Zero-Noise Extrapolation Integrated with Randomized Benchmarking for Superconducting Qubits

Abhi Muvva[1]

[1] *Department of Physics, University of Wisconsin–Madison*

Reliable characterization of single-qubit gate performance is essential for the development of scalable superconducting quantum processors. Standard randomized benchmarking (RB) provides a SPAM-robust estimate of the average error per Clifford (EPC) [1], but its compression of all underlying error sources into a single decay parameter limits its ability to distinguish between different physical error mechanisms. In this work, we integrate Zero-Noise Extrapolation (ZNE) into the RB framework [2, 3] to obtain both the intrinsic EPC and the noise-sensitivity behavior of specific gate implementations. We apply this RB+ZNE protocol to two physically distinct single-qubit gate families on a superconducting qubit platform: flux-driven Z rotations and microwave-driven X/Y rotations. By designing gate-family-appropriate noise-scaling mechanisms, we construct a unified experimental methodology capable of probing how each gate responds to controlled noise amplification. The resulting framework provides a more detailed understanding of gate robustness than standard RB alone and establishes a scalable approach for differentiating performance across multiple quantum gate implementations.

I. INTRODUCTION

The ability to execute high-fidelity quantum gates is a foundational requirement for scalable quantum computation. As devices operate in the Noisy Intermediate-Scale Quantum (NISQ) regime [4], small errors in single-qubit operations accumulate rapidly, degrading algorithmic performance long before hardware limits are reached. Reliable characterization tools are therefore essential for identifying dominant error mechanisms and guiding calibration routines. Randomized Benchmarking (RB) has become one of the most widely adopted diagnostic techniques because it yields a SPAM-robust estimate of the average error per Clifford (EPC) [1]. Its exponential decay model provides a simple, experimentally efficient way to quantify gate performance on real hardware, and is widely used in both pedagogical and applied contexts [5].

However, the very feature that makes RB attractive, the collapse of all error processes into a single decay parameter, also limits its diagnostic power. A single EPC value does not reveal how sensitive a gate is to increased circuit duration, nor does it differentiate between coherent miscalibration, decoherence-driven stochastic noise, or gate-specific control imperfections. Moreover, RB cannot predict how a given implementation behaves when noise levels fluctuate, nor can it compare the robustness of physically distinct gate types such as flux-driven Z rotations and microwave-driven X/Y rotations. As a result, standard RB is insufficient when the goal is to understand why two gate implementations differ rather than simply reporting how large their average errors are.

Zero-Noise Extrapolation (ZNE) provides a complementary tool to overcome these limitations. Instead of passively observing noise at its native level, ZNE deliberately amplifies noise in a controlled and systematic manner—through pulse stretching, gate folding, or circuit-level scaling—and then extrapolates measured performance back toward an idealized zero-noise limit. This methodology was originally proposed for mitigating noise in short-depth circuits [2] and has since been extended through digital noise-scaling techniques [3]. When integrated into RB, this approach enriches the characterization by yielding not only the intrinsic EPC but also a noise-sensitivity coefficient that quantifies how error grows as noise is increased. This provides a new axis of comparison between different gate implementations, revealing information that standard RB necessarily obscures.

In this work, we apply the RB+ZNE framework to two distinct classes of single-qubit gates used on a superconducting qubit platform. Flux-driven Z rotations, implemented through fast flux pulses that shift the qubit frequency, couple strongly to dephasing and flux noise. Microwave-driven X/Y rotations, constructed using DRAG-shaped pulses on the transverse control line, have a different noise profile dominated by amplitude and phase calibration accuracy. Because these two gate families interact with hardware and noise channels in fundamentally different ways, they require distinct ZNE scaling mechanisms. Stretching flux pulses modifies their temporal and spectral characteristics, whereas gate folding preserves the shape of microwave pulses while increasing effective circuit depth. Designing appropriate scaling for each gate family is essential to obtain meaningful extrapolation behavior.

By implementing RB+ZNE with gate-specific scaling strategies, this experiment establishes a unified methodology for comparing intrinsic fidelity and noise susceptibility across physically different gate implementations. This provides a more complete and actionable picture of gate performance than standard RB alone, and offers an experimentally efficient pathway toward understanding and improving single-qubit operations on superconducting quantum hardware.

