

Data-efficient quantum noise modeling via machine learning

Yanjun Ji^{1,*}, Marco Roth^{2,†}, David A. Kreplin³, Ilia Polian⁴, and Frank K. Wilhelm^{1,5}


¹*Institute for Quantum Computing Analytics (PQI-12), Forschungszentrum Jülich, 52425 Jülich, Germany*

²*Fraunhofer Institute for Manufacturing Engineering and Automation (IPA), Nobelstrasse 12, 70569 Stuttgart, Germany*

³*Heilbronn University of Applied Sciences, Max-Planck-Str. 39, 74081 Heilbronn, Germany*

⁴*Institute of Computer Architecture and Computer Engineering, University of Stuttgart, Pfaffenwaldring 47, 70569 Stuttgart, Germany*

⁵*Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany*

 (Received 16 September 2025; revised 2 January 2026; accepted 10 February 2026; published 16 March 2026)

Maximizing the computational utility of near-term quantum processors requires predictive noise models that inform robust, noise-aware compilation and error mitigation. Conventional models often fail to capture the complex error dynamics of real hardware or require prohibitive characterization overhead. We introduce a data-efficient framework that first constructs a physically motivated, parameterized noise model, and subsequently employs machine-learning-driven Bayesian optimization to identify its parameters. Our approach circumvents costly characterization protocols by estimating algorithm- and hardware-specific error parameters directly from readily available experimental data derived from existing application and benchmark circuit executions. The generality and robustness of the framework are demonstrated across diverse algorithms and superconducting devices, yielding high-fidelity predictions by estimating an independent parameter set tailored to each specific algorithm-hardware context. Crucially, we show that a model calibrated exclusively on small-scale circuits accurately predicts the behavior of larger validation circuits. Our data-efficient approach achieves up to a 65% improvement in model fidelity quantified by the Hellinger distance between predicted and experimental circuit output distributions, compared to standard noise models derived from device properties. This work establishes a practical paradigm for application-aware noise characterization, enabling compilation and error-mitigation strategies tailored to the specific interplay between quantum algorithms and device-specific noise dynamics.

DOI: [10.1103/5r9m-y6z6](https://doi.org/10.1103/5r9m-y6z6)

I. INTRODUCTION

Quantum computing offers a significant opportunity to address complex optimization problems. In the current noisy intermediate-scale quantum era [1], variational algorithms exemplified by the quantum approximate optimization algorithm (QAOA) [2] and the variational quantum eigensolver (VQE) [3] have been explored for diverse applications, including portfolio optimization [4], unit commitment in power systems [5], and quantum chemistry [6]. Superconducting circuits [7,8] have emerged as a leading platform because they benefit from established

microchip manufacturing techniques and offer fast operating times. However, the practical utility of these devices is constrained by noise mechanisms and hardware limitations that hinder the algorithm's performance. To enhance the reliability of quantum applications on such devices, noise-aware compilation has become an essential strategy. Quantum compilation translates quantum algorithms into sequences of physical operations that can be executed on a target backend. This process ensures that algorithms satisfy the hardware's intrinsic constraints and optimizes the circuit to minimize the impact of errors. Specifically, the compilation needs to adapt the algorithm to the device's limited qubit connectivity and small set of supported gates, also referred to as basis gates, by inserting SWAP gates and decomposing operations in algorithms into the device's basis gate set. This is enabled by compilers supported by, e.g., Qiskit [9], TKET [10], and Munich quantum toolkit [11,12]. Beyond ensuring basic executability, noise-aware compilation aims to maximize the fidelity of the compiled circuit. To achieve this, the compiler needs an accurate

*Contact author: y.ji@fz-juelich.de

†Contact author: marco.roth@ipa.fraunhofer.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

model of the device’s error characteristics, which often vary over time [13]. This model guides strategic decisions, including selecting a subset of high-fidelity qubits for execution, optimizing the routing of logical qubits to minimize noise accumulation [14–16], and choosing between different gate sequences that, while logically equivalent, may exhibit varying resilience to noise [17] and other imperfections like Trotter errors [18–20].

An accurate and predictive noise model is essential for achieving effective noise-aware compilation. However, constructing error models for real hardware presents a significant trade-off between model fidelity and experimental cost. Traditional error models, such as those offered by hardware vendors, typically assume independent Pauli noise channels that fail to capture the intricate dynamics inherent to real hardware. In practice, superconducting quantum devices exhibit spatially and temporally correlated errors [21–23], non-Markovian memory effects [24–26], and time-varying decoherence parameters [27]. Conversely, while comprehensive characterization techniques like quantum process tomography [28] can capture these intricate effects, they are resource intensive and impractical to scale for routine calibration of larger quantum processors. Various intermediate strategies have been explored to bridge this gap. For instance, some approaches use optimization techniques such as genetic algorithms to build models for specific error types [29], while others propose scalable Markovian models that often require custom hardware tuning to be accurate [30]. Further studies have shown that noise models should be tailored to specific applications, as no single model captures all noise dynamics [31]. Even integrated approaches that combine Markovian and non-Markovian effects to improve accuracy still tend to rely on resource-intensive characterizations [32]. These studies highlight the persistent trade-off between model fidelity and experimental overhead, emphasizing the necessity for a more efficient approach.

To address these limitations, we introduce a practical and data-efficient framework that circumvents resource-intensive characterization protocols by leveraging readily available experimental data, such as application benchmarks. Our core contribution is a circuit-size-independent parameterized noise model $\mathcal{N}(\theta)$, whose parameters are determined via iterative, ML-driven optimization, as depicted in Fig. 1. By minimizing the Hellinger distance [33] between simulated and experimental output distributions, we estimate optimal parameters θ^* directly from existing circuit executions. Crucially, this framework functions as an application-aware noise characterization. Rather than positing a single universal error profile, we estimate independent parameter sets tailored to specific algorithm-device contexts (e.g., QAOA, VQE, or random circuits on a target hardware). This approach captures the interplay between circuit structure and device-intrinsic error dynamics, yielding optimized models that

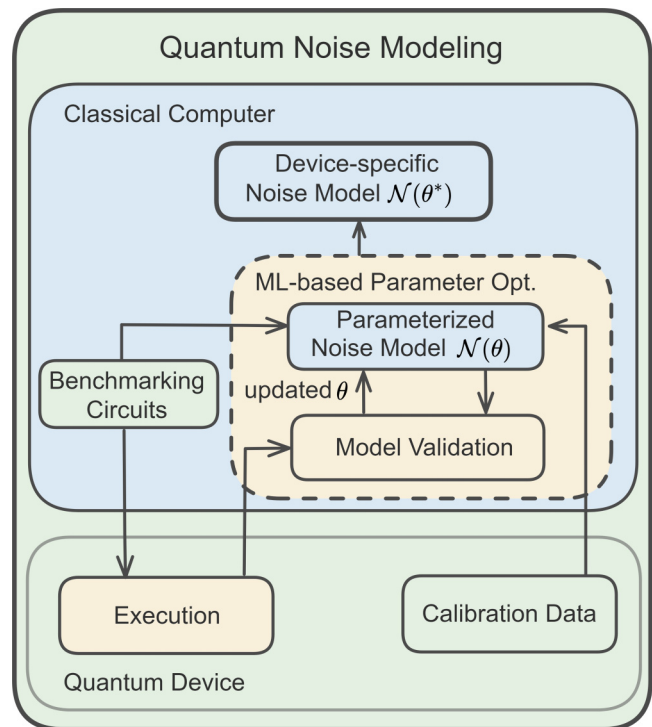


FIG. 1. Flowchart of the data-efficient quantum noise modeling framework. A parameterized noise model $\mathcal{N}(\theta)$ is initially constructed based on calibration data provided by hardware vendors. The device-specific noise model $\mathcal{N}(\theta^*)$ is then identified by executing benchmarking circuits on real quantum hardware and aligning experimental results with noisy simulations. This iterative optimization, performed on a classical computer, leverages machine-learning- (ML) based Bayesian optimization to update θ . This approach is highly flexible, designed to ingest outcome distributions from diverse pre-existing experimental data, including application-specific circuits, avoiding the need for dedicated characterization.

significantly outperform default hardware characterizations across diverse IBM quantum processors.

We demonstrate scalability by training on four–six qubit benchmarks and accurately predicting out-of-distribution behavior in seven–nine qubit circuits. This approach achieves a substantial model accuracy improvement, reducing the mean Hellinger distance to experimental results by 50% (65% peak reduction). Such predictive accuracy on larger-scale systems advances beyond foundational studies restricted to small-scale demonstrations [29], establishing that high-fidelity, device-specific noise models can be learned from compact datasets. This enables practitioners to continuously refine noise models using existing data, advancing noise-aware compilation and error mitigation [34–36] without additional experimental cost.

The paper is structured as follows. Section II presents the theoretical framework and methodology, establishing the baseline provided by device-calibrated noise models and introducing our parameterized noise model. It also

details the machine-learning- (ML) based Bayesian optimization (BO) technique employed for parameter determination and concludes with a critical analysis of the framework's scalability. Section III demonstrates a thorough experimental validation of our approach, benchmarking the model's predictive accuracy against data from multiple superconducting processors using a diverse set of applications, including the QAOA [2], VQE [3], and random circuits. This validation includes an in-depth single-device characterization, a cross-platform comparative analysis to demonstrate hardware generality, and a study of training efficiency. Finally, Sec. IV concludes the paper.

II. THEORETICAL FRAMEWORK AND METHODOLOGY

This section details our noise modeling methodology. We first introduce a baseline noise model for comparison. We then introduce our circuit-size-independent parameterized noise model, followed by parameter determination strategy. Finally, we discuss the scalability of our approach.

