Quantum Process Tomography with Digital Twins of Error Matrices

(article starts on next page)

Quantum Process Tomography with 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] Emil Hogedal,[1] Christhopher Warren,[1]
Irshad Ahmad ,[1] Janka Biznárová,[1] Amr Osman,[1] Mamta Dahiya ,[1] Marcus Rommel ,[1] Anita Fadavi Rousari,[1]
Andreas Nylander ,[1] Liangyu Chen,[1] Jonas Bylander ,[1] Gheorghe Sorin Paraoanu ,[2]
Anton Frisk Kockum ,[1] and Giovanna Tancredi[1]

[1]*Department of Microtechnology and Nanoscience, Chalmers University of Technology, 41296 Gothenburg, Sweden*
[2]*QTF Centre of Excellence and InstituteQ, Department of Applied Physics, Aalto University, 00076 Aalto, Finland*

(Received 26 May 2025; revised 18 September 2025; accepted 13 November 2025; published 1 December 2025)

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

DOI: 10.1103/dpgy-rtxr

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 multiqubit gates and accurately characterizing them for further enhancement [5–11]. Identifying errors in gate implementation and improving quantum architectures require more than a single scalar measure, such as gate fidelity from randomized benchmarking protocols [12–15]. Instead, a comprehensive characterization of the entire quantum process is essential, which is typically achieved through quantum process tomography (QPT) [16–18].

Quantum process tomography involves preparing a set of known input states $\{\rho_i\}$, applying a completely positive and trace-preserving (CPTP) quantum process $\mathcal{E}$, and measuring a set of observables $\{M_\mu\}$, typically chosen as elements of a positive operator-valued measure (POVM). This yields a collection of measurement outcomes: $p_{i,\mu} = \mathrm{Tr}[M_\mu \mathcal{E}(\rho_i)]$. In experiments, this procedure demands high-precision state preparation and measurement (SPAM) in order to faithfully characterize the underlying quantum process—a requirement that remains challenging on state-of-the-art hardware platforms [13]. Nevertheless, standard QPT assumes ideal probes ρ_i and M_μ in SPAM operations, and applies a postprocessing algorithm $\mathcal{J}$ to noisy data points $\tilde{p}_{i,\mu}$:

$$\mathrm{std-QPT}: \mathcal{J}(\rho_i, M_\mu, \tilde{p}_{i,\mu}) \to \tilde{\chi}. \tag{1}$$

This leads to internal inconsistencies [19]: the noisy measurement outcomes $\{\tilde{p}_{i,\mu}\}$ are incorrectly attributed to ideal SPAM operations, thereby distorting the reconstruction and interpretation of the resulting process matrix $\tilde{\chi}$. As a result, standard QPT frequently yields unreliable or even misleading characterizations. This fundamental issue of SPAM-induced self-inconsistency in QPT was first recognized and systematically analyzed over a decade ago [33–36], prompting the development of gate set tomography (GST) [9,37–41]. GST is a protocol that enables SPAM-error-free characterization of quantum gate sets. However, its experimental and computational overhead is substantially higher than that of QPT, rendering it impractical for systems beyond two qubits [10].

In this Letter, we propose a generic framework to realize a nearly self-consistent QPT protocol for multiqubit systems (see Fig. 1). By reconstructing effective probes

*Contact author: akshayga@chalmers.se

$\{\bar{\rho}_i, \bar{M}_\mu\}$, our SPAM-error-mitigated QPT (EM-QPT) achieves significantly higher accuracy compared to the standard, self-inconsistent QPT technique. To further improve precision and robustness, we incorporate a machine learning (ML) approach [42–44] that learns the statistical features of SPAM errors by constructing a digital twin of identity process matrices. Remarkably, we validate our method experimentally in two superconducting quantum processors, achieving at least an order-of-magnitude fidelity improvement over standard QPT. Our results establish a scalable and practical framework for high-precision quantum diagnostics, broadly applicable to quantum computing, benchmarking, and control.

Error-mitigated QPT—Our objective is to precisely estimate the process matrix of an arbitrary quantum operation $\mathcal{E}$ while accounting for SPAM errors [45–48]. We utilize identity QPT, only performing state preparation $\mathcal{E}_{\mathrm{sp}}$ and measurement $\mathcal{E}_{\mathrm{m}}$, yielding $\chi^{\mathrm{I}} \equiv \mathcal{E}_{\mathrm{m}} \circ \mathcal{E}_{\mathrm{sp}}$. Ideally, the identity process matrix will be $\chi^{\mathrm{I}}_{mn} = \delta_{m0}\delta_{n0}$, where δ_{mn} is the Kronecker delta. Deviations from the ideal χ^{I} indicate the presence of SPAM errors in the experiment, resulting in $\tilde{\chi}^I$, which we refer to as an *error matrix* [47]. By changing the argument in the QPT algorithm, we can determine the noisy input states and measurement operators [20]:

$$\mathcal{J}[\{\rho_i\}, \{M_\mu\}, \tilde{\chi}^I] \to \{\bar{\rho}_i\}, \{\bar{M}_\mu\}. \tag{2}$$

When computing $\{\bar{\rho}_i\}$, we assume ideal measurement operators, and vice versa (standard quantum state and detector tomography with the error matrix), since gauge symmetry due to unitary invariance [20] prevents simultaneously determining $\{\bar{\rho}_i, \bar{M}_\mu\}$ with arbitrary accuracy and precision [49–52]. We note that a recent theoretical work [45] proposes a strategy similar to EM-QPT, in which the probes are revised by leveraging prior knowledge of the error matrix.

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

$$\mathrm{EM-QPT}: \mathcal{J}(\bar{\rho}_i, \bar{M}_\mu, \tilde{p}_{i,\mu}; \tilde{\chi}^I). \tag{3}$$

This EM-QPT approach is resource intensive, particularly in applications where frequent process characterization is required, such as gate optimization [53,54]. It is also potentially fragile in the presence of anomaly errors, such as glitches in experiments. Since the error matrix $\tilde{\chi}^I$ is independent of the process to be characterized, it is natural to explore whether ML techniques can be leveraged to learn the statistical behavior of SPAM errors hidden in $\tilde{\chi}^I$, for more efficient error mitigation.