Beyond providing deeper insight into gate physics, RB+ZNE also supports several practical applications. By separating intrinsic control error from noise-induced

growth, the method serves as a diagnostic tool for detecting calibration drift, flux sensitivity, and gate-specific vulnerabilities. The extracted noise-sensitivity coefficients enable meaningful comparisons between different gate implementations, informing pulse design and transpiler choices. Moreover, identifying operations with high noise sensitivity can guide scheduling strategies that minimize temporal exposure to decoherence. Finally, the intrinsic EPC obtained through RB+ZNE provides a natural baseline for downstream error-mitigation pipelines such as probabilistic error cancellation or Clifford-data regression.

II. BACKGROUND THEORY

Randomized Benchmarking (RB) quantifies average gate fidelity by applying a sequence of m randomly sampled Clifford operations followed by a recovery gate designed to return the system to its initial state [1, 5]. In the presence of noise, the survival probability decays exponentially:

$$P(m) = A\,p^m + B, \tag{1}$$

where A and B absorb SPAM contributions and p is the depolarizing parameter. The corresponding error per Clifford (EPC) is

$$\mathrm{EPC} = \frac{1-p}{2}. \tag{2}$$

RB compresses a complex, gate- and time-dependent noise landscape into a single scalar p [1]. While this is powerful, it also masks how sensitive the gate implementation is to increased control duration or noise exposure. Zero-Noise Extrapolation (ZNE) resolves this by *artificially increasing the noise* in a controlled manner and extrapolating back to the zero-noise limit [2, 3].

A. Zero-Noise Extrapolation

ZNE estimates the noise-free expectation value $O(0)$ of an observable by evaluating the same circuit under different noise-scale factors λ [2, 3]. For smooth noise responses, a series expansion yields

$$O(\lambda) = O(0) + a(\lambda-1) + b(\lambda-1)^2 + \cdots. \tag{3}$$

By measuring $O(\lambda)$ at $\lambda = 1, 3, 5, \ldots$, one can fit Eq. (3) and extrapolate to $\lambda = 0$.

1. *Why Noise Scaling Works*

Noise in contemporary devices is often too weak to characterize directly from a single measurement. ZNE therefore follows a three-step philosophy [3]:

1. Increase the noise in a controlled, physically meaningful way.
2. Measure how the circuit output shifts under this amplified noise.
3. Extrapolate backwards to infer the zero-noise value.

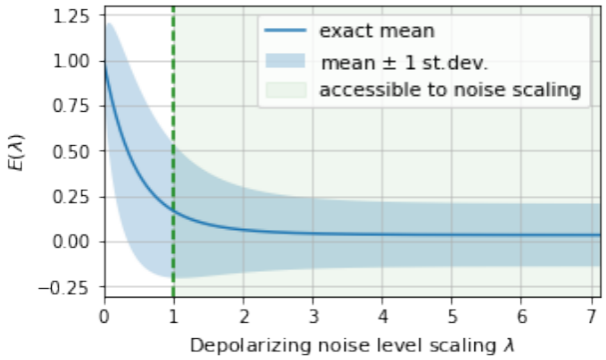

FIG. 1. Conceptual illustration of Zero-Noise Extrapolation (ZNE), adapted from Giurgica-Tiron *et al.* [3], Fig. 1. Higher noise-scale factors λ produce increasingly deviated expectation values $E(\lambda)$. Fitting across multiple λ values enables estimation of the zero-noise value at $\lambda = 0$.

Figure 1 illustrates how higher-λ curves diverge more strongly, making the extrapolated zero-noise estimate more reliable.