A. Default noise model

Most providers of superconducting quantum processing units (QPUs) offer noise models, which are often derived directly from the quantum device's calibration data. These data, provided and periodically updated by the hardware vendors, quantify the intrinsic properties of the qubits, specifically their relaxation (T_1) and dephasing (T_2) times, as well as the error rates and execution durations of the device's supported basis gates. A common example is the standard model available in Qiskit [9] for IBM devices, which can be constructed using the provided backend configurations and properties. This model isolates errors into two primary categories: gate errors and readout errors. Gate errors are modeled as a composed quantum channel $\mathcal{E}_{\text{gate}}$ applied after each ideal unitary operation. This channel is formed by the composition of a depolarizing channel $\mathcal{E}_{\text{depol}}$, followed by a thermal relaxation channel $\mathcal{E}_{\text{thermal}}$:

$$\mathcal{E}_{\text{gate}} = \mathcal{E}_{\text{thermal}} \circ \mathcal{E}_{\text{depol}}, \quad (1)$$

where the composition $\mathcal{A} \circ \mathcal{B}$ implies that channel \mathcal{B} is applied first. The thermal relaxation channel $\mathcal{E}_{\text{thermal}}$ models energy relaxation (T_1) and dephasing (T_2) processes occurring during the gate's finite duration. The depolarizing channel $\mathcal{E}_{\text{depol}}$ is not an independently measured quantity; rather, its error probability is chosen such that the infidelity of the composite channel $\mathcal{E}_{\text{gate}}$ precisely matches the experimentally characterized gate error rate e_g .

Readout errors in an n -qubit system are modeled via a $2^n \times 2^n$ stochastic matrix M , where M_{ij} denotes the conditional probability $P(\text{measure } |i\rangle \mid \text{true state } |j\rangle)$ of measuring computational basis state $|i\rangle$ given true state $|j\rangle$.

These errors, treated as a classical postprocessing step, transform the ideal probability distribution P_{ideal} into the observed distribution $P_{\text{meas}} = MP_{\text{ideal}}$, with M derived from empirical measurements.

Although this baseline model can be easily obtained, it excludes crosstalk and other correlated effects. Moreover, it captures only noise mechanisms explicitly characterized during routine device-calibration procedures, creating a static model that cannot account for the subsequent temporal evolution of the physical device. To better capture the real noise behavior of quantum algorithms executing on QPUs, we develop a tractable, data-efficient error model, which we detail below.

B. Parameterized noise model

Our model is designed to be parameterized by a vector of learnable coefficients θ , which are then chosen using ML-based optimization strategy to minimize the discrepancy between simulation and experimental data. An error channel \mathcal{E}_g for any gate operation is modeled as the composition of several distinct error channels, each targeting a specific physical error mechanism.

1. Single-qubit gate errors

For a single-qubit gate G_g (e.g., X , SX) in the bases gate set with execution duration t_g and error rate e_g , the error channel \mathcal{E}_{1Q} is modeled as a composition of four channels:

$$\mathcal{E}_{1Q} = \mathcal{E}_{\text{ph}} \circ \mathcal{E}_{\text{amp}} \circ \mathcal{E}_{\text{dep}} \circ \mathcal{E}_{\text{coh}}, \quad (2)$$

where coherent errors are applied first, followed sequentially by depolarizing, amplitude-damping, and phase-damping channels. The coherent error channel \mathcal{E}_{coh} models systematic over- or under-rotations due to control pulse distortions. This channel acts on the density matrix ρ via $\mathcal{E}_{\text{coh}}(\rho) = U_{\text{coh}}\rho U_{\text{coh}}^\dagger$, defined by the unitary operator $U_{\text{coh}} = \exp(-iH_{\text{coh}}t_g)$. The corresponding error Hamiltonian H_{coh} is a linear combination of Pauli operators with learnable amplitudes $\theta_{x,y,z}$:

$$H_{\text{coh}} = \theta_x X + \theta_y Y + \theta_z Z. \quad (3)$$

For the identity gate, this coherent error is omitted.

The depolarizing error channel \mathcal{E}_{dep} models a generic, incoherent disruption of the quantum state. For an n -qubit system, the depolarizing channel is given by

$$\mathcal{E}_{\text{dep}}^{(n)}(\rho) = (1 - \lambda_{\text{dep}}^{(n)})\rho + \lambda_{\text{dep}}^{(n)} \frac{I}{2^n}, \quad (4)$$

where $\lambda_{\text{dep}}^{(n)}$ is the depolarizing noise strength. For single-qubit, the noise strength λ_{dep} is defined as a duration-aware, saturating function

$$\lambda_{\text{dep}}(e_g, t_g) = 1 - e^{-(k_{\text{dep}}e_g + b_{\text{dep}})t_g/t_{\text{char}}}. \quad (5)$$

Here, k_{dep} and b_{dep} are learnable parameters that scale and offset the error rate, and t_{char} is a constant characteristic gate time for the device defined as the median of all single-qubit gate durations. This form captures the physical reality that error probability saturates for long-duration operations.

The amplitude damping channel \mathcal{E}_{amp} models energy relaxation. Its action on the density matrix ρ is defined by the Kraus decomposition

$$\mathcal{E}_{\text{amp}}(\rho) = \sum_{k=0}^1 K_k^{\text{amp}} \rho (K_k^{\text{amp}})^\dagger, \quad (6)$$

with Kraus operators

$$K_0^{\text{amp}} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \lambda_{\text{amp}}} \end{pmatrix}, \quad K_1^{\text{amp}} = \begin{pmatrix} 0 & \sqrt{\lambda_{\text{amp}}} \\ 0 & 0 \end{pmatrix}. \quad (7)$$

The relaxation probability λ_{amp} is governed by T_1 and t_g , with an added learnable offset b_{amp} :

$$\lambda_{\text{amp}}(t_g, T_1) = 1 - e^{-t_g/T_1} + b_{\text{amp}}. \quad (8)$$

To capture non-Markovian dephasing, potentially arising from low-frequency noise sources like two-level-system fluctuators, we model phase damping channel \mathcal{E}_{ph} using a stretched exponential decay. Similarly, the corresponding Kraus operators are

$$K_0^{\text{ph}} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \lambda_{\phi}} \end{pmatrix}, \quad K_1^{\text{ph}} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda_{\phi}} \end{pmatrix}. \quad (9)$$

The dephasing parameter λ_{ϕ} is given by

$$\lambda_{\phi}(t_g, T_{\phi}) = 1 - e^{-2(t_g/T_{\phi})^{\beta_{1Q}}}, \quad (10)$$

where T_{ϕ} is the pure dephasing time derived from the measured T_1 and T_2 , satisfying $1/T_2 = 1/(2T_1) + 1/T_{\phi}$, and β_{1Q} is a learnable exponent.

2. Two-qubit gate errors

For a two-qubit basis gate G_{2Q} acting on the qubit pair (q_0, q_1) , with gate duration t_{2Q} and error rate e_{2Q} , the corresponding error channel \mathcal{E}_{2Q} follows a similar compositional structure but includes terms for capturing crosstalk and correlated errors:

$$\mathcal{E}_{2Q} = \mathcal{E}_{zz} \circ (\mathcal{E}_{\text{ph}} \otimes \mathcal{E}_{\text{ph}}) \circ (\mathcal{E}_{\text{amp}} \otimes \mathcal{E}_{\text{amp}}) \circ \mathcal{E}_{\text{dep}}^{(2)} \circ \mathcal{E}_{\text{coh}}^{(2)}, \quad (11)$$

composing the two-qubit coherent, two-qubit depolarizing, local amplitude and phase damping, and correlated

ZZ dephasing error channels. Two-qubit coherent error $\mathcal{E}_{\text{coh}}^{(2)}$ models coherent crosstalk, such as residual ZZ coupling or control crosstalk. This error channel acts as $\mathcal{E}_{\text{coh}}^{(2)}(\rho) = U_{\text{coh}}^{(2)} \rho (U_{\text{coh}}^{(2)})^\dagger$ with $U_{\text{coh}}^{(2)} = \exp(-iH_{\text{coh}}^{(2)} t_{2Q})$. The error Hamiltonian includes two-qubit Pauli terms:

$$H_{\text{coh}}^{(2)} = \theta_{ix} I \otimes X + \theta_{zx} Z \otimes X + \theta_{zz} Z \otimes Z, \quad (12)$$

where $\theta_{ix, zx, zz}$ are learnable parameters.

Analogous to the single-qubit case in Eq. (5), the two-qubit depolarizing channel $\mathcal{E}_{\text{dep}}^{(2)}$ is characterized by a duration-aware, saturating probability $\lambda_{\text{dep}}^{(2)}$, given by

$$\lambda_{\text{dep}}^{(2)}(e_{2Q}, t_{2Q}) = 1 - e^{-(k_{\text{dep}, 2Q} e_g + b_{\text{dep}, 2Q}) t_{2Q} / t_{\text{char}, 2Q}}, \quad (13)$$

where $k_{\text{dep}, 2Q}$ and $b_{\text{dep}, 2Q}$ are learnable parameters. The characteristic timescale $t_{\text{char}, 2Q}$ is fixed as the median duration of the two-qubit gates.

