affecting quantum programs’ behaviour to support realistic testing and debugging in future tools. Our evaluation with *Qiskit* and *Cirq* under widely used hardware-informed profiles, *IBM Boston* and *Quantinuum H2*, shows fingerprints exhibit channel-specific structure and yield interpretable heatmaps. We observed systematic cross-platform discrepancies under matched noise configurations, quantified by large Frobenius distances at a fraction of full tomography cost. On 69 MQTBENCH programs, larger fingerprint differences correlate with output distributions divergences, highlighting threats for testing and cross-platform debugging tasks.

Index Terms—QSE, Noise Models, Cross-Platform Validation

I. INTRODUCTION

In contemporary Quantum Software Engineering (QSE) [1], [2], activities such as development, compilation and debugging face a unique obstacle: sensitivity to the “noise” of quantum hardware. Quantum programs are affected by fragile qubits whose behaviour is disrupted by environmental fluctuations and imperfect control [3]–[6]. In quantum ecosystems (e.g. *Qiskit* or *Cirq*), simulators are indispensable for QSE, providing controlled environments for a variety of tasks and enabling low-cost experimentation [7], [8]. Simulators such as those for *Qiskit* [9] and *Cirq* [10] must capture not only the ideal logic of algorithms but also the noisy, platform-specific imperfections of real hardware [3], [11], [12]. These imperfections are abstracted into *noise models*, which are often inaccurate and out-of-sync with respect to real hardware [13], [14]. Recent work has focused on improving these models [13], [15]–[17], highlighting their central importance for reliable and effective QSE [16], [18]–[21].

Yet, a critical *conceptual gap* exists: there is a significant lack of accurate and practically actionable noise models. Although platforms offer typical noise channels (e.g. “depolarising noise”), the underlying semantics of their implementations differ significantly. This undermines the role of simulators in QSE for evaluating quantum software behaviour in noisy environments and represents a fundamental, unaddressed challenge for building reliable and transferable quantum software.

We propose *efficient noise fingerprinting* as a new vision for addressing this challenge. We adapt *classical shadow tomography* [22], a lightweight quantum mechanics technique, as a software engineering analysis tool to generate rich, descriptive signatures of a simulator’s noise profile, thereby avoiding the exponential cost of accurately reconstructing the quantum state under noise. We investigate this paradigm empirically. As a proof of concept, we implement the pipeline, SIMSHADOW, and study two research questions:

RQ1 Do fingerprints reveal and differentiate noise behaviours within a platform?

RQ2 Can SIMSHADOW detect and quantify differences between platforms with identical noise configurations?

Investigating this enables downstream applications such as fault localisation, differential testing (distinguishing implementation bugs from noise-induced variation), and automated testing under realistic noise conditions. Preliminary *Qiskit-Cirq* comparison yields interpretable heatmaps, revealing systematic implementation differences, with fingerprints exhibiting channel-specific structure. Across six configurations (3 noise channels \times 2 hardware-informed profiles, *IBM Boston* and *Quantinuum H2*), the platform-level fingerprints Frobenius distance (≈ 0.0044 – 0.0056) is an order of magnitude greater than the bootstrap uncertainty (≈ 0.0004), while remaining substantially more efficient than full process tomography.

Our work is the first to close the aforementioned conceptual gap by leveraging classical shadow tomography to enable efficient noise fingerprinting for quantum software platforms in the QSE context. We envisage that our vision for practical noise fingerprinting will be transformative for other areas, including noise-aware compilation, transpilation, cross-platform validation, verification, and error mitigation.

Data Availability. Code, fingerprints images, results, and testing notebooks (§IV-B) are openly available at [23].

II. BACKGROUND & RELATED WORK

Basics of Quantum Measurement. [3]. A (pure) quantum state is represented by a ket $|\psi\rangle$ in a complex Hilbert space; its dual bra is $\langle\psi| = (|\psi\rangle)^\dagger$; for a single qubit with computational basis $\{|0\rangle, |1\rangle\}$, the general state is a *superposition* $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex amplitude coefficients (e.g. $|+\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$). An *observable* O is a Hermitian matrix whose eigenvalues are the possible measurement outcomes; the expected result of measuring O on $|\psi\rangle$ is $\langle\psi|O|\psi\rangle$. Density matrix $\rho = |\psi\rangle\langle\psi| \in \mathbb{C}^{2^n \times 2^n}$ represents an n -qubit system general state and scales exponentially in n .