Inspired by a recent study [44], we use a generative model as a digital twin of the error matrix to enhance EM-QPT. We find that the digital twin, a trained deep neural network, effectively captures the underlying characteristics of SPAM errors, yielding a more refined version of Eq. (2). It can potentially outperform real-time error-matrix acquisition, enabling high-precision QPT with more robust and efficient SPAM-error mitigation.

Digital twin of the error matrix—Our generative model to construct the digital twin of the error matrix is a variational autoencoder (VAE) [42], which integrates deep learning with probabilistic frameworks to learn a latent representation of training data (see Fig. 1). The 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 latent vector is a numerical representation of the essential features of the input data, usually in a lower-dimensional space. The *decoder* reconstructs the input data $\mathbf{x} \to \mathbf{x}'$ from a sampled latent vector $\mathbf{z}$. Both *encoder* and *decoder* are deep neural networks [20]. To ensure that the VAE output is CPTP, we introduce a `QProcess` layer [20] using Cholesky decomposition [55,56]. See the workflow in Fig. 1.

We characterize the SPAM errors by constructing the digital twin of the error matrix $\tilde{\chi}^I$. In practice, we first collect a training database $\mathbf{X} = \{\mathbf{x}^{(i)}\}_{i=0}^{N_x}$ of N_x independent QPT experiments for the identity process, which is implemented by applying a short idle time of a few nanoseconds in experiments [20]. The digital twin of the error matrix, $\mathbf{x}' \to \chi^I_{\mathrm{DT}}$, is expected to statistically mimic the major pattern of SPAM errors embedded in the error matrix. In this vein, we utilize the digital twin to perform the EM-QPT protocol; the digital twin is applicable to an arbitrary quantum process, as the error matrix is independent of the gate operation under test. We thus introduce the machine learning-enhanced QPT,

$$\mathrm{ML-QPT}: \mathcal{J}(\bar{\rho}_i, \bar{M}_\mu, \tilde{p}_{i,\mu}; \chi^I_{\mathrm{DT}}), \tag{4}$$

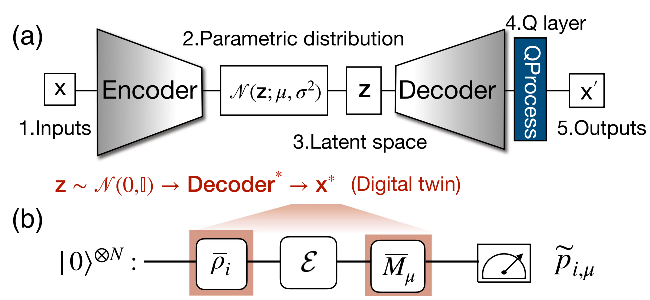

FIG. 1. Digital twin-enhanced quantum process tomography. (a) The variational autoencoder (VAE) consists of an encoder and a decoder built with deep neural networks. The input training data $\mathbf{x}$ is mapped by the encoder into a parametric probability distribution $\mathcal{N}(\mathbf{z};\mu,\sigma^2)$. The latent variable $\mathbf{z}$ is sampled from this distribution and used to reconstruct the output $\mathbf{x}'$ through the decoder and a pre-designed quantum processing layer (QProcess). (b) The digital twin is applied to reconstruct the error matrix $\mathbf{x}^* \to \chi^I_{\mathrm{DT}}$ using a trained VAE, enhancing EM-QPT for a quantum process $\mathcal{E}$.

where χ^{I}_{DT} is the digital twin of the error matrix. See Appendix D in [20] for detailed information about the model structure and learning process.

Numerical simulation—In standard N-qubit QPT, each qubit is first initialized in an initial state by applying a gate $U_1 \in \{\mathtt{I}, \mathrm{R}_x(-\pi/2), \mathrm{R}_y(-\pi/2), \mathtt{X}\}$. Then, the quantum gate $\mathcal{G}$ under investigation is applied. To enable full process characterization, a set of informationally complete rotation gates $U_2 \in \{\mathtt{I}, \mathrm{R}_x(\pi/2), \mathrm{R}_y(\pi/2)\}$ are used prior to measuring the qubits in the computational basis. In practice, the noisy readout is composed as

$$\tilde{p}_{i,\mu} = \langle\langle M_\mu | \mathcal{E}_{\mathrm{m}}^{\dagger} \circ \mathcal{E}_{\mathrm{gate}}(\mathcal{G}) \circ \mathcal{E}_{\mathrm{sp}} | \rho_i \rangle\rangle, \tag{5}$$