The single-qubit amplitude damping and phase dephasing channels \mathcal{E}_{amp} and \mathcal{E}_{ph} are applied independently to the qubit pair (q_0, q_1) , using Kraus operators defined in Eqs. (7) and (9), i.e.,

$$\mathcal{E}_{\text{amp}} \otimes \mathcal{E}_{\text{amp}}(\rho) = \sum_{k, l=0}^1 (K_k^{\text{amp}} \otimes K_l^{\text{amp}}) \rho (K_k^{\text{amp}} \otimes K_l^{\text{amp}})^\dagger, \quad (14)$$

and similarly for $\mathcal{E}_{\text{ph}} \otimes \mathcal{E}_{\text{ph}}$. Their respective error strengths $\lambda_{\text{amp}, 2Q}$ and $\lambda_{\phi, 2Q}$ are defined as

$$\lambda_{\text{amp}, 2Q}(t_{2Q}, T_1) = 1 - e^{-t_{2Q}/T_1} + b_{\text{amp}, 2Q}, \quad (15)$$

$$\lambda_{\phi, 2Q}(t_{2Q}, T_{\phi}) = 1 - e^{-2(t_{2Q}/T_{\phi})^{\beta_{2Q}}} + b_{\phi, 2Q}. \quad (16)$$

These interactions are quantified by the learnable parameters $b_{\text{amp}, 2Q}$, β_{2Q} , and $b_{\phi, 2Q}$. Here, the offset parameter $b_{\phi, 2Q}$ is additionally introduced for two-qubit gates, compared to single qubit gates [Eq. (10)], to better characterize the more complex error mechanisms of two-qubit gates. We note that $\lambda_{\text{amp}, 2Q}$ and $\lambda_{\phi, 2Q}$ are evaluated independently for q_0 and q_1 based on their respective coherence times (T_1, T_{ϕ}) , yielding asymmetric local noise channels despite the common gate duration t_{2Q} .

The correlated ZZ dephasing channel \mathcal{E}_{zz} introduces a correlated dephasing error, a common mechanism in superconducting qubits arising from spectator qubit interactions or flux noise. It is modeled as a Pauli channel that applies a $Z \otimes Z$ operator with probability λ_{zz} :

$$\mathcal{E}_{zz}(\rho) = (1 - \lambda_{zz})\rho + \lambda_{zz}(Z \otimes Z)\rho(Z \otimes Z). \quad (17)$$

The parameter λ_{zz} is also modeled as a saturating, duration-dependent function parameterized by a learnable rate k_{zz} :

$$\lambda_{zz}(t_{2Q}) = 1 - e^{-k_{zz} t_{2Q} / t_{\text{char}, 2Q}}. \quad (18)$$

3. Readout errors

We extend the standard model of independent single-qubit readout errors to include a scalable approximation of correlated errors. In superconducting processors, readout correlations often arise from shared resonators or multiplexed electronics, potentially linking multiple qubits. However, modeling the full joint readout distribution of n qubits requires a confusion matrix scaling as $2^n \times 2^n$. This exponential scaling is incompatible with our objective of a data-efficient, circuit-size-independent framework. Instead, we model correlations at the pairwise level as the leading-order correction to the independent model. We approximate the global readout confusion matrix $\mathcal{M}_{\text{global}}$ as a tensor product of independent single-qubit matrices $M^{(1)}$ and disjoint two-qubit matrices $M^{(2)}$:

$$\mathcal{M}_{\text{global}} \approx \left(\bigotimes_{(q_i, q_j) \in \mathcal{S}} M_{ij}^{(2)} \right) \otimes \left(\bigotimes_{q_k \notin \mathcal{S}} M_k^{(1)} \right), \quad (19)$$

where \mathcal{S} is a set of disjoint qubit pairs. To construct \mathcal{S} , we employ a deterministic greedy maximal matching algorithm on the device coupling graph. This algorithm iterates through the edges of the coupling map (sorted by qubit index to ensure reproducibility) and selects a pair (q_i, q_j) if neither qubit has been previously selected. This approach maximizes the coverage of correlated errors, prioritizing physically connected neighbors where crosstalk is typically strongest, while ensuring each qubit participates in at most one correlated channel.

For a selected pair (q_0, q_1) , we begin with the 4×4 confusion matrix representing independent errors, $M_{\text{ind}} = M_{q_1} \otimes M_{q_0}$. We then construct the correlated confusion matrix $M_{ij}^{(2)}$ by injecting correlations into this product. Specifically, we model the $|00\rangle \leftrightarrow |11\rangle$ correlation, corresponding to a simultaneous correlated readout misassignment on both qubits. This is achieved by transferring a probability mass $\lambda_{\text{corr}}^{00 \leftrightarrow 11}$ from the diagonal elements to the antidiagonal elements (correlated error) within the subspaces of the affected states, such as

$$M_{0,0} \rightarrow M_{0,0} - \lambda_{\text{corr}} \quad \text{and} \quad M_{0,3} \rightarrow M_{0,3} + \lambda_{\text{corr}}. \quad (20)$$

Similarly, the $|01\rangle \leftrightarrow |10\rangle$ correlation models state-swapping errors, where a probability $\lambda_{\text{corr}}^{01 \leftrightarrow 10}$ is transferred between elements corresponding to $|01\rangle$ and $|10\rangle$ states. The magnitude of the correlation is scaled by the average local error rate of the pair, $\bar{e} = (e_{q_0} + e_{q_1})/2$. The probability of a correlated error is parameterized as a linear function:

$$\lambda_{\text{corr}} = a \bar{e} + b, \quad (21)$$

where separate parameters (a, b) are learnable variables for $|00\rangle \leftrightarrow |11\rangle$ and $|01\rangle \leftrightarrow |10\rangle$ correlation types, respectively.

4. Model parameterization and calibration integration

The complete noise model, encompassing gate and readout error channels detailed above, is fully specified by a set of 20 free parameters θ . These parameters quantify the strength of each physical error mechanism, from coherent rotations and stochastic decoherence to non-Markovian dynamics and correlated readout, forming the target vector for our data-efficient optimization. For clarity, Table I summarizes these learnable model parameters, grouped by their functional role: seven for single-qubit, nine for two-qubit, and four for readout error characterization.

Crucially, our framework employs a hybrid parameterization strategy that integrates global learnable parameters with local calibration data. Instead of learning independent error rates for every qubit, which would result in an unscalable parameter space, the 20 optimization variables θ function as global modifiers applied to the spatially varying, qubit-specific calibration inputs. The error channels leverage standard device calibration parameters, including relaxation times T_1 , dephasing times T_2 , gate durations, and error rates of basis gates, which are sourced directly from routine vendor-provided characterization data. Utilizing these existing data requires no dedicated calibration circuits or additional QPU time from the user. This architecture ensures that the model naturally captures spatial heterogeneity. For instance, in the amplitude-damping channel [Eq. (8)], the relaxation time T_1 varies locally per qubit, while the offset b_{amp} is shared globally. Consequently, for a fixed gate duration t_g , a qubit with lower T_1 is assigned a higher damping probability than one with higher T_1 despite identical b_{amp} . Learning

TABLE I. The 20 learnable parameters θ of the noise model, grouped by function role.

Parameter	Description	Type
Single-qubit gate parameters (7)		
k_{dep}	Depolarizing scaling factor	Stochastic
b_{dep}	Depolarizing offset	Stochastic
b_{amp}	Amplitude damping offset	Stochastic
$\theta_{x,y,z}$	Coherent rotation angles	Coherent
β_{1Q}	Stretched dephasing exponent	Non-Markovian
Two-qubit gate parameters (9)		
$k_{\text{dep},2Q}$	Depolarizing scaling factor	Stochastic
$b_{\text{dep},2Q}$	Depolarizing offset	Stochastic
$b_{\text{amp},2Q}$	Amplitude damping offset	Stochastic
$b_{\phi,2Q}$	Phase damping offset	Stochastic
$\theta_{ix,zz}$	Coherent crosstalk angles	Coherent
β_{2Q}	Stretched dephasing exponent	Non-Markovian
k_{zz}	Correlated ZZ dephasing rate	Correlated
Correlated readout parameters (4)		
$a, b_{00 \leftrightarrow 11}$	Coefficients ($ 00\rangle \leftrightarrow 11\rangle$)	Correlated
$a, b_{01 \leftrightarrow 10}$	Coefficients ($ 01\rangle \leftrightarrow 10\rangle$)	Correlated

global corrections to local physical priors avoids the prohibitive complexity of learning unique parameter sets for every qubit and qubit pair, while successfully modeling the nonuniform error landscape. Our approach therefore complements, rather than replaces, standard device-characterization routines. It utilizes existing calibration data as a physically grounded baseline and employs ML to infer correction factors that refine the effective noise model using circuit-execution statistics captured within the specific algorithm-hardware operating regime.

C. Parameter optimization

The error model introduced in the previous section contains 20 free parameters θ . Our goal is to choose the parameters of the extended model such that the model accurately reproduces sample distributions P_i obtained from executing circuits $C_i \in \mathcal{C}$ on a specific quantum computing backend. This can be formalized as identifying the parameter configuration θ^* that minimizes the Hellinger distance [33] between simulation results and the results obtained on a real device. The Hellinger distance D_H is a bounded, symmetric metric for comparing discrete probability distributions and is numerically stable for finite-shot, sparse bitstring histograms. It ranges from 0 to 1, with 0 indicating identical distributions, and is defined as

$$D_H(P, Q) = \left(1 - \sum_j \sqrt{p_j q_j} \right)^{1/2}, \quad (22)$$

where p_j and q_j are probabilities of outcome j in distributions P and Q , respectively. Specifically, we minimize the mean Hellinger distance across all outcome distributions produced by the circuits in the training set, with respect to the circuits of interest, to obtain

$$\theta^* = \operatorname{argmin}_{\theta} \left\{ \frac{1}{n_{\text{tr}}} \sum_{i=1}^{n_{\text{tr}}} D_H[P_i, Q_i(\theta)] \right\}. \quad (23)$$

Here, $n_{\text{tr}} = |\mathcal{C}|$ is the size of the training set, $Q_i(\theta)$ is the empirical distribution obtained by simulating the circuit with a noise model with parameters θ . Each distribution is estimated from finite-shot measurement outcomes. Throughout, we use 30 000 shots (samples) per circuit in both the training and validation sets. We emphasize that Eq. (23) defines a global fit, meaning that a single parameter vector θ is identified jointly over all circuits in \mathcal{C} , with no circuit-specific parameters. This structure enforces that θ captures systematic, context-consistent error mechanisms shared by the algorithm family, rather than memorizing individual circuit outcomes, which potentially mitigates overfitting.

