Learning to Correct Errors in Quantum Circuits via Transformer-Predicted PQCs

Ashutosh Tiwari
Stony Brook University, NY

Zian Wang
Stony Brook University, NY

Himanshu Gupta
Stony Brook University, NY

***Abstract*—Practical quantum computing on Noisy Intermediate-Scale Quantum (NISQ) devices is fundamentally bottlenecked by hardware imperfections: errors accumulate quickly and can destroy the interference patterns that quantum algorithms rely on. While full quantum error correction promises fault tolerance, its overhead is prohibitive for near-term processors, and many quantum error mitigation techniques trade this limitation for substantial sampling cost, limited applicability, or per-circuit retraining.**

We propose an *active*, learning-based circuit-level correction framework that suppresses errors *during* execution. Our approach interleaves lightweight single-qubit parameterized quantum circuit (PQC) blocks into a target circuit and predicts their corrective rotation angles directly from the circuit's gate sequence. We formalize this setting as a *general parameter prediction* problem: learn a model that maps circuits to continuous corrective parameters to minimize the expected discrepancy between the ideal and corrected output distributions over the circuit domain. We implement this with a Transformer-Encoder architecture operating on a sliding-window circuit representation, enabling prediction of corrections in a single forward pass.

Across diverse random benchmark circuits, the learned predictor achieves *zero-shot* correction on unseen instances, eliminating expensive per-circuit tuning at inference time. Empirically, our interleaved corrections substantially improve output distribution fidelity, maintaining state fidelities above 0.99 in regimes where uncorrected executions average between 0.3 and 0.5.

I. Introduction

Practical, large-scale quantum computing is fundamentally bottlenecked by hardware noise. While Noisy Intermediate-Scale Quantum (NISQ) devices have achieved remarkable milestones [28], they inherently accumulate errors during execution [26]. Because quantum algorithms rely on precise interference, unchecked noise rapidly degrades output fidelity, rendering deep computations useless. Consequently, developing efficient and scalable error-suppression strategies remains a critical prerequisite for unlocking quantum advantage.

Prior Work. To address hardware errors, we primarily rely on Quantum Error Correction (QEC) and Quantum Error Mitigation (QEM). QEC enables long-term fault tolerance by redundantly encoding logical information and actively correcting errors via continuous syndrome measurements [33, 40]. However, the substantial spatial and temporal overheads of this encoding render full QEC infeasible for current NISQ processors [14], necessitating advanced techniques for ancilla resource optimization [41]. To bridge this gap, QEM strategies offer near-term alternatives that bypass logical qubits in favor of classical post-processing or additional circuit sampling [6]. Unfortunately, existing QEM techniques introduce critical bottlenecks, including exponential sampling overheads, instability in deep circuits, limited applicability to specific physical invariants, and, for learning-based models, a strict reliance on classically-simulatable training circuits for each target circuit [11, 20, 34, 39].

This work was partly supported by the National Science Foundation award 2504621.

Our Approach. To overcome the limitations of massive QEC overheads and passive QEM post-processing, we present an active, machine-learning-driven framework that neutralizes errors directly within the circuit execution. We accomplish this by interleaving lightweight Parameterized Quantum Circuits (PQCs) [3, 9] into the target sequence. By framing error correction and mitigation as a continuous parameter prediction problem, we utilize a Transformer architecture to read the sequence of gates and directly output the optimal corrective angles for these PQC blocks.

Contributions. This work shifts the paradigm from passive post-processing to active, structure-aware circuit correction. Our specific contributions are fourfold:

1) *General Parameter Prediction Problem.* We formulate *learned circuit-level correction/mitigation* as a regression problem: Given a circuit $\mathcal{C}$, predict continuous corrective parameters (e.g., rotation angles).
2) *Interleaved PQC Correction Strategy.* We introduce a method to actively neutralize errors during execution by interleaving lightweight, single-qubit Parameterized Quantum Circuit (PQC) blocks directly into the target sequence.
3) *Transformer Training Methodology.* We design a training pipeline using a Transformer-Encoder architecture to predict the optimal PQC angles in a single forward pass. This enables zero-shot error correction for entirely unseen circuits, completely eliminating the need for expensive per-circuit tuning at inference time.
4) *Evaluation.* We validate our approach across a range of random circuits, demonstrating that our active correction significantly improves the fidelity of the output distribution. Specifically, our model ensures that the fidelity of the output state is largely above 0.99, whereas non-corrected states average between 0.3 and 0.5.

Paper Organization. The remainder of the paper is organized as follows. Section II reviews basic notions of quantum circuits, parameterized quantum circuits and quantum noise. Section III formalizes the *general parameter prediction* problem. Section IV discusses related work on quantum error correction as well as quantum error mitigation. Section V introduces the sliding-window noise model and the layerized circuit representation. Section VI details the single-qubit corrective block architecture and its insertion pattern. Section VII presents the sliding-window training algorithm and the Transformer-based predictor. Finally, Section VIII reports empirical results, and Section IX concludes with a discussion of limitations and future directions.

II. Background

In this section, we review notation for quantum states and circuits, describe parameterized quantum circuits, and summarize the distinction between incoherent and coherent noise mechanisms.

Quantum States and Measurements. A single-qubit state is represented as a unit vector

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \quad \alpha,\beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1, \tag{1}$$

within the Hilbert space $\mathbb{C}^2$. A multi-qubit register comprising n qubits resides in the tensor-product space $(\mathbb{C}^2)^{\otimes n}$ [26]. Given two pure n-qubit states $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$, we define their state fidelity as

$$F(|\psi\rangle, |\phi\rangle) := |\langle\psi|\phi\rangle|^2 \in [0,1]. \tag{2}$$

Since all states in our simulations are pure, this notion of fidelity suffices for our purposes. Later, in our Transformer evaluations (Section VIII), we will report the mean global fidelity between the ideal output state and the noisy or corrected output states.

While the state vector $|\psi\rangle$ fully describes the system, information about the final quantum state is extracted through measurement. Measurement in the computational basis yields a discrete probability distribution over bitstrings $x \in \{0,1\}^n$. The Born rule gives the probability of observing outcome x:

$$P(x) = \big|\,\langle x|\psi\rangle\,\big|^2. \tag{3}$$

Circuit Representations. An n-qubit quantum circuit is described as an ordered sequence of gates $\mathcal{C} = (g_1, g_2, \ldots, g_N)$. Each gate $g_k = (U_k, \mathbf{q}_k)$ is specified by a unitary U_k acting on a subset of qubits $\mathbf{q}_k \subseteq \{1,\ldots,n\}$. The logical unitary implemented by the ideal circuit is $U_{\text{ideal}}(\mathcal{C}) = U_N \cdots U_2 U_1$. Common elementary gates include single-qubit Pauli rotations (e.g., X, Y, Z or their corresponding rotations) and two-qubit entangling gates such as the CONTROL-NOT (CNOT) gate.

Parameterized Quantum Circuits (PQC). Parameterized quantum circuits [43] (PQCs) are *tunable* quantum circuits whose action depends on a set of parameters $\boldsymbol{\theta} \in \mathbb{R}^m$, where m is the number of parameters of the PQC. By adjusting these parameters, a PQC can approximate a family of unitaries or probability distributions [3, 9, 32]. Figure 1 shows an example of a PQC.