where $|\hat{O}\rangle\rangle$ denotes the column-vector form of the operator $\hat{O}$, and $\mathcal{E}_{\mathrm{m}}$, $\mathcal{E}_{\mathrm{gate}}$, $\mathcal{E}_{\mathrm{sp}}$ represent the error channels acting on the measurement, gate, and initial states, respectively. To simulate incoherent SPAM errors, we use a depolarizing error channel $\mathcal{E}_{\mathrm{dep}}(\rho;\lambda) = (1-\lambda)\rho + (\lambda/2^N)I$, where I is the identity operator. We randomly sample the error strength for state preparation $\mathcal{E}_{\mathrm{sp}} = \mathcal{E}_{\mathrm{dep}}(\rho;\lambda_{\mathrm{sp}})$ and measurements $\mathcal{E}_{\mathrm{m}} = \mathcal{E}_{\mathrm{dep}}(\rho;\lambda_{\mathrm{m}})$ in terms of a given error rate $\lambda_1 = \lambda_{\mathrm{m}} + \lambda_{\mathrm{sp}}$. We also introduce coherent errors with a unitary channel $\mathcal{E}_{\mathrm{uni}}(\rho) = U\rho U^{\dagger}$ by adding a rotation shift $\Delta\theta = \theta' - \theta_0$ on a rotation gate $\mathrm{R}_{\mathrm{i}\in[\mathrm{x,y,z}]}(\theta_0) \to \mathrm{R}_{\mathrm{i}\in[\mathrm{x,y,z}]}(\theta')$ in SPAM; we uniformly sample the deviation $\Delta\theta/\pi \in [-\lambda_2, \lambda_2]$ $(\lambda_2 \in [0,1])$. Therefore, we express the total SPAM error as $\lambda_{\mathrm{tot}} = (\lambda_1 + \lambda_2)/2 \in [0,1]$. The numerical experiment consists of three steps: (i) set $\mathcal{G} = \mathtt{I}$, the N-qubit identity gate, and perform std-QPT to obtain the error matrix $\tilde{\chi}^{I}$; (ii) reconstruct 4^N noisy quantum states and 6^N observables using Eq. (2); (iii) perform std-QPT on a randomly selected unitary gate $\mathcal{G}$ using the error-mitigated probes, giving a corrected process matrix according to Eq. (3). Here, we focus on unitary operations, but the EM-QPT approach is valid for any general CPTP process [20].

In Fig. 2, we present numerical results for std-QPT, EM-QPT, and ML-QPT under the influence of both incoherent and coherent errors. Each data point represents the average process infidelity [57] computed over 10^2 randomly chosen unitary gates. Particularly, in Fig. 2(a), we analyze the gate infidelity as a function of coherent error λ_2 and incoherent error λ_1 for a single-qubit system. EM-QPT outperforms std-QPT with significant fidelity improvement. Furthermore, we investigate how the infidelity scales with the total error λ_{tot} in the case of evenly mixed contributions, i.e., $\lambda_1 = \lambda_2$, with the corresponding results for one- and two-qubit gates presented in Figs. 2(b) and 2(c), respectively. For ML-QPT, we collected 10^3 error matrices of each data point to train a digital twin model across a range of λ_{tot}, and achieved better performance than with EM-QPT, as seen in Figs. 2(b) and 2(c). Moreover, we also demonstrate that our method maintains high performance even in the presence of extremely biased SPAM errors $[(\lambda_{\mathrm{m}}/\lambda_{\mathrm{sp}}) \to \infty/0]$; see Appendix H in [20].

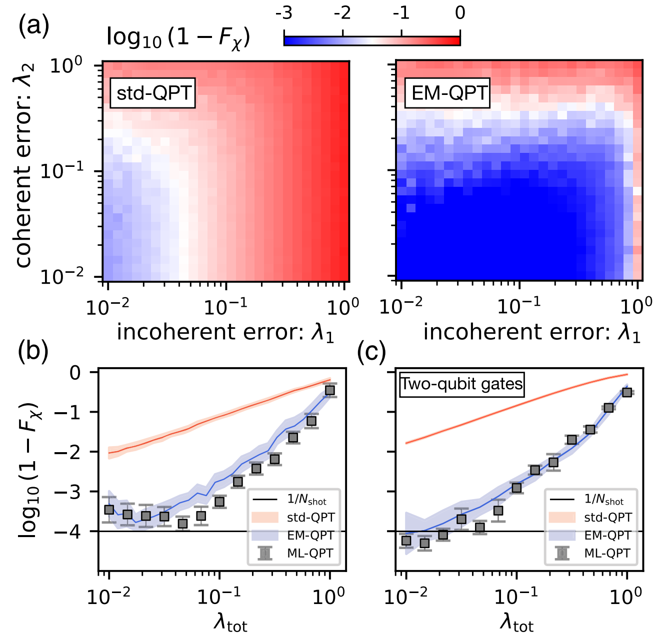

FIG. 2. Numerical results for EM-QPT of one- and two-qubit gates. (a) Average process infidelity of single-qubit gates using std-QPT (left panel) and EM-QPT (right panel), as a function of coherent and incoherent errors. (b),(c) Fidelity for the evenly mixed error regime for single- and two-qubit gates, respectively. Solid curves and squares are the average results over 10^2 gate samples, with shadow and error bar showing one standard deviation. The horizontal line is the statistical error $1/N_{\mathrm{shot}}$ with the shot number $N_{\mathrm{shot}} = 10^4$.

Next, we verify EM-QPT in experiments with single-qubit Clifford gates. We consider the average gate fidelity [59–61]

$$\mathcal{F}_{\mathrm{gate}} = \frac{d\mathcal{F}_\chi + 1}{d+1}, \tag{6}$$

where $d = 2^N$ is the Hilbert-space dimension of the N-qubit system; the process fidelity $\mathcal{F}_\chi$ [57] is obtained through QPT. For small gate errors, randomized benchmarking (RB) [62] statistically captures the average gate error over Clifford gates, implying that the RB fidelity then approximates the average gate fidelity: $\mathcal{F}_{\mathrm{rb}} \approx \mathcal{F}_{\mathrm{gate}}$. In the following, we demonstrate the experimental implementations of our method.

Experiments on single-qubit gates—We implement our method on 24 single-qubit Clifford gates on a superconducting quantum processor (see Device A in [20]). In experiments, we calibrate both single-qubit gates, achieving fidelity of 99.96% using RB measurements, and readout performance [20]. We then introduce coherent and incoherent errors. Incoherent errors are introduced by biasing the optimized amplitude of square pulses A_{r0}, used for readout, by λ_1: $A_r = (1 - \lambda_1)A_{r0}$ ($\lambda_1 \in \mathbb{R}$). Coherent errors are introduced by adding rotation uncertainties to the single-qubit gates through amplitude fluctuations: the target amplitude A_0 is modified to $A_U = (1 + r)A_0$, with the offset r uniformly sampled as $r \sim \mathcal{U}(-\lambda_2, \lambda_2)$ ($\lambda_2 \in [0, 1]$). See the End Matter for details about the error setup.