B. Gate Folding

Gate folding increases the noise exposure of a circuit without altering its ideal unitary action [3]. A gate G is replaced by

$$G \longrightarrow G\,(G^\dagger G)^k, \tag{4}$$

forcing the hardware to execute additional pulses while the ideal evolution is unchanged.

a. Why $\lambda = 2k+1$. The folded sequence contains:

$$\underbrace{1}_{G} + \underbrace{2k}_{G^\dagger G \text{ pairs}},$$

giving a noise-scale factor

$$\lambda = 2k + 1. \tag{5}$$

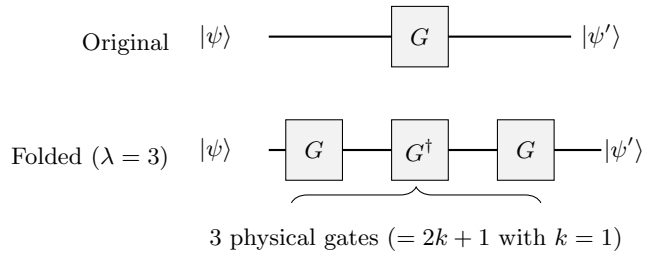

FIG. 2. Schematic illustration of digital gate folding. The original circuit (top) applies G once. In the folded circuit (bottom), G is replaced by $G(G^\dagger G)$, increasing the number of physical gates to $\lambda = 2k + 1 = 3$.

1. *How Folding Modifies RB Decay*

Under a depolarizing approximation,

$$\Lambda(\rho) = p\rho + (1-p)\frac{I}{2},$$

λ-folding yields

$$\Lambda_\lambda = \Lambda^\lambda,$$

with modified decay parameter

$$p(\lambda) = p^\lambda. \tag{6}$$

Thus the RB curve becomes

$$P(m, \lambda) = A\, p^{\lambda m} + B, \tag{7}$$

and the EPC at scale λ is

$$\mathrm{EPC}(\lambda) = \frac{1 - p^\lambda}{2}. \tag{8}$$

This steepening of the RB curve under increased λ improves sensitivity to noise strength and enables a more accurate extrapolation to the zero-noise limit [2].

C. Microwave X/Y vs Flux-Driven Z

Microwave-driven X/Y rotations admit clean inverses and well-behaved pulse envelopes, making them compatible with digital folding. Flux-driven Z rotations, however, rely on nonlinear frequency detuning trajectories whose inverse is not given by negating the waveform. Hysteresis, distortions, and flux noise break the $G^\dagger G$ symmetry, so digital folding injects uncontrolled distortions [3].

Therefore ZNE for flux gates must rely on alternatives such as waveform stretching, trajectory-duration scaling, or amplitude scaling.

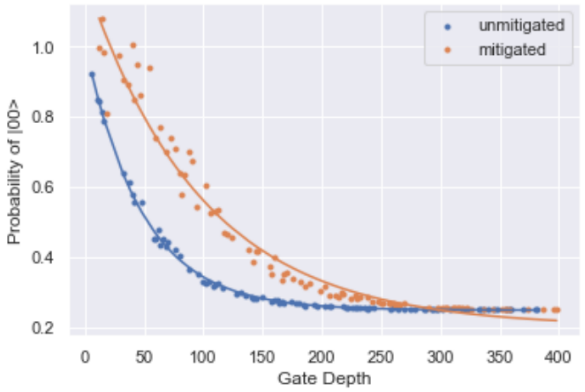

FIG. 3. Example of how digital noise scaling modifies an RB decay, adapted from Giurgica-Tiron *et al.* [3], Fig. 2. Although the original data correspond to a two-qubit RB experiment, the same functional dependence $P(m, \lambda) = Ap^{\lambda m} + B$ arises in single-qubit RB under gate folding.

III. EXPERIMENTAL PROCEDURE

All experiments were performed on a fixed-frequency superconducting qubit using the Qolab control platform. Pulse generation, sequencing, and measurement acquisition were handled through the Qualibrate environment. Each experimental trial consisted of initialization into the $|0\rangle$ state, application of a randomized sequence of Clifford operations, and single-shot readout in the computational basis.

To enable noise-scaling for zero-noise extrapolation (ZNE), the default Qualibrate single-qubit randomized benchmarking (RB) script was modified to (i) save the complete Clifford sequences used in each run, (ii) execute each sequence under multiple noise-scaling factors λ, and (iii) record raw measurement outcomes for all depths m and all noise scales without performing any on-board fitting or normalization.

The modified script produces two data sets per experiment: (1) a $(K \times m_{\max})$ array of Clifford labels describing each random sequence up to its maximum depth $m_{\max}$, and (2) a $(K \times D)$ array of binary measurement outcomes for the truncated depths $m \in \{m_1, \ldots, m_D\}$ for each noise scale λ. Here K is the number of random sequences, and the prefix lengths $\{m_i\}$ are obtained by truncating each sequence so that a single random sequence contributes measurement data to multiple depths.