Since the configuration space of the model is large and evaluating the Hellinger distance requires running noisy

simulations, identifying θ^* is not tractable using brute-force search. Moreover, gradient-based optimization faces practical limitations in this setting. Efficient gradient-based optimization via automatic differentiation would require end-to-end differentiability of the noisy-simulation stack with respect to the noise parameters. In our workflow based on standard simulators (e.g., Qiskit Aer), such gradients are not natively available; obtaining them would require substantial reimplementations or numerical gradient estimation via finite differences, incurring $\mathcal{O}(d)$ additional objective evaluations per update. With $d = 20$ parameters in our model, this imposes a substantial computational overhead given that the noisy simulation itself is the primary bottleneck. Additionally, the noise landscape is often nonconvex and can be multimodal due to the complex interplay between coherent and incoherent error channels that challenges local search methods. We thus employ Bayesian optimization to efficiently obtain estimates of θ^* . BO is a ML-based strategy well suited for expensive-to-evaluate blackbox objective functions [37]. Unlike standard gradient-based local search, which determines step directions based on local derivatives, BO employs a probabilistic surrogate model to intelligently select parameter evaluations by maximizing an acquisition function that balances exploration of uncertain regions with exploitation of promising areas.

As a surrogate model, we use a tree-structured Parzen estimator [38], which uses adaptive, nonparametric Parzen window density models to distinguish promising from poor regions of the search space based on past evaluations. Then sample new configurations from the most promising areas to enable efficient hyperparameter optimization in complex, hierarchical domains. We benchmark the BO against random search (RS) in which parameter points are randomly selected from the search space to provide unbiased model estimates. Both methods are implemented using Optuna [39].

D. Scalability analysis

A key advantage of our approach is that we determine a fixed set of parameters within the noise model, irrespective of the circuit size or qubit count, leading to a circuit-size-independent complexity. This design choice decouples the computational complexity from the scale of the quantum hardware or the dimensions of the quantum system under study, enabling applicability to large-scale systems without prohibitive resource demands.

The resource requirements of our framework consist of experimental and computational costs. Experimentally, our data-efficient approach circumvents the prohibitive data overhead associated with full process tomography, which scales exponentially with qubit count due to the necessity of exhaustive state or process reconstructions. Instead, it relies solely on circuit-level outcome

distributions obtained from a compact, predefined set of benchmark circuits executed on the target quantum hardware. Computationally, the primary cost arises from the classical simulation of these benchmark circuits, required for each evaluation of the objective function during parameter determination. Notably, the cost per determination step remains constant, as it involves simulating the entire fixed set of training circuits to compute the discrepancy against the experimental dataset. Each iteration evaluates an objective function that quantifies this mismatch, here, via the Hellinger distance. The selection of the ML is pivotal for minimizing the number of such evaluations, given their computational intensity. We adopt BO, which excels in handling expensive blackbox functions, accelerating convergence and enhancing the framework’s scalability. Moreover, the entire characterization process is conducted as an offline, classical postprocessing step, obviating any need for real-time interaction with the QPU.

III. EXPERIMENTAL RESULTS

We benchmark our method using data collected from multiple IBM QPUs by executing quantum algorithms with varying qubit counts. The chosen systems `ibmq_kolkata`, `ibmq_mumbai`, and `ibmq_ehningen` comprise 27 qubits with CNOT as the basis gate, physically implemented via echoed cross-resonance and single-qubit rotations. Notably, `ibmq_ehningen` supports two distinct physical implementations of the CNOT gate [40]. In contrast, `ibmq_kingston` features 156 qubits and utilizes native CZ gates implemented via tunable couplers. Our noise model characterizes these basis gates, as all circuits must be decomposed into the basis gate set for execution. The benchmarks include the QAOA and VQE, which are critical classes of variational algorithms designed for near-term hardware. We further include random circuits to evaluate the framework’s suitability for unstructured quantum computations. Moreover, we partition our benchmark circuits by size into a small-size training set (four–six qubits) and a large-size validation set (seven–nine qubits) for testing

TABLE II. Properties of QAOA circuits executed on `ibmq_kolkata`, `ibmq_mumbai`, and `ibmq_ehningen`. Metrics include circuit depth and gate counts (CNOT and single qubit).

Property	Training set	Prediction set
Depth	27.0 ± 3.4	39.0 ± 3.4
CNOT gates	27.0 ± 10.7	78.0 ± 18.4
Single-qubit gates	50.3 ± 10.7	92.3 ± 13.2

the model’s ability to extrapolate from smaller systems to predict the behavior of larger, more complex systems.

A. QAOA

The QAOA constructs a parameterized quantum circuit consisting of alternating unitaries generated by a problem-dependent cost Hamiltonian H_C and a mixing Hamiltonian H_M :

$$U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{l=1}^L e^{-i\beta_l H_M} e^{-i\gamma_l H_C}, \quad (24)$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are variational parameters determined classically to minimize the expectation value of H_C . The mixing Hamiltonian is commonly chosen as $H_M = \sum_i X_i$, which promotes exploration of the Hilbert space. The cost Hamiltonian encodes the objective function of the optimization problem, typically taking the form of a diagonal operator in the computational basis

$$H_C = \sum_i c_i Z_i + \sum_{i<j} c_{ij} Z_i Z_j, \quad (25)$$

where Z_i are Pauli- Z operators and the coefficients c_i, c_{ij} reflect the problem structure. In this work, we employ QAOA applied to portfolio optimization problems, which involves fully connected two-qubit gates; further details can be found in Refs. [4,15,17,40].

For each QAOA circuit, we use the constrained optimization by linear approximation (COBYLA) [41] to

TABLE III. Experiment details of QAOA circuits executed on `ibmq_kolkata`, `ibmq_mumbai`, and `ibmq_ehningen`. The circuit fidelity F is defined in Eq. (26). The physical qubits represent the active qubits selected for circuit execution within each benchmark set.

Property	ibmq_kolkata		ibmq_mumbai		ibmq_ehningen	
	Training	Prediction	Training	Prediction	Training	Prediction
Default D_H	0.175 ± 0.000	0.275 ± 0.000	0.179 ± 0.001	0.277 ± 0.001	0.230 ± 0.000	0.328 ± 0.000
Parameterized D_H	0.096 ± 0.002	0.141 ± 0.004	0.089 ± 0.001	0.137 ± 0.001	0.151 ± 0.001	0.225 ± 0.004
D_H reduction (%)	45.3 ± 0.8	48.5 ± 1.7	50.4 ± 0.7	50.5 ± 0.5	34.4 ± 0.6	31.3 ± 1.2
Circuit fidelity F (%)	77.6 ± 8.1	47.9 ± 8.3	77.4 ± 7.1	51.5 ± 6.7	79.7 ± 6.9	53.2 ± 8.0
Physical qubits	{12, 15, 18, 21, 23, 24}	{4, 7, 10, 12, 15, 18, 21, 23, 24}	{14, 16, 19, 20, 22, 24, 25}	{12–16, 19, 22, 24, 25}	{15, 18–26}	{12–15, 18–26}
Physical qubit count	6	9	7	9	10	13

choose angles γ and β . Additionally, we use SWAP networks [15,16,42] on linear topologies to address the connectivity constraints of the QPU, which enables all compiled circuits to share the same fundamental gate structure and qubit connectivity pattern, regardless of the target device. This standardization allows us to control for variability in circuit design and compilation strategies.

We quantify circuit depth and gate counts of benchmark circuits executed on `ibmq_kolkata`, `ibmq_mumbai`, and `ibmq_ehningen` to assess their complexity and noise sensitivity. As detailed in Table II, the training set comprises circuits with four–six qubits and depths of 27.0 ± 3.4 , while the prediction set exhibits seven–nine qubits and depths of 39.0 ± 3.4 . The scaling of CNOT gates, from 27.0 ± 10.7 in training to 78.0 ± 18.4 in prediction, reflects the increased entanglement complexity and noise accumulation in larger circuits. While the increase in circuit width (four–six to seven–nine qubits) is moderate, this nearly $3\times$ increase in entangling gates drives the validation circuits into a significantly lower circuit fidelity regime (see Table III), ensuring that the model is rigorously tested on its ability to generalize error dynamics into a more noise-affected operating regime distinct from the high circuit fidelity training regime.

1. Single-device characterization

We evaluate the method on `ibmq_kolkata`. On the training set, our parameterized noise model optimized with BO and RS achieves an average Hellinger distance of 0.095 and 0.118, respectively. This represents a significant reduction in predictive error of 45.71% and 32.57% compared to the default model’s distance of 0.175, as shown in Fig. 2(a). Crucially, the model’s performance generalizes robustly to the validation set (seven–nine qubits) [Fig. 2(b)]. For these larger circuits, our model achieves an average Hellinger distance of 0.135 (BO) and 0.148 (RS), corresponding to a reduction of 50.91% and 46.18% over the default model’s distance of 0.275. These results also show that BO consistently outperforms RS in both the training and prediction processes, demonstrating its high optimization efficiency.

The contour plots in Figs. 3(a) and 3(b) compare the optimization landscapes explored by BO and RS, respectively, within a two-dimensional parameter space, which represent the most important parameters determined by Optuna. We observe that BO exhibits a structured and adaptive pattern, with sampled points increasingly concentrated in regions of minimal Hellinger distance, demonstrating that it efficiently identifies promising parameter regions. In contrast, RS yields a uniform and inefficient distribution of trials across the parameter space, resulting in a diffuse exploration that fails to converge effectively on the global minimum. This demonstrates the advantage of employing BO to achieve data-efficient parameterization.

