

Enhanced Quantum Process Tomography by Digital Twins of Error Matrices

Tangyou Huang,^{1,2} Akshay Gaikwad,^{1,*} Ilya Moskalenko,² Anuj Aggarwal,¹ Tahereh Abad,¹ Marko Kuzmanović,² Yu-Han Chang,² Ognjen Stanisavljević,² Gheorghe Sorin Paraoanu,² Anton Frisk Kockum,¹ and Giovanna Tancredi¹

¹*Department of Microtechnology and Nanoscience (MC2),*

Chalmers University of Technology, SE-41296 Göteborg, Sweden

²*QTF Centre of Excellence and InstituteQ, Department of Applied Physics, Aalto University, FI-00076 Aalto, Finland*
(Dated: April 26, 2025)

Accurate and robust quantum process tomography (QPT) is crucial for verifying quantum gates and diagnosing implementation faults in the experiments aimed at building a universal quantum computer. However, the reliability of the QPT protocols is often compromised by faulty probes, particularly state preparation and measurement (SPAM) errors, which introduce fundamental inconsistencies into the traditional QPT algorithms. In this work, we investigate an enhanced version of QPT for multi-qubit systems by integrating the error matrix within a digital twin of the identity process matrix, enabling the statistical refinement of SPAM error learning and improving the precision of QPT. Through numerical simulations, we demonstrate that this approach enables highly accurate and faithful process characterization, which is further experimentally validated using superconducting quantum gates, achieving an order-of-magnitude fidelity improvement over the standard QPT. Our results provide a practical and precise method for assessing quantum gate fidelity and enhancing QPT on a given hardware.

Introduction. Significant advancements have been made in building large-scale quantum processors using diverse physical platforms [1–4]. Although a higher qubit count provides exponential computational benefits, it also brings major challenges in implementing high-fidelity multi-qubit gates and accurately characterizing them for further enhancement [5–10]. Identifying errors in gate implementation and improving quantum architectures require more than a single scalar measure, such as gate fidelity from randomized benchmarking protocols [11–14]. Instead, a comprehensive characterization of the entire quantum process is essential, which is usually done using quantum process tomography (QPT) [15–17].

Quantum process tomography basically reconstructs an unknown quantum process, represented by completely positive (CP) and trace preserving (TP) map \mathcal{E} , by determining its action on the set of probe input states $\{\rho_i\}$ as, $\mathcal{E}(\rho_i) \rightarrow \rho'_i$. This requires a state preparation and measurement (SPAM) step just before and after the action of \mathcal{E} respectively. Mathematically, the prime objective of QPT is to calculate the process matrix χ characterizing the quantum map \mathcal{E} acting on d dimensional quantum state using the operator-sum representation [18–22]:

$$\mathcal{E}(\rho_i) = \sum_{m,n=1}^{d^2} \chi_{mn} E_m \rho_i E_n^\dagger, \quad (1)$$

where χ_{mn} is the matrix element of $\chi \in \mathbb{C}^{d^2 \times d^2}$ which provides a unique description of \mathcal{E} in a chosen operator basis $\{E_j \in \mathbb{C}^{d \times d}\}$ satisfying CPTP criteria as $\chi \geq 0$ and $\sum_{m,n} \chi_{mn} E_m^\dagger E_n = I$ respectively. The straightforward way to compute the χ -matrix is as follows: We first prepare a set of known input states $\{\rho_i\}$, then apply the quantum map \mathcal{E} (evolution under target process), followed by measurement of a set of observables

$\{M_\mu\}$, typically chosen as positive operator-valued measures (POVMs). This yields a collection of data points $\{p_{i,\mu}\}$ as

$$p_{i,\mu} = \text{Tr}(M_\mu \mathcal{E}(\rho_i)) = \sum_{m,n=1}^{d^2} \chi_{mn} \text{Tr}(M_\mu E_m \rho_i E_n^\dagger). \quad (2)$$

Using the collected data together with appropriate data post-processing QPT algorithm \mathcal{J} , the estimated process matrix χ_e can be calculated as,

$$\text{C0} : \mathcal{J}(\rho_i, M_\mu, p_{i,\mu}) \rightarrow \chi_e \quad (\text{ideal case}). \quad (3)$$

Note that the case C0, as defined in Eq. (3), represents the ideal scenario of SPAM-error-free QPT. However, this situation is never realized in practice due to unavoidable imperfections in the SPAM step. Due to the faulty SPAM step we end up collecting noisy data $\{\tilde{p}_{i,\mu}\}$ leading to incorrect estimation of process matrix as

$$\text{C1} : \mathcal{J}(\cdot, \cdot, \tilde{p}_{i,\mu}) \rightarrow \tilde{\chi}_e \quad (\text{noisy case}). \quad (4)$$

This intrinsic inconsistency in QPT, as evident from Eqs. (3) and (4), leads to the discrepancy $\chi_e \neq \tilde{\chi}_e$, despite both matrices representing the same quantum process \mathcal{E} . Consequently, the noisy case C1 as described in Eq. (4) can be further categorized into three distinct cases:

$$\text{C11} : \mathcal{J}(\tilde{\rho}_i, \tilde{M}_\mu, \tilde{p}_{i,\mu}) \rightarrow \tilde{\chi}_e^1 \quad (\text{self-consistent}), \quad (5)$$

$$\text{C12} : \mathcal{J}(\rho_i, M_\mu, \tilde{p}_{i,\mu}) \rightarrow \tilde{\chi}_e^2 \quad (\text{self-inconsistent}), \quad (6)$$

$$\text{C13} : \mathcal{J}(\tilde{\rho}_i, \tilde{M}_\mu, \tilde{p}_{i,\mu}) \rightarrow \tilde{\chi}_e^3 \quad (\approx \text{self-consistent}). \quad (7)$$

Case C11, C12 and C13, given in Eq. (5), (6), and (7), respectively are contained within Eq. (4) where C11 is self-consistent, as the noisy data $\{\tilde{p}_{i,\mu}\}$ is associated with exact faulty probes $\{\tilde{\rho}_i, \tilde{M}_\mu\}$. Yet, this case remains nearly

unattainable since in experiments they are typically inaccessible.

In contrast, case C12, defined in Eq. (6), represents a practical but self-inconsistent approach. We refer to this case as standard QPT (std-QPT). Here, noisy data $\{\tilde{p}_{i,\mu}\}$ is erroneously associated with ideal probes $\{\rho_i, M_\mu\}$, a flawed assumption that researchers often adopt blindly in their QPT calculation and analysis. This unrealistic presumption—that SPAM operations are perfect—not only leads to inconsistencies in QPT but also distorts the interpretation of results, rendering the corresponding $\tilde{\chi}_e^2$ unreliable. On the other hand, case C13, as defined in Eq. (7), represents a practical yet nearly self-consistent version of QPT [23]: $(\bar{\rho}_i \approx \tilde{\rho}_i, \bar{M}_\mu \approx \tilde{M}_\mu) \rightarrow \tilde{\chi}_e^3 \approx \tilde{\chi}_e^1$. Here, $\{\bar{\rho}_i, \bar{M}_\mu\}$ correspond to experimentally tomographed noisy probes, obtained with certain assumptions and approximations. Due to gauge symmetry, it is inherently impossible to tomograph states and POVMs independently with arbitrary accuracy and precision, as quantum state tomography (QST) relies on POVMs as probes and vice versa [24–26]. For practical applications, Case C13 emerges as the most promising and reliable alternative to Case C12, as it effectively mitigates SPAM errors in QPT and enables faithful reconstruction in a self-consistent manner.

This problem was initially introduced and explored a decade ago in the papers [27–30], which later inspired the development of gate set tomography [8, 31–35]—a sophisticated protocol for SPAM-error-free characterization of quantum gate set. However, this method comes at the expense of significantly higher experimental and computational complexity compared to QPT, making it impractical to implement beyond two qubits [9].

In this Letter, we numerically and experimentally investigate and compare C12 with C13—practically feasible versions of QPT—using C0 as a reference for a multi-qubit system. Our results show that the revised probes $\{\bar{\rho}_i, \bar{M}_\mu\}$ in C13 achieve greater accuracy than the standard QPT in C12. To enhance the efficiency and reliability of error-mitigated QPT (EM-QPT), we use a machine-learning technique [36–38] to learn the statistical behavior of SPAM errors by constructing the digital twin of identity process matrices, thereby enabling high-precision and practical applications on a given hardware. Moreover, we employ CVX tool [39] to solve convex optimization problems given in Eqs. (3) and (4) and compute process matrices in our numerical analysis. We use the Uhlmann-Jozsa metric to compute the process fidelity between χ_1 and χ_2 as [40],

$$\mathcal{F}(\chi_1, \chi_2) = \left(\text{Tr} \sqrt{\sqrt{\chi_1} \chi_2 \sqrt{\chi_1}} \right)^2. \quad (8)$$

Error-mitigated QPT. Our objective is to precisely estimate the process matrix of an arbitrary quantum operation $\mathcal{E}(\cdot)$ while accounting for SPAM errors [23, 41–43].

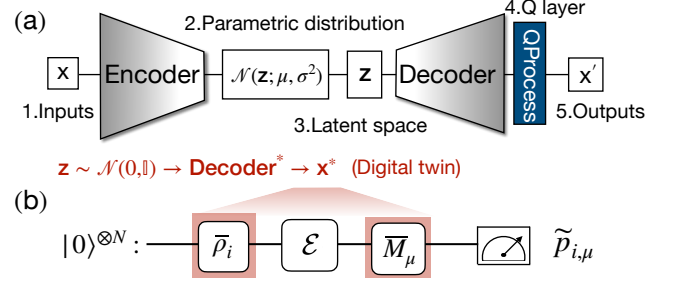


FIG. 1. **Schematic diagram of digital twin-assisted quantum process tomography.** (a) The variational autoencoder (VAE) model consists of an encoder and a decoder built with deep neural networks. The input training data x is mapped by the encoder into a parametric probability distribution $\mathcal{N}(z; \mu, \sigma^2)$. The latent variable z is sampled from this distribution and used to reconstruct the output x' through the decoder and a pre-designed quantum processing layer (QProcess). This enables the generation of a *digital twin*, where synthetic data is produced from a standard normal distribution $z \sim \mathcal{N}(z; 0, \mathbb{I})$ and decoded into x' . (b) The digital twin is applied to reconstruct the error matrix $x = \chi_{\text{err}}$ using a well-trained VAE model. This allows for the correction of input states and measurement probes, leading to SPAM (state preparation and measurement) error-mitigated quantum process tomography of quantum process $\mathcal{E}(\cdot)$.