A. Random Sequence Generation and Gate Decomposition

Randomized benchmarking requires sampling uniformly from the single-qubit Clifford group (24 elements). For each experiment, K random sequences of maximum length $m_{\max}$ were generated. Each Clifford operation was

decomposed into the hardware-native pulse primitives:

- microwave-driven X and Y rotations, and
- flux-driven Z rotations.

This decomposition is necessary because noise amplification is applied at the level of these primitive operations via gate folding.

Each random sequence $g = (g_1, \ldots, g_{m_{\max}})$ is evaluated at multiple depths by truncation:

$$g_{1:m} = (g_1, g_2, \ldots, g_m),$$

allowing the RB decay curve to be estimated while reusing randomness across depths, reducing statistical variance.

B. Noise Scaling via Gate Folding

Zero-noise extrapolation requires scaling the physical noise experienced by a gate while preserving its ideal unitary action. This is achieved by replacing each primitive gate with a longer, algebraically equivalent folded implementation.

Flux-pulse Z gates. A rotation $Z(\theta)$ is replaced with

$$Z_\lambda(\theta) = Z(\theta)\,[Z(-\theta)Z(\theta)]^{(\lambda-1)/2}, \tag{9}$$

which increases the flux-pulse duration and therefore the exposure to flux noise, while cancelling coherent overrotations.

Microwave X/Y rotations. For a primitive rotation $G \in \{X, Y\}$,

$$G_\lambda = G\left(G^\dagger G\right)^{(\lambda-1)/2}, \tag{10}$$

where $G^\dagger G = I$ ideally. The additional pulse blocks extend the gate duration, scaling amplitude/phase noise without altering the logical operation.

Noise scales used in this experiment were $\lambda \in \{1, 3, 5\}$, corresponding to no-folding, one-fold, and two-fold extensions.

C. Measurement and Survival Probability Estimation

For each noise scale λ and each depth m, the experiment outputs a single-shot binary measurement outcome for each of the K random sequences. The empirical survival probability at depth m is computed as

$$\overline{s}(m, \lambda) = \frac{1}{K}\sum_{k=1}^{K} s_k(m, \lambda), \tag{11}$$

where $s_k(m, \lambda) = 1$ if the k-th sequence returns outcome "0" and 0 otherwise.

Algorithm 1 Unified RB+ZNE Procedure

Require: Noise scales $\Lambda = \{1, 3, 5\}$; RB lengths $\{m_\ell\}$;
number of sequences N_{seq}.
for $\lambda \in \Lambda$ **do**
  for $m \in \{m_\ell\}$ **do**
    for $s = 1$ to N_{seq} **do**
      Sample random Clifford sequence $(C_1, \ldots, C_m)$
      Compute recovery Clifford C_{rec}
      Decompose all Cliffords into primitive gates
      for each gate G **do**
        if G is Flux-Z **then**
          Replace with $Z_\lambda(\theta)$
        else if G is X or Y **then**
          Replace with G_λ
        end if
      end for
      Execute folded sequence and record survival outcome
    end for
    Fit $P(m; \lambda) = A\,p(\lambda)^m + B$
    Extract EPC(λ) from $p(\lambda)$
  end for
end for
Fit EPC(λ) vs. λ to obtain intrinsic EPC and noise sensitivity

To remove SPAM offsets and isolate the RB decay, the curve is normalized relative to the smallest non-zero depth m_0:

$$G(m, \lambda) = \frac{\overline{s}(m, \lambda)}{\overline{s}(m_0, \lambda)}. \tag{12}$$

D. Extraction of EPC(λ)

For each noise scale λ, the normalized decay is fit to the standard exponential RB model,

$$G(m, \lambda) = A_\lambda\, p_\lambda^{m-1}, \tag{13}$$

using a least-squares fit in either the linear or logarithmic domain. The extracted decay parameter p_λ gives the average error per Clifford at scale λ:

$$\text{EPC}(\lambda) = \frac{1 - p_\lambda}{2}. \tag{14}$$

This procedure yields EPC(1), EPC(3), and EPC(5), which are subsequently used for zero-noise extrapolation in Sec. IV.

