Predicting Non-Markovian Superconducting-Qubit Dynamics from Tomographic Reconstruction

Haimeng Zhang[1,2,*] Bibek Pokharel[2,3] E.M. Levenson-Falk[2,3] and Daniel Lidar[1,2,3,4]

[1]*Department of Electrical & Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*

[2]*Center for Quantum Information Science & Technology, University of Southern California, Los Angeles, California 90089, USA*

[3]*Department of Physics & Astronomy, University of Southern California, Los Angeles, California 90089, USA*

[4]*Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA*

Non-Markovian noise presents a particularly relevant challenge in understanding and combating decoherence in quantum computers, yet is challenging to capture in terms of simple models. Here we show that a simple phenomenological dynamical model known as the post-Markovian master equation (PMME) accurately captures and predicts non-Markovian noise in a superconducting qubit system. The PMME is constructed using experimentally measured state dynamics of an IBM Quantum Experience cloud-based quantum processor, and the model thus constructed successfully predicts the non-Markovian dynamics observed in later experiments. The model also allows the extraction of information about crosstalk and measures of non-Markovianity. We demonstrate definitively that the PMME model predicts subsequent dynamics of the processor better than the standard Markovian master equation.

I. INTRODUCTION

A central challenge in controlling and programming quantum processors is to overcome noise. Open quantum system dynamics are often modeled using the Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) master equation [1,2], also commonly known as the Lindblad master equation (LME). The LME is completely positive and is formally easily solvable. However, the LME is derived under the assumption of Markovianity. Loosely, this assumption amounts to an environment that is "memoryless" and is only valid when the system is weakly coupled to a bath whose characteristic timescale is much shorter than that of the system dynamics [3,4]. Although the Markovian assumption allows for significant simplifications, it is only an approximation and in reality it is often desirable to account for non-Markovian effects [5]. This is true in particular in the case of the dynamics of superconducting qubit systems [6,7]. For example, it has been observed on the IBM Quantum Experience (IBMQE) processors [8] that the fidelity of a gate operation is conditional on the gate operation that preceded it in a gate sequence [9]. This is an example of non-Markovian noise, which introduces temporally correlated errors. Non-Markovian effects may arise from, for example, spatially correlated noise that arises from nonlocal external pulse controls, coherent errors caused by residual Hamiltonian terms, or stochastic errors due to slow environmental fluctuations. Such correlations, as well as correlated errors on multiple qubits, have been shown to be a leading source of failure in achieving quantum error correction [10–12], and also in other near-term quantum applications [13]. In other words, dealing with non-Markovianity will be vital to achieving fault-tolerant quantum computation [14–17].

Unfortunately, most device characterization and validation methods do not fully capture non-Markovian effects, as these methods either implicitly or explicitly make the Markovian approximation. For instance, the standard T_1 and T_2 measurements that quantify qubit lifetime assume exponential decay of the excited-state population or the qubit coherence. Similarly, randomized benchmarking and gate-set tomography [18] consider circuits of varying length and assume that the fidelity of corresponding operations decay as circuits become longer. However, on real quantum processors, recent studies [19–21] have observed deviations of the qubit dynamics from the prediction of a purely Markovian treatment.

In this work we focus on noise processes that govern the free (undriven) evolution of a superconducting quantum system. Here, the non-Markovian effects can be both coherent, e.g., due to unintentional crosstalk with neighboring qubits, or incoherent, e.g., due to coupling

---

*haimeng@usc.edu

to magnetic impurities [22]. However, a complete first-principles model for the system-bath interaction [for example, the formally exact Nakajima-Zwanzig (NZ) master equation [23]] may be infeasible to construct or too hard to solve numerically [3]. There has been extensive work on developing a set of master equations that are both easily solvable and account for non-Markovian effects, e.g., the Gaussian collapse model [24], quantum collisional models [25], and the time-convolutionless master equations [3]. Here, we choose to focus on the post-Markovian master equation (PMME) [26], which includes bath memory effects via a phenomenological memory kernel $k(t)$ (see also Ref. [27] for an updated derivation). We choose the PMME for its conceptual and computational simplicity and because it has a closed form analytical solution in terms of a Laplace transform. It naturally interpolates between the exact dynamics (a completely positive map [28]) and the Markovian Lindblad equation, and at the same time, retains complete positivity with an appropriate choice of the form and parameters for $k(t)$ [26,29,30], and remains analytically solvable.

The PMME we consider takes the following form:

$$
\begin{aligned}
\frac{d}{dt}\rho(t) = {}& \mathcal{L}_0\rho(t) \\
& + \mathcal{L}_1 \int_0^t dt' k\left(t'\right) \exp\left[(\mathcal{L}_0 + \mathcal{L}_1)\, t'\right] \rho\left(t - t'\right).
\end{aligned}
\tag{1}
$$

Here $\rho(t)$ is the reduced system state and $\mathcal{L}_0$ and $\mathcal{L}_1$ are Markovian (super)generators in Lindblad form that describe the dissipative dynamics, where $\mathcal{L}_0$ can have additional Hermitian (i.e., unitary evolution generating) components. Non-Markovian effects in the evolution under the Lindblad superoperator $\mathcal{L}_1$ are introduced via a phenomenological memory kernel $k(t)$ to assign weights to the previous "history" of the system state. We note that Eq. (1) differs from the original PMME [26], in that the latter did not contain $\mathcal{L}_0$ inside the integral. The reason behind this choice will become apparent below; in essence, it allows us more flexibility in partitioning the various terms in the Lindbladian.

In our protocol, the PMME model is constructed by fitting to an ensemble of time-domain tomography measurements. Hence, we call our protocol *PMME tomography*. We demonstrate PMME tomography on an IBMQE device and then use the PMME model thus constructed to quantify the degree of non-Markovianity. Our procedure predicts the non-Markovian effects we observe in future measurements and can model the information backflow from the environment to the system on the device we test. Our protocol provides a robust estimation method for a continuous dynamical model beyond the commonly assumed Markovian approximation, paving the way to more accurate

modeling of noisy intermediate-scale quantum (NISQ) devices.

The structure of this paper is as follows. In Sec. II, we start by constructing the closed system model for the qubit evolution, introduce the non-Markovian open system model by proposing a Lindbladian and different memory kernel terms for Eq. (1), describe the data-collection procedure of the state-tomography experiments, and finally use the estimated qubit states sampled during the evolution to find the best-fit PMME model parameters. In Sec. III we apply the PMME model construction protocol to an IBMQE processor, discuss the descriptive power of the PMME models on the fitting dataset, and discuss the models' predictive power by using them to predict qubit dynamics in a previously unseen testing dataset. We also quantify the degree of non-Markovianity based on the constructed PMME model. In Sec. IV we contrast our method with previous work, and provide a discussion of the results and conclusions. The Appendix contains various additional technical details.