In Figs. 3(a) and 3(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; EM-QPT yields at least an order-of-magnitude improvement in gate fidelity. Notably, RB outperforms EM-QPT because SPAM errors cannot be explicitly separated from the error matrix. We leave further optimization of the QPT method [45] through adjusting the weight function between state preparation and measurements for future work.

Two-qubit CZ *gates*—We start with a well-tuned adiabatic CZ gate with $\mathcal{F}_{\rm rb}^{\rm CZ} \sim 99.23\%$ (see Device B in [20]). After fine-tuned calibration with single-qubit gate fidelity $\mathcal{F}_{\rm rb} \approx 99.95\%$ [20], we observed a 6% process fidelity reduction due to SPAM errors. We perform identity QPT followed by QPT of this gate 10^2 times. We can thus obtain the process fidelity of the CZ gate with both EM-QPT and ML-QPT. In Fig. 3(c), we present the probability distribution of gate infidelity estimated by Eq. (6) using std-QPT, EM-QPT, and ML-QPT; the ML-QPT results are based on a digital twin trained on the 10^2 error matrices. As a result, the gate fidelities estimated by EM-QPT and ML-QPT are significantly closer to the RB fidelity than those from standard QPT, with the overestimation in ML-QPT attributed to the limited size of the training dataset.

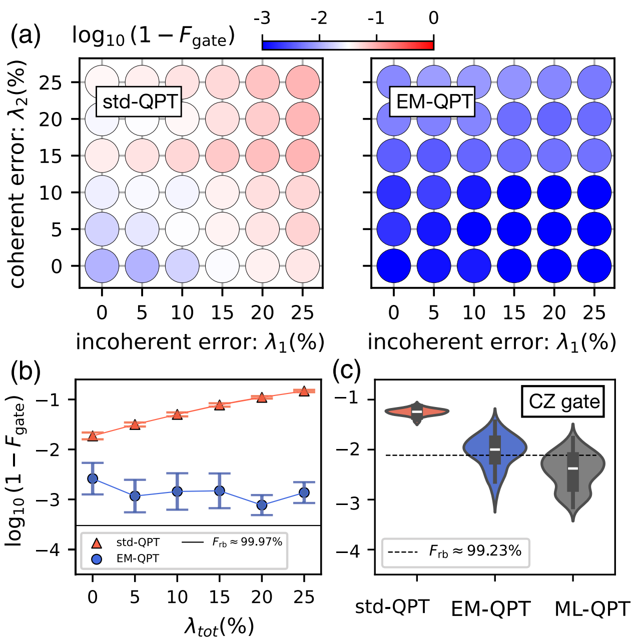

FIG. 3. Experimental results for one- and two-qubit gates. (a) Performance of std-QPT (left) and EM-QPT (right) as a function of coherent and incoherent errors. (b) The points from (a) with $\lambda_1 = \lambda_2$. Each data point in (a) and (b) is averaged over 15 and 10^2 QPT experiments, respectively, for all 24 single-qubit Clifford gates. (c) Infidelity distribution over 10^2 QPT experiments of a CZ gate estimated with std-QPT, EM-QPT, and ML-QPT. The inner box plot indicates the median (white horizontal line) and the interquartile range (black box). Here, SPAM errors ~6% and the RB fidelity is 99.23%.

Precision and sensitivity—In experiments with our method, the gate fidelity of an unknown gate $\mathcal{G}$ can be statistically estimated over the set $\mathcal{S}_I = \{\tilde{\chi}_1^I, \tilde{\chi}_2^I, \ldots, \tilde{\chi}_{N_{\rm err}}^I\}$, forming a probability distribution $\mathbb{P}_{\rm EM}(\mathcal{F}(G; \tilde{\chi}^I))|_{\tilde{\chi}^I \in \mathcal{S}_I}$, where the variance is primarily induced by SPAM errors. However, averaging fidelity over all error matrices may reduce EM-QPT precision due to experimental anomalies. To address this, we use ML to extract the dominant SPAM error patterns, creating a digital twin that reconstructs them. Therefore, the gate fidelity relying on the digital twin admits the distribution $\mathbb{P}_{\rm ML}(\mathcal{F})|_{\mathbb{D}^*(\mathbf{z}) \to \chi^I}$, where $\mathbb{D}^*(\mathbf{z})|_{\mathbf{z} \sim \mathcal{N}(0, \mathbb{I})}$ is the decoder from the trained VAE model. We next evaluate the performance of our method through empirical information theory [63,64], which focuses on the behavior of information measures in practical, finite-sample settings. To quantify the precision and sensitivity of gate characterization, we calculate the distance between the empirical $\mathbb{P}_{(i)}(q)$ and reference $\mathbb{P}_{\rm ref}(q)$ probability distributions by the one-dimensional Wasserstein (earth mover's) distance [65]

$$W_1(\mathbb{P}_{(i)}, \mathbb{P}_{\rm ref}) = \int_{\mathbb{R}} |\mathcal{C}_{(i)}(q) - \mathcal{C}_{\rm ref}(q)| dq, \qquad (7)$$

where $\mathcal{C}_P(q') = \mathrm{Prob}[q \leq q']$ is the cumulative distribution function (CDF) of the probability distribution P. The W_1 distance is the area between the CDF curve P and the reference; see the lower panel in Fig. 4(a) as an example. This distance directly captures first-moment deviations and provides an informative proxy for the second moment.

In our case, the statistical variable in Eq. (7) is the logarithm of infidelity: $q = \log_{10}(1 - \mathcal{F}) \in [-\xi_{\max}, -\xi_{\min}]$ with constants $\xi_{\max/\min} \in \mathbb{R}^+$. For simplicity, we set the reference probability distribution as a delta function $\mathbb{P}_{\rm ref} = \Delta(-\xi_0)$ referring to the ideal measurement protocol that always perfectly estimates the gate fidelity $\mathcal{F}_0 = 1 - 10^{-\xi_0}$ fulfilling $\xi_0 \in [\xi_{\min}, \xi_{\max}]$. The W_1 distance then obeys $W_1 \in [0, \xi_{\max} - \xi_{\min}]$, where $W_1 = 0$ is the ideal measurement scheme that gives the exact gate fidelity. The larger the W_1 distance, the further away from the perfect characterization.