Error Mitigation. Quantum computations are affected by noise from imperfect control, decoherence, and measurement errors [3]. Noise is studied on *quantum hardware*, where device-specific effects can be time-dependent, correlated, and non-Markovian or in *simulation*, where errors are modelled via *abstract quantum channels* (e.g. depolarising or amplitude-damping) but fail to reproduce accurately the statistical behaviour observed on devices. *Error mitigation* seeks to reduce noise impact without full quantum *error-correction* overhead, e.g. [24], [25]. Shadow tomography and classical shadows, used alongside error mitigation methods, utilise randomised measurements to estimate many observables without full state reconstruction. *Shadow tomography* [26], [27] retains a list of measurement vectors for implicit density matrix approximation. *Classical shadows* [22] compacts this list into *classical* data structures for prediction of observables. SIMSHADOW uses shadow-inspired estimation but differs from both: it stores no shadow representation and maps outcomes directly to deviation signatures tailored specifically to QSE.

Quantum Software Testing is challenging, with techniques such as property-based [28], [29], mutation [30] and metamorphic [31], [32]. Benchmarks such as Bugs4Q [33] and MQTBench [34] support testing and debugging studies. Most approaches assume ideal or simplified noise models [32], [35], compromising test reliability. Noise-aware testing [36] demonstrates the need for practical, noise characterisation to support robust testing, debugging, and cross-platform analysis.

Quantum Fingerprinting was introduced by Buhrman et al. [37] to encode inputs into exponentially small quantum states for equality testing. We intend to use their framework in the future to establish formal properties of our approach.

III. LIVE EMPIRICAL NOISE FINGERPRINTING

We study the problem of undocumented discrepancies in quantum noise models by introducing SIMSHADOW to create lightweight noise fingerprints to characterise noisy behaviour of a quantum simulator (e.g. *Qiskit's AerSimulator*) as a live, dynamic signature for QSE tasks using two core ideas: 1) noise models fingerprinting and 2) quantum measurements.

Defining the Fingerprint. (§III-A with formalism in §III-B) To analyse these discrepancies, we introduce the notion of a noise fingerprint as a compact, rapidly updatable matrix of a simulator's observed outcome deviations from ideal outcomes, distinguishing it from physics-oriented noise models for device-level diagnostics. By probing reference states with

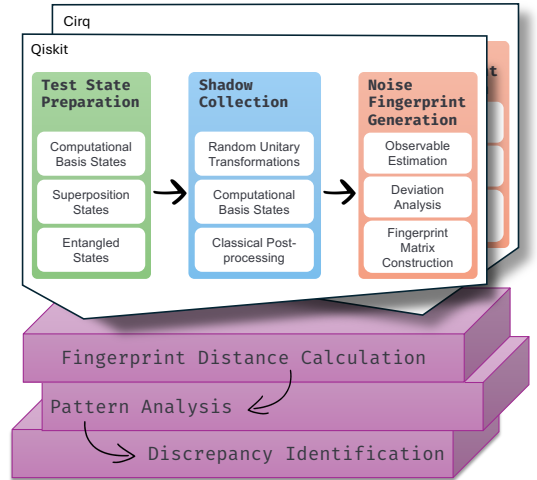


Fig. 1: SIMSHADOW Architecture

a fixed set of Pauli observables and aggregating deviations, SIMSHADOW obtains a descriptive matrix, making simulator differences explicit, measurable, and comparable across noises, simulators and platforms.

Analysis & Application. Fingerprints are then used in the analysis loop (§III-C). We applied it to *Qiskit* and *Cirq*, revealing systematic discrepancies in their simulator behaviour (§IV); we outline broader applications in §V.

A. The Noise Fingerprinting Paradigm

Existing vendor-supplied noise models provide the closest available characterisation of device behaviour. Still, they are necessarily simplified, offer limited transparency and detail, and are updated only periodically (e.g. every 24 hours). We advocate for a new paradigm centred on live, empirical and descriptive noise fingerprinting, designed to be transparent and accessible, with the following 3 key properties.

Empirical. A fingerprint is generated from *empirical data* by actively probing the target system (simulator or hardware) through measurement, not from a theoretical model or vendor's datasheet. Here, fingerprints represent the *actual*, “de facto” behaviour of the system, not the “de jure” behaviour described in its documentation. Our approach grounds noise understanding in the observable reality, “de facto” system behaviour, unlike (e.g.) process tomography [38] and gate set tomography [39], which provide physically complete but exponentially costly noise characterisations impractical beyond a few qubits.