II. METHODS

A. Closed system model

We consider a single qubit described by the effective Hamiltonian

$$
H = -\frac{1}{2}\omega_z\sigma_z,
\tag{2}
$$

which is written in the rotating frame of the qubit drive, where ω_z accounts for the detuning between the qubit frequency and the drive frequency. In practice, the drive frequency is typically set to be the qubit frequency. The latter is determined via a calibration procedure, typically carried out on a single qubit with the rest of the neighboring "spectator" qubits all in their ground states. A shift in qubit frequency can lead to a nonzero detuning between the qubit frequency and the drive, thus a nonzero ω_z in the effective Hamiltonian. In addition, the sign and the magnitude of ω_z can change depending on the initial state of the spectator qubits due to the presence of an always-on *ZZ* interaction, which arises from unintended coupling between the qubit and its neighbors [21]. For these reasons, we include a Hamiltonian term in our model.

B. Open system models and their physical motivation

Our task is to find a model that best describes a time series of state-tomography observations and find the best-fit parameters of that model. In this case the model itself is represented by the functional form of the memory kernel $k(t)$, while the model parameters are the parameters of the kernel function. We consider a sequence of models in order of increasing model complexity: the Lindblad model $\mathcal{M}_0$ and the PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$. In general, we denote

the models by $\mathcal{M}_i(\theta)$ where θ is a list of p_i free model parameters, with p increasing monotonically with i.

We take as our simplest model ($\mathcal{M}_0$) the Lindblad master equation

$$\frac{d\rho}{dt} = \mathcal{L}(\rho) = \mathcal{L}_0(\rho) + \mathcal{L}_1(\rho) \tag{3a}$$

$$= -i[H, \rho] + \sum_k \gamma_k \left(V_k \rho V_k^\dagger - \frac{1}{2}\left\{ V_k^\dagger V_k, \rho \right\} \right). \tag{3b}$$

As is clear from Eq. (1), this is equivalent to a PMME with a δ-function kernel, $k_0(t) = \delta(t)$. In the notation of Eq. (1), we choose the first generator as

$$\mathcal{L}_0 = \mathcal{H} + \mathcal{L}_{\text{GAD}}, \tag{4}$$

which has a Hermitian component $\mathcal{H}(\rho) = -i[H, \rho]$ and a generalized amplitude-damping Lindbladian $\mathcal{L}_{\text{GAD}}$ with the Lindblad operators $V_k \in \{\sigma_+, \sigma_-\}$. We choose the second generator as a pure dephasing Lindbladian:

$$\mathcal{L}_1(\rho) = \gamma_z \left(\sigma_z \rho \sigma_z - \rho \right). \tag{5}$$

With this choice, which is motivated by our experiments with the IBMQE devices, the population decay (T_1) is essentially Markovian, as it is dominated by the $\mathcal{L}_0\rho(t)$ term outside the integral in Eq. (1).

The parameters of interest to us are the following: ω_z, the amplitude of the static z field due to the always-on ZZ coupling with the neighboring qubit(s); γ_z, the pure dephasing rate; $\Gamma_s = \gamma_+ + \gamma_-$ and $\Gamma_r = \gamma_+/\gamma_-$, respectively, the sum and ratio of the excitation rate and the relaxation rate. The Kubo-Martin-Schwinger (KMS) condition [3], which states that the rate of excitation in a system is exponentially suppressed relative to the rate of relaxation at the same frequency [$\gamma(-\omega) = e^{-\beta\omega}\gamma(\omega)$, where β is the inverse temperature and $\omega > 0$], implies that $\Gamma_r < 1$. This Lindblad model has a total of $p_0 = 4$ parameters.

To go beyond the Lindblad model $\mathcal{M}_0$, we consider two PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ with extra parameters for the memory kernel. We consider a family of kernels whose Laplace transform can be written as rational functions, i.e.,

$$\tilde{k}_1(s) = \text{Lap}[k(t)] = P(s)/Q(s), \tag{6}$$

with polynomials $P(s) = a_m s^m + \cdots + a_1 s + a_0$ and $Q(s) = b_n s^n + \cdots + b_1 s + b_0$ of degree m and n, respectively. This include a large class of kernels, which can be expressed as linear combinations of functions of the form $t^d e^{ct}$ for complex c and integer d. For the PMME model $\mathcal{M}_1$ we choose the simplest kernel in this family:

an exponentially decaying kernel

$$k_1(t) = a_0 \exp(-b_0 t) \iff \tilde{k}_1(s) = a_0/(s + b_0), \tag{7}$$

where henceforth we impose the constraint $k(0) = 1$ (since the normalization can always be absorbed into $\mathcal{L}_1$), which leads to $a_0 = 1$. This model thus has a total of $p_1 = 5$ free parameters: the Lindblad model parameters and b_0.

The more complex PMME model $\mathcal{M}_2$ has two additional free parameters a_0 and b_1:

$$k_2(t) = \begin{cases} e^{(-b_1/2)t}\left[\frac{2a_0 - b_1}{\mu}\sinh(\frac{\mu}{2}t) + \cosh(\frac{\mu}{2}t)\right] & \text{if } B \geq 0 \\ e^{-\frac{b_1}{2}t}\left[\frac{2a_0 - b_1}{\mu}\sin(\frac{\mu}{2}t) + \cos(\frac{\mu}{2}t)\right] & \text{if } B < 0 \end{cases}$$

$$\iff \tilde{k}_2(s) = (s + a_0)/(s^2 + b_1 s + b_0), \tag{8}$$

where $B = b_1^2 - 4b_0$ and $\mu = \sqrt{|B|}$. This model has a total of $p_2 = 7$ free parameters (the Lindblad parameters and a_0, b_0, and b_1). The sign of B specifies whether the kernel is overdamped or underdamped.

As illustrated in Fig. 1, the models $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ are a sequence of nested models of increasing complexity; for example, $\mathcal{M}_2$ reduces to $\mathcal{M}_1$ with the identifications $a_0 = 0, b_0 = 0$, and with a renaming of the parameter $b_1 \to b_0$ [since now the kernel function $\tilde{k}(s)$ has a lower degree].