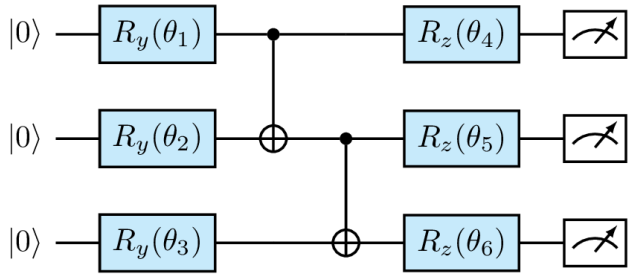


Fig. 1: **An Arbitrary Parameterized Quantum Circuit (PQC).** The circuit consists of layers of parameterized single-qubit rotations (highlighted in blue) interleaved with fixed entangling operations (CNOTs). The vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_6)$ represents the tunable parameters optimized during the learning process to approximate a target distribution or unitary.

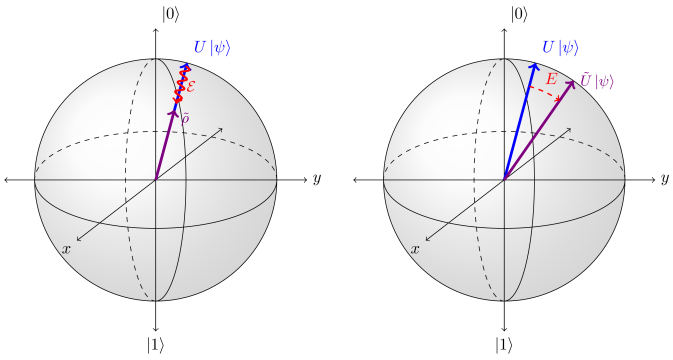


Fig. 2: **Incoherent and coherent errors on the Bloch sphere.** The ideal pure state $U|\psi\rangle$ (blue) lies exactly on the surface. (a) An incoherent channel $\mathcal{E}$ shrinks the state vector into the interior, yielding a mixed state $\tilde{\rho}$ (violet) with reduced purity. (b) A coherent error E systematically rotates the state to $\tilde{U}|\psi\rangle$ (violet), preserving its purity.

Coherent vs. Incoherent Noise. Ideal unitaries U_{ideal} are difficult to realize on physical hardware due to environmental coupling and control imperfections. It is often useful to separate noise into incoherent, non-unitary effects and coherent, unitary miscalibrations, since they induce qualitatively different transformations on the Bloch sphere (Figure 2). *Incoherent errors* include relaxation T_1 and dephasing T_2 and are modeled by completely positive trace-preserving channels $\mathcal{E}$. These represent the physical process of *decoherence*, where the system's interaction with its surroundings causes an irreversible loss of quantum information [12, 16, 35]. On a single qubit, these channels contract the Bloch vector from the surface into the interior, so a pure input state is driven toward a mixed output state $\tilde{\rho}$ [26]. In contrast, *coherent errors* arise from systematic unitary miscalibrations such as consistent over- or under-rotations, axis misalignment, or residual couplings [17]. Mathematically, an intended unitary U is implemented as a

perturbed unitary

$$\tilde{U} = E\,U, \tag{4}$$

where $E \in \mathrm{U}(2^k)$ is a small unitary deviation that is approximately fixed over a calibration interval. Because E is unitary, coherent errors preserve state purity: the output $\tilde{U}\,|\psi\rangle$ remains pure and, in the single-qubit picture, stays on the Bloch sphere surface but is rotated to an unintended orientation. In the absence of environmental coupling, such coherent miscalibrations can interfere constructively or destructively with circuit depth, producing accumulation and worst-case behavior that can differ markedly from predictions based on stochastic Pauli-noise models [17, 24].

III. Problem Formulation

In this section, we define the parameter prediction problem. We specify the ideal circuit domain, motivate the focus on coherent errors, define the theoretical per-circuit corrective objective, and state the main learning objective.

Circuit Ensemble and Ideal Outputs. Let $\mathbf{C}$ denote a domain of quantum circuits. Each instance $\mathcal{C} \in \mathbf{C}$ is an n-qubit gate sequence of depth T. In the ideal, noise-free setting, $\mathcal{C}$ implements a unitary $U_{\mathrm{ideal}}(\mathcal{C})$, producing the state

$$|\psi_{\mathrm{ideal}}(\mathcal{C})\rangle = U_{\mathrm{ideal}}(\mathcal{C})\,|\psi_{\mathrm{in}}\rangle \tag{5}$$

for an arbitrary input $|\psi_{\mathrm{in}}\rangle$. Measurement in the computational basis yields a probability distribution $P_{\mathrm{ideal}}(\mathcal{C}; x)$ over bitstrings $x \in \{0,1\}^n$, representing the ideal probability of observing outcome x.

Per-Circuit Correction. We denote the noisy evolution by $U_{\mathrm{noisy}}(\mathcal{C})$. To counteract this noise, we augment $\mathcal{C}$ with a fixed[1] parameterized quantum circuit (PQC) architecture $V(\boldsymbol{\theta})$. Executing this augmented circuit yields a corrected output distribution $P_{\mathrm{corr}}(\mathcal{C}, \boldsymbol{\theta}; x)$.

For a fixed circuit $\mathcal{C}$, the ideal corrective parameters $\boldsymbol{\theta}^\star(\mathcal{C})$ are those that completely minimize the discrepancy between the ideal and corrected distributions:

$$\boldsymbol{\theta}^\star(\mathcal{C}) = \arg\min_{\boldsymbol{\theta}} D\Big(P_{\mathrm{ideal}}(\mathcal{C}), P_{\mathrm{corr}}(\mathcal{C}, \boldsymbol{\theta})\Big), \tag{6}$$

where $D(\cdot,\cdot)$ is a chosen probability distance metric. Solving Eq. (6) separately for every $\mathcal{C}$ gives good corrections for those specific circuits, but does not by itself provide a way to handle new circuits without repeating the computationally expensive optimization.

Predictive Parameter Model. To achieve scalable correction, we introduce a parameter prediction model F_ϕ with weights ϕ that maps a quantum circuit directly to the required corrective parameters:

$$\hat{\boldsymbol{\theta}}(\mathcal{C}) = F_\phi(\mathcal{C}). \tag{7}$$

The goal of F_ϕ is to approximate the per-circuit optimum $\boldsymbol{\theta}^\star(\mathcal{C})$ in a single forward pass, bypassing the need for repeated device calibration or exhaustive simulation at inference time.

[1]We target coherent (unitary) errors; thus we fix the PQC structure and optimize only its parameters.

General Parameter Prediction Problem. We now state the main problem addressed in this work.

Generalized-Correction **Problem.** Find model weights $\phi^\star$ such that the predicted parameters minimize the expected distance between the ideal and corrected output distributions across the entire circuit domain $\mathbf{C}$:

$$\phi^\star = \arg\min_{\phi}\ \mathbb{E}_{\mathcal{C}\sim\mathbf{C}}\left[D\Big(P_{\mathrm{ideal}}(\mathcal{C}), P_{\mathrm{corr}}\big(\mathcal{C}, F_\phi(\mathcal{C})\big)\Big)\right] \tag{8}$$

In Section VII, we detail the methodology used to approximate this expectation, including the generation of instance-specific surrogate labels that replace $\boldsymbol{\theta}^\star(\mathcal{C})$ during model training.