Live and Updatable. Our fingerprint is *live*: it can be generated quickly and updated repeatedly. This is vital for hardware, where calibration drifts within hours, and for software, where fingerprints can be generated on demand for specific versions or commits in a CI/CD pipeline. Prior QSE methodologies mainly focused on testing and benchmarking [2], reporting on execution speed [35], [40], [41]. Our approach captures fidelity and consistency by transforming noise from a static property into a dynamic, observable state of the system.

Descriptive and Actionable. Randomised benchmarking [42] yields only an aggregate fidelity score, revealing nothing about *how* or *why* simulators differ. We instead propose fingerprints as richer signatures: a *descriptive, structured signature* as a

matrix of deviations. Its patterns may reveal the *type, strength* and *asymmetries* of noise, making fingerprints *actionable*.

B. Fingerprint as a Function

A *configuration* c specifies the ecosystem, backend (or hardware-informed profile), noise channel family and parameters, and software version/seed policy. Given a fixed probe suite of k reference states $\{|\psi_i\rangle\}$ and m Pauli observables $\{O_j\}$, SIMSHADOW defines the fingerprint $F(c) \in \mathbb{R}^{k \times m}$:

$$F_{i,j}(c) = \hat{\mu}_{i,j}(c) - \mu_{i,j}^{\text{ideal}}, \quad \mu_{i,j}^{\text{ideal}} = \langle \psi_i | O_j | \psi_i \rangle. \quad (1)$$

Here $\hat{\mu}_{i,j}(c)$ is the empirical estimator of $\mathbb{E}[O_j] = \text{Tr}(\rho_c^{(i)} O_j)$, obtained by executing the noisy circuit for $|\psi_i\rangle$, rotating into the measurement basis of O_j , sampling N shots, mapping outcomes to ± 1 , and averaging. This approach is *shadow-inspired*, as it efficiently estimates selected Pauli expectations without reconstructing the full density matrix or storing a classical shadow representation (§II).

C. Shadow-Inspired Estimation

SIMSHADOW targets deviation estimates rather than full physical reconstruction. Its fingerprints are intentionally less precise than physics-oriented tomography but are lightweight, interpretable, and well-suited for software analysis and testing.

Our *protocol for generating and analysing a fingerprint* instantiates this paradigm through the following stages, illustrated in Figure 1.

1. Prepare Reference States. We prepare a small but diverse, well-characterised quantum states set $\{|\psi_i\rangle\}$, chosen so that different error mechanisms manifest distinctly in the fingerprint:

- *Computational basis states* (e.g. $|00\rangle, |11\rangle$), primarily sensitive to population errors such as amplitude damping.
- *Superposition states* (e.g. $|+\rangle|1\rangle, |0\rangle|+\rangle$), sensitive to coherence loss such as phase damping.
- *Entangled states* (e.g. a two-qubit GHZ state), necessary to probe correlated multi-qubit errors such as crosstalk.

2. Estimate a Selected Set of Observables. After evolving each reference state under a configuration, c , we estimate expectation values for a selected set of Pauli observables $\{O_j\}$, including two-qubit correlators. Rather than reconstructing a density matrix, we directly estimate observables required for the fingerprint, capturing correlated noise effects efficiently.

3. Construct the Deviation Fingerprint. For each reference state $|\psi_i\rangle$ (rows) and Pauli observable O_j (columns), we assemble the fingerprint matrix $F(c) \in \mathbb{R}^{k \times m}$, where entry $F_{i,j}(c)$ corresponds to measuring O_j on $|\psi_i\rangle$. Thus, $F(c)$ forms a state–observable grid of deviations from ideal behaviour, not a density matrix. Ideal expectations are computed analytically; empirical estimates $\hat{\mu}_{i,j}(c)$ are obtained without density matrix reconstruction. We compare configurations using the Frobenius distance $\|F(c_A) - F(c_B)\|_F$, yielding a compact, quantitative measure of cross-platform divergence.

4. Algorithmic Analysis. We analyse the resulting fingerprint matrices empirically using summary statistics and visual comparisons, as detailed in the next section.

IV. EMERGING RESULTS

We generated 12 fingerprints, one per configuration, and compared them visually and through statistical distances and divergences to answer RQ1 and RQ2 (§I).