The models $\{\mathcal{M}_i\}$ predict a functional form of the dynamics depending on the model variables θ, so the predicted evolution of a state can be written as

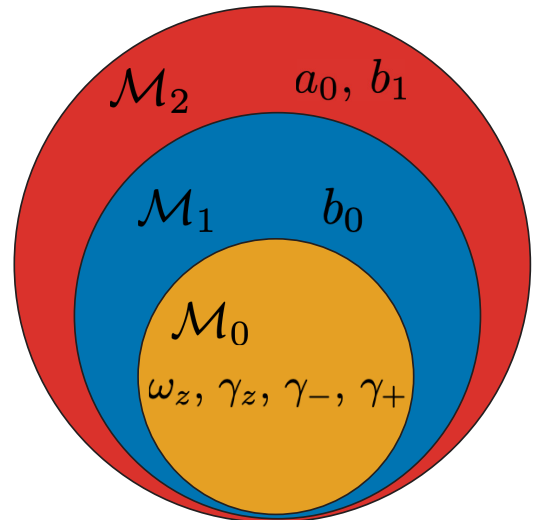$$\rho^{\text{prd}}(t) = f(t|\theta), \tag{9}$$

FIG. 1. The nested candidate models we use to describe the qubit free evolution: the Lindblad model $\mathcal{M}_0$, parameterized by a Hamiltonian term and the Lindbladian, and the PMME models $\mathcal{M}_1, \mathcal{M}_2$ with their additional kernel parameters.

where $\theta = \{\omega_z, \gamma_z, \gamma_-, \gamma_+, \vec{a}, \vec{b}\}$ is the list of model parameters. The kernel parameters are $\vec{a} = \{a_0, \ldots, a_{m-1}\}$, and $\vec{b} = \{b_0, \ldots, b_{n-1}\}$, some of which may be constrained, in addition to the positivity constraint $\{\gamma_z, \gamma_-, \gamma_+\} > 0$ and the KMS constraint $\gamma_+/\gamma_- < 1$.

The goal of this procedure is to specify the master equation that governs the dynamics of the system. We formulate this in terms of the inverse problem: given a discrete time series of measurement records of the state, we seek the dynamical model $\mathcal{M}_i(\theta)$ that most closely matches the observations.

C. Quantum state tomography

To get complete qubit-state information, we perform state tomography on a single qubit of the *ibmq_athens* processor (see Appendix A). The data is collected with the main qubit state ρ_S initialized in one of the five states in the preparation set $\mathbf{P} = \{|\psi_i(0)\rangle\}_{i=0}^{i=4}$; and the rest of the processor's spectator qubits initialized in the ground state $|0\rangle$. The states in $\mathbf{P}$ are illustrated in Fig. 2 and specified in Table I. After the qubit initialization, the main qubit undergoes free evolution for a variable time t_j, and then state tomography is performed to construct the density matrix, augmented by measurement error mitigation (see Appendix B). Specifically, the circuits of the tomography experiment contain the following steps:

1. State preparation: the qubit is initialized in the ground state, and a state-preparation gate is applied to initialize the qubit in one of the five fiducial states $|\psi_i(0)\rangle$.

TABLE I. The set of initial states $\mathbf{P}$ used in our experiments, corresponding to the states shown in Fig. 2.

$|\psi_0(0)\rangle$	$(\sqrt{8/9}, 0, -1/3)$
$|\psi_1(0)\rangle$	$(0, 0, -1)$
$|\psi_2(0)\rangle$	$(-\sqrt{2/9}, \sqrt{2/3}, -1/3)$
$|\psi_3(0)\rangle$	$(-\sqrt{2/9}, -\sqrt{2/3}, -1/3)$
$|\psi_4(0)\rangle$	$(0.50, -0.75, -0.41)$

2. Evolution: the qubit undergoes free evolution, with a variable evolution time. This corresponds to applying a sequence of identity gates $\mathbb{I}$ and sweeping the number of gates.

3. Measurement: one of the three single-qubit gates ($\mathbb{I}$, $\mathsf{M_X}$, or $\mathsf{M_Y}$) is applied before the measurement, corresponding to measurement in the Pauli z, x, and y bases, respectively. Here $\mathsf{M_X} = H$ is a Hadamard gate and $\mathsf{M_Y} = HS^\dagger$ where S is a phase gate. We record the measurement outcomes 0 or 1.

The steps above are repeated for all combinations of initial states, free evolution duration, and measurement basis, and each combination is repeated for $N_s = 8192$ shots. Letting N_{jk} denote the number of times out of N_s shots that outcome 1 occurred at time t_j in measurement basis k, the state-tomography raw data are the recorded outcome counts,

$$\{N_{jk} | j = 1, \ldots, n_t, k \in \{x, y, z\}\}, \qquad (10)$$

where n_t is the total number of time points. We perform Bayesian measurement error mitigation on the raw data (see the Appendix for details) and the measurement mitigated data is then fed into a maximum-likelihood estimation (MLE) routine. This routine determines the qubit state $\hat{\rho}(t_j)$ at time t_j, represented by the experimentally measured Bloch vector $\vec{v}^{\exp}(t_j)$ as $\hat{\rho}(t_j) = \frac{1}{2}[\mathbb{I} + \vec{v}^{\exp}(t_j) \cdot \vec{\sigma}]$. The uncertainties associated with the tomographically constructed states, i.e., the standard deviations of the corresponding Bloch vector components, denoted as σ_{jk}, $k \in \{x, y, z\}$, are estimated by Bayesian bootstrapping. The collected datasets are divided into the model fitting set $\{\hat{\rho}\}_{\rm fit}$ and the model testing set $\{\hat{\rho}\}_{\rm test}$. The former contains a time series of qubit dynamics with a single initial state $|\psi_0(0)\rangle$, the latter contains the other four time series with four different initial states $\{|\psi_i(0)\rangle\}_{i=1}^{i=4}$. Finally, we use the constructed qubit states $\{\hat{\rho}(t_j)\}_{j=0}^{j=n_t-1}$ from the fitting dataset to construct the PMME description of the free-evolution channel.
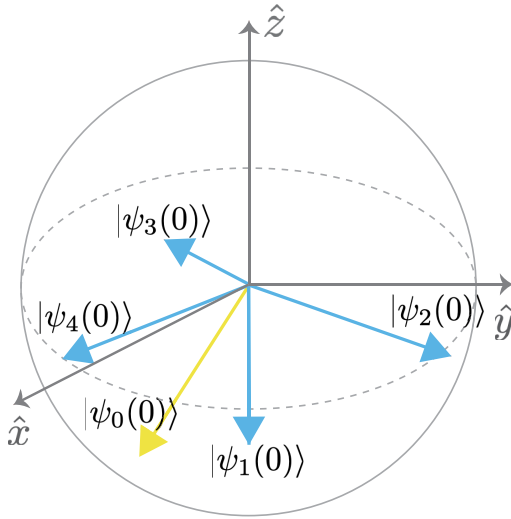

FIG. 2. The initial states used for the fitting dataset (yellow) and the testing dataset (blue). The initial states are chosen such that the states $|\psi_i(0)\rangle$, $i = 0, 1, 2, 3$ form a tetrahedron on the Bloch sphere, where $|\psi_1(0)\rangle = |1\rangle$ is the excited state. The state $|\psi_4(0)\rangle$ is a fixed, randomly chosen state on the Bloch sphere. The same set of five states are used in all our experiments.