IV. Related Work

There are two primary paradigms for addressing noise in quantum hardware: *Quantum Error Correction* (QEC), which seeks to eliminate errors during execution via encoding, and *Quantum Error Mitigation* (QEM), which manages errors through statistical or structural adjustments. We discuss these below to contextualize our proposed methodology.

Quantum Error Correction (QEC). Quantum error correction (QEC) enables fault-tolerant computation by encoding logical qubits within a larger subspace of physical qubits [33, 40]. By performing periodic *syndrome measurements*, QEC allows for the detection and correction of errors without collapsing the underlying quantum state. Foundational architectures, such as the surface code [14] and color codes [4], demonstrate that logical error rates can be suppressed exponentially, provided physical noise remains below a strict fault-tolerance threshold [1]. However, implementing QEC remains prohibitively resource-intensive for current-generation hardware. The requirement for massive spatial overhead, low-latency classical feedback loops, and continuous syndrome extraction creates a "resource gap" that Noisy Intermediate-Scale Quantum (NISQ) devices cannot yet bridge [28]. This motivates the search for active error-suppression strategies that provide error reduction without the overhead of formal logical encoding.

Quantum Error Mitigation (QEM). Unlike QEC, QEM strategies suppress hardware noise by combining additional circuit executions with classical post-processing [6]. These methods generally fall into four categories:

- *Noise Scaling and Extrapolation:* Zero-noise extrapolation (ZNE) [23, 39] executes circuits at amplified noise levels—typically via pulse stretching [21] or gate folding [19]—to extrapolate back to the zero-noise limit. While hardware-efficient, ZNE can become unstable for deep circuits where the noise model deviates from the extrapolation fit.
- *Probabilistic and Inversion Methods:* Probabilistic error cancellation (PEC) [11, 39] represents ideal unitaries as a quasi-probability mixture of noisy operations. Similarly, measurement error mitigation (MEM) [15] inverts a calibrated readout confusion matrix. While theoretically

rigorous, these methods suffer from an exponential sampling overhead as circuit depth increases. Furthermore, they are *model-dependent*, requiring a precise a priori characterization of the processor's noise profile (e.g., via gate-set tomography).

- *Symmetry and Projection Constraints:* These techniques leverage physical invariants—such as symmetry verification [5], virtual distillation [20, 22], or N-representability [25, 30]—to project noisy states onto a valid subspace. However, their utility is *narrowly restricted* to specific physical problems where symmetries are known a priori. Furthermore, these methods often incur high resource costs by requiring multi-copy collective measurements or the discarding of a significant portion of samples via post-selection. Most critically, they remain blind to *symmetry-preserving errors*, which can degrade the computation without triggering the projection or verification criteria.
- *Learning-Based Mitigation:* Recent research has explored training classical models to learn a correction function $f(E(P); \vec{\theta})$ that maps noisy expectation values to noise-free targets [10, 34]. These methods typically rely on *circuit-specific* training sets—ensembles of classically simulable circuits that structurally resemble the target circuit P to ensure the error model generalizes. While effective for fixed workloads, this approach assumes the existence of sufficiently large, simulable training sets for every target application. In contrast, our methodology utilizes *generic* training sequences (e.g., identity circuits) whose ideal outputs are known a priori, regardless of the target circuit's complexity. By employing a transformer architecture to capture sequence-dependent gate errors and context-heavy noise, we move beyond passive post-processing. Instead, our model learns the underlying processor noise profile from these generic sequences and actively interleaves parameterized corrective gates, providing a general-purpose mitigation layer that does not require classically simulable analogues of the target circuit.

V. Noise Model

In this section, we formalize the structure of our noise model targeted by our correction framework. We begin by justifying our focus on *coherent errors*, highlighting their characteristic scaling and fundamental invertibility compared to stochastic noise. Next, we introduce a *sliding-window formulation* that captures context-dependent coherent errors. Finally, we adopt this formulation for multi-qubit circuits using a *local decomposition strategy* in which single-qubit unitaries model the dominant error on the participating qubits.

Coherent Noise. We assume hardware noise is primarily coherent rather than stochastic [8, 18]. Unlike memoryless Pauli channels, coherent errors act as deterministic, context-dependent unitaries [37] that interfere constructively. This results in a worst-case error accumulation of $O(T)$ over circuit depth T—a quadratic increase over the $O(\sqrt{T})$ scaling typical of stochastic noise [17, 42].

We focus on correcting coherent errors because they are inherently unitary operations that a properly tuned PQC can invert. We do not address incoherent noise (e.g., relaxation or dephasing), as it represents an irreversible leakage of quantum information to the environment that unitary gates alone cannot recover.

Single-Qubit Context-Aware Errors. Consider a single-qubit circuit comprising a sequence of gates $(g_1, g_2, \ldots, g_N)$. Rather than assuming memoryless operations, we model the time-correlated mechanisms inherent to real hardware—such as control imperfections and environmental couplings—which manifest as context-dependent deviations from the ideal gate sequence [2, 18, 29, 31].

We model this behavior by inserting an explicitly deterministic, coherent error unitary $E_t \in \mathrm{SU}(2)$ into the sequence at time step t. Unlike fully general non-Markovian models that depend on the entire execution history, we assume the error is dominated by recent interactions, effectively imposing a finite quantum Markov order [38]. We define E_t based on a *sliding window* of a maximum depth L. For operations where the context window is fully populated ($t \geq L$), the error is determined by a context-aware function f. During the initial operations ($t < L$), the error is governed by a partial-context function p acting on the available gate history:

$$E_t = \begin{cases} f(g_{t-L+1}, \ldots, g_t) & \text{if } t \geq L, \\ p(g_1, \ldots, g_t) & \text{if } t < L. \end{cases} \tag{9}$$

This ensures that coherent deviations are modeled continuously from the very first gate, accurately reflecting the buildup of transient dynamics before the full memory depth L is reached. For example, with a maximum depth of $L = 3$, the fully contextual error E_5 depends strictly on (g_3, g_4, g_5), while E_2 is evaluated via $p(g_1, g_2)$. This windowed formulation explicitly captures several physically distinct, coherent noise regimes:

1) *Global Miscalibration ($L = 1$):* Systematic calibration offsets dependent only on the current gate type (e.g., static over-rotations or axis misalignments) [17, 27].
2) *Pattern Dependence ($L > 1$):* Errors conditioned on specific subsequences (e.g., a distortion triggered specifically by a $\mathrm{Z}-\mathrm{X}-\mathrm{Z}$ pattern), accounting for pulse distortions or local spectral crowding [31].
3) *Temporal Drift:* Parameter evolution of f and p over time t, capturing heating or control drift that systematically degrades fidelity with circuit depth [2, 29].

The exact functional forms of f and p depend heavily on the specific architecture and control electronics of the target quantum processor. Because our correction framework is inherently general and hardware-agnostic, we defer the formal definition of a specific, hardware-motivated error generator to Section VIII, where it is utilized to benchmark our approach.