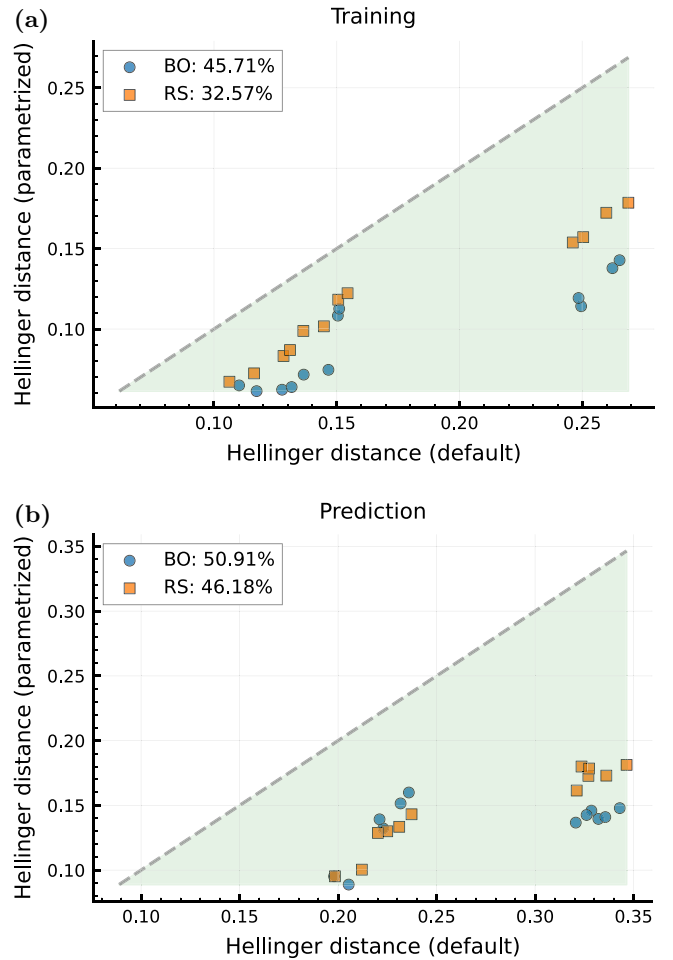


FIG. 2. Hellinger distances between outcomes from simulations using default device noise model (x axis) and our parameterized noise models (y axis) versus real executions on `ibmq_kolkata`. Blue circles represent models optimized via BO, while orange squares denote those optimized through RS. All points below the dashed diagonal ($y = x$) indicate an improvement in model fidelity over the default model. The percentages shown in the legend represent the average reduction in Hellinger distance achieved by each optimization method relative to the default model. (a) Training phase using QAOA circuits on four-, five-, and six-qubit instances, and (b) predictive performance on seven-, eight-, and nine-qubit circuits. Each circuit size was executed at four distinct times, yielding a total of 12 circuits for both training and prediction sets.

2. Cross-platform comparative analysis

To assess the robustness of our data-efficient framework, we repeat the full training and prediction pipeline on `ibmq_kolkata` three times, comparing BO to RS and plotting results with error bars in Fig. 4. As shown in Fig. 4(a), BO (Kolkata) outperforms RS (Kolkata-RS) in both training and prediction, exhibiting better performance. Moreover, Fig. 4(b) demonstrates that BO achieves a lower Hellinger distance than RS, confirming faster convergence to high-fidelity noise models.

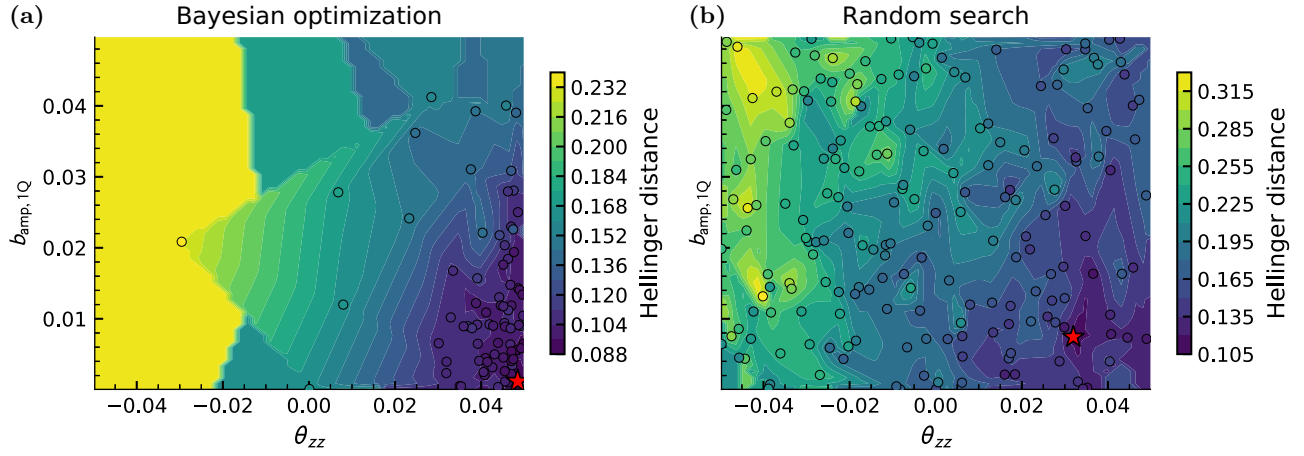


FIG. 3. Optimization landscape of the mean Hellinger distance, projected onto the two most significant noise parameters, as identified by Optuna, with marginalization over other noise model parameters. (a) Bayesian optimization and (b) random search. Each point represents a trial, and the red star marks the optimal configuration yielding the minimum value.

In the following, we employ BO for determining the parameters of noise models and extend our analysis across two other QPUs (`ibmq_mumbai` and `ibmq_ehningen`) featuring varying noise characteristics to evaluate the generalizability and robustness of our approach. For each device, we apply the same parameterized model architecture but train it independently on device-specific data. This process yields a unique model tailored to the specific error profile of each hardware instance. As shown in Fig. 4, our approach consistently yields a substantial reduction in D_H across all backends, and the optimization process converges to a low final D_H value. The framework’s effectiveness is most pronounced on `ibmq_mumbai`, which exhibits an average D_H reduction of over 50% for both training and prediction sets, with the reduction for individual prediction circuits peaking at 65%. Strong performance is also observed on `ibmq_kolkata` (45% training, 49% prediction) and `ibmq_ehningen` (35% training, 32% prediction). We summarize these achievements in Table III. This high-fidelity performance across multiple systems, each with distinct error characteristics, confirms the robustness and broad applicability of our methodology for capturing hardware-specific noise profiles.

We use device calibration data to calculate the circuit fidelity F to analyze the noise level of executing circuits on each device:

$$F = \prod_i (1 - \epsilon_i), \quad (26)$$

where ϵ_i are gate and readout errors. The increase in complexity from training set to prediction set corresponds to a significant circuit fidelity degradation from about 78% to about 50%, as shown in Table III, underscoring the challenge of extrapolating noise models to larger circuit regimes where errors become more pronounced. Notably,

the physical qubit allocation, driven by device-specific transpilation strategies, reveals that prediction circuits span a broader device region compared to training circuits, potentially introducing greater variability from spatial heterogeneity in device error rates. Moreover, the subgraph of active qubits is not fixed from circuit to circuit, further increasing the diversity of noise conditions encountered during evaluation. We observe that `ibmq_ehningen`, compared to other QPUs, maps circuits across a significantly wider physical qubit topology across four different benchmark executions. The increased spatial complexity is probably responsible for the more modest performance gains. A potential reason is that our amount of training data is insufficient to fully characterize the more complex error landscape of an expanded qubit ensemble. Nevertheless, the systematic increase in circuit complexity from training to prediction provides a meaningful test of our noise model’s ability to extrapolate beyond its training regime. This key extrapolation is enabled because our framework avoids qubit-specific error-rate mappings. Instead, it constructs a generative model of the device’s noise, optimizing global hyperparameters θ that parameterize statistical distributions for gate errors from which individual qubit and gate parameters are drawn.

In our experiments, we maintain circuit fidelities $F \gtrsim 50\%$ in both the training phase (four–six qubits) and the larger-scale prediction phase (seven–nine qubits), so their output distributions remain far from uniform and retain structure that is sensitive to coherent and correlated error mechanisms. By contrast, circuits with substantially larger width (e.g., 20–30 qubits) at comparable depth on the same devices are expected to operate close to the noise floor, yielding nearly uniform outcome distributions that obscure detailed differences between noise models and strongly reduce the discriminative power of metrics such as the Hellinger distance. Our conclusions should

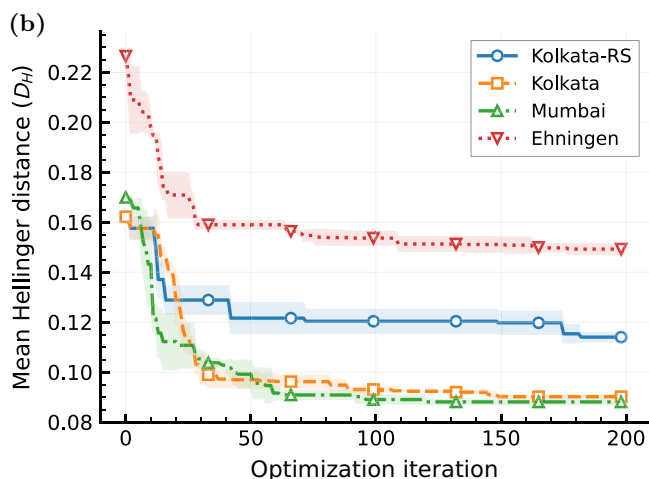
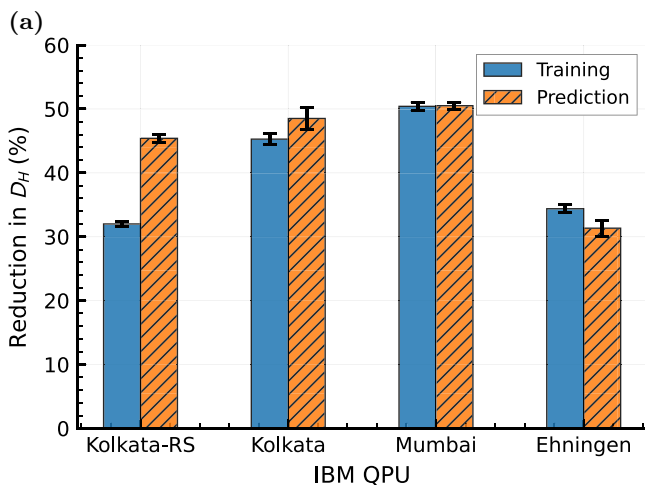