D. Model fitting

Given the observations from the state-tomography experiment, our task is to find the best-fit model parameters θ, as parametrized in Sec. II B, of the dynamical model

$\mathcal{M}_i(\theta)$. To tackle the problem, we perform a maximum-likelihood estimation, a well-studied method of determining unknown parameters of a model from a set of data. The input data of this MLE procedure is a time series of the qubit evolution $\{t_j, \hat{\rho}^{\text{exp}}(t_j)\}_{j=0}^{j=N_t-1}$, initialized in one of the states in the set **P**. The qubit state $\hat{\rho}^{\text{exp}}$ at time t_j is constructed from tomography experiment outcomes via a Qiskit state-tomography MLE routine [31]. In the fitting procedure, we seek to minimize the distance between the observed state $\hat{\rho}^{\text{exp}}(t_j)$ and the model predicted state $\rho^{\text{prd}}(t_j)$ for all the sampled time instances t_j. We define the following standard objective function in the least-squares form [32]:

$$\chi^2(\theta) = \sum_{j=0}^{N_t-1} \sum_{k=x,y,z} \frac{[\hat{v}_{jk}^{\text{exp}}(t_j) - v_{jk}^{\text{prd}}(t_j;\theta)]^2}{\sigma_{jk}^2}, \qquad (11)$$

where $\vec{v}_{jk}^{\text{exp}}$ and $\vec{v}_{jk}^{\text{prd}}$ denote the kth Bloch vector component of the qubit state at time t_j constructed from the experimental observations and predicted by the model, respectively. The latter is obtained by solving the model $\mathcal{M}_i(\theta)$ for $\rho^{\text{prd}}(t)$. Since the PMME admits a closed-form analytical solution, the evaluation of the objective function Eq. (11) is efficient to compute under the assumption that the noise associated with each data point follows a normal distribution. Minimizing the least-squares function $\chi^2(\theta)$ for the model parameters θ is equivalent to maximizing the likelihood of observing the dataset given that the underlying model is true.

We would like the dynamical model to generate a completely positive and trace-preserving (CPTP) map (though this is not strictly necessary, as non-CP maps are also valid physical models; see, e.g., Ref. [33]). For this reason, we restrict the Lindblad rates $\{\gamma_z, \gamma_+, \gamma_-\}$ to be positive. For the PMME models with different choices of kernel functions, we derive the condition that guarantees CPTP dynamics in Appendix E. We find that the conditions $|\xi_4| < 1$ and $|\xi_2| = |\xi_3| < \sqrt{(\Gamma_r + \xi_4)(1 + \Gamma_r\xi_4)}/1 + \Gamma_r < 1$ are necessary and sufficient for complete positivity, where $\xi_i(t) = \text{Lap}^{-1}\left[1/s - \lambda_i^0 - \lambda_i^1 k(s - \lambda_i)\right]$ (the inverse Laplace transform), and λ_i^0, λ_i^1, and λ_i are the eigenvalues of the matrix representation of $\mathcal{L}_0$, $\mathcal{L}_1$, and $\mathcal{L}$, respectively, in the Pauli basis (see Appendix C).

The best-fit PMME parameters are found by solving the following minimization problem:

$$\text{minimize } \chi^2\left(\omega_z, \gamma_z, \gamma_+, \gamma_-, \vec{a}, \vec{b}\right)$$

subject to $\gamma_z, \gamma_+, \gamma_- > 0$, $\Gamma_r = \gamma_+/\gamma_- < 1$, $k(0) = 1$. (12)

After we fit the models by solving Eq. (12), we compute the trace-norm distance

$$\mathcal{D}(t_j) = \frac{1}{2}\|\hat{\rho}^{\text{exp}}(t_j) - \rho^{\text{prd}}(t_j)\|_1, \qquad (13)$$

where $\|A\|_1 = \text{Tr}[\sqrt{A^\dagger A}]$ is the trace norm. $\mathcal{D}(t_j)$ quantifies the probability with which one can optimally discriminate the experimentally observed $\hat{\rho}(t_j)$ from the predicted state $\rho(t_j)$ [34], and thus a small $\mathcal{D}(t_j)$ indicates an accurate prediction.

III. RESULTS

We test our approach for tomographic PMME construction in three increasingly challenging settings, summarized in Fig. 3. We start with time-series state-tomography data for a single initial state on a transmon qubit [the fitting dataset on $|\psi_0(0)\rangle$] and show that we can construct a sequence of nested PMME models to represent this evolution. We then compare the PMME models with the Markovian Lindblad counterpart and show that the PMME models provide a measurably better fit. Secondly, we test the PMME's predicted evolution for quantum states that are not used to construct the PMME model. Again, we show that the PMME model faithfully predicts the evolution even for these new states without requiring full process tomography, while the Lindblad model does not. Lastly, we compute the degree of non-Markovianity in the evolution using the tomographically constructed PMME model and show that it correctly approximates the observed non-Markovianity in the qubit free-evolution dynamics, which the Lindblad model fails to do.

A. Fitting set

First, we construct the dynamical models $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ using the time-series tomography data of a single initial state. Our results are displayed in the first column of Fig. 4, which shows the observed evolution $\vec{v}^{\text{exp}}$ (gray circles with error bars) and the predicted evolution $\vec{v}^{\text{prd}}$ (solid
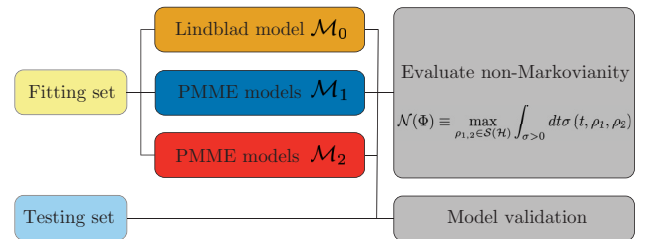