To suppress these errors, one can first identify the noisy input states and measurement operators, then integrate them into the QPT of the target channel. This approach, outlined in Eq. (7), represents a nearly self-consistent refinement of standard QPT. To achieve this, we utilize identity QPT—where no operation is applied, only state preparation (\mathcal{E}_{sp}) and measurement (\mathcal{E}_{m}) are performed,

$$\chi^{\text{I}} \equiv \mathcal{E}_{\text{m}} \circ \mathcal{E}_{\text{sp}}. \quad (9)$$

The χ^{I} represents identity process matrix. Ideally, $\chi_{mn}^{\text{I}} = \delta_{m0}\delta_{n0}$, where δ_{mn} denotes the Kronecker delta function. Any deviation from the ideal χ^{I} directly indicates the presence of SPAM errors in the experiment, resulting in $\tilde{\chi}^{\text{I}}$ which we refer to as an ‘*error matrix*’ [42]. By partially inverting the QPT algorithm, one can readily determine the noisy input states and measurement operators as follows [44]:

$$\mathcal{J}^{-1}(\tilde{\chi}^{\text{I}}, \tilde{p}_{i,\mu}^{\text{I}}, M_\mu) \rightarrow \bar{\rho}_i, \quad (10)$$

$$\mathcal{J}^{-1}(\tilde{\chi}^{\text{I}}, \tilde{p}_{i,\mu}^{\text{I}}, \rho_i) \rightarrow \bar{M}_\mu, \quad (11)$$

where $\{\tilde{p}_{i,\mu}^{\text{I}}\}$ represents noisy data obtained in identity QPT experiment. Additionally, while computing $\{\bar{\rho}_i\}$, we assume ideal measurement operators, and vice versa. This is because the gauge symmetry inherently prevents the simultaneous determination of $\{\bar{\rho}_i, \bar{M}_\mu\}$ with arbitrary accuracy and precision.

In this paper, we benchmark the practical, error-mitigated and nearly self-consistent version of QPT:

$$\text{EM-QPT} : \mathcal{J}(\bar{\rho}_i, \bar{M}_\mu, \tilde{p}_{i,\mu}; \tilde{\chi}^{\text{I}}).$$

This EM-QPT approach, while effective, is highly resource-intensive, particularly in applications where frequent process characterization is required, such as gate optimization [45, 46]. It is also potentially fragile in the presence of anomaly errors, such as glitch effects in realistic experiments. Since the error matrix $\tilde{\chi}^I$ is independent of gate operation, it is natural to explore whether machine learning techniques can be leveraged to learn the statistical behavior of SPAM error hidden in $\tilde{\chi}^I$, thereby enabling a more practical and efficient approach to assist in error mitigation. Inspired by a recent study [38], we propose using a generative model as a digital twin of the error matrix to enhance the feasibility and practicality of EM-QPT method. Remarkably, we found that the digital twin, trained via a deep neural network on a given hardware, effectively captures the underlying characteristics of SPAM errors, yielding a more refined version of Eq. (10) and (11). Surprisingly, it can potentially outperform real-time error matrix acquisition, enabling high-precision QPT and offering a more robust and efficient approach to SPAM error mitigation.

The Variational Autoencoder. A Variational Autoencoder (VAE) [36] is a generative model that integrates deep learning with probabilistic frameworks to learn a latent representation of data by mapping inputs to a continuous latent space (see the schematic diagram in Fig.1). Unlike traditional autoencoders [47], a VAE models the latent space as a probability distribution rather than as fixed embeddings. Specifically, VAE consists of an *encoder* that maps the input \mathbf{x} to a latent vector \mathbf{z} obeying a probability distribution $\mathbb{Q}(\mathbf{z}|\mathbf{x})$. The *decoder* reconstructs the input data $\mathbf{x} \rightarrow \mathbf{x}'$ from a sampled latent vector \mathbf{z} . To enable backpropagation through the network, the latent variable \mathbf{z} is sampled using the reparameterization trick [48]: $\mathbf{z} \sim \mathcal{N}(\mu(\mathbf{x}), \epsilon \cdot \sigma(\mathbf{x}))$, where an auxiliary noise variable is sampled from a normal distribution $\epsilon \sim \mathcal{N}(0, \mathbb{I})$. The total loss function consists of a reconstruction loss and a KL divergence loss

$$\text{Loss} = \mathcal{L}_{\text{recon}}(\mathbf{x}', \mathbf{x}) + \beta D_{\text{KL}}[\mathbb{Q}(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(\mathbf{z}|\mathbf{x})] \quad (12)$$

where the mean squared error (MSE) $\mathcal{L}_{\text{recon}} = \sum_{n,m} (\mathbf{x}_{n,m} - \mathbf{x}'_{n,m})^2$ ensures the generated \mathbf{x}' is similar to \mathbf{x} , and the Kullback-Leibler (KL) divergence D_{KL} regularizes $\mathbb{Q}(\mathbf{z}|\mathbf{x})$ to match prior normal distribution $\mathcal{N}(\mathbf{z}|\mathbf{x})$ and β is a hyperparameter balancing the two losses. Since the learned posterior and target prior are assumed to be two Gaussian distributions, the KL divergence has a closed-form solution [36]: $D_{\text{KL}}[\mathbb{Q}(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(\mathbf{z}|\mathbf{x})] = \frac{1}{2} (\sigma^2 + \mu^2 - 1 - \log \sigma^2)$, making it computationally efficient and mathematically well-defined.

The digital twinning of error matrix ($\tilde{\chi}^I$). We characterize the SPAM errors by constructing the digital twin of the error matrix $\tilde{\chi}^I$. To achieve this, we first collect a training database $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=0}^{N_x}$, consisting of N_x independent QPT experiments for the identity quantum process, which is implemented by applying a short idle time

of a few nanoseconds in experiments. We then decompose each χ -matrix into two channels of neural networks, corresponding to the real and imaginary parts, forming the input training data points $\chi_i \rightarrow \mathbf{x}_i \in \mathbb{R}^{2 \times d^2 \times d^2}$. To ensure that the VAE output preserves the CPTP constraint, we introduce an additional layer called **QProcess** [44] relying on the Cholesky decomposition [49, 50]. In this framework, the digital twin of the error matrix, $\mathbf{x}'_i \rightarrow \chi'_i$, is expected to statistically mimic the major pattern of SPAM errors embedded in the error matrix, enabling enhanced QPT against the abnormal error interference. In this vein, the EM-QPT is further enhanced by machine learning (ML) as,

$$\text{ML-QPT} : \mathcal{J}_{\text{ML}}(\bar{\rho}_i, \bar{M}_\mu, \tilde{\rho}_{i,\mu}; \chi_{\text{DT}}^I),$$

where χ_{DT}^I is the digital twin of error matrix.

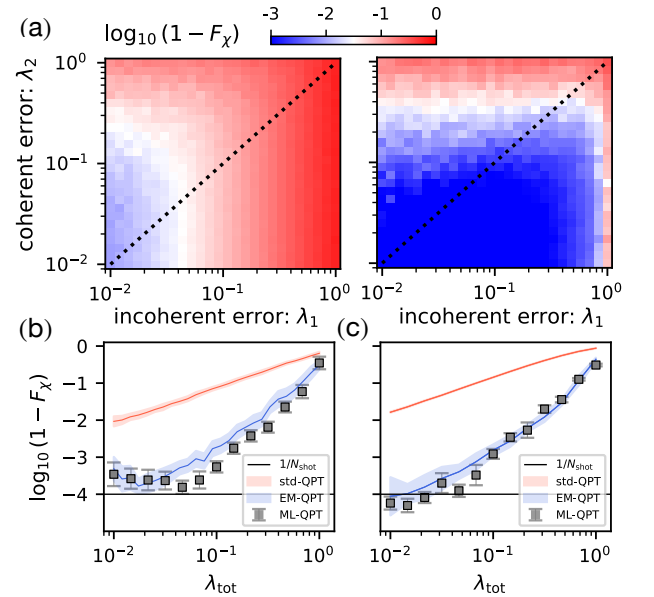


FIG. 2. The EM-QPT for one/two-qubit gates. In (a), the average process infidelity of single-qubit gates by using the standard QPT and EM-QPT is compared with respect to both coherent and incoherent error. In explicit, we present the fidelity behavior of a mixed error regime shown by the dashed line in (b) and (c) for single-qubit and two-qubit gates, respectively. The solid curve and square symbol refers to the average result over 10^2 gate samples with the standard deviation in shadow and errorbar. The horizontal line denotes the statistical error $1/N_{\text{shot}}$ with the shot number $N_{\text{shot}} = 10^4$ in numeric experiments.

Numerical simulation. In standard N -qubit QPT, each qubit is first initialized in one of the states $\{|0\rangle, |1\rangle, |+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, |-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}\}$. This is achieved by applying one of the preparation gates $U_1 \in \{\mathbb{I}, \mathbf{X}, \mathbf{R}_y(\pi/2), \mathbf{R}_x(-\pi/2)\}$, respectively. After state preparation, the quantum gate \mathcal{G} under investigation is applied. To enable full process characterization, we then apply a set of informationally complete rotation

gates $U_2 \in \{I, R_y(-\pi/2), R_x(\pi/2)\}$ prior to measuring the qubits in the computational basis. In practice, the noisy readout is composed by

$$\tilde{\rho}_{i,\mu} = \langle\langle M_\mu | \mathcal{E}_m \circ \mathcal{E}_{\text{gate}}(\mathcal{G}) \circ \mathcal{E}_{\text{sp}} | \rho_i \rangle\rangle, \quad (13)$$

where $\mathcal{E}_m(\cdot), \mathcal{E}_{\text{gate}}(\cdot), \mathcal{E}_{\text{sp}}(\cdot)$ represent the error channel upon the ideal measurement, gate and initial states, respectively. To simulate the incoherent SPAM errors, we add the error channel upon the qubit reset and measurement with the depolarizing error $\mathcal{E}_{\text{dep}}(\rho; \lambda) = (1 - \lambda)\rho + \frac{\lambda}{2^N}I$ with a real number $\lambda \in [0, 1]$. In this vein, we randomly sample the error strength for state preparation $\mathcal{E}_{\text{sp}} = \mathcal{E}_{\text{dep}}(\rho; \lambda_{\text{sp}})$ and measurements $\mathcal{E}_m = \mathcal{E}_{\text{dep}}(\rho; \lambda_m)$ in terms of a given accumulate error rate $\lambda_1 = \lambda_m + \lambda_{\text{sp}}$. In addition, we also introduce the coherent error with the unitary channel $\mathcal{E}_{\text{uni}}(\rho) = U\rho U^\dagger$ by adding a rotation shift $\Delta\theta = \theta' - \theta_0$ on a rotating gate $R_{i \in [x,y,z]}(\theta_0) \rightarrow R_{i \in [x,y,z]}(\theta')$ in SPAM, where we uniformly sample the deviation $\Delta\theta/\pi \in [-\lambda_2, \lambda_2]$ ($\lambda_2 \leq [0, 1]$). Therefore, the total SPAM error is composed of coherent and incoherent errors as $\lambda_{\text{tot}} = (\lambda_1 + \lambda_2)/2$, fulfilling the condition $\lambda_{\text{tot}} \in [0, 1]$.