FIG. 4. Comparison of BO and RS for training the parameterized noise model on `ibmq_kolkata` (Kolkata-RS and Kolkata), alongside BO results on `ibmq_mumbai` (Mumbai) and `ibmq_ehningen` (Ehningen). (a) The average percentage reduction in Hellinger distance D_H relative to the default noise model, for both training (blue, left) and prediction (orange, right) sets, evaluated on QAOA circuits with four–six qubits (training) and seven–nine qubits (prediction), each executed at four calibration instances. Error bars denote means and standard deviations derived from three independent repetitions across the 12 circuits per set. (b) The convergence of the optimization process corresponding to the training results presented in (a), which displays the best-achieved mean Hellinger distance of the training circuits over iterations.

therefore be understood as applying to this intermediate-fidelity regime; extending the analysis to significantly larger mapped circuits will require dedicated experiments and is left for future work.

3. Training efficiency study

This section examines the interplay between model performance and training dataset size to identify the minimal data volume needed for target accuracy with reduced

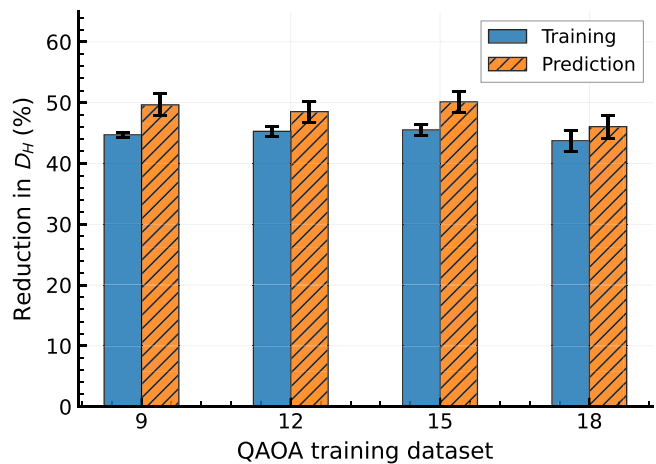


FIG. 5. Impact of different training dataset sizes on the performance of BO-optimized noise model on `ibmq_kolkata`. Each training set is denoted by the number of QAOA circuits. The experimental setting is the same as in Fig. 4.

resource demands. To assess our framework’s data requirements, we train and validate our noise model using four distinct QAOA datasets of varying sizes (9, 12, 15, and 18 circuits). The results reveal strong robustness and data efficiency. Across training-set sizes from 9 to 18 circuits, circuit properties remain stable. The preserved complexity distribution across all dataset sizes, combined with consistent fidelity estimates, suggests that our smallest nine-circuit training set sufficiently constrains the model for this benchmark; more details can be found in Table VI in the Appendix. This structural invariance explains the robust performance shown in Fig. 5, where the framework achieves a consistently similar relative reduction in the mean D_H of approximately 45% to 50% across all tested datasets, despite a $2\times$ difference in training data volume. Moreover, the optimization convergence profiles for 12, 15, and 18 circuit datasets are nearly indistinguishable, reaching the same optimal D_H value at a similar rate, highlighting stability (see Fig. 6 in the Appendix).

These results demonstrate that the framework is data efficient and provides empirical evidence of performance saturation with respect to training-set size, with no indication of overfitting in the tested regime. Within the tested range (9–18 training circuits), performance stabilizes: increasing the training set from 9 to 18 circuits yields no appreciable change in the optimal D_H and produces nearly indistinguishable convergence behavior (Fig. 6). This is consistent with adequate constraint diversity: circuits with varying depths and two-qubit-gate densities probe distinct error channels, providing complementary constraints on the shared 20-parameter model, while physics-informed functional forms restrict effective capacity by confining θ to physically realistic mechanisms. Out-of-distribution validation, where improvements match or in several cases

exceed training across devices, further confirms robust generalization.

B. VQE

To verify the efficacy of the method beyond QAOA, we perform a similar optimization for VQE circuits on `ibm_kingston`, a 156-qubit Heron R2 processor featuring IBM’s heavy-hex topology. We consider the task of minimizing the ground-state energy of the cluster Hamiltonian

$$H_n = \alpha_0 Z_0 + \alpha_n Z_n + \sum_{j=0}^{n-3} \beta_j Z_j X_{j+1} Z_{j+2}, \quad (27)$$

where α_i and β_i are real coefficients. In the following, we choose $\alpha_k = 1/n$ ($\beta_k = 1/n$) for all k for $n = 4, \dots, 9$. Using a simple, hardware-efficient ansatz corresponding to the `EfficientSU2` circuit in Qiskit, we optimize the parameters using the sequential least-squares quadratic programming (SLSQP) [43] in a noiseless simulation and verify that the optimized circuits converge to the true ground-state energy. We then execute the optimized circuits on the 156-qubit Heron R2 QPU `ibm_kingston`. For the execution, we group the circuits for each n into two qubit-wise commuting groups and sample from each group with 30 000 shots. This results in a total of 12 circuits, which we partition into six circuits ($n = 4, 5, 6$) for training and six circuits for validation ($n = 7, 8, 9$).

As in the QAOA study, we apply the same parameterized noise model architecture to the VQE experiments on `ibm_kingston`; however, the parameters θ are learned solely from the VQE training data specific to `ibm_kingston`. Performing three independent optimizations using 200 iterations each, we obtain optimized parameters, which result in an improvement of $43.99\% \pm 0.14\%$ in model fidelity for the training set and $18.58\% \pm 0.90\%$ for the validation set [up to 27% for a circuit prediction, see Fig. 7(a) in the Appendix]. While the parameterized model aligns more closely with experiment, especially on the training circuits, the improvement on the validation set is less pronounced than in the QAOA case. This likely reflects differences between the devices from which the data were sampled. As detailed in Table IV, VQE circuits contain significantly fewer entangling gates (CZs) than QAOA circuits (CNOT). Moreover, the `ibm_kingston` processor itself exhibits lower error rates. These factors contribute to inherently higher baseline circuit fidelities, resulting in substantially smaller Hellinger distances between the baseline model and experiment. On the validation set, the default model’s Hellinger distance is less than half of that observed for the comparable model on the older generation devices used for QAOA. Consequently, the margins of improvement over the baseline are smaller. Nevertheless, these results demonstrate that our framework

TABLE IV. Properties of VQE circuits executed on `ibm_kingston`.

Property	Training set	Prediction set
Default D_H	0.075 ± 0.000	0.124 ± 0.000
Parameterized D_H	0.042 ± 0.000	0.101 ± 0.001
D_H reduction (%)	$43.99\% \pm 0.14\%$	$18.58\% \pm 0.90\%$
Depth	41.2 ± 3.2	53.7 ± 3.4
CZ gates	12.0 ± 2.7	21.0 ± 2.7
Single-qubit gates	107.2 ± 21.9	184.7 ± 19.3
Circuit fidelity F (%)	91.0 ± 1.5	83.8 ± 3.9
Physical qubits	{0–5}	{0–8}
Physical qubit count	6	9

is robust across diverse hardware architectures, accommodating distinct qubit counts and native gate sets, establishing a clear path forward for modeling noise in more complex VQE circuits.

C. Random circuits

We further benchmark the method through random circuits on `ibmq_kolkata`. Random quantum circuits were generated using Qiskit’s `random_circuit` function with systematically varied parameters: qubit count $n \in \{4, \dots, 10\}$ and depth $d = n + 3$. The function constructs circuits by randomly sampling single-qubit gates from SU(2) and placing two-qubit gates between random qubit pairs within the depth constraint, with terminal measurements on all qubits. The results show a substantial mean D_H improvement of 43.60% in model fidelity for training and 45.90% for prediction [see Fig. 7(b) in the Appendix]. We perform three independent optimization processes and achieve a mean Hellinger distance reduction of $42.9\% \pm 1.5\%$ for the training set and $46.8\% \pm 2.1\%$ for the prediction set, with a maximum reduction of 63% observed for a circuit in the prediction set. This demonstrates that our method’s high performance is a statistically significant result, making it a reliable tool for practical application-aware hardware characterization.

Table V highlights the markedly distinct complexity profile of the random circuits. Compared to QAOA benchmarks, these random circuits exhibit approximately $3\times$ greater depth and employ a broader, more fragmented qubit allocation across 17 physical qubits on the device. This escalation in depth and spatial distribution leads to substantially degraded fidelities, $64.0\% \pm 14.0\%$ for the training circuits, and $24.8\% \pm 7.8$ for the prediction sets. The $3.5\times$ scaling in CNOT gates from training to prediction surpasses the $2.9\times$ factor observed in QAOA, underscoring the amplified accumulation of noise in deeper, less structured entangling operations. Remarkably, under these stringent conditions with fidelities roughly 25%, our approach achieves a performance improvement of around 44%, affirming its exceptional robustness across varied

TABLE V. Properties of random circuits executed on `ibmq_kolkata`.