In Figs. 4(a) and 4(b), we compare the normalized $W_1/(\xi_{\max} - \xi_{\min})$ distance of the EM-QPT and ML-QPT methods for 10^2 QPT experiments of an X gate, using 10^3 realistic error matrices and their corresponding digital twin, respectively. We take the distribution of std-QPT and RB as

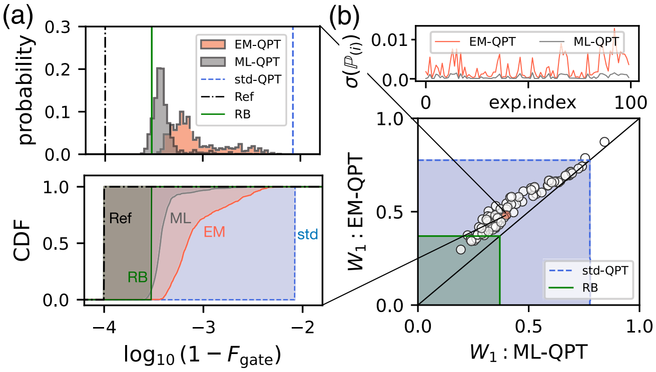

FIG. 4. Precision and sensitivity. (a) Top: infidelity distribution of EM-QPT and ML-QPT, based on the training dataset and digital twin, compared with std-QPT (dashed blue), reference (dot-dashed black), and RB (solid green). Bottom: CDFs used to calculate the W_1 distance from the reference distribution. (b) Top: standard deviations $\sigma(\mathbb{P}_{(i)})$ of the fidelity distributions based on EM-QPT and ML-QPT for a testing dataset comprising 10^2 QPT experiments on X gates. Bottom: normalized W_1 distance of EM-QPT and ML-QPT. The blue and green shaded regions indicate cases where the W_1 distance is smaller than that of std-QPT and RB, respectively. Parameters: $\xi_{\max} = 5, \xi_{\min} = 1, \xi_0 = 4$.

delta functions, since the uncertainty for an individual QPT is negligible. In Fig. 4(a), the probability distribution of EM-QPT (pink) contains all types of errors in error matrices, implying large variance in fidelity estimation due to abnormal errors. Consequently, ML-QPT yields more reliable fidelity estimates, leading to a smaller W_1 distance than EM-QPT in Fig. 4(b).

Discussion and conclusion—We have investigated nearly self-consistent and SPAM-error-mitigated quantum process tomography (EM-QPT) by constructing noisy probes from identity process matrices. Moreover, we proposed machine-learning-assisted QPT (ML-QPT), further enhancing EM-QPT by fully leveraging knowledge of SPAM errors hidden in identity process matrices, enabling accurate and high-precision QPT for reliable and practical applications across a variety of quantum devices. The 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. Furthermore, we have discussed the experimental feasibility of our approach: the model demonstrates stability without time drift in practical implementations, and the training exhibits reliable convergence [20]. Compared to gate set tomography [9,10], our method offers advantages in terms of experimental complexity and generality; see the End Matter for details. In particular, ML-QPT is more resilient to anomalous errors than other methods (see Appendix G in Supplemental Material [20]). To further improve upon our method, one could leverage prior knowledge of SPAM errors [45] (although ML-QPT already performs well without such knowledge; see Appendix H in Supplemental Material [20]), or advance the generative model [66] to find a higher-performance digital twin of the error matrix.

A possible extension of our method is to diagnose the type of SPAM error; for example, the particular behavior of coherent and incoherent errors, providing a useful reference for experimental design. The digital twin of SPAM errors can also serve as a sensitive sensor to detect anomalies in realistic experiments [44]. More broadly, our approach can be directly extended to N-qubit ($N > 2$) quantum processes [8,67], providing an efficient toolkit in quantum technology, e.g., for gate optimization [53,54]. Furthermore, our EM-QPT (ML-QPT) protocol has great potential in ancilla-assisted QPT, where input states are often entangled and the measurement schemes involve intricate global measurements with complex unitary operations [68,69].

Acknowledgments—This work has received funding from the EU Flagship on Quantum Technology through HORIZON-CL4-2022-QUANTUM-01-SGA Project No. 101113946 OpenSuperQPlus100. The Chalmers team acknowledges financial support by the Knut and Alice Wallenberg through the Wallenberg Center for Quantum Technology (WACQT). A. F. K. is also supported by the Swedish Foundation for Strategic Research (Grants No. FFL21-0279 and No. FUS21-0063). The quantum chips used in this work were fabricated at Myfab Chalmers. 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.

T. H. and A. G. conceived the idea. I. M., A. A., M. K., Y.-H. C., O. S., and T. H. performed the experiments and analyzed the data. T. H. also contributed to the numerical simulations and the development of the machine learning algorithms. T. H., A. G., and I. M. wrote the manuscript, and A. F. K., T. A., G. S. P., and G. T. contributed to its revision. G. S. P., A. F. K., and G. T. provided supervision and guidance throughout the project. E. H., C. W., I. A., J. B., A. O., M. D., M. R., A. F. R., A. N., L. C., and J. B. participated in the device fabrication. All authors contributed to discussions and interpretation of the results.

Data availability—The data that support the findings of this Letter are openly available in the repository [70].

[1] H. Aghaee Rad *et al.*, Scaling and networking a modular photonic quantum computer, Nature (London) **638**, 912 (2025).

[2] D. Bluvstein *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature (London) **626**, 58 (2024).

[3] S. A. Moses *et al.*, A race-track trapped-ion quantum processor, Phys. Rev. X **13**, 041052 (2023).