A. Experimental Setup

Configurations. The evaluation used 12 configurations (3 noise channels \times 2 simulators \times 2 profiles). We evaluated SIMSHADOW on 2 simulator ecosystems: *Qiskit* and *Cirq*, with 2 hardware-informed profiles: *IBM Boston* (1Q gate-error proxy with damping mapped per ~ 1 Q gate step; see IBM’s Aer noisy simulation guidance [43], [44]) and *Quantinuum H2* (datasheet-anchored 1Q gate infidelity and depth-1 memory error; Quantinuum reports memory error rather than T_1/T_2 [45]–[47]). We used 3 noise channels: *depolarising*, *amplitude damping*, and *phase damping*, modelling respectively generic random errors, energy relaxation (a T_1 -like decay $|1\rangle \rightarrow |0\rangle$), and pure dephasing (a T_2 -like loss of coherence) [48]–[50].

Probe suite. We designed a protocol adapted from standard practices in quantum system characterisation [42], [51]. We used $k = 13$ reference quantum states prepared with standard minimal circuits, strategically chosen to be sensitive to a wide range of error mechanisms (§III-C): 4 *computational basis states*, 8 *superposition states* and 1 *entangled state*. We measured $m = 9$ two-qubit Pauli observables, covering all correlated errors per state, utilising a total of 117 state–observable pairs (13 states \times 9 observables, i.e. size of each fingerprint).

Statistical Robustness. We ran $n=1000$ independent repetitions of the probe suite per configuration. We reported the mean fingerprint across repetitions with $\text{shots}=10,000$ per state–observable measurement to reduce finite-shot sampling error, and as shot count being simulators divergence’s prime factor [35]. Furthermore, we bootstrapped n runs: 2,000 resamples with replacement, and per resample computed mean *Qiskit* and *Cirq* fingerprints and their Frobenius distance, with std. dev. being std. err. of our point estimate (distance between original mean matrices).

B. Threats to Validity

To reduce the risk of discrepancies stemming from implementation or experimental environment, we developed the pipeline incrementally and validated intermediate steps via dedicated test notebooks before constructing the full flow. To minimise environment-induced variation, we pinned dependencies and ran all experiments in the same virtual environment (CloudLab c6525-25g [52], [53]). Also, we utilised statistically robust setup (§IV-A): Expectation values lie in $[-1, 1]$, hence fingerprint differences are in $[-2, 2]$; the observed run-to-run variability is small relative to this range (std. dev. ≤ 0.01076). We used two fixed-parameter synthetic noise profiles, *Low*¹ and *High*², to verify that the *High* profile induces stronger decoherence and correspondingly larger cross-platform (*Qiskit* vs *Cirq*) discrepancies than the *Low* profile; this was confirmed.

1. i.e. $p_{\text{dep}} = 5 \times 10^{-4}$, $\gamma_{\text{AD}} = 1 \times 10^{-4}$, $\gamma_{\text{PD}} = 2 \times 10^{-4}$.
2. i.e. $p_{\text{dep}} = 5 \times 10^{-3}$, $\gamma_{\text{AD}} = 2 \times 10^{-3}$, $\gamma_{\text{PD}} = 5 \times 10^{-3}$.

C. Results

Figure 2 visualises mean fingerprints (13×9 matrix) with reference states as rows and two-qubit Pauli observables as columns (e.g. XX, XY, \dots). Entries report deviations from the corresponding ideal expectations. Rows group configurations by hardware-informed profile and platform: *IBM Boston Qiskit* (A–C) and *Cirq* (D–F); *Quantinuum H2 Qiskit* (G–I) and *Cirq* (J–L). Columns correspond to depolarising, amplitude-damping, and phase-damping noise channels.

RQ1: Each noise channel exhibits a distinct visual structure, supporting noise-channel identification in Figure 2. Within a fixed (profile, channel) pair, *Qiskit* and *Cirq* show clear, systematic differences. **In (D)**, $|\Phi^+\rangle$ shows the largest deviations in XX, YY , and ZZ , expectedly as these correlators are ideally ± 1 for $|\Phi^+\rangle$ and depolarisation uniformly shrinks all non-identity Pauli expectations towards zero; $|+0\rangle, |-0\rangle, |+1\rangle, |-1\rangle$ deviate strongly under XZ and $|0+\rangle, |0-\rangle, |1+\rangle, |1-\rangle$ under ZX , as these are exactly the pairs for which $\langle XZ \rangle$ or $\langle ZX \rangle$ is ideally ± 1 . **In (B) and (E)**, largest ZZ deviations occur for $|01\rangle$ and $|10\rangle$, consistent with relaxation driving single excitation states toward $|00\rangle$ and shifting $\langle ZZ \rangle$ from -1 toward $+1$; we observe pronounced changes for $|11\rangle$, as decay on either qubit first transfers population to $|01\rangle$ and $|10\rangle$ (which yield $ZZ = -1$) before relaxing to $|00\rangle$, producing a sizable net shift; for the remaining results, the largest deviations occur in $*Z$ and $Z*$ (3rd and last four col.), consistent with relaxation biasing Z populations (captured by ZZ) and suppresses coherences when a qubit is probed in the X/Y basis (captured by XZ, ZX, YZ, ZY); observables without a Z component are less effected. **In (H) and (K)**, *Qiskit* and *Cirq* show consistent behaviour for the ZZ observable on $|01\rangle, |10\rangle$, and $|11\rangle$, with deviations only for states more sensitive to amplitude damping.