Generalization to Multi-Qubit Circuits via Local Decomposition. General multi-qubit noise often introduces unwanted

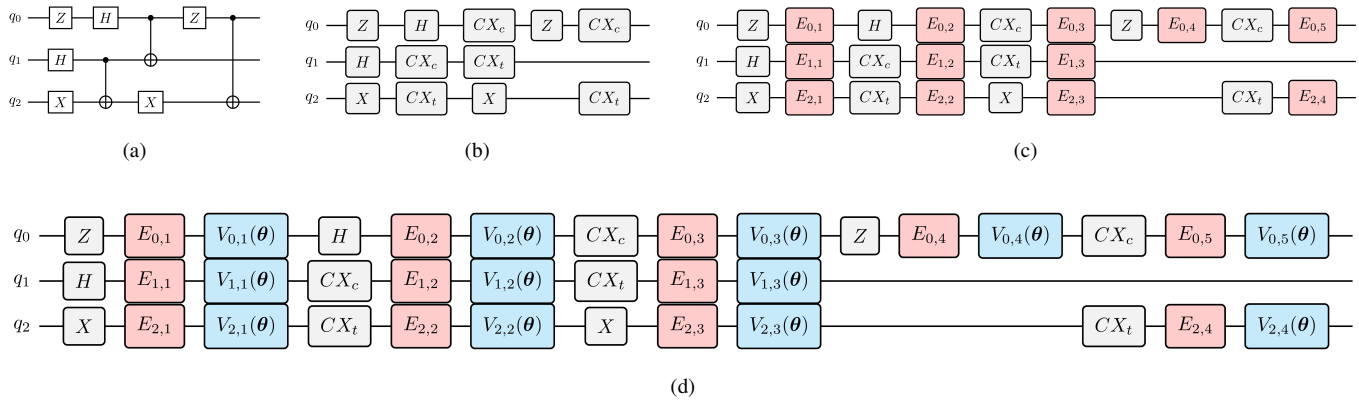


Fig. 3: **Workflow for circuit tokenization, error insertion, and interleaved correction. (a)** An arbitrary 3-qubit, 9 gate circuit. **(b)** Tokenization of the circuit into parallel per-qubit sequences. Entangling gates are converted into their role-specific tokens. **(c)** Error insertion: a single-qubit error unitary $E_{i,t}$ (red) is applied after each gate t on qubit i. **(d)** Interleaved correction: a tunable R_z–R_x–R_z block $V(\boldsymbol{\theta}_{i,t})$ (blue) is placed after each $E_{i,t}$ to provide local, layer-by-layer correction.

entanglement (e.g., residual ZZ couplings) that ordinarily requires complex multi-qubit corrections. However, learning full $\mathrm{SU}(4)$ corrections for every entangling gate is computationally prohibitive. To ensure scalability, we rely on a *local decomposition strategy*. If an entangling gate is of reasonably high fidelity, its dominant residual errors can be accurately modeled as local coherent distortions. Mathematically, we approximate the noisy two-qubit gate as:

$$\tilde{U}_{\mathrm{CNOT}} \approx (E_{\mathrm{c}} \otimes E_{\mathrm{t}})\, U_{\mathrm{CNOT}}^{\mathrm{ideal}} \tag{10}$$

where E_{c} and E_{t} are single-qubit error unitaries applied to the control and target qubits, respectively. This assumption is firmly supported by recent hardware studies on both superconducting and silicon processors [7, 8, 36, 37] which show that during entangling operations such as cross-resonance, the most prominent residual errors manifest as local distortions, including axis skews and spectator phase shifts.

To capture these effects, we decouple entangling operations into role-specific single-qubit representations (e.g., CX_c for the control and CX_t for the target). We then apply the sliding-window error model from Eq. (9) to each one-dimensional sequence $\mathbf{s}_q$ individually, yielding per-qubit errors $E_{q,t}$. Since single-qubit operations can correct these dominant errors, our framework adopts the same strategy: we apply per-qubit corrections layer-by-layer, reducing the impact of primary noise sources without the heavy overhead of modeling multi-qubit crosstalk. The resulting per-qubit sequences and the placement of error unitaries are illustrated in Figures 3(b) and 3(c).

VI. PQC Architecture

We now describe the parameterized quantum circuit (PQC) used to counteract the coherent, sequence-dependent noise introduced in Section V. Our design follows a *predict-and-insert* approach: we augment the base circuit with lightweight *single-qubit* corrective unitaries placed at a fine (layer-wise) granularity. The goal is to provide enough expressivity to represent general local coherent corrections while keeping the inserted structure fixed and hardware-friendly.

Single-Qubit Corrective Block. For each qubit q and layer t, we insert a corrective unitary $V_{q,t}$ parameterized by $\boldsymbol{\theta} = (\alpha_{q,t}, \beta_{q,t}, \gamma_{q,t}) \in \mathbb{R}^3$. We use the standard $Z-X-Z$ Euler form

$$V(\boldsymbol{\theta}) = \mathrm{R}_z(\gamma)\, \mathrm{R}_x(\beta)\, \mathrm{R}_z(\alpha), \tag{11}$$

which can represent any element of $\mathrm{SU}(2)$ up to a global phase [26]. Accordingly, for any local coherent error $E_{q,t} \in \mathrm{SU}(2)$, there exists a choice $\boldsymbol{\theta}^\star$ such that $V(\boldsymbol{\theta}^\star) = E_{q,t}^{-1}$.

Composition with the Error Model. At layer t, qubit q ideally undergoes $U_{q,t}$ but is followed by a coherent deviation $E_{q,t}$. After inserting $V_{q,t}(\boldsymbol{\theta})$ immediately after the deviation, the effective operation becomes

$$U_{q,t}^{\mathrm{eff}} = V_{q,t}(\boldsymbol{\theta})\, E_{q,t}\, U_{q,t}. \tag{12}$$

Our objective is to learn a predictor that outputs parameters satisfying $V_{q,t}(\boldsymbol{\theta}) \approx E_{q,t}^{-1}$ across layers, so that $U_{q,t}^{\mathrm{eff}} \approx U_{q,t}$ and the corrected output distribution tracks the ideal one.

Layer-wise Interleaving. We interleave the PQC blocks directly into the gate sequence. Specifically, for each qubit q and each layer t, we insert $V_{q,t}(\boldsymbol{\theta})$ after the layer token $g_{q,t}$:

$$\widetilde{\mathbf{s}}_q = (g_{q,1}, V_{q,1}, g_{q,2}, V_{q,2}, \ldots, g_{q,T}, V_{q,T}). \tag{13}$$

When the context window is not yet fully populated ($t < L$), the corresponding parameters are predicted from the available prefix context (as defined in Section V); for $t \geq L$, they are predicted from the full length-L window.

VII. Training Algorithm

This section details the training algorithm for our predictive model, which learns to map circuit structure directly to corrective parameters. We assume the sliding-window noise model defined in Section V and the interleaved single-qubit PQC architecture from Section VI are fixed.

Overview. Our training pipeline consists of two primary stages: (i) *Dataset construction:* We sample a collection of fixed-length window circuits whose ideal evolution can be

efficiently simulated, and then assign an instance-specific correction label to each window by estimating the corrective parameters (as detailed below). (ii) *Supervised learning:* We train a Transformer model to predict these continuous correction labels directly from the discrete circuit tokens.

At inference time, the trained model outputs precise parameter corrections for unseen circuits in a single forward pass, completely bypassing the need for per-circuit optimization. Figure 4 illustrates the overall pipeline above.