We then carry out the numerical experiment in the following three steps: (i) Set the N -qubit quantum channel to the identity $\mathcal{G} = I$ with the idling process, performing standard QPT to obtain the error matrix $\tilde{\chi}^I$. (ii) Reconstruct 4^N noisy quantum states and 6^N observables according to Eqs. (10–11). (iii) Finally, perform standard QPT on a randomly selected unitary gate \mathcal{G} using the error-mitigated probes, resulting in a corrected process matrix $\tilde{\chi}_e^3$ according to Eq.(7). Here, we mainly focus on the unitary operation, however, without loss of generality, the EM-QPT approach is also valid for general CPTP processes [44]. In Fig. (2), we present numerical results for standard QPT, EM-QPT and ML-QPT under the influence of both incoherent and coherent errors. Each data point represents the average process infidelity computed over 10^2 randomly chosen unitary gates. In particular, we investigate how the infidelity varies with the total error strength in the case of evenly mixed errors, i.e., $\lambda_1 = \lambda_2$ (dotted line in (a)), and show the results for one- and two-qubit unitary gates in panels (b) and (c), respectively. To implement ML-QPT, we collected 10^3 error matrices of each datapoint to train a digital twin model across a range of total error strengths. The resulting average infidelities over 10^2 random gates, obtained via ML-QPT, are also shown in panels (b) and (c).

Experiment: (1) Single-qubit Clifford gates. Next, we verify EM-QPT of single-qubit Clifford gates with experiments on superconducting qubit gates. The average gate fidelity is estimated by [51–53]:

$$\mathcal{F}_{\text{gate}} = \frac{d\mathcal{F}_\chi + 1}{d + 1}, \quad (14)$$

where $d = 2^N$ is the Hilbert space dimension of an N -qubit system, and the process fidelity \mathcal{F}_χ is given by

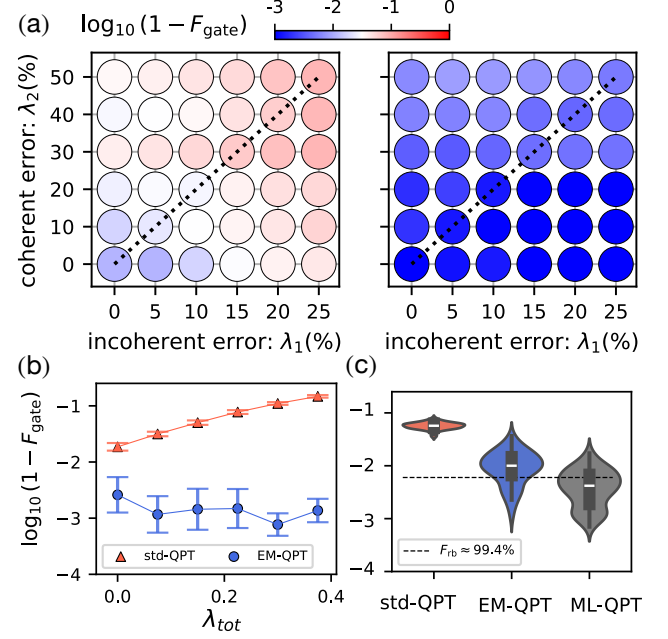


FIG. 3. The experimental results for multi-qubit gates. The standard QPT and EM-QPT as a function of coherent and incoherent errors in the left and right subplots of (a), respectively. The dashed line is explicitly illustrated in (b). Each data point in (a) and (b) is respectively averaged over 15 and 10^2 QPT for all 24 Clifford-typed gates, h. In (c), we compare the infidelity distribution over 10^2 QPT experiments of two-qubit CZ gate estimated by means of the std-QPT, EM-QPT, and ML-QPT, while the SPAM errors $\sim 6\%$ and the RB fidelity is 99.4% .

Eq. (8), as obtained through QPT. In the small error regime ($\lambda_{\text{tot}} \rightarrow 0$), the randomized benchmarking (RB) technique [54] statistically captures the average gate error over Clifford gates, implying that the RB fidelity approximates the average gate fidelity $F_{\text{rb}} \approx F_{\text{gate}}$ in small error limits. In experiments, we implement our method on 24 single-qubit Clifford gates on a superconducting quantum processor [44]. Coherent errors are introduced by adding rotation uncertainties to the single-qubit gates, modeled as amplitude fluctuations in the cosine-shaped DRAG pulses [55]. Specifically, the amplitude offset δ is uniformly sampled as $\delta \sim \mathcal{U}(0, \lambda_1)$, resulting in a modified amplitude $A' = (1 + \delta)A_0$, where $\lambda_2 \in [0, 1]$. Incoherent errors are introduced by biasing the amplitude of reset and readout pulses weighted by a bias percentage λ_1 (%). We refer the reader to the SM (appendix 4) [44] for detailed description about the error setup in experiments. In Fig. (3).(a-b), we present experimental results for std-QPT and EM-QPT for single-qubit Clifford gates under varying levels of coherent and incoherent SPAM errors, achieving at least an order-of-magnitude improvement in gate fidelity.

(2) Two-qubit CZ gates. We start with a well-tuned adiabatic CZ gate, achieving RB fidelity $\mathcal{F}_{\text{rb}}^{\text{CZ}} \sim 99.4\%$

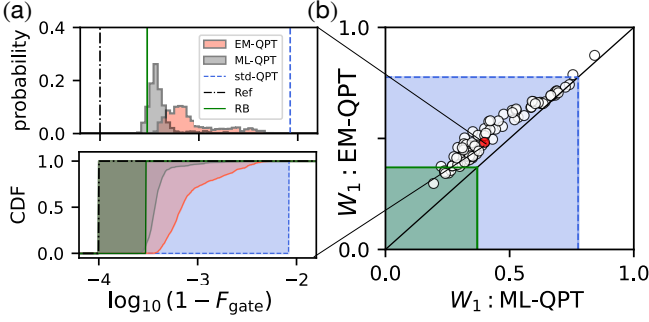


FIG. 4. **Precision and sensitivity.** As an example in (a), we illustrate the fidelity distribution of EM-QPT and ML-QPT relying on the training dataset and digital twin compared with std-QPT, reference, and RB infidelity. Accordingly, the W_1 distance from the reference distribution is calculated based on the area in CDF representation. In (b), the normalized W_1 distance of EM-/ML-QPT for 10^2 QPT experiments of X gates, where the blue and green areas represent the W_1 distance less than std-QPT and RB protocol, respectively. Note that the dot-dashed black, solid green, and dashed blue lines represent the reference infidelity ($10^{-\xi_0}$), RB infidelity (3×10^{-4}), and averaged infidelity (10^{-2}) of std-QPT over 10^2 samples, respectively. Parameters in (a-b): $\xi_{\max} = 5$, $\xi_{\min} = 1$, $\xi_0 = 4$.

(see setup in [56]). In our experiments, we performed QPT of the CZ gate 10^2 times. Prior to each CZ QPT measurement, an identity QPT was conducted. Consequently, one can obtain the fidelity of the CZ gate based on both EM-QPT and ML-QPT. In Fig. (3).(c), we present the probability distribution of gate infidelity estimated by Eq. (14) using standard QPT, EM-QPT and ML-QPT, where the ML-QPT results based on digital twin trained only on the 10^2 error matrix.

Precision and sensitivity. In the EM-QPT method, to achieve reliable gate fidelity, one needs to measure a set of error matrices estimated from N_{err} QPT experiments of identity processes, i.e., $\mathcal{S}_{\text{idn}} = \{\tilde{\chi}_1^I, \tilde{\chi}_2^I, \dots, \tilde{\chi}_{(N_{\text{err}})}^I\}$. For an unknown gate G , the gate fidelity can be statistically estimated over the prior set \mathcal{S}_{idn} , forming a probability distribution $\mathbb{P}_{\text{EM}}(\mathcal{F}(G; \tilde{\chi}^I))|_{\tilde{\chi}^I \in \mathcal{S}_{\text{idn}}}$, where the variance is primarily induced by SPAM error. However, the fidelity averaged over the whole set of error matrices may weaken the precision and stability of the EM-QPT, since some anomaly errors may occur in experiments. To better characterize the SPAM error, we here utilize a machine learning technique to learn the major pattern of SPAM errors, resulting in constructing the digital twin of error metrics. It is expected that the digital twin is capable of reconstructing the error metrics by learning the most significant features among the training data. Therefore, one can estimate the gate fidelity relying on the digital twin of error matrix by a well-trained neural networks, the corresponding probability distribution is $\mathbb{P}_{\text{ML}}(\mathcal{F})|_{\mathcal{D}^*(\mathbf{z}) \rightarrow \chi^I}$, where $\mathcal{D}^*(\mathbf{z})|_{\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})}$ is the decoder

from the trained VAE model.

We next evaluate the performance of our method through *Empirical Information Theory* [57, 58], which focuses on the behavior of information measures in practical, finite-sample settings. To quantify the precision and sensitivity of gate characterization among different strategies, we calculate the distance between the empirical $\mathbb{P}_{(i)}(q)$ and reference probability distribution (PD) $\mathbb{P}_{\text{ref}}(q)$ by the one-dimensional Wasserstein (W_1) distance (also known as Earth Mover’s Distance) [59]:

$$W_1(\mathbb{P}_{(i)}, \mathbb{P}_{\text{ref}}) = \int_{\mathbb{R}} |\mathcal{C}_{(i)}(q) - \mathcal{C}_{\text{ref}}(q)| dq, \quad (15)$$

where $\mathcal{C}_P(q') = \text{Prob}[q \leq q']$ denotes the cumulative distribution function (CDF) of the probability distribution P , representing the probability that the random variable q takes on a value less than or equal to q' . In this vein, the W_1 distance is the area between the CDF curve of P and the reference; see the lower panel in Fig.4.(c). The Wasserstein distance measures the “cost” of transporting probability “mass” from one distribution to another. While it directly captures first-moment deviations, it also provides an informative proxy for the second moment (i.e., the spread or variance), especially when compared against a delta distribution.