E. Unified RB+ZNE Workflow

For clarity, the full experimental workflow is summarized in Algorithm 1, which combines random sequence generation, gate folding, and EPC extraction into a single RB+ZNE protocol.

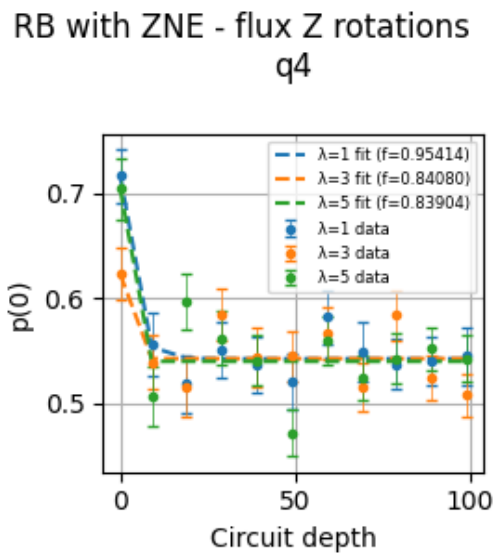

FIG. 4. Randomized benchmarking decay curves for flux-driven Z rotations on qubit q4 for noise scales $\lambda \in \{1, 3, 5\}$. Points show normalized survival probabilities $G(m, \lambda)$; dashed lines denote fitted exponential decays.

IV. RESULTS

A. Flux-Z Gates

Randomized benchmarking with zero-noise extrapolation (RB+ZNE) was performed on a fixed-frequency superconducting qubit (q4) using noise scales $\lambda \in \{1, 3, 5\}$ applied exclusively to flux-driven Z rotations. A total of $N_{\text{seq}} = 100$ random Clifford sequences were generated, each evaluated at RB lengths

$$m \in \{1, 10, 20, \ldots, 100\},$$

with 10 averages per sequence. As described in Sec. III, this procedure yields survival probabilities $\overline{s}(m, \lambda)$ and normalized decays $G(m, \lambda)$ for each noise scale.

Figure 4 shows the measured decay curves. All three exhibit an initial drop in survival probability followed by a shallow plateau for $m \gtrsim 20$, characteristic of coherence-limited RB behavior. The ordering of the curves reflects the intended noise amplification: larger λ produces a visibly stronger decay.

For each λ, the decay parameter p_λ was obtained using the fitting procedure described in Sec. III, and converted to error-per-Clifford (EPC) and error-per-gate (EPG) values. These results are listed in Table I. As expected, introducing folded flux pulses increases the effective noise rate, with a clear separation between the unfolded case ($\lambda = 1$) and the noise-scaled cases ($\lambda = 3, 5$).

To estimate the intrinsic control error of the flux-driven Z gate, a linear zero-noise extrapolation was performed on these three values:

$$\text{EPG}(\lambda) \approx a + b\lambda.$$

TABLE I. Extracted EPC and EPG for flux-driven Z rotations on q4 at noise scales $\lambda \in \{1, 3, 5\}$. The gate fidelity is defined as $f = 1 - \text{EPG}$.

λ	EPC	EPG	Fidelity $1 - \text{EPG}$
1	0.1318	0.04586	0.9541
3	0.4577	0.15920	0.8408
5	0.4628	0.16096	0.8390

The extrapolated zero-noise intercept yields

$$\text{EPG}_0 \approx 0.036, \qquad f_0 = 1 - \text{EPG}_0 \approx 0.964.$$

This intrinsic fidelity represents the control-limited performance of the flux-Z implementation, separating it from the decoherence accumulated during the longer folded gate sequences.

[1] E. Magesan, J. M. Gambetta, and J. Emerson, Physical Review A **85**, 042311 (2012).
[2] K. Temme, S. Bravyi, and J. M. Gambetta, Physical Review Letters **119**, 180509 (2017).
[3] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, arXiv preprint arXiv:2005.10921 (2020).
[4] J. Preskill, Quantum **2**, 79 (2018).
[5] A. Silva and E. Greplova, Hands-on introduction to randomized benchmarking (2022), lecture notes.