Stage 1: Dataset construction. We begin by assembling a dataset $\mathbf{C}_{\text{dataset}} \subset \mathbf{C}$ consisting of length-L window circuits. Because our framework decomposes multi-qubit circuits into localized, per-qubit sequences (Section V), computing the ideal ground truth for any window is inherently efficient. For every $\mathcal{C} \in \mathbf{C}_{\text{dataset}}$, we calculate its ideal target unitary $U_{\text{ideal}}(\mathcal{C})$ simply by multiplying the sequence of L ideal single-qubit gate matrices within that window. When a window contains a binary-gate marker (e.g., CX_c or CX_t in the per-qubit unary view), we treat it as the identity gate in this multiplication, since it functions only as a role/error context marker rather than a single-qubit operation on the tracked qubit. This convention matches our $SU(2)$ error model and window-level labeling abstraction, and later in Stage 1.2, we can see that any treatment would be equivalent as long as it is applied consistently when constructing the corresponding training labels. We then partition $\mathbf{C}_{\text{dataset}}$ into strictly disjoint training, validation, and test subsets ($\mathbf{C}_{\text{train}}$, $\mathbf{C}_{\text{val}}$, and $\mathbf{C}_{\text{test}}$). To prevent the model from seeing the test data during training, we populate these splits using independent randomized streams, guaranteeing no accidental overlap.

Stage 1.1: Estimate per-window coherent error. For each circuit $\mathcal{C} \in \mathbf{C}_{\text{dataset}}$, we treat its noisy implementation as a black box and execute it using a fixed set of known input probe states: $\{|0\rangle, |1\rangle, |+\rangle\}$. We collect *probe responses*, defined as the measurement outcome statistics in the Pauli (X, Y, Z) basis. Under the local-decomposition and coherent-error assumptions of Section V, we use these responses to construct an effective single-qubit unitary $U_{\text{noisy}}(\mathcal{C}) \in \text{SU}(2)$.

This reconstruction follows a deterministic, step-by-step procedure. First, for each probe input $k \in \{0, 1, +\}$, we convert the measured Pauli expectations into an observable output Bloch vector:

$$\widehat{\mathbf{r}}_{\text{out},k} \triangleq \big(\langle X\rangle_k, \langle Y\rangle_k, \langle Z\rangle_k\big) \tag{14}$$

where each expectation value is estimated directly from the empirical shot probabilities.

Next, under the assumption that dominant errors are coherent, we first map each measured Bloch vector $\widehat{\mathbf{r}}_{\text{out},k}$ to a corresponding normalized pure state $y_k \in \mathbb{C}^2$. However, because quantum states are only unique up to an arbitrary global phase, we cannot simply combine the computational basis states (y_0 and y_1) as they are. They must share a consistent relative phase to form a valid transformation matrix.

To establish this relative phase, we use the superposition state y_+ as a reference bridge. By calculating the inner products between this reference and our two basis states, we compute their phase difference δ:

$$\alpha = \langle y_0, y_+\rangle, \qquad \beta = \langle y_1, y_+\rangle, \qquad \delta = \arg\left(\frac{\beta}{\alpha}\right) \tag{15}$$

Here, $\arg(\cdot)$ denotes the complex argument (i.e., the phase angle) of its input; thus δ is the relative phase between β and α.

We apply this calculated phase shift directly to the second basis state. With both states now strictly phase-aligned, we assemble the raw unitary matrix by stacking them as columns:

$$\widetilde{U}_{\text{noisy}}(\mathcal{C}) = \begin{bmatrix} y_0 & e^{-i\delta} y_1 \end{bmatrix} \tag{16}$$

Finally, because statistical sampling noise can cause this raw matrix to deviate slightly from a perfect quantum operation, we project $\widetilde{U}_{\text{noisy}}(\mathcal{C})$ to the nearest valid unitary matrix using Singular Value Decomposition (SVD). This yields our final estimated coherent error $U_{\text{noisy}}(\mathcal{C}) \in \text{SU}(2)$. This ground-truth label is entirely data-driven, relying strictly on observable measurement statistics rather than any hidden simulator parameters.

Stage 1.2: Obtaining correction label from estimated error. We define the estimated error block using the reconstructed noisy unitary as:

$$\widehat{N}(\mathcal{C}) \triangleq U_{\text{noisy}}(\mathcal{C})\, U_{\text{ideal}}(\mathcal{C})^{\dagger} \tag{17}$$

Because this error block represents an arbitrary single-qubit rotation in $\text{SU}(2)$, we decompose it into standard Euler angles. We parameterize it using the $Z-X-Z$ form to exactly match the physical architecture of our parameterized quantum circuit (PQC) block, $V(\boldsymbol{\theta})$, defined in Section VI:

$$\widehat{N}(\mathcal{C}) = \text{R}_z(\gamma)\,\text{R}_x(\beta)\,\text{R}_z(\alpha), \quad \boldsymbol{\theta}_{\text{noise}}(\mathcal{C}) \triangleq (\alpha, \beta, \gamma) \tag{18}$$

To correct this noise, we must apply its inverse operation, $\widehat{N}(\mathcal{C})^{-1}$. We define our optimal target parameters, $\boldsymbol{\theta}^{\star}(\mathcal{C})$, by negating the extracted error angles. We then pass them through a bounding function, $\text{canon}(\cdot)$, which maps the angles back to a standard periodic domain (e.g., $[-\pi, \pi]$):

$$\boldsymbol{\theta}^{\star}(\mathcal{C}) = \text{canon}\big(-\boldsymbol{\theta}_{\text{noise}}(\mathcal{C})\big) \tag{19}$$

Appending the PQC block $V\big(\boldsymbol{\theta}^{\star}(\mathcal{C})\big)$ will now actively neutralize the coherent error for this specific circuit window.

We apply this process to all sampled circuits and obtain our final dataset of paired circuits and their target corrections:

$$\mathcal{D}_{\text{dataset}} = \Big\{ \big(\mathcal{C}_i, \boldsymbol{\theta}^{\star}(\mathcal{C}_i)\big) \mid \mathcal{C}_i \in \mathbf{C}_{\text{dataset}} \Big\} \tag{20}$$

We create our training, validation, and test datasets ($\mathcal{D}_{\text{train}}$, $\mathcal{D}_{\text{val}}$, $\mathcal{D}_{\text{test}}$) by applying Eq. (20) to the strictly disjoint circuit subsets established at the beginning of Stage 1.

Stage 2: Learning the circuit-to-parameter map. We train a Transformer F_ϕ as a predictor that maps the tokenized circuit representation to corrective parameters:

$$\widehat{\boldsymbol{\theta}}(\mathcal{C}) = F_\phi(\mathcal{C}). \tag{21}$$

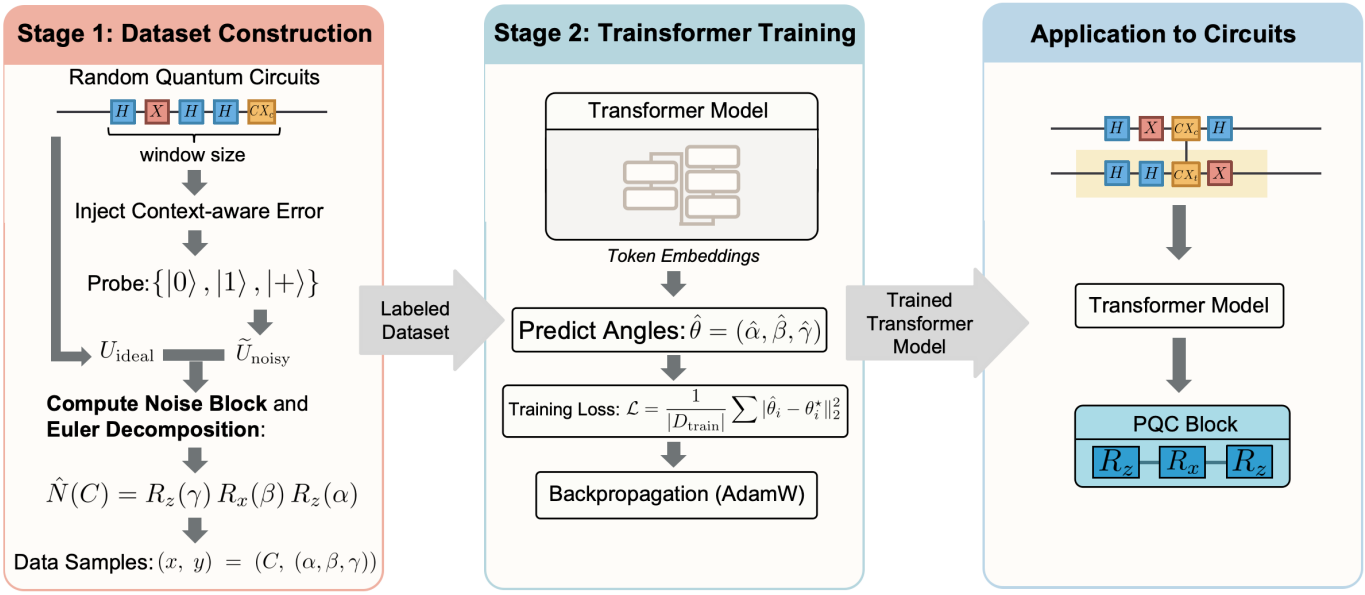


Fig. 4: **Overview of our data generation, training, and deployment pipeline.** (**Stage 1**) We sample length-L window circuits, inject context-aware coherent errors, and compare $U_{\text{noisy}}(C)$ with $U_{\text{ideal}}(C)$ to extract Euler angles (α, β, γ) as training labels. (**Stage 2**) A Transformer takes the circuit token sequence as input and regresses the corrective angles $\hat{\theta}$ using an MSE objective (optimized with AdamW). (**Application**) For unseen circuits, we apply the model in a sliding-window manner and insert the predicted R_z–R_x–R_z PQC block to mitigate sequence-dependent errors.

The model is optimized on $\mathcal{D}_{\text{train}}$ with a supervised regression objective:

$$\mathcal{L}(\phi) = \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(\mathcal{C}_i, \boldsymbol{\theta}^\star(\mathcal{C}_i)) \in \mathcal{D}_{\text{train}}} \left\| F_\phi(\mathcal{C}_i) - \boldsymbol{\theta}^\star(\mathcal{C}_i) \right\|_2^2. \quad (22)$$

We use $\mathcal{D}_{\text{val}}$ for model selection and report final performance on the held-out $\mathcal{D}_{\text{test}}$.

Application to new circuits. At inference time, we apply F_ϕ to a new circuit $\mathcal{C}_{\text{new}}$ by (i) converting it to the layerized per-qubit token representation (Section V), (ii) feeding this representation to F_ϕ to obtain predicted parameters $\widehat{\boldsymbol{\theta}}(\mathcal{C}_{\text{new}})$, and (iii) inserting the corresponding PQC blocks at the designated locations as described in Section VI.

VIII. Evaluation

This section evaluates the performance and generalization of our proposed framework. We begin by defining the context-aware error functions used to characterize the noise model (Section V). We then specify the experimental setup, including the gate alphabet and the parameters used to generate the training and validation circuit datasets. After detailing the Transformer-Encoder configuration, we assess the model's performance through two primary lenses: (i) an analysis of training and validation loss, and (ii) an evaluation of zero-shot inference on unseen, multi-qubit circuits of varying depths. Our results indicate that the model consistently restores circuit fidelity to approximately 0.99, maintaining a worst-case fidelity of 0.96 across diverse benchmark instances.

Context-Aware Error Function. To generate the context-dependent coherent error parameters associated with a gate sequence as described in Equation 9, we use the following *hidden-states model*. Rather than treating errors as memoryless stochastic channels, we physically motivate our error generation by capturing sequence-dependent coherent hardware drifts. Motivated by observations that coherent errors and slowly varying drifts can be strongly context dependent in certain operating regimes [18], we consider a drift model capturing two representative mechanisms: drive-induced accumulation effects (e.g., heating) [37] and low-frequency parameter detuning (e.g., $1/f$-like fluctuations) [2].

Inspired by autoregressive approaches to tracking quantum trajectories [13], we model the unobservable environment using two hidden states that update gate-by-gate: an amplitude drift tracker a_n and a phase detuning tracker Δ_n. In our synthetic generator, each gate symbol $g \in \mathcal{G}$ has fixed metadata given by lookup tables: $\vartheta(g)$ (rotation-angle scale), $\text{power}(g)$ (drive-strength proxy), and $\text{duration}(g)$ (relative gate time). These are treated as deterministic functions of g throughout. For a sequence of operations $g_1, g_2, \ldots, g_L$, the states update as:

$$a_n = \lambda_A a_{n-1} + \eta_A \,\text{power}(g_n), \quad (23)$$

$$\Delta_n = \lambda_D \Delta_{n-1} + \eta_D \,\text{power}(g_n)\,\text{duration}(g_n). \quad (24)$$

In this work, the drift evolution and the resulting error parameters are deterministic given the gate sequence and the fixed metadata tables, serving as a controlled context-dependent drift benchmark rather than a fully stochastic noise model. Here, λ_A and λ_D act as leaky integrators governing the physical relaxation of the environment, ensuring a bounded, finite fading memory of past operations [38]. The terms η_A and

Gate g	$\vartheta(g)$	$\mathrm{power}(g)$	$\mathrm{duration}(g)$
H	$\pi/2$	0.8	1.2
X	π	1.0	1.0
CXc	0	1.6	2.5
CXt	π	1.3	2.5

TABLE I: **Gate metadata for the hidden-states model.** Normalized units representing the physical profile of each gate. These values serve as inputs to Equations 23 and 24.

η_D scale the system's sensitivity to the applied microwave power and total pulse-energy proxy. At the final gate in the window (g_L), the accumulated history maps to an amplitude error factor ϵ_L and a phase shift ϕ_L:

$$\epsilon_L = K_A a_L, \quad \phi_L = K_D \Delta_L \,\mathrm{duration}(g_L). \tag{25}$$

Finally, we inject a deterministic coherent unitary error parameterized by the Euler angles (α, β, γ):

$$(\alpha, \beta, \gamma) \equiv (\alpha_L, \beta_L, \gamma_L) = \left(\tfrac{1}{2}\phi_L,\ \epsilon_L\, \vartheta(g_L),\ \tfrac{1}{2}\phi_L\right). \tag{26}$$

We use an R_Z–R_X–R_Z parameterization because it provides a standard, compact representation of generic single-qubit coherent miscalibrations. The symmetric choice $\alpha = \gamma = \phi_L/2$ simply splits the accumulated Z-phase ϕ_L evenly before and after the X-overrotation component, matching a convenient canonical form used by our generator.

Unless otherwise stated, we use fixed hyperparameters $\lambda_A = 0.98$, $\lambda_D = 0.99$, $\eta_A = 0.03$, $\eta_D = 0.02$, and $K_A = K_D = 1.0$ across all experiments, and the per-gate metadata in Table I.