[4] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[5] P. Cerfontaine, R. Otten, and H. Bluhm, Self-consistent calibration of quantum-gate sets, Phys. Rev. Appl. **13**, 044071 (2020).

[6] T. Abad, J. Fernández-Pendás, A. Frisk Kockum, and G. Johansson, Universal fidelity reduction of quantum operations from weak dissipation, Phys. Rev. Lett. **129**, 150504 (2022).

[7] T. Abad, Y. Schattner, A. F. Kockum, and G. Johansson, Impact of decoherence on the fidelity of quantum gates leaving the computational subspace, Quantum **9**, 1684 (2025).

[8] C. W. Warren, J. Fernández-Pendás, S. Ahmed, T. Abad, A. Bengtsson, J. Biznárová, K. Debnath, X. Gu, C. Križan, A. Osman, A. Fadavi Roudsari, P. Delsing, G. Johansson, A. Frisk Kockum, G. Tancredi, and J. Bylander, Extensive characterization and implementation of a family of three-qubit gates at the coherence limit, npj Quantum Inf. **9**, 44 (2023).

[9] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[10] R. Brieger, I. Roth, and M. Kliesch, Compressive gate set tomography, PRX Quantum **4**, 010325 (2023).

[11] C. A. Riofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, Experimental quantum compressed sensing for a seven-qubit system, Nat. Commun. **8**, 15305 (2017).

[12] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, What randomized benchmarking actually measures, Phys. Rev. Lett. **119**, 130502 (2017).

[13] T. Proctor, K. Young, A. D. Baczewski, and R. Blume-Kohout, Benchmarking quantum computers, Nat. Rev. Phys. **7**, 105 (2025).

[14] R. Blume-Kohout, T. Proctor, and K. Young, Quantum characterization, verification, and validation, arXiv:2503.16383.

[15] A. Hashim, Long B. Nguyen, N. Goss, B. Marinelli, Ravi 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, Kevin C. Young, R. Blume-Kohout, and I. Siddiqi, A practical introduction to benchmarking and characterization of quantum computers, PRX Quantum **6**, 030202 (2025).

[16] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[17] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum process tomography of a controlled-NOT gate, Phys. Rev. Lett. **93**, 080502 (2004).

[18] 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).

[19] We refer the reader to Appendix A of Supplemental Material [20] for a detailed description of the interpretation and consequences of self-inconsistency in standard QPT.

[20] See Supplemental Material http://link.aps.org/supplemental/10.1103/dpgy-rtxr for analytical derivations and additional numerical simulations, which include Refs. [21–32].

[21] Andrey V. Rodionov, A. Veitia, R. Barends, J. Kelly, D. Sank, J. Wenner, John M. Martinis, Robert L. Kosut, and Alexander N. Korotkov, Compressed sensing quantum process tomography for superconducting quantum gates, Phys. Rev. B **90**, 144504 (2014).

[22] S. Diamond and S. Boyd, CVXPY: A Python-embedded modeling language for convex optimization, J. Mach. Learn. Res. **17**, 2909 (2016).

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[24] D. P. Kingma, T. Salimans, and M. Welling, Variational dropout and the local reparameterization trick, 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., New York, 2015).

[25] A. Gaikwad, M. S. Torres, S. Ahmed, and A. F. Kockum, Gradient-descent methods for fast quantum state tomography, Quantum Sci. Technol. **10**, 045055 (2025).

[26] D. M. Bashtannyk and R. J. Hyndman, Bandwidth selection for kernel conditional density estimation, Comput. Stat. Data Anal. **36**, 279 (2001).

[27] A. Paszke *et al.*, PyTorch: An imperative style, high-performance deep learning library, arXiv:1912.01703.

[28] J. Koch, Terri M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the Cooper pair box, Phys. Rev. A **76**, 042319 (2007).

[29] M. Kuzmanović, I. Moskalenko, Y.-H. Chang, O. Stanisavljević, C. Warren, E. Hogedal, A. Aggarwal, I. Ahmad, J. Biznárová, M. Dahiya, M. Rommel, A. Nylander, G. Tancredi, and G. S. Paraoanu, Neural-network-based design and implementation of fast and robust quantum gates, arXiv:2505.02054.

[30] 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, arXiv:2503.08645.

[31] Z. Chen, Metrology of quantum control and measurement in superconducting qubits, Ph.D. thesis, University of California, Santa Barbara, 2018.

[32] Z. Chen *et al.*, Measuring and suppressing quantum state leakage in a superconducting qubit, Phys. Rev. Lett. **116**, 020501 (2016).

[33] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, S. Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 062119 (2013).

[34] C. Stark, Self-consistent tomography of the state-measurement Gram matrix, Phys. Rev. A **89**, 052109 (2014).