In our case, we take the logarithm of infidelity as the basis $q = \log_{10}(1 - \mathcal{F}) \in [-\xi_{\max}, -\xi_{\min}]$ bounded by real and positive constant ξ_{\max}/ξ_{\min} . For the sake of simplicity, we set the reference PD as a delta function $\mathbb{P}_{\text{ref}} = \Delta(-\xi_0)$ at $q = -\xi_0$ referring to the ideal measurement protocol that always perfectly estimates the gate fidelity $\mathcal{F}_0 = 1 - 10^{-\xi_0}$ fulfilling the condition $\xi_0 \in [\xi_{\min}, \xi_{\max}]$. In this vein, the W_1 distance is bounded by $W_1 \in [0, \xi_{\max} - \xi_{\min}]$, where $W_1 = 0$ indicate the ideal measurement scheme that gate fidelity can be perfectly revealed. Otherwise, the larger W_1 distance is the one that far away from the perfect characterization. We interpret that this geometric metric quantifies the precision and sensitivity of gate fidelity estimation, which can be used to benchmark the performance of various QPT techniques.

In Fig. 4.(a-b), we compare the W_1 distance of the ML-QPT and EM-QPT methods, where the digital twin used in ML-QPT is trained on 10^3 identity matrices. For 10^2 QPT experiments of X gate, we apply the EM-QPT and ML-QPT to estimate the gate fidelity relying on the 10^3 realistic error matrices and corresponding digital twin, respectively. In comparison, we take the distribution of std-QPT and RB methods as two delta functions, since the uncertainty for an individual QPT is negligible. As an example, in Fig. 4.(a), the probability distribution of EM-QPT contains all types of errors in error matrices, implying large variance (distribution in pink) in fidelity estimation in the presence of abnormal errors. Consequently, the ML-QPT method yields more reliable fidelity

estimates, leading to a smaller W_1 distance compared to EM-QPT in panel (b).

Discussion and conclusion. To faithfully estimate the gate fidelity in the presence of SPAM errors, a commonly used technique is *randomized benchmarking* (RB), which statistically estimates the average gate error over a group of Clifford gates in terms of quantum twirling [11, 54]. However, this method lacks gate-specific information and assumes sufficiently small coherent errors, which limits its applicability in practical quantum computing. An alternative approach for SPAM-insensitive gate characterization is gate set tomography (GST), which constructs a gauge-invariant description by fitting a large set of experimental data [8]. While GST can capture more detailed noise properties, it is computationally expensive and sensitive to time-dependent drift.

In this letter, we have proposed SPAM-error mitigated quantum process tomography (EM-QPT) by constructing nearly self-consistent probes based on identity process matrices. Moreover, the machine-learning-assisted QPT (ML-QPT) enhances EM-QPT by fully leveraging knowledge of SPAM error hidden in identity process matrices, enabling reliable and practical applications across a variety of quantum devices. To benchmark the performance, we introduce a metric, the one-dimensional Wasserstein (W_1) distance, to quantify the precision and sensitivity among different gate characterization protocols. The resulting SPAM-aware digital twin improves gate characterization beyond standard methods, allowing accurate fidelity estimation up to the second moment. Both numerical simulations and experimental results demonstrate that our method achieves at least an order-of-magnitude improvement in precision over standard QPT.

The following extension of our method is to diagnose the type of SPAM error; for example, the particular behavior of coherent and incoherent errors can be classified, providing a useful reference for experimental design. The digital twin of SPAM error can also serve as a sensitive sensor to detect anomalies in realistic experiments [38]. More broadly, our approach is generic, which can be directly extended to N -qubit ($N > 2$) quantum processes [7] and continuous-variable quantum computing. Furthermore, our EM-QPT (ML-QPT) protocol holds greater potential for effectiveness in the context of ancilla-assisted QPT, where input states are typically more complex—often entangled—and the measurement schemes involve intricate global measurements with complex unitary operations [60, 61]. This scenario warrants further investigation to fully realize the advantages of our approach.

Acknowledgements. This work has received funding from the EU Flagship on Quantum Technology through HORIZON-CL4-2022-QUANTUM-01-SGA Project No. 101113946 OpenSuperQPlus100. T.Y.H acknowledges financial support by the Knut and Alice Wallenberg

through the Wallenberg Center for Quantum Technology (WACQT). The work by the Aalto researchers has been done under the Academy of Finland Centre of Excellence program (Project No. 352925). The Aalto team acknowledges the provision of facilities and technical support by the Aalto University at the national research infrastructure OtaNano - Low Temperature Laboratory.

* akshayga@chalmers.se

- [1] H. Aghaee Rad and et al, *Nature* (2025), [10.1038/s41586-024-08406-9](https://doi.org/10.1038/s41586-024-08406-9).
- [2] D. Bluvstein and et al, *Nature* **626**, 58 (2024).
- [3] S. A. Moses and et al, *Phys. Rev. X* **13**, 041052 (2023).
- [4] F. Arute and et al, *Nature* **574**, 505 (2019).
- [5] P. Cerfontaine, R. Otten, and H. Bluhm, *Phys. Rev. Appl.* **13**, 044071 (2020).
- [6] T. Abad, J. Fernández-Pendás, A. Frisk Kockum, and G. Johansson, *Phys. Rev. Lett.* **129**, 150504 (2022).
- [7] C. W. Warren, J. Fernández-Pendás, S. Ahmed, T. Abad, A. Bengtsson, J. Biznárová, K. Debnath, X. Gu, C. Krizan, A. Osman, A. Fadavi Roudsari, P. Delsing, G. Johansson, A. Frisk Kockum, G. Tancredi, and J. Bylander, *npj Quantum Information* **9**, 44 (2023).
- [8] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, *Quantum* **5**, 557 (2021).
- [9] R. Brieger, I. Roth, and M. Kliesch, *PRX Quantum* **4**, 010325 (2023).
- [10] C. A. Riofrio, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, *Nature Communications* **8**, 15305 (2017).
- [11] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *Phys. Rev. Lett.* **119**, 130502 (2017).
- [12] T. Proctor, K. Young, A. D. Baczewski, and R. Blume-Kohout, *Nature Reviews Physics* **7**, 105 (2025).
- [13] R. Blume-Kohout, T. Proctor, and K. Young, “Quantum characterization, verification, and validation,” (2025), [arXiv:2503.16383 \[quant-ph\]](https://arxiv.org/abs/2503.16383).
- [14] A. Hashim, L. B. Nguyen, N. Goss, B. Marinelli, R. K. Naik, T. Chistolini, J. Hines, J. P. Marceaux, Y. Kim, P. Gokhale, T. Tomesh, S. Chen, L. Jiang, S. Ferracin, K. Rudinger, T. Proctor, K. C. Young, R. Blume-Kohout, and I. Siddiqi, “A practical introduction to benchmarking and characterization of quantum computers,” (2024), [arXiv:2408.12064 \[quant-ph\]](https://arxiv.org/abs/2408.12064).
- [15] I. L. Chuang and M. A. Nielsen, *Journal of Modern Optics* **44**, 2455 (1997).
- [16] J. L. O’Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, *Physical Review Letters* **93**, 080502 (2004).
- [17] J. F. Poyatos, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **78**, 390 (1997).
- [18] K. Kraus, A. Bohm, J. Dollard, and W. Wootters, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag Berlin Heidelberg, 1983).
- [19] C. H. Baldwin, A. Kalev, and I. H. Deutsch, *Phys. Rev. A* **90**, 012110 (2014).
- [20] A. V. Rodionov, A. Veitia, R. Barends, J. Kelly, D. Sank, J. Wenner, J. M. Martinis, R. L. Kosut, and A. N. Ko-

- rotkov, *Phys. Rev. B* **90**, 144504 (2014).
- [21] Y. Kim, Y.-S. Kim, S.-Y. Lee, S.-W. Han, S. Moon, Y.-H. Kim, and Y.-W. Cho, *Nature Communications* **9**, 192 (2018).
- [22] A. Gaikwad, K. Shende, Arvind, and K. Dorai, *Scientific Reports* **12**, 3688 (2022).
- [23] R. Blume-Kohout and T. Proctor, “Easy better quantum process tomography,” (2024), [arXiv:2412.16293 \[quant-ph\]](https://arxiv.org/abs/2412.16293).
- [24] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, *Nature Communications* **8**, 14485 (2017).
- [25] L. Rudnicki, Z. Puchała, and K. Życzkowski, *Quantum* **2**, 60 (2018).
- [26] J. Lin, B. Buonacorsi, R. Laflamme, and J. J. Wallman, *New Journal of Physics* **21**, 023006 (2019).
- [27] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, *Phys. Rev. A* **87**, 062119 (2013).
- [28] C. Stark, *Phys. Rev. A* **89**, 052109 (2014).
- [29] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, “Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit,” (2013), [arXiv:1310.4492 \[quant-ph\]](https://arxiv.org/abs/1310.4492).
- [30] N. Quesada, A. M. Brańczyk, and D. F. James, in *The Rochester Conferences on Coherence and Quantum Optics and the Quantum Information and Measurement meeting* (Optica Publishing Group, 2013) p. W6.38.
- [31] D. Greenbaum, “Introduction to quantum gate set tomography,” (2015), [arXiv:1509.02921 \[quant-ph\]](https://arxiv.org/abs/1509.02921).
- [32] Z.-T. Li, C.-C. Zheng, F.-X. Meng, H. Zeng, T. Luan, Z.-C. Zhang, and X.-T. Yu, *Quantum Science and Technology* **9**, 035027 (2024).
- [33] S. Cao, D. Lall, M. Bakr, G. Campanaro, S. D. Fasciati, J. Wills, V. Chidambaram, B. Shteynas, I. Rungger, and P. J. Leek, *Phys. Rev. Lett.* **133**, 120802 (2024).
- [34] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, *Quantum Science and Technology* **5**, 044002 (2020).
- [35] P. Viñas and A. Bermudez, *npj Quantum Information* **11**, 23 (2025).
- [36] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” (2022), [arXiv:1312.6114 \[stat.ML\]](https://arxiv.org/abs/1312.6114).
- [37] T. Huang, Y. Ban, E. Y. Sherman, and X. Chen, *Phys. Rev. Appl.* **17**, 024040 (2022).
- [38] T. Huang, Z. Yu, Z. Ni, X. Zhou, and X. Li, *Communications Physics* **7**, 172 (2024).
- [39] S. Diamond and S. Boyd, *Journal of Machine Learning Research* **17**, 1 (2016).
- [40] R. Jozsa, *Journal of Modern Optics* **41**, 2315 (1994).
- [41] A. G. Kofman and A. N. Korotkov, *Phys. Rev. A* **80**, 042103 (2009).
- [42] A. N. Korotkov, “Error matrices in quantum process tomography,” (2013), [arXiv:1309.6405 \[quant-ph\]](https://arxiv.org/abs/1309.6405).
- [43] X. Y. Han, T. Q. Cai, X. G. Li, Y. K. Wu, Y. W. Ma, Y. L. Ma, J. H. Wang, H. Y. Zhang, Y. P. Song, and L. M. Duan, *Phys. Rev. A* **102**, 022619 (2020).
- [44] “Supplemental material,” (2025).
- [45] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O’Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, *Phys. Rev. Lett.* **112**, 240504 (2014).
- [46] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, A. R. Carvalho, and M. J. Biercuk, *PRX Quantum* **2**, 040324 (2021).
- [47] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, edited by L. Rokach, O. Maimon, and E. Shmueli (Springer International Publishing, Cham, 2023) pp. 353–374.
- [48] D. P. Kingma, T. Salimans, and M. Welling, in *Advances in Neural Information Processing Systems*, Vol. 28, edited by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Curran Associates, Inc., 2015).
- [49] K. Banaszek, G. M. D’Ariano, M. G. A. Paris, and M. F. Sacchi, *Phys. Rev. A* **61**, 010304 (1999).
- [50] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, *Phys. Rev. Lett.* **127**, 140502 (2021).
- [51] M. Horodecki, P. Horodecki, and R. Horodecki, *Phys. Rev. A* **60**, 1888 (1999).
- [52] M. A. Nielsen, *Physics Letters A* **303**, 249 (2002).
- [53] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. Lett.* **102**, 090502 (2009).
- [54] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [55] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, *Phys. Rev. Lett.* **103**, 110501 (2009).
- [56] A. Aggarwal, J. Fernández-Pendás, T. Abad, D. Shiri, H. Jakobsson, M. Rommel, A. Nylander, E. Hogedal, A. Osman, J. Biznárová, R. Rehammar, M. F. Giannelli, A. F. Roudsari, J. Bylander, and G. Tancredi, “Mitigating transients in flux-control signals in a superconducting quantum processor,” (2025), [arXiv:2503.08645 \[quant-ph\]](https://arxiv.org/abs/2503.08645).
- [57] E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- [58] A. J. Majda and B. Gershgorin, *Proceedings of the National Academy of Sciences* **108**, 10044 (2011).
- [59] V. M. Panaretos and Y. Zemel, *Annual review of statistics and its application* **6**, 405 (2019).
- [60] S. Xue, Y. Wang, J. Zhan, Y. Wang, R. Zeng, J. Ding, W. Shi, Y. Liu, Y. Liu, A. Huang, G. Huang, C. Yu, D. Wang, X. Fu, X. Qiang, P. Xu, M. Deng, X. Yang, and J. Wu, *Phys. Rev. Lett.* **129**, 133601 (2022).
- [61] A. Patel, A. Gaikwad, T. Huang, A. F. Kockum, and T. Abad, “Selective and efficient quantum state tomography for multi-qubit systems,” (2025), [arXiv:2503.20979 \[quant-ph\]](https://arxiv.org/abs/2503.20979).
- [62] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016) <http://www.deeplearningbook.org>.
- [63] D. M. Bashtannyk and R. J. Hyndman, *Computational Statistics & Data Analysis* **36**, 279 (2001).
- [64] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” (2019), [arXiv:1912.01703 \[cs.LG\]](https://arxiv.org/abs/1912.01703).
- [65] A. Bengtsson, P. Vikstål, C. Warren, M. Svensson, X. Gu, A. F. Kockum, P. Krantz, C. Križan, D. Shiri, I.-M. Svensson, G. Tancredi, G. Johansson, P. Delsing, G. Ferrini, and J. Bylander, *Phys. Rev. Appl.* **14**, 034010