This structure guarantees that thermal accumulation (a_L) systematically scales the target rotation $\vartheta(g_L)$ representing a coherent amplitude calibration error, while energy accumulation (Δ_L) induces a systematic Z-axis twist representing continuous phase integration. This construction yields a history-dependent error model that is non-Markovian with respect to a gate-only error description over the sequence window L.

For this hidden states form, we set the partial-context function to identity (i.e., $p(\cdots) = I$) for $t < L$, and we do not apply noise when a qubit is idle in the circuit.

Evaluation Setup. We evaluate our model using window sizes $L \in \{1, 5\} \cup \{10, 20, \ldots, 100\}$, and restrict the gate alphabet of the circuit to $\{H, X, \mathrm{CX}_c, \mathrm{CX}_t\}$. To generate the ground-truth labels $\boldsymbol{\theta}^\star$ for supervised learning, we follow the procedure in Section VII in which we execute the noisy sequence using three probe states $\{|0\rangle, |1\rangle, |+\rangle\}$ on each circuit window.

Transformer Configuration and Training. We use a Transformer encoder to predict the three corrective angles from tokenized single-qubit windows. Each window is embedded with a 128-dimensional token embedding and a learned positional embedding. The encoder has 4 layers, 4 self-attention heads per layer, a 256-dimensional feed-forward sublayer, and dropout of 0.1. Masked mean pooling over valid time steps yields a fixed-size representation, followed by a linear head that outputs the three angles.

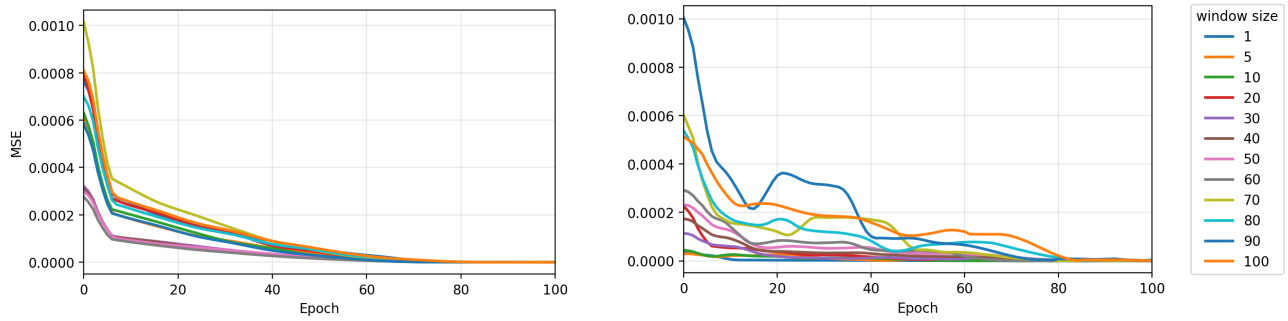
We train the model with AdamW (learning rate 3×10^{-4}, weight decay 0.01), gradient clipping at 1.0, and a cosine learning-rate schedule, using the MSE loss in Eq. 22. Unless otherwise stated, we use a batch size of $10{,}000$, train for $1{,}000$ epochs, and split the window dataset into 60% training, 20% validation, and 20% testing.

Window-level Behavior. Figure 5 shows the training and validation losses for $L \in \{1, 5\} \cup \{10, 20, \ldots, 100\}$. For all window sizes, training loss decreases over epochs, and validation loss follows closely, with no significant gap. Even though the validation loss may exhibit some fluctuations due to the limited number of validation samples, it converges to the same level as training progresses and the model converges. This indicates that the model generalizes to unseen windows rather than overfitting the training set.

Circuit-level Evaluation. We next apply the learned predictors to randomly generated multi-qubit circuits not used in training. We consider circuits of lengths $T \in \{100, 200, 300, 400, 500\}$ over 1–5 qubits drawn from the same gate alphabet. For each circuit, we: (i) convert it to the layerized per-qubit token representation; and (ii) slide a length-L window along each qubit sequence, query F_ϕ to obtain predicted angles, and insert the corresponding PQC blocks. For each randomly generated n-qubit circuit C and random input state $|\psi_{\text{in}}\rangle$, we simulate: (i) the ideal circuit to obtain $|\psi_{\text{ideal}}\rangle$, (ii) the noisy circuit under the windowed coherent error model to obtain $|\psi_{\text{noisy}}\rangle$, and (iii) the corrected circuit with our learned parameters to obtain $|\psi_{\text{corr}}\rangle$. We then compute the state fidelity as described by Equation 2. Figure 6 reports the mean fidelity, averaged over circuits and random input states. Figure 7 shows the mean global state fidelity without error correction across circuit lengths T for one to five qubits, which we use as the baseline.

Summary of Results. We evaluate the framework's performance across varying circuit depths, qubit counts, and window sizes. The results demonstrate that the Transformer-based predictor generalizes effectively to unseen circuits, consistently restoring high fidelity even under significant coherent drift. In contrast, without error correction, the mean fidelity can drop to as low as 0.5, and for four- and five-qubit circuits it can fall below 0.1. Figures 5, 6, and 7 demonstrate the effectiveness of our approach:

- *Training Stability.* Both training and validation Mean Squared Error (MSE) consistently converge below 10^{-4} within 100 epochs. This stability confirms that the model reliably learns the mapping between circuit context and corrective rotations across all window sizes.
- *Large Circuit Performance.* The model maintains high mean fidelity even as circuit depth increases. At a depth of $T = 500$, where uncorrected noise typically causes significant degradation, corrected fidelities remain above 0.96 for all qubit configurations.
- *Zero-Shot Generalization.* The predictor transfers to new circuits without per-circuit tuning or access to ideal outputs. Mean fidelity remains near 0.99 for depths up to $T = 300$, showing that the model generalizes well to unseen sequences.



(a) Training Set (b) Validation Set

Fig. 5: **Learning curves for the Transformer predictor. (a)** The training loss decreases steadily, indicating successful convergence. **(b)** Although the validation loss exhibits some fluctuations, it decreases alongside the training loss, demonstrating that the model generalizes to predict corrective parameters for unseen circuits rather than merely memorizing the training data.

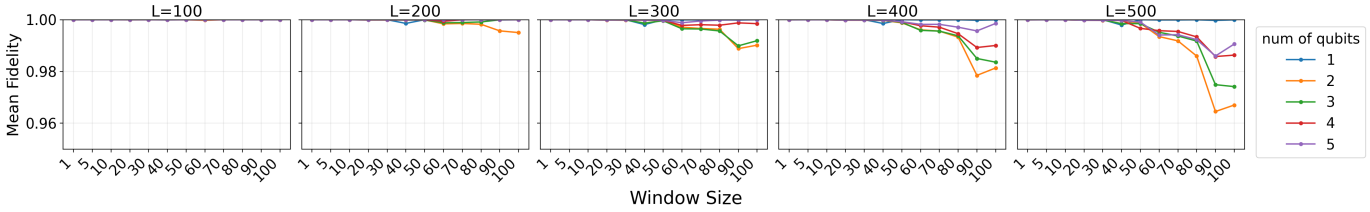


Fig. 6: **Mean global state fidelity on long multi-qubit circuits for different correction schemes.** For each circuit and random input state, we compute the state fidelity between the ideal output and the uncorrected noisy circuit, or learned windowed correctors, with $L \in \{1, 5\} \cup \{10, 20, \ldots, 100\}$. Results are averaged over randomly generated circuits of lengths 100, 200, 300, 400, and 500.