[35] 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, arXiv:1310.4492.
[36] N. Quesada, A. M. Brańczyk, and D. F. James, Self-calibrating tomography for non-unitary processes, in *The Rochester Conferences on Coherence and Quantum Optics and the Quantum Information and Measurement meeting* (Optica Publishing Group, Washington, 2013), p. W6.38.
[37] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921.
[38] Z.-T. Li, C.-C. Zheng, F.-X. Meng, H. Zeng, T. Luan, Z.-C. Zhang, and X.-T. Yu, Non-Markovian quantum gate set tomography, Quantum Sci. Technol. **9**, 035027 (2024).
[39] S. Cao, D. Lall, M. Bakr, G. Campanaro, Simone D. Fasciati, J. Wills, V. Chidambaram, B. Shteynas, I. Rungger, and Peter J. Leek, Efficient characterization of qudit logical gates with gate set tomography using an error-free virtual z gate model, Phys. Rev. Lett. **133**, 120802 (2024).
[40] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, Probing quantum processor performance with pyGSTi, Quantum Sci. Technol. **5**, 044002 (2020).
[41] P. Viñas and A. Bermudez, Microscopic parametrizations for gate set tomography under coloured noise, npj Quantum Inf. **11**, 23 (2025).
[42] D. P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv:1312.6114.
[43] T. Huang, Y. Ban, E. Y. Sherman, and X. Chen, Machine-learning-assisted quantum control in a random environment, Phys. Rev. Appl. **17**, 024040 (2022).
[44] T. Huang, Z. Yu, Z. Ni, X. Zhou, and X. Li, Quantum force sensing by digital twinning of atomic Bose-Einstein condensates, Commun. Phys. **7**, 172 (2024).
[45] R. Blume-Kohout and T. Proctor, Easy better quantum process tomography, arXiv:2412.16293.
[46] A. G. Kofman and A. N. Korotkov, Two-qubit decoherence mechanisms revealed via quantum process tomography, Phys. Rev. A **80**, 042103 (2009).
[47] A. N. Korotkov, Error matrices in quantum process tomography, arXiv:1309.6405.
[48] 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, Error analysis in suppression of unwanted qubit interactions for a parametric gate in a tunable superconducting circuit, Phys. Rev. A **102**, 022619 (2020).
[49] J. Lin, Joel J. Wallman, I. Hincks, and R. Laflamme, Independent state and measurement characterization for quantum computers, Phys. Rev. Res. **3**, 033285 (2021).
[50] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 14485 (2017).
[51] Ł. Rudnicki, Z. Puchała, and K. Zyczkowski, Gauge invariant information concerning quantum channels, Quantum **2**, 60 (2018).
[52] J. Lin, B. Buonacorsi, R. Laflamme, and J. J. Wallman, On the freedom in representing quantum operations, New J. Phys. **21**, 023006 (2019).
[53] J. Kelly *et al.*, Optimal quantum control using randomized benchmarking, Phys. Rev. Lett. **112**, 240504 (2014).
[54] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, Andre R. Carvalho, and Michael J. Biercuk, Experimental deep reinforcement learning for error-robust gate-set design on a superconducting quantum computer, PRX Quantum **2**, 040324 (2021).
[55] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Maximum-likelihood estimation of the density matrix, Phys. Rev. A **61**, 010304(R) (1999).
[56] S. Ahmed, C. Sánchez Muñoz, F. Nori, and A. F. Kockum, Quantum state tomography with conditional generative adversarial networks, Phys. Rev. Lett. **127**, 140502 (2021).
[57] We use the Uhlmann-Jozsa metric to compute the process fidelity between χ_1 and χ_2 as [58] $\mathcal{F}_\chi(\chi_1,\chi_2) = (\mathrm{Tr}\sqrt{\sqrt{\chi_1}\chi_2\sqrt{\chi_1}})^2$.
[58] R. Jozsa, Fidelity for mixed quantum states, J. Mod. Opt. **41**, 2315 (1994).
[59] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).
[60] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).
[61] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, Lev S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Randomized benchmarking and process tomography for gate errors in a solid-state qubit, Phys. Rev. Lett. **102**, 090502 (2009).
[62] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).
[63] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).
[64] A. J. Majda and B. Gershgorin, Improving model fidelity and sensitivity for complex systems through empirical information theory, Proc. Natl. Acad. Sci. U.S.A. **108**, 10044 (2011).
[65] V. M. Panaretos and Y. Zemel, Statistical aspects of Wasserstein distances, Annu. Rev. Stat. Appl. **6**, 405 (2019).
[66] M. Mirza and S. Osindero, Conditional generative adversarial nets, arXiv:1411.1784 (2014).
[67] M. T. Mądzik *et al.*, Precision tomography of a three-qubit donor quantum processor in silicon, Nature (London) **601**, 348 (2022).
[68] 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, Variational entanglement-assisted quantum process tomography with arbitrary ancillary qubits, Phys. Rev. Lett. **129**, 133601 (2022).
[69] A. Patel, A. Gaikwad, T. Huang, A. F. Kockum, and T. Abad, Selective and efficient quantum state tomography for multi-qubit systems, arXiv:2503.20979.
[70] https://github.com/huangtangy/EM-QPT.
[71] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, Phys. Rev. Lett. **103**, 110501 (2009).

End Matter

Experimental setup—Here, we detail the experimental design for varying SPAM errors. Specifically, we introduce both incoherent and coherent SPAM errors by modifying the optimized pulse envelopes in the experiments. Before proceeding, we calibrate the single-qubit gates, which rely on DRAG pulses [71], to achieve a randomized benchmarking (RB) fidelity of $\mathcal{F}_{\mathrm{rb}} \approx 99.96\%$. In QPT experiments for an N-qubit gate, one needs to implement 12^N circuits, corresponding to the preparation of four initial states and three measurement rotations per qubit. Each circuit execution consists of five steps: (1) active reset of the qubits; (2) apply U_1 for state preparation; (3) perform the target gate operation $\mathcal{E}$; (4) apply the rotation gate U_2; (5) read out all qubit states. Here, both U_1 and U_2 are composed of single-qubit gates. We refer the reader to Supplemental Material [20] for more details about experimental setups.

Incoherent error: We introduce an additional incoherent noise channel by reducing the amplitude of the readout pulse according to the expression $A_r = A_{r0}(1-\lambda_1)$, while simultaneously scaling assignment threshold by the same factor, $(1-\lambda_1)$. Here, A_{r0} denotes the optimal readout amplitude, calibrated in the absence of additional noise ($\lambda_1 = 0$). In each experiment, we simultaneously scale the readout pulses from the first step (active reset) and the fifth step (readout) 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. 5(b) presents a comparison of one-dimensional histograms at two different noise levels: $\lambda_1 = 0$ (top graph) and $\lambda_1 = 0.4$ (bottom graph). In the absence of noise, the two-state mean assignment fidelity is $F_{\mathrm{assign}(|0\rangle,|1\rangle)} = 95.65\%$ and it reduces to 87.40% for $\lambda_1 = 0.4$.