- (2020).
- [66] Z. Chen, “Metrology of quantum control and measurement in superconducting qubits,” (2018).
 - [67] Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O’Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and J. M. Martinis, *Phys. Rev. Lett.* **116**, 020501 (2016).
 - [68] K. V. Rosales, “Derivative removal by adiabatic gate (drag) and ac stark-shift calibration,” (2025), [qua-platform/Qua-lib/Quantum-Control-Applications/Superconducting/Single-Flux-Tunable-Transmon/Use Case 2 - DRAG coefficient calibration](#).
 - [69] K. V. Rosales, “Optimized readout with optimal weights qua-lib,” (2024), [qua-platform/Qua-lib/Quantum-Control-Applications/Superconducting/Single-Fixed-Transmon/Use Case 2 - Optimized readout with optimal weights](#).

Supplemental Material:

xxxxxxx

For the sake of simplicity and clarity in our derivations, we introduce the Pauli-Liouville representation as a foundational framework. The Pauli set reads:

$$\mathcal{P} := \left\{ \bigotimes_{v=1}^N E_v : E_v \in \{I, \sigma_x, \sigma_y, \sigma_z\} \right\}, \quad (16)$$

where I represents the identity matrix, and σ_x, σ_y and σ_z are the standard Pauli matrices. Within this framework, the density matrix ρ and a quantum channel $\mathcal{E}(\rho)$ can be expressed as:

$$\rho = \sum_{j=1}^{4^N} a_j E_j, \quad \mathcal{E}(\rho) = \sum_{\alpha, \beta=1}^{4^N} \chi_{\alpha\beta} E_\alpha \rho E_\beta^\dagger, \quad (17)$$

where $E_j \in \mathcal{P}$, and the coefficients a_j are given by $a_j = \text{Tr}(E_j^\dagger \rho) / 2^N$. To facilitate further analysis, we vectorize the density operator as:

$$\text{vec}(\rho) = |\rho\rangle\rangle = (\rho_1, \rho_2, \dots, \rho_{4^N})^T, \quad \mathcal{E}(\rho) = \chi \circ |\rho\rangle\rangle. \quad (18)$$

Here, the symbol $A \circ B$ denotes the *completely positive trace-preserving* (CPTP) mapping of A acting on B . Additionally, we note the following trace property:

$$\text{Tr}[\hat{O}\rho] = \text{vec}(\hat{O}^T) |\rho\rangle\rangle = \langle\langle \hat{O} | \rho \rangle\rangle. \quad (19)$$

Appendix A: Estimation of Noisy Probes

The objective of this section is to estimate the noisy probes, including both quantum states and measurements, based on the process tomography of the idling process. Specifically, we aim to perform state and detector tomography by leveraging the error matrix.

To achieve this, we define a loss function that incorporates the state ρ , measurement M , and process matrix χ :

$$J(M, \rho, \chi) = \|\mathbf{A}|\chi\rangle\rangle - \langle\langle M | \rho \rangle\rangle\|, \quad (20)$$

where \mathbf{A} is a prior known matrix defined as:

$$\mathbf{A} = \sum_{m,n,i,\mu} \text{Tr}(E_m \rho_i^{\text{ideal}} E_n^\dagger M_\mu^{\text{ideal}}), \quad (21)$$

where the Pauli basis $E_{m,n}$ is defined in Eq.(16). This loss function in (20) quantifies the discrepancy between the ideal readout and the experimental outcomes obtained for given initial states and measurement setups. Given the process matrix χ_{err} obtained from idle process tomography, the noisy initial states can be estimated by solving the following optimization problem:

$$\begin{aligned} \bar{\rho}_i &= \underset{\rho}{\text{argmin}} \sum_{\mu} J(M_\mu, \rho_i, \chi_{\text{err}}) \\ \text{subject to } & \text{Tr}(\rho) = 1, \quad \rho^\dagger = \rho, \quad \rho \geq 0. \end{aligned} \quad (22)$$

Here, we assume that the measurement probes $\{M_\mu\}$ are perfect, and the optimized states satisfy the trace condition ($\text{Tr}(\rho) = 1$), Hermiticity ($\rho^\dagger = \rho$), and positive semi-definiteness ($\rho \geq 0$). Similarly, the noisy measurement probes can be estimated by:

$$\begin{aligned} \{\bar{M}_\mu\} &= \underset{\{M_\mu\}}{\text{argmin}} \sum_i J(\{M_\mu\}, \rho_i, \chi_{\text{err}}) \\ \text{subject to } & \sum_{\mu} M_\mu = \mathbb{I}, \quad M_\mu \geq 0, \end{aligned} \quad (23)$$

where the initial states are assumed to be perfect, and the estimated positive operator-valued measures (POVMs) satisfy the properties of positivity and completeness. These optimizations are performed using the Python-based library `cvxpy` [39].

1. The EM-QPT on an arbitrary N -qubit quantum channel

In this section, we shall investigate the performance of our method for a general CPTP quantum channel, which can be expressed on Pauli basis. The Kraus operators $\{\mathcal{K}^{(i)}\}$ can be expressed in the Pauli basis $\{E_k\}$ (defined in Eq. (16)) as

$$\mathcal{K}^{(i)} = \sum_k c_k^{(i)} E_k \quad (24)$$

where $c_k^{(i)} \in \mathbb{C}$ are complex coefficients. For an N -qubit system, any CPTP map can be constructed using up to $r \leq 4^N$ Kraus operators:

$$\mathcal{E}_{\text{rand}}(\rho) = \sum_{i=1}^{r \leq 4^N} \mathcal{K}^{(i)} \rho (\mathcal{K}^{(i)})^\dagger, \quad (25)$$

with the trace-preserving condition requiring $\sum_{i=1}^r (\mathcal{K}^{(i)})^\dagger \mathcal{K}^{(i)} = \mathbb{I}$. To numerically generate valid CPTP maps, we first sample random complex coefficients $\{c_k^{(i)}\} \in \mathbb{C}^{4^N \times r}$ to form unnormalized Kraus operators $\{\mathcal{K}'^{(i)} = \sum_k c_k^{(i)} E_k\}_{i=1}^r$, then normalize them to satisfy the TP condition:

$$\mathcal{K}^{(i)} = \frac{\mathcal{K}'^{(i)}}{\sqrt{\sum_{i=1}^r \mathcal{K}'^{(i)} (\mathcal{K}'^{(i)})^\dagger}}. \quad (26)$$

The resulting process matrix χ , defined through

$$\chi_{\alpha\beta} = \sum_i c_\alpha^{(i)} (c_\beta^{(i)})^*, \quad (27)$$

is guaranteed to be Hermitian and completely positive by construction. In this vein, we insert the quantum operation $\mathcal{E}_{\text{rand}}(\cdot)$ in the quantum circuit by using the function `Kraus({K})` in `qiskit` platform. To investigate

the EM-QPT for a general CPTP process, we randomly sample the Kraus number $r \in [1, 4^N]$ for N -qubit quantum channel. The process fidelity as a function of total SPAM error are presented in Fig. 5 for one- and two-qubit channels, respectively.