FIG. 3. Analysis protocol for PMME tomography. The tomography datasets for qubit free evolution with different initial states are divided into a fitting set [in our case with a single initial state $|\psi_0(0)\rangle$] and a testing set [in our case with initial states $|\psi_{1,2,3,4}(0)\rangle$]. The fitting set is used to fit the dynamical models $\{\mathcal{M}_i\}$ using a classical optimizer based on the MLE method.
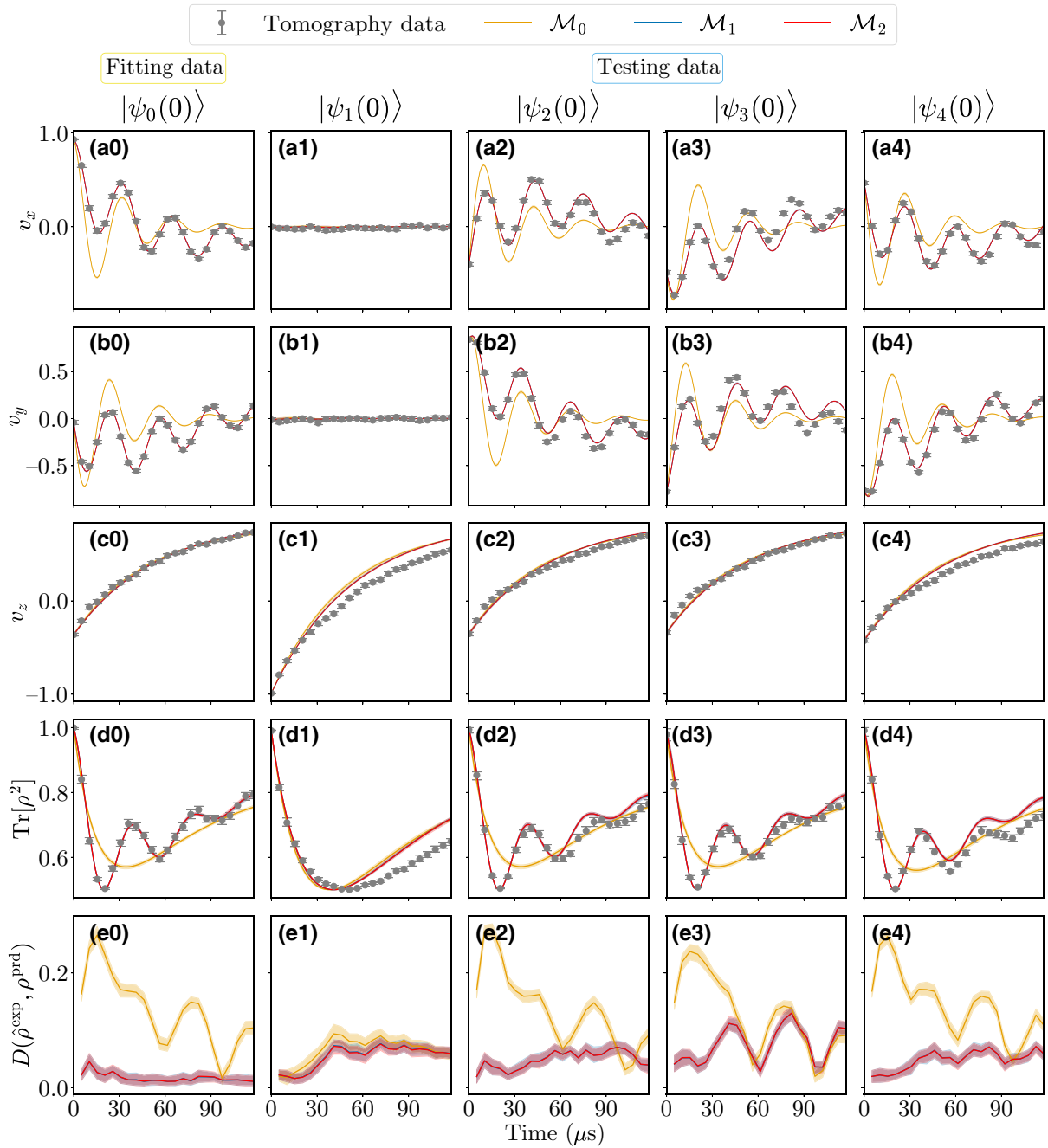
FIG. 4. PMME tomography protocol applied to single-qubit free evolution, with the spectator qubits all in their ground state. (a0)–(c0) The free-evolution tomography data $\hat{v}^{\mathrm{exp}}(t)$ of the fitting dataset with the qubit initialized in $|\psi_0(0)\rangle$ and the best-fit models for the Lindbladian model $\mathcal{M}_0$ (orange lines), the PMME model with kernel type 1, $\mathcal{M}_1$ (blue lines), and the PMME model with kernel type 2, $\mathcal{M}_2$ (red lines). The shaded regions denote the 95% confidence region of the model predictions. (a1)–(c4) The free-evolution tomography data in the testing dataset with the qubit initialized in $\{|\psi_i(0)\rangle\}_{i=1}^4$, and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). (d0)–(d4) The empirical purity $\mathrm{Tr}[\rho^2]$ of the qubit and that predicted by the best-fit models. (e0)–(e4) The distance between the tomographically constructed state and the state predicted by the best-fit models. All models perform equally well at predicting the dynamics of the excited state $|\psi_1\rangle = |1\rangle$ ($\mathcal{M}_0$ is obscured by $\mathcal{M}_1$ and $\mathcal{M}_2$), while the PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ predict the dynamics of $\{|\psi_i(0)\rangle\}_{i=2}^4$ better than the Lindbladian model.

lines) from the constructed dynamical models. The best-fit parameters of the constructed models $\mathcal{M}_i$ are reported in Table V in the Appendix. We plot the Bloch vector components in rows a, b, and c, the state purity $\mathrm{Tr}[\rho^2]$ in row d, and the trace-norm distance between the measured and the predicted state $\mathcal{D}(t_j)$ [Eq. (13)] in row e. The PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ both faithfully capture the evolution of the system. To compare our models, we
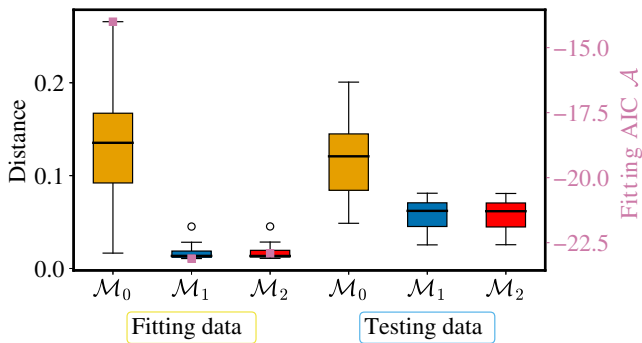
FIG. 5. The predictive ability of different models for the fitting dataset (left) and the testing dataset (right) for *ibmq_athens* data. For both datasets, we show the trace-norm distance between the tomographically constructed state $\{\hat{\rho}^{\text{exp}}\}$ and the model predicted state $\{\rho^{\text{prd}}\}$. The box plot shows the 5th, 50th (median), and 95th percentiles for this distance over t_j and respective initial states. The open circles denote extremal outliers. The PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ describe both the fitting dataset and the testing dataset better than the Lindblad model $\mathcal{M}_0$ (see Table II). We also report the AIC of the models on the fitting data (purple squares, right axis), again with better performance by the non-Markovian models; note that all AIC values are negative.

use the Akaike information criterion (AIC) [35], which is a model selection metric (see Appendix F for more details). AIC accounts for the goodness of fit of a model and penalizes models with more parameters. As shown by the purple squares in Fig. 5, $\mathcal{M}_1$ and $\mathcal{M}_2$ have much lower AIC than $\mathcal{M}_0$ and going from $\mathcal{M}_1$ to $\mathcal{M}_2$ decreases the AIC slightly. Overall, $\mathcal{M}_1$ and $\mathcal{M}_2$ are better models even after being penalized for utilizing more parameters.