Coherent error: We introduce a coherent noise channel by adding amplitude uncertainties to single-qubit DRAG pulses $\Omega(t) = A_0[1-\cos(2\pi t/t_g)]$ [71] with the gate length $t_g = 40$ ns, see Fig. 5(c), left graph. Specifically, we use the amplitude $A_U = (1+r)A_0$ with $r \sim \mathcal{U}(-\lambda_2, \lambda_2)$, where A_0 is the optimal amplitude (calibrated as described in [20]) at $\lambda_2 = 0$. For clarity, the probability densities of the modified unitary amplitude are presented in Fig. 5(c) (right graph) for different values of λ_2. In the experiment, the random variable r is sampled 10^4 times within the outer averaging loop, and the corresponding amplitude correction is applied to both U_1 and U_2.

Experimental complexity—Here, we compare experimental complexity in various QPT methods: std-QPT, EM-QPT, ML-QPT, and long-sequence gate set tomography (LSGST) [9]. Here, experimental complexity refers to the total number of experiments required to perform tomography of an unknown quantum process on a given hardware platform.

The gate sets used in QPT for state preparation and measurement are $\{\mathtt{I}, \mathrm{R}_x(-\pi/2), \mathrm{R}_y(-\pi/2), \mathtt{X}\}$ and $\{\mathtt{I}, \mathrm{R}_x(\pi/2), \mathrm{R}_y(\pi/2)\}$, respectively. Together, these form a comprehensive gate set: $\mathbb{G} = \{\mathtt{I}, \mathrm{R}_x(\pm\pi/2), \mathrm{R}_y(\pm\pi/2), \mathtt{X}\}$, comprising six distinct gates.

Note that EM-QPT targets only the first moment (accuracy), whereas ML-QPT enhances both accuracy and precision. Their scopes of performance differ

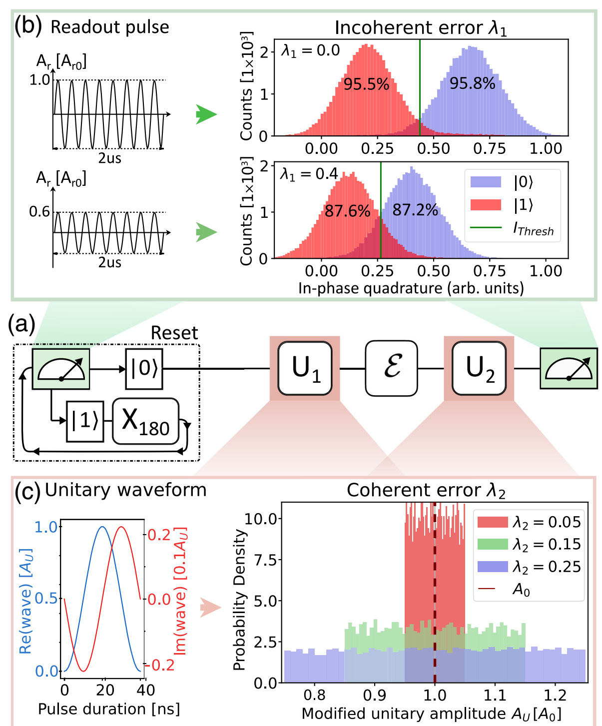

FIG. 5. Schematic diagram of QPT and implementation of noise channels. (a) Quantum process tomography with active reset (dashed rectangle) for the initialization of the qubit in the ground state $|0\rangle$. U_1 and U_2 are the sets of unitary rotations responsible for initial state preparation and measurement projectors, respectively. $\mathcal{E}$ is the process under study. (b) Incoherent noise channel, where the amplitude of the readout pulse (left graph) are scaled by a factor of $(1-\lambda_1)$, resulting in a biased readout threshold (vertical green line in the right graph). Readout signal histograms visualize the separation between the peaks from the ground $|0\rangle$ (blue) and excited $|1\rangle$ (red) states for different λ_1 values. (c) Coherent noise channel, the amplitude A_U of the unitary waveforms (left graph) is modified according to $A_U = A_0(1+r)$ with a uniformly sampled factor $r \sim \mathcal{U}(-\lambda_2, \lambda_2)$ (right graph).

fundamentally, and a direct complexity comparison would therefore be misleading. We thus benchmark EM-QPT against std-QPT, and ML-QPT against LSGST in the following analysis.

In Table I, we list the experimental complexities we find. We see that EM-QPT only doubles the complexity of std-QPT for an N-qubit gate, requiring 2×12^N circuits, while achieving significantly higher accuracy. The complexity for ML-QPT is estimated based on $N_x = 10^2$ error matrices, along with an additional experiment for QPT of the target process, resulting in $(N_x + 1) \times 12^N$ experiments for an N-qubit gate. For LSGST of single-qubit gate, a maximum sequence length of 16 is used, resulting in 2904 experiments [8] generated using pyGSTi [40] to perform GST on the SPAM gate set $\mathbb{G}$, plus 12 additional experiments for QPT of the target process. For a two-qubit gate, the LSGST requires 15925 circuits when using the gate set $\{\mathtt{I}, \mathtt{R}_x(\pi/2), \mathtt{R}_y(\pi/2)\}^{\otimes 2}$, which is generated by the predefined module `smq2Q_XXYYII` [40]. In principle, the complete gate set $\mathbb{G}^{\otimes 2}$ should be implemented for a full two-qubit gate characterization within LSGST, which is not defined in pyGSTi [40]. Consequently, the total number of required experiments is expected to be significantly larger than the example presented here.

We also emphasize that, in the case of ML-QPT, once the digital twin is trained, the experimental cost for performing QPT on a given N-qubit gate is only 12^N experimental circuits. In contrast, in the presence of an experimental anomaly, the entire GST might fail to provide a faithful fidelity estimation since it lacks the statistical precision and robustness offered by ML-QPT on a given hardware.

TABLE I. Experimental complexity comparison for different QPT methods at $N = 1$ and $N = 2$.

	Experimental complexity	
Methods	$N = 1$	$N = 2$
std-QPT	12	144
EM-QPT	24	288
ML-QPT	1212	14544
LSGST	2916	> 16069