RQ1 Answer: Yes. Fingerprints show channel-specific, theory-consistent structure across states and observables.

RQ2: Frobenius [54] distances between *Qiskit* and *Cirq* mean fingerprints were (approx.) for *IBM Boston* profile 0.0056 (depolarising), 0.0051 (amplitude damping), and 0.0049 (phase damping) and *Quantinuum H2*, 0.0046, 0.0051, and 0.0044, respectively. Bootstrap resampling yields std. err. of approx. 0.0004 for these distances, indicating that the observed differences are well above the uncertainty in estimating the mean distance (e.g. $0.0056 \gg 0.0004$), confirming simulators exhibit distinct behaviours despite identical noise configurations.

For two configuration pairs (F–I and A–D, with the former exhibiting smaller fingerprint deviations than the latter), we evaluated portability on 69 MQTBENCH QASM programs [34] of 5–11 qubits (see list in [23]). All 4 configurations were executed with `shots=10,000`; *Jensen–Shannon divergence* (JSD) was computed over 30 repeats per circuit (mean & std.). A–D exhibits slightly higher cross-configuration divergence on benchmark circuits (JSD 0.0728, std. dev. 0.1559) than I–F (JSD 0.0693, std. dev. 0.1364), supporting the link between fingerprint discrepancy and portability impact.

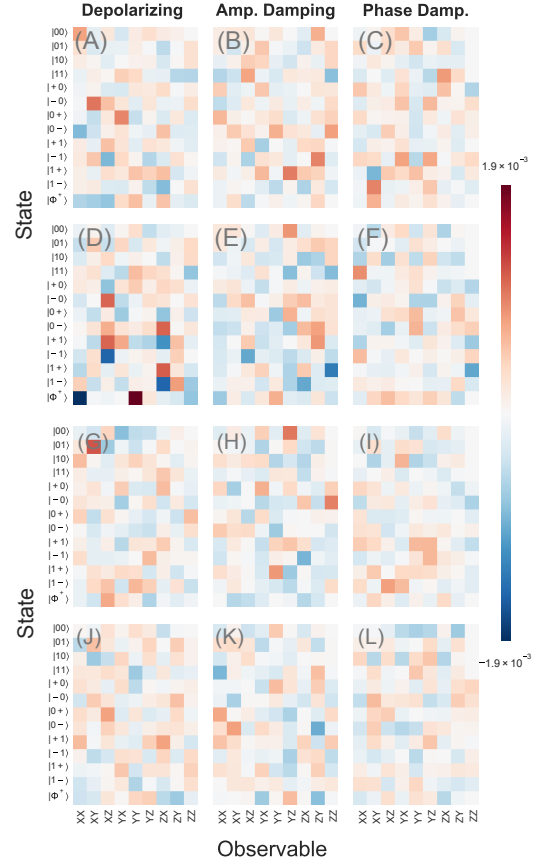


Fig. 2: *IBM Boston vs Quantinuum H2*.

RQ2 Answer: Matched *Qiskit-Cirq* configurations yield systematically different fingerprints quantified beyond bootstrap uncertainty and consistent with output divergence on MQTBENCH programs.

V. CONCLUSION AND OUTLOOK

Lightweight empirical noise signatures open new QSE tooling directions, supporting noise-aware debuggers and testing frameworks, aiding compiler optimisation, and educational artefacts, necessary for mature QSE ecosystems. We took the first step toward cross-platform portability analysis by generating a fingerprint for each simulator–backend–noise configuration and quantifying deviations from ideal expectations. Empirically, we observe gaps between configurations across ecosystems, providing developers a concrete diagnostic artefact to reason about when a quantum program may perform well under one configuration yet degrade when ported to another, and alternatively, near-identical fingerprints imply (at the measurement resolution) low portability risk. Future work will expand to additional platforms and hardware backends, refine estimation precision, integrate fingerprints into development and testing tools, and expand the probe set with additional observables and measurement bases to capture a wider range of noise effects and produce richer fingerprints.