Property	Training set	Prediction set
Default D_H	0.162 ± 0.000	0.250 ± 0.000
Parameterized D_H	0.093 ± 0.002	0.133 ± 0.005
D_H reduction (%)	42.9 ± 1.3	46.8 ± 1.9
Depth	87.2 ± 18.9	182.6 ± 23.8
CNOT gates	38.3 ± 17.8	136.6 ± 20.4
Single-qubit gates	106.8 ± 32.0	271.7 ± 34.3
Circuit fidelity F (%)	64.0 ± 14.0	24.8 ± 7.8
Physical qubits	{0–7, 10, 12, 15, 17, 18, 21, 23, 24}	{0, 1, 4, 6, 7, 10, 12–19, 21, 23, 24}
Physical qubit count	16	17

circuit topologies and noise regimes inherent to superconducting quantum hardware.

D. Application-aware noise characterization

The results show that the proposed framework achieves high predictive accuracy across three qualitatively different algorithmic families (QAOA, VQE, and random circuits). Crucially, rather than seeking a single universal parameter vector θ that simultaneously describes all algorithm classes on a target device, we adopt an application-aware philosophy: we estimate independent parameter sets for each algorithm-hardware context while retaining a fixed model architecture. This design is physically motivated by the context-dependent nature of noise in superconducting processors. Different algorithms can emphasize different effective error mechanisms through their circuit structure (e.g., depth, routing, and parallelism). For instance, QAOA applied to portfolio-optimization problems can require extensive SWAP-network routing to achieve all-to-all connectivity, whereas VQE with hardware-efficient ansätze typically conforms to the native coupling map, resulting in significantly shallower depths and fewer entangling gates. Since our parameterization captures duration-aware dephasing and correlated errors, we do not expect a parameter set calibrated on high-depth QAOA circuits to accurately predict low-depth VQE fidelity, or vice versa. Generality is thus achieved at the framework level: the same workflow applies across devices and algorithms, while the effective parameters adapt to the specific error dynamics of each context.

While cross-algorithm parameter transfer is not a prerequisite for deployment, quantifying it would elucidate the fundamental trade-off between intrinsic device physics and application-specific error manifestations. Investigating the predictive validity of a model trained on one algorithm family (e.g., QAOA) when applied to another (e.g., VQE) helps distinguish between static hardware

properties and the effective error dynamics induced by specific circuit designs. Future work utilizing transfer learning could quantify the adaptation cost between these contexts, ultimately determining whether the fidelity benefits of application-aware calibration justify the overhead of independent training or if approximate universality offers a viable compromise for large-scale systems.

IV. DISCUSSION AND CONCLUSION

This work develops an alternative parameterized noise model with circuit-size-independent complexity. Leveraging ML-driven optimization, we repurpose data from a few routine circuit executions to construct data-efficient and prediction accurate noise models for superconducting quantum processors. Crucially, our framework prioritizes application-aware characterization: we estimate effective noise parameters independently for each algorithm family on the target hardware rather than seeking a universal parameter set across all algorithm families. This design captures the distinct interaction between the algorithm and hardware, achieving high predictive accuracy with no additional dedicated calibration circuits or QPU jobs beyond the workload executions themselves. By systematically searching complex, multiparameter error spaces, our approach consistently produces models that are significantly more predictive than default ones obtained from hardware vendors. We demonstrate a reduction of up to 65% in the Hellinger distance, a metric that quantifies the divergence between predicted and experimental measurement-outcome (bitstring) probability distributions, confirming that our optimized models produce output distributions that more accurately represent the behavior of algorithms executed on real quantum hardware. This significant improvement in predictive accuracy is essential for a reliable simulation of quantum circuits on noisy quantum devices, providing a foundation for more accurate noise-aware quantum circuit compilation by enabling informed optimization of qubit selection, SWAP gate insertion, and other compilation passes.

A key advantage of our approach is its efficiency and practicality. It repurposes existing benchmarking or application-specific execution data, avoiding the need for extensive characterization protocols such as quantum process tomography. This reusability reduces experimental overhead, enabling practitioners to leverage prior executions to improve algorithm testing and error-mitigation protocols while iteratively refining noise models as additional data become available. Comprehensive benchmarking across multiple IBM quantum devices and quantum algorithms validates the framework’s adaptability to different types of superconducting hardware, highlighting its generality and potential for broad adoption in applications. By bridging the gap between theoretical models and experimental reality, our data-efficient

approach provides a critical enabling technology for developing the next generation of effective error mitigation techniques and noise-aware compilers, accelerating the realization of practical quantum computing applications.

Future work should address scaling challenges through hierarchical noise modeling. As system size grows, achieving a target accuracy may require more expressive parameterizations, increasing both the ML optimization cost and the amount of circuit-execution data. One promising direction is to partition the device coupling graph into regional subgraphs and assign region-specific parameter sets, capturing spatial inhomogeneity in larger systems while avoiding per-component overparameterization. Moreover, as differentiable or surrogate-based simulators mature, it will be valuable to investigate gradient-based and hybrid strategies, such as BO for initialization followed by gradient refinement, to further improve scalability in higher-dimensional parameter spaces. Additionally, investigating transfer learning across algorithmic families could quantify the adaptation cost between algorithmic contexts, helping determine if approximate universality is a viable compromise for large-scale systems. Moreover, incorporating leakage errors, where qubits excite outside of the computational subspace, represents a critical extension. This requires higher-dimensional simulations and often specialized characterization protocols, substantially increasing the computational burden of the ML loop. Although beyond the scope of this initial study, incorporating leakage is essential for further fidelity gains.

Our pairwise readout correlation model captures nearest-neighbor effects but neglects long-range correlations; small cluster-wise confusion matrices tied to resonators or readout lines could provide systematic improvements. Benchmarking dense algorithms like QAOA on utility-scale processors would complement our VQE results on 156-qubit hardware, revealing how the framework interacts with large-scale qubit mapping strategies. Additionally, extending to alternative qubit modalities represents essential future work. While our experiments focus on superconducting devices, the methodology is architecture-agnostic. Adaptation requires replacing the physics-informed ansatz $\mathcal{N}(\theta)$ with platform-specific mechanisms. For trapped-ion processors [44,45], for instance, a tailored ansatz would capture dominant error sources in Mølmer-Sørensen gates, such as motional mode heating and laser-intensity fluctuations. The ultimate goal is a full integration into quantum compiler toolchains, enabling adaptive, algorithm-specific compilation that can dynamically tailor noise models to both the hardware’s current state and circuit structure. This would optimize execution by capturing the unique interplay between algorithms and device-specific error channels.

ACKNOWLEDGMENTS

The paper was written as part of the project “Quantum-based Energy Grids (QuGrids),” which is receiving funding from the programme “Profilbildung 2022”, an initiative of the Ministry of Culture and Science of the State of North Rhine-Westphalia. The contribution of M.R. and D.A.K. has been supported by the Fraunhofer Heilbronn Research and Innovation Centers HNFIZ. Parts of the data used for this study have been generated in the course of project QORA funded by by Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg in the frame of the Competence Center Quantum Computing Baden-Württemberg. We acknowledge the use of IBM Quantum services for this work and to advanced services provided by the IBM Quantum Researchers Program. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain commercially sensitive information. The data are available from the authors upon reasonable request.

APPENDIX: EXPERIMENTAL DETAILS

We present the experimental details. Figure 6 shows the optimization convergence across different QAOA circuit sizes on `ibmq_kolkata`, corresponding to Fig. 5. Table VI quantifies the structural consistency of our benchmark circuits across different training set sizes. The physical qubits used are identical across different circuit sizes: {12, 15, 18, 21, 23, 24} for the training set and {4, 7, 10, 12, 15, 18, 21, 23, 24} for the prediction set.

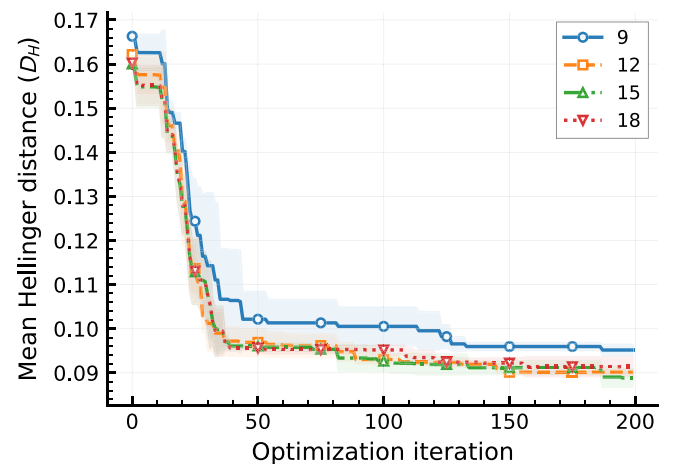


FIG. 6. Convergence of the optimization process across different circuit sizes on `ibmq_kolkata` in Fig. 5.

TABLE VI. Properties of benchmark circuit sets on `ibmq_kolkata`, analyzed for different numbers of QAOA circuits in Fig. 5.