Both the Lindblad model $\mathcal{M}_0$ and the PMME models can account for a spurious longitudinal field component in the qubit Hamiltonian due to the always-on *ZZ* interaction between neighboring qubits [36]. This spurious field effectively shifts the qubit frequency [Eq. (2)] and manifests as the oscillations in the off-diagonal elements $v_x(t)$ and $v_y(t)$ of the density matrices of the qubit state (the first and second rows of Fig. 4). However, a Hamiltonian term does not modify the purity $p = \text{Tr}[\rho^2]$ since $\dot{p} = -i\text{Tr}[\rho[H,\rho]] = 0$, so purity oscillations must have a different origin. The fact that $\mathcal{M}_0$ is unable to capture the purity oscillations seen in the fourth row of Fig. 4, while in contrast both $\mathcal{M}_1$ and $\mathcal{M}_2$ do display purity oscillations, is evidence of non-Markovianity, as we explore below in greater depth. The nonmonotonic envelope of the purity oscillations is consistent with the nonunitality of the Lindbladian model [37]. Overall, in comparison to the Lindbladian model $\mathcal{M}_0$, the PMME-predicted evolution is significantly closer to the empirical data, as quantified by $\mathcal{D}(\hat{\rho}^{\text{exp}}, \rho^{\text{prd}})$ (last row of Fig. 4).

B. Testing set

Next, we test how well the fitted models predict the evolution of states that are not used to fit these models.

The last four columns of Fig. 4 represent four time series of quantum-state evolutions in the testing dataset with each of the different initial states $\{|\psi_1(0)\rangle\}_{i=1}^4$. The goal of this "testing set" is to validate whether the dynamical model is capable of describing the evolution for an *arbitrary* single-qubit state. Rather than doing this by selecting random states, we choose a set of four states that is maximally separated on the Bloch sphere, i.e., four pure states on the vertices of a tetrahedron (see Fig. 2). We emphasize that no fitting is done on this testing data; instead, we use the fits from the fitting dataset to *predict* the state evolution in the testing dataset.

Once again, we find that the Lindbladian model's prediction provides a crude approximation to the evolution. In particular, for $|\psi_1(t)\rangle$, where the initial qubit state is the excited state, the amplitude damping process dominates, and the Markovian Lindblad model is sufficient to describe the dynamics. However, $|\psi_1(t)\rangle$ is an exception. For all other states, the PMME models are far more accurate, as is clear from Fig. 4(a2)–(d4). This suggests that the non-Markovian effects mainly manifest in the evolution of the qubit phase coherence but not in the state populations, so the need for the more complex PMME models arises when dealing with states with coherence in the computational basis (this pointer basis—the ground and excited states—is einselected [38] due to thermal relaxation). Another observation visible from Fig. 4 (row d) is that the Markovian model's predictions become more accurate at relatively long evolution times, i.e., Markovian effects become more dominant on a timescale of approximately 100 μs.

Overall, as shown in Fig. 5 and Table II, the median and worst-case prediction distances of the PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ are very close, and substantially better than those of the Lindbladian model $\mathcal{M}_0$. In particular, consider the box plot in Fig. 5 reporting the statistics for the trace-norm distance across all sampling points t_j and initial states of the fitting and testing datasets, respectively. For the testing data, while the Lindblad model $\mathcal{M}_0$ has a median trace-norm distance of 0.12 and worst-case fidelity of 0.18, both the PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ have a median distance of 0.06 and the worst-case fidelity of 0.08.

Notably, the distance does increase slightly when going from the fitting data to the testing data, which can be attributed to the qubit's environment changing during the

TABLE II. Trace-norm distances corresponding to the results shown in Fig. 5.

	Fitting data		Testing data	
	Median	95th percentile	Median	95th percentile
$\mathcal{M}_0$	0.14	0.24	0.12	0.19
$\mathcal{M}_1$	0.01	0.03	0.06	0.08
$\mathcal{M}_2$	0.01	0.03	0.06	0.08

time between measurements; for example, the qubit relaxation time T_1 fluctuates [39]. Still, the PMME models provide a significantly closer correspondence to the fitting dataset and more accurate predictions of the testing datasets than the Lindblad model does in both the average and worst cases. In this case the best-fit model for $\mathcal{M}_2$ is approximately equal to $\mathcal{M}_1$ (see Appendix F), indicating that the simpler model is sufficient to capture the behavior of the fitting dataset. However, this does not hold true for all data, as shown in Appendix F.

Figure 4 clearly shows that these differences in distance between $\mathcal{M}_0$ and $\mathcal{M}_{1,2}$ come from $\mathcal{M}_0$'s inability to capture oscillations about a nonzero mean in the Bloch vector components $v_x(t)$ and $v_y(t)$, which $\mathcal{M}_1$ and $\mathcal{M}_2$ do capture. Our results validate that while $\mathcal{M}_0$ is a good first Markovian approximation, the PMME models account for non-Markovian nuances.

C. Quantifying non-Markovianity

Lastly, we test whether the PMME model can capture non-Markovianity during the qubit free evolution. To quantify this, we adopt the measure in Ref. [40], which uses the rate of change of the trace-norm distance between two quantum states under some noise channel Φ_t:

$$\sigma\left(t, \rho_1, \rho_2\right) \equiv \frac{d}{dt} \mathcal{D}\left\{\Phi_t\left[\rho_1(0)\right], \Phi_t\left[\rho_2(0)\right]\right\}. \quad (14)$$

Under Markovian dynamics [Eq. (3a)], the trace-norm distance between two quantum states is monotonically decreasing as a function of time, whereas non-Markovian dynamics violates this contractive property, i.e., there can be an increase in the trace-norm distance. In other words, non-Markovianity leads to revival of distinguishability between two states at some point during the evolution, and a process is non-Markovian if there exists any pair of initial states $\rho_1(0), \rho_2(0)$ and a time t for which $\sigma\left(t, \rho_1, \rho_2\right) > 0$ [40].