Acknowledgement. Authors are listed in alphabetical order. We thank CloudLab [52] for providing the infrastructure that enabled our evaluation.

REFERENCES

- [1] J. M. Murillo, J. Garcia-Alonso, E. Moguel, J. Barzen, F. Leymann, S. Ali, T. Yue, P. Arcaini, R. Pérez-Castillo, I. García-Rodríguez de Guzmán, M. Piattini, A. Ruiz-Cortés, A. Brogi, J. Zhao, A. Miranskyy, and M. Wimmer, “Quantum software engineering: Roadmap and challenges ahead,” *ACM Trans. Softw. Eng. Methodol.*, vol. 34, no. 5, May 2025. [Online]. Available: <https://doi.org/10.1145/3712002>
- [2] M. De Stefano, F. Pecorelli, D. Di Nucci, F. Palomba, and A. De Lucia, “The quantum frontier of software engineering: A systematic mapping study,” *Information and Software Technology*, vol. 175, p. 107525, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584924001307>
- [3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information (10th Anniversary edition)*. UK: Cambridge University Press, 2010.
- [4] Z. Pan, Y. Feng, Z. Li, Y. Liu, and Y. Li, “Understanding the impact of quantum noise on quantum programs,” in *2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. Los Alamitos, CA, USA: IEEE, 2023, pp. 426–437.
- [5] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, “Noisy intermediate-scale quantum (NISQ) algorithms,” *Rev. Mod. Phys.*, vol. 94, p. 015004, Feb 2022. [Online]. Available: <https://link.aps.org/doi/10.1103/RevModPhys.94.015004>
- [6] T. Piskor, M. Schöndorf, M. Bauer, D. Smith, T. Ayril, S. Pogorzalek, A. Auer, and M. Papič, “Simulation and benchmarking of real quantum hardware,” 2025. [Online]. Available: <https://arxiv.org/abs/2508.04483>
- [7] IBM Quantum, “Ibm quantum pricing at, e.g., \$48 per minute.” <https://www.ibm.com/quantum/pricing>, 2025, accessed: 2025-09-16.
- [8] Microsoft Azure, “Azure quantum pricing,” <https://learn.microsoft.com/en-us/azure/quantum/pricing>, 2025, accessed: 2025-09-16.
- [9] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, “Quantum computing with Qiskit,” 2024.
- [10] C. Developers, “Cirq,” Apr. 2025. [Online]. Available: <https://zenodo.org/doi/10.5281/zenodo.4062499>
- [11] IBM Quantum, “Exact and noisy simulation with qiskit aer primitives,” <https://quantum.cloud.ibm.com/docs/en/guides/simulate-with-qiskit-aer>, Accessed: 2025-09-16.
- [12] Google Quantum AI, “Representing noise,” https://quantumai.google/cirq/noise/representing_noise, 2025.
- [13] S. R. Maschek, J. Schwitalla, M. Franz, and W. Mauerer, “Make some noise! measuring noise model quality in real-world quantum software,” in *Proceedings of the IEEE International Conference on Quantum Software (QSW)*. Los Alamitos, CA, USA: IEEE, jul 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/11134320>
- [14] J. A. Bravo-Montes, M. Bastante, G. Botella, A. del Barrio, and F. García-Herrero, “A methodology to select and adjust quantum noise models through emulators: benchmarking against real backends,” *EPJ Quantum Technology*, vol. 11, no. 1, p. 71, 2024. [Online]. Available: <https://doi.org/10.1140/epjqt/s40507-024-00284-4>
- [15] Y. Sung, D. Hangleiter, and J. Eisert, “Characterizing the effects of noise on quantum algorithms with classical shadows,” *Physical Review Research*, vol. 5, no. 4, p. 043210, 2023. [Online]. Available: <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.5.043210>
- [16] M. Heinrich, F. Ullrich, and C. Schmiegelow, “Noise-aware simulation and benchmarking for nisq devices,” *EPJ Quantum Technology*, vol. 11, p. 24, 2024. [Online]. Available: <https://epjquantumtechnology.springeropen.com/articles/10.1140/epjqt/s40507-024-00284-4>