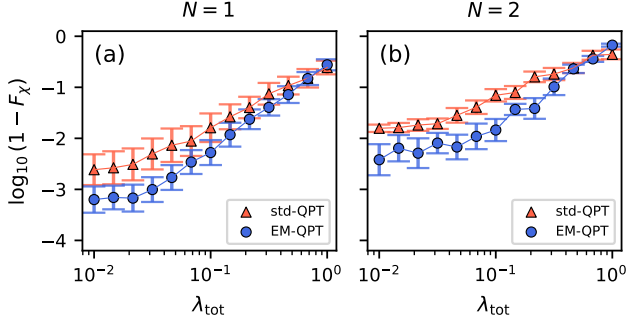


FIG. 5. **The EM-QPT of general CPTP process for one-/two- qubit quantum process.** The logarithm of infidelity versus the total SPAM error for a single- and two-qubit CPTP process with randomly sampled Kraus number. The orange triangle and blue circle refer to the result averaged over 10^2 random CPTP with one-sigma error bar.

Appendix B: Digital Twinning of the Error Matrix

1. The Generative Model: Variational Autoencoder

A Variational Autoencoder (VAE) [36] is a generative model that combines deep learning with probabilistic frameworks to learn a latent representation of data. Unlike traditional autoencoders [62], a VAE models the latent space as a probability distribution rather than fixed embeddings. Specifically, a VAE consists of an *encoder* that maps the input \mathbf{x} to a latent distribution $\mathbb{Q}(\mathbf{z}|\mathbf{x}) \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are learned through separate fully connected layers. The *decoder* reconstructs the input data \mathbf{x}' from a sampled latent vector \mathbf{z} .

To enable backpropagation through the network, the latent variable \mathbf{z} is sampled using the reparameterization trick [48]:

$$\mathbf{z} = \mu(\mathbf{x}) + \epsilon \cdot \sigma(\mathbf{x}), \quad (28)$$

where $\epsilon \sim \mathcal{N}(0, 1)$ is an auxiliary noise variable sampled from a standard normal distribution. The total loss function of the VAE consists of two components: a reconstruction loss and a Kullback-Leibler (KL) divergence loss:

$$\text{Loss} = \mathcal{L}_{\text{recon}}(\mathbf{x}', \mathbf{x}) + \beta D_{\text{KL}}[\mathbb{Q}(\mathbf{z}|\mathbf{x}) \parallel \mathbb{P}(\mathbf{z})], \quad (29)$$

where $\mathcal{L}_{\text{recon}} = \sum_{n,m} (\mathbf{x}_{n,m} - \mathbf{x}'_{n,m})^2$ ensures that the reconstructed data \mathbf{x}' closely matches the original input \mathbf{x} , and D_{KL} regularizes the learned posterior $\mathbb{Q}(\mathbf{z}|\mathbf{x})$ to

match the prior $\mathbb{P}(\mathbf{z}) = \mathcal{N}(0, \mathbb{I})$. The hyperparameter β balances the two loss terms. Since both the learned posterior and the target prior are Gaussian distributions, the KL divergence has a closed-form solution [36]:

$$D_{\text{KL}}[\mathbb{Q}(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(0, \mathbb{I})] = \frac{1}{2} (\sigma^2 + \mu^2 - 1 - \log \sigma^2). \quad (30)$$

2. Unsupervised Learning of Error Matrices

The goal of this section is to learn the state preparation and measurement (SPAM) errors by training a generative model on realistic experimental data. We hypothesize that a deep neural network (NN)-based model can statistically generate the idle process matrix, capturing the features of SPAM errors. This NN serves as a *digital twin* [38] of the SPAM error.

To achieve this, we first construct a training database $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=0}^{N_x}$, consisting of N_x independent quantum process tomography (QPT) experiments for the idling quantum process. For an N -qubit quantum channel, the error matrix is a complex matrix $\chi_i \in \mathbb{C}^{4^N \times 4^N}$. Each χ -matrix is decomposed into real and imaginary parts, forming the input training data points $\mathbf{x}_i \in \mathbb{R}^{2 \times 4^N \times 4^N}$.

In this framework, the digital twin of the error matrix, $\mathbf{x}'_i \rightarrow \chi'_i$, is expected to statistically mitigate SPAM errors for an arbitrary quantum process $\mathcal{E}(\cdot)$. The neural network learns the underlying features of SPAM errors embedded in the error matrix, enabling enhanced QPT with the assistance of the digital twin.

The *encoder* extracts key features from the input data \mathbf{x}_i using a series of convolutional layers [62], progressively reducing the spatial dimensions while increasing the number of channels. At the end of this process, the high-dimensional input is transformed into a low-dimensional feature representation. Instead of encoding the input into a fixed lower-dimensional representation $\mathbf{x} \rightarrow \mathbf{z}$, the VAE learns a probabilistic distribution characterized by two variational parameters $\{\mu, \sigma^2\}$, such that $\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2)$.

QProcess layer. To ensure that the VAE output satisfies the CPTP constraint, we introduce an additional layer called **QProcess**, which relies on the Cholesky decomposition [49]. Cholesky decomposition is a powerful mathematical tool that plays a significant role in quantum information science, particularly in Quantum Process Tomography (QPT) [50]. This layer ensures that the decoded output forms a valid positive definite matrix using Cholesky decomposition: $\mathbf{A} = \mathbf{L}\mathbf{L}^T$, where \mathbf{L} is a lower triangular matrix and \mathbf{L}^T is its transpose. To ensure numerical stability, the transformed output is $\mathbf{x} = \mathbf{L}\mathbf{L}^T + \epsilon\mathbf{I}$, where ϵ is a small positive number (e.g., 10^{-5}). This layer constructs a valid quantum process χ , which satisfies: (i). *Hermitian Property*: $\chi = \chi^\dagger$;

(ii). *Positive Semi-definiteness*: All eigenvalues $\lambda_i \geq 0$ and (ii) *Trace preserving*: $\text{Tr}(\chi \circ |\rho\rangle\rangle) = 1$.

As a result, the forward data flow of the model is as follows:

$$\mathbf{x} \rightarrow \text{Encoder} \rightarrow \mathbf{z} \rightarrow \text{Decoder} \rightarrow \text{QProcess} \rightarrow \mathbf{x}'. \quad (31)$$

For a well-trained model (denoted as $*$), there is a straightforward approach to evaluating the VAE model. As mentioned earlier, the goal of VAE is to map the training dataset to a Gaussian distribution that closely approximates the standard normal distribution. In this context, one can verify the distribution of the latent vectors by passing the training data through the trained encoder as

$$\mathbf{x} \rightarrow \text{Encoder}^* \rightarrow \mathbf{z}', \quad (32)$$

where, ideally, $\mathbf{z}' \sim \mathcal{N}(0, \mathbb{I})$.

In Fig. 10(a-b), we illustrate the evolution of the loss value during training and the distribution of the latent vectors \mathbf{z}' across the entire training dataset. The probability density function (PDF) of \mathbf{z}' is estimated using Gaussian kernel density estimation [63] and is depicted by the blue curves in Fig. 10(a-b). The results indicate that the two-dimensional latent vectors \mathbf{z}' produced by the trained encoder are close to a standard normal distribution.

Following this, digital twins can be automatically generated by sampling the latent space from a normal distribution:

$$\mathbf{z} \sim \mathcal{N}(0, \mathbb{I}) \rightarrow \text{Decoder}^* \rightarrow \text{QProcess}^* \rightarrow \mathbf{x}^*, \quad (33)$$

where \mathbf{x}^* denotes the digital twin of error matrix. We now present the VAE-generated digital twin of error matrices for two-qubit unitary gates with an error rate of $\lambda_{\text{tot}} = 0.01$. In Fig. 10(c-d), we compare the reconstructed data for the real (first channel) and imaginary components (second channel) of the error matrices with those sampled from the training dataset.

This approach enables the generation of error matrices that statistically capture the characteristics of SPAM errors, providing an efficient tool for constructing the error-mitigated quantum process tomography.

The whole training process is performed using the PyTorch [64] framework.

Appendix C: Experimental setups

In this section, we detail the device, experimental setup, and QPT protocol together with our implementation of the noise channels.

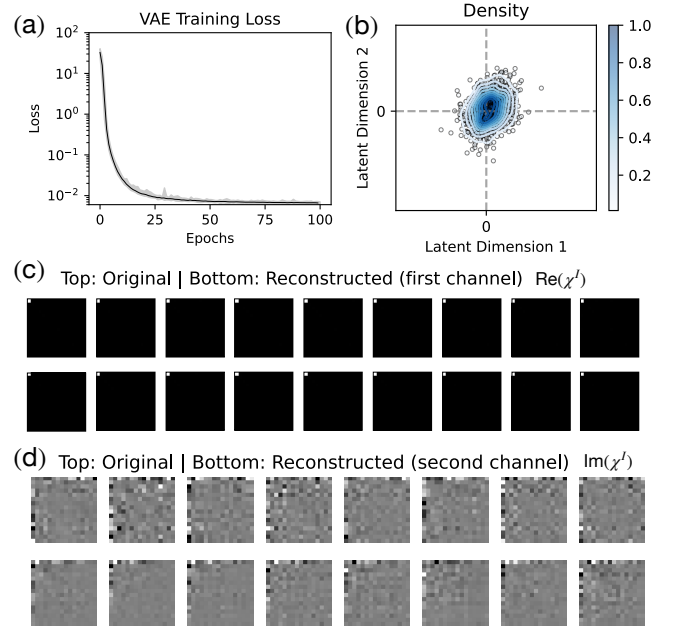


FIG. 6. **The training outcomes of the VAE model for generating digital twins of two-qubit error matrices.** (a) Evolution of the training loss over 100 epochs, showing a rapid initial decrease and eventual convergence, indicating successful training. (b) The distribution of the latent vectors \mathbf{z}' obtained by encoding the training dataset using the trained encoder. The probability density function (PDF) is estimated via Gaussian kernel density estimation (blue curves), demonstrating that the learned latent representation closely follows a standard normal distribution. (c) Comparison between original (top) and reconstructed (bottom) error matrices for the real component (first channel) of the two-qubit gate errors. The reconstructed matrices are generated by decoding latent vectors sampled from the trained encoder. (d) Similar comparison for the imaginary component (second channel) of the error matrices. The results indicate that the VAE successfully captures the statistical properties of the error matrices, enabling the generation of realistic digital twins for quantum process tomography and error mitigation. The total error rate is $\lambda_{\text{tot}} = 0.01$ and size of training datasets is 10^3 .