The measure for the degree of non-Markovianity of a quantum process is thus defined as:

$$\mathcal{N}(\Phi) \equiv \max_{\rho_{1,2} \in \mathcal{S}(\mathcal{H})} \int_{\sigma > 0} dt\, \sigma\left(t, \rho_1, \rho_2\right). \quad (15)$$

Here, the time-integration is over time intervals where σ is positive, and the maximum is over all pairs of initial states. Thus, this quantity measures the total increase of distinguishability over the whole evolution time. We note that it is not normalized, and hence its actual numerical values are difficult to interpret. Its main significance is in the the fact that a non-zero non-Markovian measure means non-Markovian dynamics, and, as shown in Table III, the experimentally obtained values are in fair agreement with the values predicted by our model. The non-Markovianity of the PMME is discussed in [30] using this measure.

TABLE III. Degree of non-Markovianity for three different initial states of the spectator qubits (column 1) as computed from the experimental data (column 2) and the two non-Markovian models (columns 3 and 4).

Spectator qubit initial state	$\mathcal{N}^{\mathrm{exp}}$	$\mathcal{N}^{\mathrm{prd}}_{\mathcal{M}_1}$	$\mathcal{N}^{\mathrm{prd}}_{\mathcal{M}_2}$
$|0\rangle^{\otimes 4}$	1.06 ± 0.02	1.11 ± 0.02	1.10 ± 0.02
$|1\rangle^{\otimes 4}$	0.18 ± 0.03	0.28 ± 0.02	0.25 ± 0.01
$|+\rangle^{\otimes 4}$	0.13 ± 0.02	0.49 ± 0.03	0.47 ± 0.03

We performed state tomography on free evolution with the qubit initialized in the following two pairs of states: $\{\rho_1(0) = |+\rangle\langle+|, \rho_2(0) = |-\rangle\langle-|\}$, and $\{\rho_1(0) = |+i\rangle\langle+i|, \rho_2(0) = |-i\rangle\langle-i|\}$. From the time-series state tomography data, we see how the trace-norm distance $\mathcal{D}\{\Phi_t[\rho_1(0)], \Phi_t[\rho_2(0)]\}$, starting from $\mathcal{D} = 1$ (maximally distinguishable), evolves as a function of time, as plotted in the top row of Fig. 6. Using forward differences to approximate the time derivative of our experimental data, we then approximate the quantity in Eq. (14), plotted in the bottom row of Fig. 6. By performing a linear interpolation on the discrete samples shown there, we further estimate the observed degree of non-Markovianity as defined in Eq. (15) (but excluding the maximization it requires). The results we find for the degree of non-Markovianity are summarized in Table III.

Of course, $\mathcal{N}^{\mathrm{prd}}_{\mathcal{M}_0} \equiv 0$ for the Markovian Lindblad model. The experimental measure is estimated from the
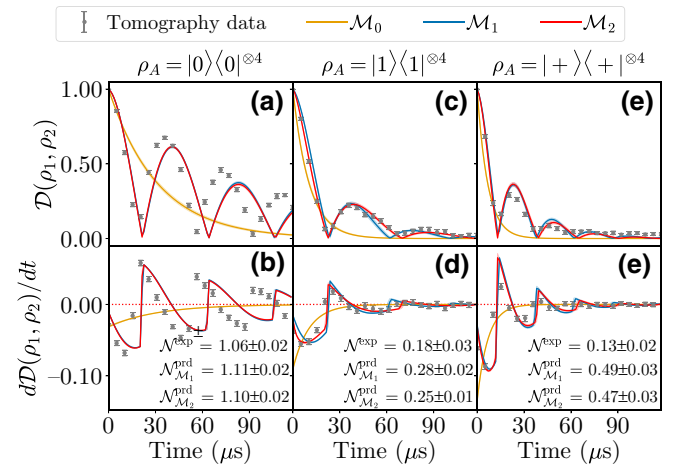

FIG. 6. Non-Markovianity of qubit free-evolution dynamics for spectator qubits in the ground state (a),(b), the excited state (c),(d) and the $|+\rangle$ state (e),(f). (a),(c),(e) The trace-norm distance $\mathcal{D}[\rho_1(t), \rho_2(t)]$ predicted by the best-fit models (solid lines) and experimentally measured by performing free-evolution tomography with a pair of initial states $\rho_1(0) = |+\rangle\langle+|$ and $\rho_2(0) = |-\rangle\langle-|$ (gray circles). (b),(d),(f) The derivative $\sigma(t)$, defined in Eq. (14), predicted by the best-fit models (solid lines), and approximated experimentally using forward differencing based on the tomography data in (a),(c),(e) (gray circles).

state-tomography data, while the predicted non-Markovian measure is calculated exactly from our dynamical models, since they provide a continuous description of the qubit-state evolution. We train these models on a new fitting dataset with the same initial state $|\psi_0(0)\rangle$ as the prior fitting data (Fig. 2); the new fitting data is taken in the same batch of jobs as the trace-norm distance data in order to minimize the time delay between them, thus reducing the effect of systematic errors due to differences between batches. As seen in Table III, the two PMME models converge to almost the same quantity in their predictions, despite the different forms of their kernels. The tomography experiments confirm that there is indeed an increase of distinguishability between quantum states during evolution. The predictions of the PMME models adequately match the observed quantity $\sigma(t_j)$, with some deviations, which are due to numerical errors arising from using the finite difference formula to approximate Eq. (14), and due to the run-to-run system fluctuations of the quantum processor [41].

IV. DISCUSSION

We develop and implement a procedure to fit a family of phenomenological quantum master equations to time-series state-tomography data. We demonstrate this method by characterizing the free evolution of a single qubit on an IBMQE processor. From the constructed models, we conclude that the qubit Hamiltonian accounts for a residual longitudinal field due to the crosstalk with the neighboring qubits, and find the Lindblad rates that correspond to dephasing, spontaneous emission, and thermal excitation. However, a purely Markovian model provides a relatively poor fit to the data. We thus construct post-Markovian master equation models that contain a phenomenological bath memory kernel, which account for the non-Markovianity of the dynamics—something a Markovian Lindblad model cannot do. These PMME models provide a closer fit to the tomography data, and also much more accurately predict the future dynamics for new initial states that the models are not already fitted to.

Our PMME construction procedure is an alternative to other methods characterizing qubit noise processes such as process tomography [42–46], machine learning (ML) [47–51], and shadow tomography [52–54]. Compared to the first of these methods, our approach is less demanding in terms of both data collection and analysis, since it relies on state tomography, which generally (but not always [45]) requires a number of measurements, which is quadratically smaller in the Hilbert-space dimension [55]. The ML methods usually require a large training dataset for the solution to converge since the ML model is usually overparametrized. In contrast, at least in the single-qubit setting, we demonstrate that our method requires doing state tomography with only one initial state at multiple time points. Another advantage of our method is that the evaluation of the cost function in our problem is straightforward because the PMME is analytically solvable (see Appendix D). In contrast, the noise models that have been considered in ML so far are numerically challenging because they involve solving the Nakajima-Zwanzig master equation, whose complexity is proportional to the kernel's length [48], or solving a stochastic master equation [49]. In regards to shadow tomography methods, the comparison is less direct since the goal of these methods is not to reconstruct the dynamics at every time point, but rather to use a small set of measurements to predict a much larger set of specific observables.