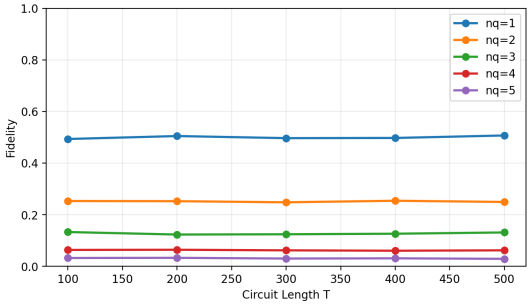


Fig. 7: Mean global state fidelity versus circuit length T for number of qubits in $\{1, \ldots, 5\}$. Because we average over many random input states, per-instance fidelities span a wide range, and the mean concentrates around a qubit-dependent baseline (a practical lower envelope for the average). By $T = 100$, the circuits are already sufficiently noisy that the mean has essentially reached this baseline, so further increasing T shows little additional change.

Varying Number of Circuit Qubits. In Fig. 6, we observe that the mean fidelity actually increases as the number of qubits increases. We believe this trend is driven by the "cascading" effect of the "residual" errors (i.e., errors remaining after corrective measures)—this cascading effect is more pronounced in circuits with a lower number of qubits (for the same number of *total circuit gates*)–due to higher gate density per qubit. Based on this trend, we focus our evaluation on circuits with up to five qubits since the model's performance will only improve with the number of qubits.

IX. Conclusion

In this work, we introduced an *in-circuit* correction framework that goes beyond post-processing by inserting learned corrective unitaries directly into the circuit. A Transformer encoder maps a circuit's gate sequence (via sliding-window tokenization) to the rotation angles of lightweight, interleaved single-qubit PQC blocks. Once trained, the model performs *zero-shot* inference on unseen multi-qubit circuits, producing corrective parameters in a single forward pass without per-circuit optimization.

On held-out benchmarks, our method consistently restores output-state fidelity to ≈ 0.99, compared to 0.3–0.5 without correction in deep-circuit settings where coherent miscalibrations accumulate. These results indicate that direct parameter prediction can provide a scalable correction layer for coherent errors.

Future work will expand this framework by developing multi-qubit correction blocks to address hardware crosstalk. We also plan to implement adaptive schemes that learn exactly where to place corrections to reduce circuit overhead, along with "confidence checks" to handle sudden changes in hardware calibration. Finally, we aim to validate the noise model and training pipeline on physical quantum processors to demonstrate its effectiveness against real-world noise.

References

[1] D. Aharonov and M. Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of*

the twenty-ninth annual ACM symposium on Theory of computing, 1997.

[2] H. Ball et al. The role of master clock stability in quantum information processing. *npj Quantum Information*, 2016.

[3] M. Benedetti et al. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, November 2019.

[4] H. Bombin and M. A. Martin-Delgado. Topological quantum distillation. *Physical Review Letters*, 2006.

[5] X. Bonet-Monroig et al. Low-cost error mitigation by symmetry verification. *Physical Review A*, 2018.

[6] Z. Cai et al. Quantum error mitigation. *Rev. Mod. Phys.*, December 2023.

[7] Z. Cai et al. Tackling coherent noise in quantum computing via cross-layer compiler optimization. *arXiv preprint arXiv:2410.09664*, 2024.

[8] G. Cenedese et al. Correcting coherent errors by random operation on actual quantum hardware. *Entropy*, 2023.

[9] M. Cerezo et al. Variational quantum algorithms. *Nature Reviews Physics*, September 2021.

[10] P. Czarnik et al. Error mitigation with clifford quantum-circuit data. *Quantum*, 2021.

[11] S. Endo et al. Practical quantum error mitigation for near-future applications. *Physical Review X*, 2018.

[12] X. Fan et al. Distribution and purification of entanglement states in quantum networks. In *2025 International Conference on Quantum Communications, Networking, and Computing (QCNC)*. IEEE, 2025.

[13] E. Flurin et al. Using a recurrent neural network to reconstruct quantum dynamics of a superconducting qubit from physical observations. *Physical Review X*, 2020.

[14] A. G. Fowler et al. Surface codes: Towards practical large-scale quantum computation. *Physical Review A—Atomic, Molecular, and Optical Physics*, 2012.

[15] M. R. Geller. Rigorous measurement error correction. *Quantum Science & Technology*, 2020.

[16] M. Ghaderibaneh et al. Efficient quantum network communication using optimized entanglement swapping trees. *IEEE Transactions on Quantum Engineering*, 2022.

[17] D. Greenbaum and Z. Dutton. Modeling coherent errors in quantum error correction. *Quantum Science and Technology*, January 2018.

[18] M. J. Gullans et al. Compressed gate characterization for quantum devices with time-correlated noise. *PRX Quantum*, 2024.

[19] A. He et al. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Physical Review A*, 2020.

[20] W. Huggins et al. Virtual distillation for quantum error mitigation. *Physical Review X*, 2021.

[21] A. Kandala et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 2019.

[22] B. Koczor. Exponential error suppression for near-term quantum devices. *Physical Review X*, 2021.

[23] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, 2017.

[24] Á. Márton and J. K. Asbóth. Coherent errors and readout errors in the surface code. *Quantum*, September 2023.

[25] D. A Mazziotti. Pure-n-representability conditions of two-fermion reduced density matrices. *Physical Review A*, 2016.

[26] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition, 2010.

[27] T. Pandit and R. Uzdin. Over-rotation coherent error induced by pseudotwirling of quantum gates. *Physical Review A*, 2025.

[28] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2018.

[29] T. Proctor et al. Detecting and tracking drift in quantum information processors. *Physical Review Letters*, 2020.

[30] N. C. Rubin et al. Application of fermionic marginal constraints to hybrid quantum algorithms. *New Journal of Physics*, 2018.

[31] K. Rudinger et al. Probing context-dependent errors in quantum processors. *Physical Review A*, 2019.

[32] M. Schuld et al. Circuit-centric quantum classifiers. *Phys. Rev. A*, March 2020.

[33] P. W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical review A*, 1995.

[34] A. Strikis et al. Learning-based quantum error mitigation. *PRX Quantum*, 2021.

[35] R. G. Sundaram et al. Distributing quantum circuits with minimum circuit execution time over quantum networks. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2024.

[36] N. Sundaresan et al. Reducing unitary and spectator errors in cross resonance with optimized rotary echoes. *PRX Quantum*, 2020.

[37] T. Tanttu et al. Assessment of the errors of high-fidelity two-qubit gates in silicon quantum dots. *Nature Physics*, 2024.

[38] P. Taranto et al. Quantum markov order. *Physical Review Letters*, 2019.

[39] K. Temme et al. Error mitigation for short-depth quantum circuits. *Physical review letters*, 2017.

[40] B. M. Terhal. Quantum error correction for quantum memories. *Reviews of Modern Physics*, 2015.

[41] A. Tiwari et al. Uncomputing ancilla qubits in quantum circuits. In *2025 International Conference on Quantum Communications, Networking, and Computing (QCNC)*. IEEE, 2025.

[42] J. Wallman et al. Estimating the coherence of noise. *New Journal of Physics*, 2015.

[43] C. Zhan and H. Gupta. Quantum sensor network algorithms for transmitter localization. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2023.