- [17] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, “Machine learning of noisy quantum hardware,” *npj Quantum Information*, vol. 10, p. 96, 2024. [Online]. Available: <https://www.nature.com/articles/s41534-024-00955-1>
- [18] H. Kurniawan, L. Rodríguez-Soriano, D. Cuomo, C. G. Almudever, and F. G. Herrero, “On the use of calibration data in error-aware compilation techniques for nisq devices,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 01. Los Alamitos, CA, USA: IEEE, 2024, pp. 338–348.
- [19] P. Murali, J. M. Baker, A. Javadi-Abhari, F. T. Chong, and M. Martonosi, “Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers,” in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS ’19. New York, NY, USA: ACM, 2019, p. 1015–1029. [Online]. Available: <https://doi.org/10.1145/3297858.3304075>
- [20] G. Meuli, M. Soeken, M. Roetteler, and T. Häner, “Enabling accuracy-aware quantum compilers using symbolic resource estimation,” *Proc. ACM Program. Lang.*, vol. 4, no. OOPSLA, Nov. 2020. [Online]. Available: <https://doi.org/10.1145/3428198>
- [21] A. Muqet, T. Yue, S. Ali, and P. Arcaini, “Mitigating noise in quantum software testing using machine learning,” *IEEE Transactions on Software Engineering*, vol. 50, no. 11, pp. 2947–2961, 2024.
- [22] H.-Y. Huang, R. Kueng, and J. Preskill, “Predicting many properties of a quantum system from very few measurements,” *Nature Physics*, vol. 16, no. 10, pp. 1050–1057, 2020.
- [23] A. Bensoussan, E. CHACHKAROVA, K. Even-Mendoza, S. Fortz, V. Klimis, and M. R. Mousavi, “Artifact of the paper: Noise fingerprints for cross-platform quantum simulator discrepancy analysis,” Feb. 2026. [Online]. Available: <https://doi.org/10.5281/zenodo.18551509>
- [24] K. Temme, S. Bravyi, and J. M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Phys. Rev. Lett.*, vol. 119, p. 180509, Nov 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.119.180509>
- [25] S. Endo, S. C. Benjamin, and Y. Li, “Practical quantum error mitigation for near-future applications,” *Phys. Rev. X*, vol. 8, p. 031027, Jul 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.8.031027>
- [26] S. Aaronson, “Shadow tomography of quantum states,” in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2018. New York, NY, USA: ACM, 2018, p. 325–338. [Online]. Available: <https://doi.org/10.1145/3188745.3188802>
- [27] S. Aaronson and G. N. Rothblum, “Gentle measurement of quantum states and differential privacy,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2019. New York, NY, USA: ACM, 2019, p. 322–333. [Online]. Available: <https://doi.org/10.1145/3313276.3316378>
- [28] S. Honarvar, M. R. Mousavi, and R. Nagarajan, “Property-based testing of quantum programs in q#,” in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops (ICSEW)*. IEEE/ACM, 2020, pp. 430–435.
- [29] G. Pontolillo, M. R. Mousavi, and M. Grzesiuk, “Qcheck: A property-based testing framework for quantum programs in qiskit,” *CoRR*, vol. abs/2503.22641, 2025. [Online]. Available: <https://arxiv.org/abs/2503.22641>
- [30] X. Wang, T. Yu, P. Arcaini, T. Yue, and S. Ali, “Mutation-based test generation for quantum programs with multi-objective search,” in *Genetic and Evolutionary Computation Conference (GECCO)*. ACM, 2022, pp. 1345–1353.
- [31] R. Abreu, J. P. Fernandes, L. Llana, and G. Tavares, “Metamorphic Testing of Oracle Quantum Programs,” in *2022 IEEE/ACM 3rd International Workshop on Quantum Software Engineering (Q-SE)*, May 2022, pp. 16–23. [Online]. Available: <https://ieeexplore.ieee.org/document/9814906>
- [32] M. Paltenghi and M. Pradel, “MorphQ: Metamorphic Testing of the Qiskit Quantum Computing Platform,” in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, May 2023, pp. 2413–2424, arXiv:2206.01111 [cs]. [Online]. Available: <http://arxiv.org/abs/2206.01111>