Appendix c1: Experimental setup

The parameters of the sample used in the experiment are listed in Table I. The sample qubit is a subsystem of a device comprising two fixed-frequency xmon-style transmon qubits capacitively coupled via a frequency-tunable anharmonic oscillator (the sample design is similar to the quantum processor studied in Ref. [65]). The two qubits are capacitively coupled to individual control lines and quarter-wavelength resonators for readout. The tunability of the coupler is provided by two Josephson junctions in a superconducting quantum interference device (SQUID) configuration. Our experiments were performed at the zero flux sweet spot where the coupler had the maximal frequency to suppress the interaction with

Parameter		Qubit
Qubit frequency,	$\omega_q/2\pi$ (GHz)	3.772
Qubit anharmonicity,	$\alpha_q/2\pi$ (MHz)	-260
Resonator frequency,	$\omega_r/2\pi$ (GHz)	6.439
Dispersive shift,	$\chi_r/2\pi$ (MHz)	0.350
Relaxation time,	T_1 (μ s)	131
Decoherence time,	T_2^* (μ s)	64
Spin Echo,	T_2^E (μ s)	141
Single-qubit gate fidelity,	F_{1Q} (%)	99.963
Two-state assignment fidelity,	$F_{Assign(0\rangle, 1\rangle)}$ (%)	95.65

TABLE I. Measured qubit parameters, coherence properties, and single-qubit performance.

the qubit under investigation.

All experiments were conducted in a Bluefors XLD400 dilution refrigerator maintaining a base temperature stabilized at ~ 15 mK. The room temperature setup together with the full wiring diagram in our dilution refrigerator are shown in Fig. 7. The drive pulses for the qubits and resonators are generated by a combination of Quantum Machines operator-X (OPX+) quantum controller and Quantum Machines Octave. The IF's (intermediate frequency) signals from the OPX+'s arbitrary waveform generators are fed to the built-in IQ mixers of the Octave and combined with signals from the internal local oscillators. Then the outputs of the Octave are fed into qubit drive line which is capacitively coupled to the transmon circuits or into the readout feedline capacitively coupled to the resonators on chip.

The transmitted readout signal propagates through the sample to a double junction circulator (4-8 GHz from Low Noise Factory), directional coupler (4-12 GHz) and a traveling wave parametric amplifier (TWPA) fabricated at VTT Technical Research Center of Finland. After TWPA amplification the signal transmits through the single junction circulator (4-8 GHz from QuinStar Technology), series of low-pass, infrared, and high pass filters, as well as additional double junction circulator (4-8 GHz from Low Noise Factory) before being amplified by around 40 dB at 4 K with a high-electron-mobility transistor (HEMT) from Low Noise Factory. In addition, the room-temperature microwave amplifier from Narda-MITEQ is connected to the readout chain outside of the dilution refrigerator. The pump signal and DC bias current for the TWPA are provided by Anritsu MG3692 signal generator and Yokogawa GS200, through a current limiting resistor. Circulators in the output line are essential to prevent strong signals from reflecting back into the sample. Finally, the output signal is down-converted inside the Octave to the hundreds of MHz range to be recorded and integrated by the OPX+ internal digitizer.

Appendix c2: Gate calibration

Calibrations of the fundamental transition frequencies and coherence times provided in Table I mainly follow well-established parameter estimation methods used for superconducting quantum computing [66]. Once the qubit frequency and coherence times are determined, we calibrate the amplitudes of $\pi/2$ ($\pm X90$ and $\pm Y90$) and π -pulses ($X180$ and $Y180$) for the later use in standard quantum process tomography (std-QPT) and randomized benchmarking (RB). Our single-qubit pulses use a cosine envelope with 40 ns duration combined with the Derivative Removal by Adiabatic Gate (DRAG) technique.

The goal of the calibration is to determine the correct amplitudes of the $\pi/2$ and π pulses, and then the corresponding DRAG coefficients. Amplitude tune up is performed via repeatedly executing the pseudo identity pulse sequences (such as two π -pulses or four $\pi/2$ -pulses) 'N' times and measuring the state of the qubit across different pulse amplitudes and number of pulses. For the DRAG calibration we utilized "Google" method [67]. Both protocols are implemented on the Quantum Machines OPX+ and Octave based on the open source libraries for pulse level control of quantum bits built with the QUA programming language. The programming code is similar to the example described in [68].

Appendix c3: Interleaved Randomized Benchmarking

The performance of the single-qubit gates is characterized with Clifford RB [54]. To implement this technique, we generate random sequences of single-qubit Clifford gates from the Table II and measure the state of the qubit afterwards. Each random sequence ends with the recovery gate that will bring the qubit back to its ground state. It should be noted here that the average number of single-qubit gates in Clifford operation is 1.875.

In case of the interleaved RB we are aiming to characterize the fidelity of a specific gate. Here, we use the gate of interest (I, X90, Y90, X180) to construct the pseudo identity gate (implemented as a 40 ns delay, four X90 gates, four Y90 gates or two X180 gates, respectively) which is interleaved between each Clifford gate in the random sequence. Experimental results are shown in Fig. 8. For each circuit depth, measurements are performed for a large number (50) of different random sequences. It is clearly seen that the average gate fidelity exceeds 99.96%.

Appendix c4: QPT protocol and noise channels

In our experiments we performed std-QPT as shown in Fig. 9(a) with minor modifications allowing us to introduce different noise channels as desired. We detail the

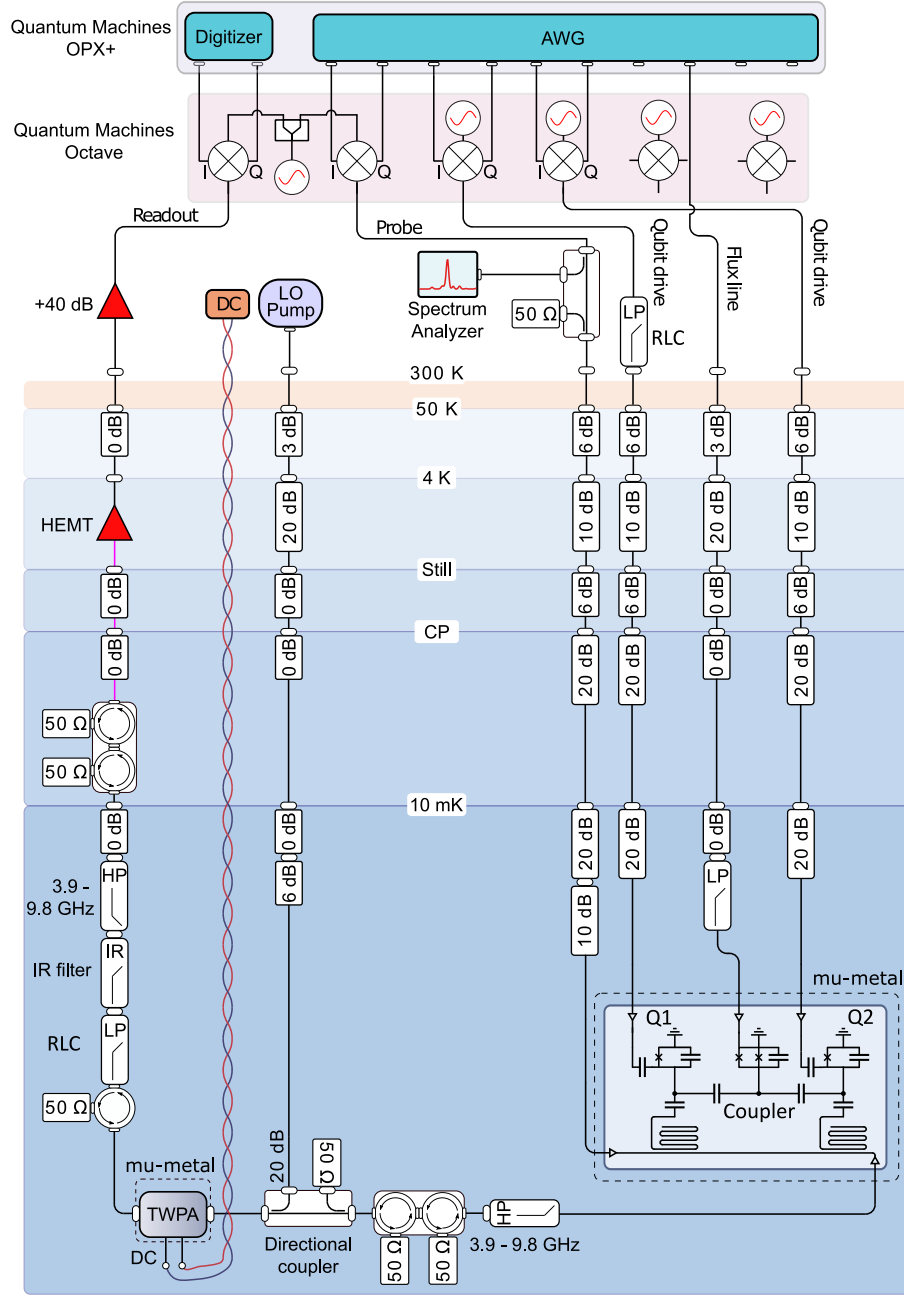


FIG. 7. Wiring schematic of the experimental setup.

QPT protocol below.

1. **Active reset.** Each experiment begins with an active reset procedure to initialize the qubit in the ground state $|0\rangle$. Here we start by measuring the qubit state with readout pulse of $2\mu s$. After digitization, the state is encoded in the in-phase component according to an initial calibration similar to [69], where we optimize the information obtained from the readout signal by deriving the optimal integration weights. The aim is to maximize the separation of the IQ blobs when the ground and excited state are measured. In 1D Histogram, which

is shown in Fig.9.(b) right top graph, we can visualize the peaks from the ground $|0\rangle$ (blue) and excited $|1\rangle$ (red) states are separated from each other along in-phase quadrature (the histogram comprises one million single-shot measurements). This approach simplifies on-the-fly discrimination between the two quantum states by requiring only a single threshold I_{Thresh} (vertical solid green line in Fig.9.(b)).

Thus, based on the in-phase component the qubit state is estimated. If the OPX+ detects the ground state $|0\rangle$, no pulse follows. If the excited state $|1\rangle$ is recognized, a

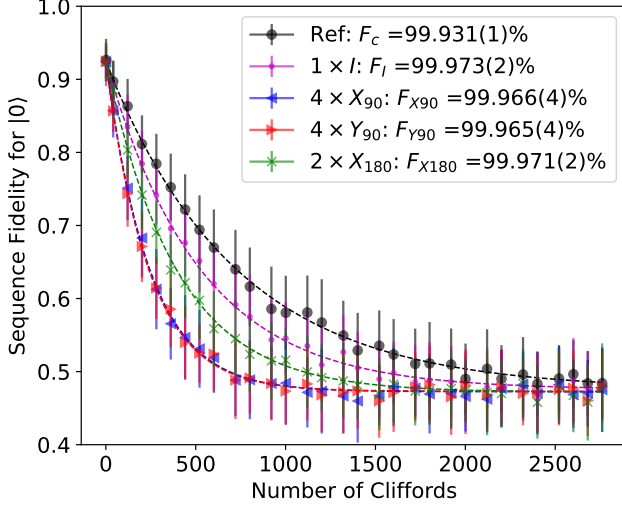


FIG. 8. **Interleaved randomized benchmarking (RB) of single-qubit gates.** All curves are averaged over 50 random seeds.