Property	Training set				Prediction set			
	9 circuits	12 circuits	15 circuits	18 circuits	9 circuits	12 circuits	15 circuits	18 circuits
Depth	27.0 ± 3.5	27.0 ± 3.4	27.0 ± 3.4	27.0 ± 3.4	39.0 ± 3.5	39.0 ± 3.4	39.0 ± 3.4	39.0 ± 3.4
CNOT gates	27.0 ± 10.9	27.0 ± 10.7	27.0 ± 10.6	27.0 ± 10.5	78.0 ± 18.6	78.0 ± 18.4	78.0 ± 18.2	78.0 ± 18.1
Single-qubit gates	50.3 ± 10.8	50.3 ± 10.7	50.3 ± 10.6	50.3 ± 10.5	92.3 ± 13.4	92.3 ± 13.2	92.3 ± 13.1	92.3 ± 13.0
Circuit fidelity F (%)	77.3 ± 8.1	77.6 ± 8.1	77.7 ± 8.1	77.8 ± 8.1	48.3 ± 8.2	47.9 ± 8.3	47.7 ± 8.3	47.5 ± 8.3

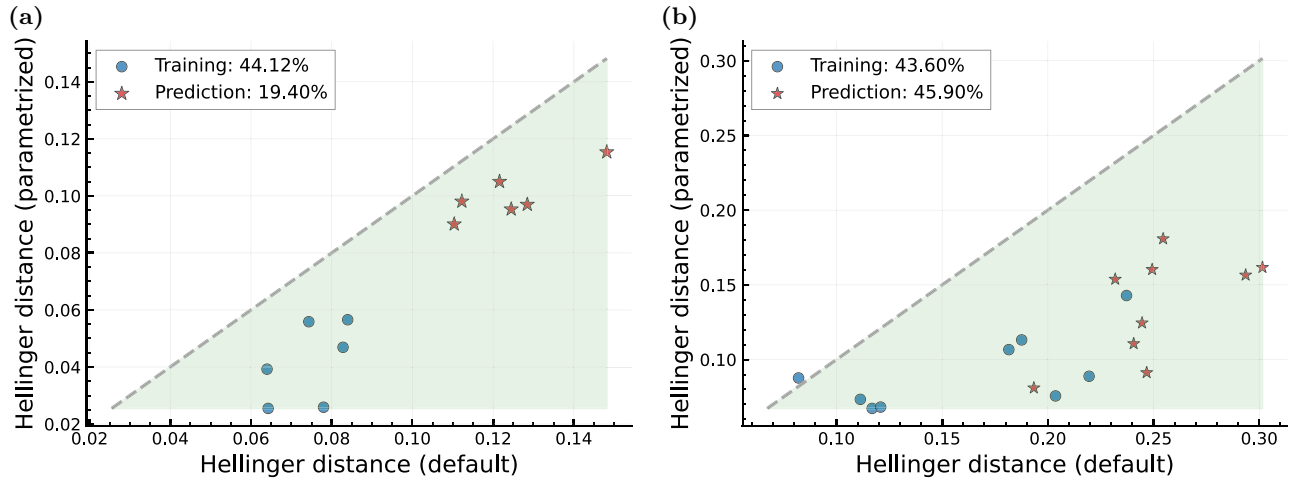
FIG. 7. Performance comparison of noise models using (a) VQE on `ibmq_kingston` and (b) random circuits on `ibmq_kolkata`.

Figure 7(a) compares our parameterized model against the default model using VQE obtained on `ibmq_kingston` in Sec. III B, while Fig. 7(b) presents results using random circuits on `ibmq_kolkata` in Sec. III C.

- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [2] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
- [3] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 1 (2014).
- [4] S. Brandhofer, D. Braun, V. Dehn, G. Hellstern, M. Hüls, Y. Ji, I. Polian, A. S. Bhatia, and T. Wellens, Benchmarking the performance of portfolio optimization with QAOA, *Quantum Inf. Process.* **22**, 1 (2023).
- [5] S. Koretsky, P. Gokhale, J. M. Baker, J. Viszlai, H. Zheng, N. Gurung, R. Burg, E. A. Paaso, A. Khodaei, R. Eskandarpour, and F. T. Chong, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE Computer Society, Los Alamitos, CA, USA, 2021), pp. 181–187.
- [6] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).

- [7] J. Clarke and F. K. Wilhelm, Superconducting quantum bits, *Nature* **453**, 1031 (2008).
- [8] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer’s guide to superconducting qubits, *Appl. Phys. Rev.* **6**, 021318 (2019).
- [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, [arXiv:2405.08810](https://arxiv.org/abs/2405.08810).
- [10] S. Sivarajah, S. Dilkes, A. Cowtan, W. Simmons, A. Edgington, and R. Duncan, `t|ketra`: a retargetable compiler for NISQ devices, *Quantum Sci. Technol.* **6**, 014003 (2021).
- [11] L. Burgholzer, J. Echavarria, P. Hopf, Y. Stade, D. Rovara, L. Schmid, E. Kaya, B. Mete, M. N. Farooqi, M. Chung, *et al.*, in *Proceedings of the Supercomputing Asia and International Conference on High Performance Computing in Asia Pacific Region (SCA/HPCAsia ’26)* (ACM, New York, NY, USA, 2026), pp. 55–67.
- [12] L. Burgholzer, Y. Stade, T. Peham, and R. Wille, MQT Core: The backbone of the Munich Quantum Toolkit (MQT), *J. Open Source Software* **10**, 7478 (2025).
- [13] Y. Ji, S. Brandhofer, and I. Polian, in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE Computer Society, Los Alamitos, CA, USA, 2022), pp. 204–214.
- [14] N. Quetschlich, L. Burgholzer, and R. Wille, Mqt predictor: Automatic device selection with device-specific circuit compilation for quantum computing, *ACM Trans. Quantum Comput.* **6**, 1 (2025).

- [15] Y. Ji, X. Chen, I. Polian, and Y. Ban, Algorithm-oriented qubit mapping for variational quantum algorithms, *Phys. Rev. Appl.* **23**, 034022 (2025).
- [16] J. Montanez-Barrera, Y. Ji, M. R. von Spakovsky, D. E. B. Neira, and K. Michielsen, Optimizing QAOA circuit transpilation with parity twine and SWAP network encodings, [arXiv:2505.17944](https://arxiv.org/abs/2505.17944).
- [17] Y. Ji, K. F. Koenig, and I. Polian, Improving the performance of digitized counterdiabatic quantum optimization via algorithm-oriented qubit mapping, *Phys. Rev. A* **110**, 032421 (2024).
- [18] P. K. Faehrmann, M. Steudtner, R. Kueng, M. Kieferova, and J. Eisert, Randomizing multi-product formulas for hamiltonian simulation, *Quantum* **6**, 806 (2022).
- [19] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of trotter error with commutator scaling, *Phys. Rev. X* **11**, 011020 (2021).
- [20] A. M. Childs, A. Ostrander, and Y. Su, Faster quantum simulation by randomization, *Quantum* **3**, 182 (2019).
- [21] F. B. Maciejewski, F. Baccari, Z. Zimborás, and M. Ozmaniec, Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm, *Quantum* **5**, 464 (2021).
- [22] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
- [23] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of addressability by simultaneous randomized benchmarking, *Phys. Rev. Lett.* **109**, 240504 (2012).
- [24] A. Agarwal, L. P. Lindoy, D. Lall, F. Jamet, and I. Rungger, Modelling non-Markovian noise in driven superconducting qubits, *Quantum Sci. Technol.* **9**, 035017 (2024).
- [25] H. Zhang, B. Pokharel, E.M. Levenson-Falk, and D. Lidar, Predicting non-Markovian superconducting-qubit dynamics from tomographic reconstruction, *Phys. Rev. Appl.* **17**, 054018 (2022).
- [26] G. A. White, C. D. Hill, F. A. Pollock, L. C. Hollenberg, and K. Modi, Demonstration of non-Markovian process characterisation and control on a quantum processor, *Nat. Commun.* **11**, 6301 (2020).
- [27] J. Etxezarreta Martinez, P. Fuentes, P. Crespo, and J. Garcia-Frias, Time-varying quantum channel models for superconducting qubits, *npj Quantum Inf.* **7**, 115 (2021).
- [28] J. F. Poyatos, J. I. Cirac, and P. Zoller, Complete characterization of a quantum process: the two-bit quantum gate, *Phys. Rev. Lett.* **78**, 390 (1997).
- [29] K. Georgopoulos, C. Emary, and P. Zuliani, Modeling and simulating the noisy behavior of near-term quantum computers, *Phys. Rev. A* **104**, 062432 (2021).
- [30] D. Brand, I. Sinayskiy, and F. Petruccione, Markovian noise modelling and parameter extraction framework for quantum devices, *Sci. Rep.* **14**, 4769 (2024).
- [31] M. L. Dahlhauser and T. S. Humble, Modeling noisy quantum circuits using experimental characterization, *Phys. Rev. A* **103**, 042603 (2021).
- [32] Y. Oda, O. Shehab, and G. Quiroz, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Bellevue, WA, USA, 2023), Vol. 2, pp. 219–222.
- [33] E. Hellinger, Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen, *J. Reine Angew. Math.* **1909**, 210 (1909).
- [34] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggens, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [35] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* **567**, 491 (2019).
- [36] Y. Ji and I. Polian, Synergistic dynamical decoupling and circuit design for enhanced algorithm performance on near-term quantum devices, *Entropy* **26**, 586 (2024).
- [37] F. Archetti and A. Candelieri, *Bayesian Optimization and Data Science* (Springer Publishing Company, Cham, Switzerland, Incorporated, 2019), 1st ed., ISBN/ID: 10.5555/3379057.
- [38] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., Granada, Spain, 2011), Vol. 24.
- [39] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, NY, USA, 2019).
- [40] Y. Ji, K. F. Koenig, and I. Polian, Optimizing quantum algorithms on bipotent architectures, *Phys. Rev. A* **108**, 022610 (2023).
- [41] M. J. Powell, *A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation* (Springer, Dordrecht, The Netherlands, 1994).
- [42] J. Weidenfeller, L. C. Valor, J. Gacon, C. Tornow, L. Bello, S. Woerner, and D. J. Egger, Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware, *Quantum* **6**, 870 (2022).
- [43] D. Kraft, A software package for sequential quadratic programming, Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt (1988).
- [44] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, *Appl. Phys. Rev.* **6**, 021314 (2019).
- [45] S. A. Moses *et al.*, A race-track trapped-ion quantum processor, *Phys. Rev. X* **13**, 041052 (2023).