- [33] P. Zhao, Z. Miao, S. Lan, and J. Zhao, “Bugs4q: A benchmark of existing bugs to enable controlled testing and debugging studies for quantum programs,” *Journal of Systems and Software*, vol. 205, p. 111805, 2023.
- [34] N. Quetschlich, L. Burgholzer, and R. Wille, “MQT Bench: Benchmarking software and design automation tools for quantum computing,” *Quantum*, 2023, MQT Bench is available at <https://www.cda.cit.tum.de/mqtbench/>.
- [35] V. Klimis, A. Bensoussan, E. Chachkarova, K. Even-Mendoza, S. Fortz, and C. Lenihan, “Shaking up quantum simulators with fuzzing and rigour,” *Proc. ACM Program. Lang.*, vol. 9, no. OOPSLA, Oct. 2025. [Online]. Available: <https://doi.org/10.1145/3763100>
- [36] G. J. Pontolillo, A. Muqet, S. Ali, and M. R. Mousavi, “From ideal to noisy: Adapting property-based testing for real-world noisy quantum computers,” in *Proceedings of the IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2025, pp. 405–416. [Online]. Available: https://web-backend.simula.no/sites/default/files/2025-08/Noise_aware_property_based_testing-8.pdf
- [37] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, “Quantum fingerprinting,” *Phys. Rev. Lett.*, vol. 87, p. 167902, Sep 2001. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.87.167902>
- [38] I. L. Chuang and M. A. Nielsen, “Prescription for experimental determination of the dynamics of a quantum black box,” *Journal of Modern Optics*, vol. 44, no. 11-12, pp. 2455–2467, 1997.
- [39] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography,” *Nature Communications*, vol. 8, no. 1, p. 14485, Feb 2017. [Online]. Available: <https://doi.org/10.1038/ncomms14485>
- [40] A. Miranskyy, L. Zhang, and J. Doliskani, “On testing and debugging quantum software,” 2021. [Online]. Available: <https://arxiv.org/abs/2103.09172>
- [41] S. Honarvar, M. R. Mousavi, and R. Nagarajan, “Property-based testing of quantum programs in q#,” in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, ser. ICSEW’20. New York, NY, USA: ACM, 2020, p. 430–435. [Online]. Available: <https://doi.org/10.1145/3387940.3391459>
- [42] E. Magesan, J. M. Gambetta, and J. Emerson, “Scalable and robust randomized benchmarking of quantum processes,” *Physical Review Letters*, vol. 106, no. 18, p. 180504, 2011. [Online]. Available: <https://doi.org/10.1103/PhysRevLett.106.180504>
- [43] IBM Quantum, “Exact and noisy simulation with Qiskit Aer primitives,” <https://quantum.cloud.ibm.com/docs/guides/simulate-with-qiskit-aer>, accessed: 2026-02-08.
- [44] —, “Building noise models,” <https://quantum.cloud.ibm.com/docs/guides/build-noise-models>, accessed: 2026-02-08.
- [45] Quantinuum, “Quantinuum system model H2 product data sheet,” https://docs.quantinuum.com/systems/data_sheets/Quantinuum%20H2%20Product%20Data%20Sheet.pdf, 2025, accessed: 2026-02-08.
- [46] —, “FAQs,” <https://docs.quantinuum.com/systems/support/faqs.html>, accessed: 2026-02-08.
- [47] —, “Performance validation,” https://docs.quantinuum.com/systems/user_guide/hardware_user_guide/performance_validation.html, accessed: 2026-02-08.
- [48] Qiskit, “depolarizing_error — Qiskit Aer documentation,” https://qiskit.github.io/qiskit-aer/stubs/qiskit_aer.noise.depolarizing_error.html, accessed: 2026-02-08.
- [49] —, “amplitude_damping_error — Qiskit Aer documentation,” https://qiskit.github.io/qiskit-aer/stubs/qiskit_aer.noise.amplitude_damping_error.html, accessed: 2026-02-08.
- [50] —, “phase_damping_error — Qiskit Aer documentation,” https://qiskit.github.io/qiskit-aer/stubs/qiskit_aer.noise.phase_damping_error.html, accessed: 2026-02-08.
- [51] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, “Validating quantum computers using randomized model circuits,” *Phys. Rev. A*, vol. 100, p. 032328, Sep 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.100.032328>
- [52] D. Duplyakin, R. Ricci, A. Maricq, G. Wong, J. Duerig, E. Eide, L. Stoller, M. Hibler, D. Johnson, K. Webb, A. Akella, K. Wang, G. Ricart, L. Landweber, C. Elliott, M. Zink, E. Cecchet, S. Kar, and P. Mishra, “The design and operation of cloudlab,” in *Proceedings of the 2019 USENIX Conference on Usenix Annual Technical Conference*, ser. USENIX ATC ’19. USA: USENIX Association, 2019, p. 1–14.
- [53] CloudLab, “Hardware (The CloudLab Manual),” <https://docs.cloudlab.us/hardware.html>, accessed: 2026-02-09.
- [54] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.