Index	Combinations
1	[I]
2	[X180]
3	[Y180]
4	[Y180, X180]
5	[X90, Y90]
6	[X90, -Y90]
7	[-X90, Y90]
8	[-X90, -Y90]
9	[Y90, X90]
10	[Y90, -X90]
11	[-Y90, X90]
12	[-Y90, -X90]
13	[X90]
14	[-X90]
15	[Y90]
16	[-Y90]
17	[-X90, Y90, X90]
18	[-X90, -Y90, X90]
19	[X180, Y90]
20	[X180, -Y90]
21	[Y180, X90]
22	[Y180, -X90]
23	[X90, Y90, X90]
24	[-X90, Y90, -X90]

TABLE II. The table of single-qubit Clifford gates.

X180 pulse is executed to rotate it to the ground state $|0\rangle$, after which the measurement repeats. As a result, the qubit is initialized in its ground state. It is worth noting, that each measurement in std-QPT is followed by a buffer time of $10\mu\text{s}$ before applying qubit pulses, in order to avoid dephasing due to photon shot noise.

2. **Apply U_1 .** Here U_1 is a set of known controlled unitary operations (I, $-X90$, $-Y90$, X180) applied to the ground $|0\rangle$ state. This prepares the qubit in different

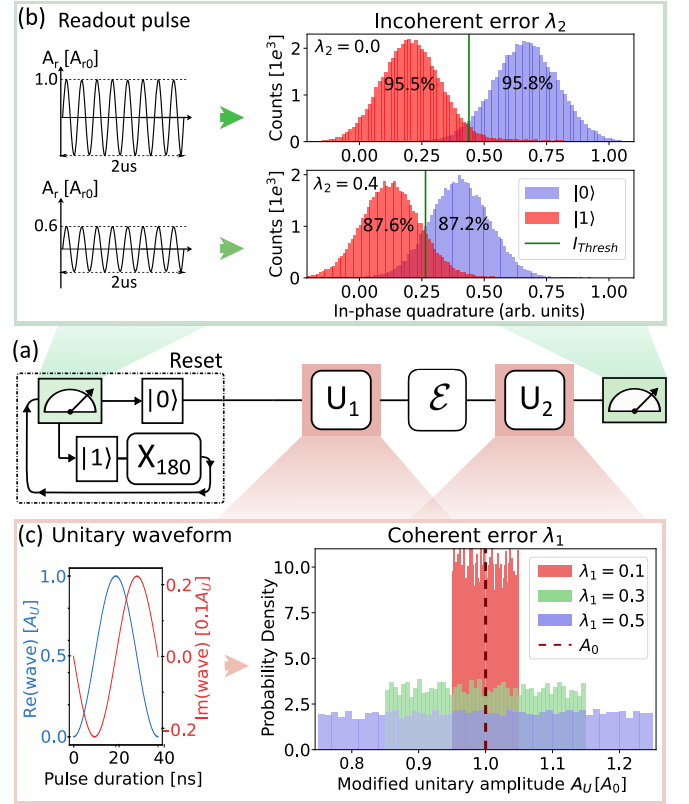


FIG. 9. **Schematic diagram of quantum process tomography and implementation of noise channels.** (a) Quantum process tomography with active reset procedure (dashed rectangle) for the initialization of the qubit in the ground state $|0\rangle$. U_1 and U_2 are the sets of unitary rotations responsible for initial state preparation and measurement projectors, respectively. \mathcal{E} is the process under study. (b) Incoherent noise channel, where amplitude of the readout pulse (left graph) as well as I_{Thresh} (vertical green line in the right graph) are scaled by a factor of $(1 - \lambda_2)$. Readout signal histograms visualize the separation between the peaks from the ground $|0\rangle$ (blue) and excited $|1\rangle$ (red) states for different λ_2 values. The state assignment fidelities are written on top of the corresponding histograms. (c) Coherent noise channel, where amplitude A_U of the unitary waveforms (left graph) are modified according to $A_U = A_0[1 + r \cdot \lambda_1]$, where $r \sim \mathcal{U}(-0.5, 0.5)$ is a uniformly distributed random number. Probability density (right graph) of the modified unitary amplitude for different coherent errors λ_1 . In the experiment, for each value of λ_1 the amplitude is sampled 10,000 times to obtain the corresponding distribution.

input states. As mentioned above, all single-qubit pulses have the same duration of 40 ns, with the real part of the pulse waveform shaped as a cosine envelope and the imaginary part corresponding to the DRAG correction, as shown in the left graph of Fig.9(c). For consistency, the I gate is implemented here as 40 ns delay.

3. **Apply \mathcal{E} .** The quantum process under study (generally unknown). Experimental results for std-QPT, EM-QPT and ML-QPT on the 24 single-qubit Clifford gates

presented in the main text were obtained simultaneously by implementing a sweeping loop through the entire gate set listed in Table II. The tomography data obtained for the identity operation I is then used in EM-QPT to define the ‘error matrix’ $\tilde{\chi}^I$, as described in the main text.

4. **Apply U_2 .** Here U_2 is a set of three measurement projectors (I , $X90$, $Y90$) to read out z , y and x components, respectively. Unitary waveforms are implemented in the same way as for U_1 .

5. **Readout operation.** For the std-QPT the readout operation is a constant pulse with calibrated readout length and amplitude A_{r0} followed by $10\ \mu\text{s}$ buffer time. The measurement records consist of single-shot data (indicating whether the final state of the qubit is $|0\rangle$ or $|1\rangle$) obtained on-fly using the threshold I_{Thresh} , as well as the raw IQ data for further post-processing if necessary.

The switching between state preparations, measurement projections and processes under study, as well as outer loop for averaging, is implemented via four nested QUA “for” loops directly on the OPX+. This enables precise real-time control, efficient parallelism, and strict preservation of experimental timing. Considering the duration of the readout pulse, buffer time and the duration of a single unitary rotation, we estimate the total time required for one QPT experiment across the whole Clifford group with 10,000 averages. Thus, the time for a single QPT is approximately $t_{\text{QPT}} \approx 10,000 \times 4 \times 24 \times 3 \times [2\ \mu\text{s} + 10\ \mu\text{s} + 0.12\ \mu\text{s} + 2\ \mu\text{s} + 10\ \mu\text{s}] = 70\ \text{s}$, which enables us to acquire the 2D data plot shown in the main text in only 10 hours, using 15 QPT repetitions for ref error value. For comparison, the total duration of the RB experiment shown in Fig.8 is approximately 435 s, with data for all five plots acquired in parallel using only 100 averages per random seed.

Incoherent error λ_2 . In the experiment, we introduce an additional incoherent noise channel by reducing the amplitude of the readout pulse according to the expression $A_r = A_{r0}(1 - \lambda_2)$, while simultaneously scaling I_{Thresh} by the same factor, $(1 - \lambda_2)$. Here, A_{r0} denotes the optimal readout amplitude, calibrated in the absence of additional noise ($\lambda_2 = 0$). In each experiment, we simultaneously scale the readout pulses from the first step (**Active reset**) and the fifth step (**Readout operation**) in the same manner. The reduced amplitude of the readout pulses leads to poor separation between the histograms corresponding to the $|0\rangle$ and $|1\rangle$ states, resulting in lower readout fidelity and less reliable ground-state initialization. To illustrate the impact of noise, Fig.9.(c) presents a comparison of one-dimensional histograms at two different noise levels: $\lambda_2 = 0$ (top graph) and $\lambda_2 = 0.4$ (bottom graph). In the absence of noise, the two-state mean assignment fidelity is $F_{\text{Assign}(|0\rangle, |1\rangle)} = 95.65\%$ and it reduces to 87.40% for $\lambda_2 = 0.4$.

Coherent error λ_1 . A coherent noise channel is introduced by adding uncertainties to the amplitude A_U of the single-qubit gate waveforms (see Fig.9.(c), left graph,

where the real and imaginary waveform components are scaled by A_U). Specifically, the amplitude A_U is modified according to $A_U = A_0(1 + r \cdot \lambda_1)$, where A_0 is the optimal amplitude calibrated as described in at $\lambda_1 = 0$, and the random number r is sampled from the uniform distribution $\mathcal{U}(-0.5, 0.5)$. For clarity, the probability densities of the modified unitary amplitude are presented in Fig.9.(c) (right graph) for different values of λ_1 . In the experiment, the random variable r is sampled 10,000 times within the outer averaging loop, and the corresponding amplitude correction is applied to both unitary operations U_1 and U_2 .

Appendix D: The model stability and training convergence

In this section, we analyze the stability of the trained model. A digital twin was first trained using a dataset of 10^3 samples. We then performed quantum process tomography (QPT) six times over 18 hours, with each experiment comprising 10^2 X-gate measurements and corresponding identity gate measurements. Fig. 6.(a) compares the performance of three different methods over time. The results demonstrate that ML-QPT exhibits stability comparable to EM-QPT while achieving even higher accuracy. Fig. 6(b) examines the training data requirements. Non-repeating subsets of 10^2 to 10^3 points were randomly sampled from the full dataset ($N = 10^3$) to train the digital twin. Each model was evaluated on the same 10^2 X-gates via EM-QPT, with fidelities averaged over 10^2 digital twin simulations of error matrices. The infidelity converges to $\sim 10^{-3}$ when the training set exceeds 5×10^2 samples (error bars: standard deviation across 10^2 X-gate measurements).

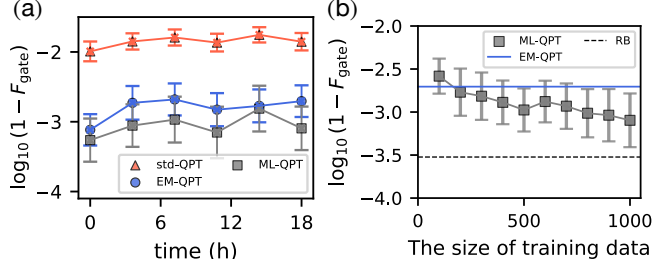


FIG. 10. **The stability and convergence condition.** (a) The QPT infidelity as a function of time (over 18 hours) across various methods. For the ML-QPT, the digital twin is trained on a dataset of 10^3 error matrices. (b) The infidelity of ML-QPT versus the size of the training data, where the training data is randomly sampled from the 10^3 dataset. Each data point is averaged over the same measurement dataset comprising 10^2 X gates. For comparison, we include the average gate fidelity of EM-QPT (solid blue line) and the RB fidelity (dashed line). Error bars in (a) and (b) correspond to the standard deviation of the QPT process for 10^2 experimentally implemented X gates.