A similar approach to ours for constructing a dynamical model of the noise channel using the Markovian Lindblad master equation was considered in Ref. [56]. Here, in order to go beyond the Markovian approximation, a model based on a class of time-convolution-less master equations is considered, and a damping-rate function describes the non-Markovian dynamics of the system at the sampling time point [57]. This method provides a discrete description of the noise dynamics but does not directly provide physical insight into quantifying the timescale of the memory effect. Likewise, a more sophisticated and systematic approach, the process-tensor (PT) framework has been proposed [58] and experimentally demonstrated [59,60] to characterize non-Markovian dynamics on actual quantum processors. The PT model, once fully characterized, can be used to interpolate the dynamics between discrete times due to the containment property of the PT map. However, its construction and the interpretation of the memory effect are fairly involved.

Our method provides a continuous dynamical description of the channel beyond the Markov approximation, and does not require full process tomography to construct the PMME model. However, a disadvantage of our method is that it requires one to construct a parametric model; thus, it does not account for the most general noise process. Moreover, because of the specific parametrization that the PMME demands, the resulting optimization problem is no longer convex, i.e., a unique global minimum of the optimization problem is not guaranteed. Additionally, our method requires the consideration of a hierarchy of kernels to find a model that is complex enough to describe the data accurately. But since the PMME is straightforward to solve, model estimation under different kernels is straightforward so long as there is a model selection metric under which the best model among those candidates can be selected. A promising future approach for retaining the physical interpretability and analytical solvability that come with using the PMME is the use of machine learning, especially in light of the recent development of neural ordinary differential equation solvers [61], in order to avoid overfitting and structural errors of the PMME model.

Our work is a proof-of-principle demonstration of the PMME tomography protocol on a fixed-frequency qubit. For future work, we are interested in applying our protocol to characterize the dynamics of a frequency-tunable qubit where dephasing noise is more pronounced, and the qubit's sensitivity to dephasing noise can be varied as a function of the qubit frequency. An extension to the multiqubit case is also a natural next step. This will require a convergence analysis to decide how the fitting dataset scales as the number of qubits increases. In this work, we focus on the free-evolution channel of a single IBMQE qubit, which we show to be highly susceptible to non-Markovian noise. It would be interesting to extend the protocol to characterize non-Markovian effects during computation, in order to understand whether such effects are significant beyond qubit idle times. By explicitly including non-Markovianity in the dynamical characterization and modeling, we expect that it will become possible to use these realistic noise models to improve and tailor error suppression and correction techniques, and ultimately realize high-fidelity quantum control and computation.

ACKNOWLEDGMENTS

APPENDIX A: SYSTEM INFORMATION

The IBMQE device used in this work is Athens, which is a five-qubit processor consisting of superconducting transmon qubits. The main qubit we perform PMME tomography on is qubit 0 (Q0). The relevant device calibration details on the date of data collection are provided in Table IV.

TABLE IV. Qubit calibration information of the *ibmq_athens* processor on the date of data collection.

Dataset	Fig. 4	Fig. 7, Fig. 8	Fig. 9
Date collected	6/25/2021	6/30/2021	7/1/2021
T_1 (μs)	72.6	70.8	75.2
T_2 (μs)	93.4	82.6	62.9
Readout error [10^{-2}]	1.9	0.99	1.00

APPENDIX B: MEASUREMENT ERROR MITIGATION

Measurement error mitigation is performed by using information from calibration experiments to remove any systematic bias in the measurement results [62]. The calibration experiments involve preparing of computational basis states $|j\rangle$, which are then used to learn the response matrix M. The entries $m_{kj} = $ probability(prepare $|j\rangle$|measure bitstring k) represent conditional probabilities. Any subsequent experiment gives us the measured probability vector $\vec{p}(E)$, which is used to infer the true probability of vector $\vec{t}(E) = f[\vec{p}(E), M]$. The most commonly used MEM method—called response matrix-inversion method—defines $\vec{t} = M^{-1}\vec{p}$. Crucially, M^{-1} is not stochastic, so $\vec{t}(E)$ can have negative entries. Recently, a Bayesian solution to the nonstochasticity problem was proposed [63]. In this method, inspired from similar unfolding methods in high-energy physics, we start with a prior truth spectrum $\vec{t}^0$ and update it using Bayes' rule to get

$$t_i^{n+1} = \sum_j \frac{M_{ji}t_i^n}{\sum_k M_{jk}t_k^n} \times p_j.$$

The prior t_i^n is updated using the response matrix M and gives the posterior t_i^{n+1}, and the process proceeds for 100 iterations (in practice this was found to always be sufficient for convergence). After each tomography experiment, the probabilities of each measurement outcome are updated using measurement error mitigation.

APPENDIX C: DAMPING-BASIS CONSTRUCTION

We present a systematic construction of the damping basis, sketched originally in Ref. [64]. For our purposes, the damping basis is simply the basis of left and right eigenoperators (or eigenmatrices) $\{L_i\}$ and $\{R_i\}$, respectively, of the superoperator (e.g., Lindbladian) $\mathcal{L}$.

Let $\mathcal{H}$ denote a d-dimensional Hilbert space, $\mathcal{B}(\mathcal{H})$ the space of linear operators acting on $\mathcal{H}$, and consider a superoperator $\mathcal{L} : \mathcal{B}(\mathcal{H}) \mapsto \mathcal{B}(\mathcal{H})$. We are interested in particular in superoperators in Lindblad form, i.e., Eq. (3a):

$$\mathcal{L}(A) = -i[H, A] + \sum_k \gamma_k \left(V_k A V_k^\dagger - \frac{1}{2}\left\{ V_k^\dagger V_k, A \right\} \right),$$
(C1)

where $A, V_k \in \mathcal{B}(\mathcal{H})$, $H = H^\dagger$, and $\gamma_k > 0 \; \forall k$.

Let $\{F_i\}_{i=0}^{d^2-1}$ denote an orthonormal, Hermitian operator basis for $\mathcal{B}(\mathcal{H})$:

$$\mathrm{Tr}(F_i F_j) = \delta_{ij}, \quad F_i = F_i^\dagger, \quad F_0 = I/\sqrt{d}. \quad \text{(C2)}$$

For example, when $\mathcal{H} = \mathbb{C}^2$ (a qubit), we can choose the normalized Pauli matrices as the operator basis, i.e., $\{F_i\} = \{I, \sigma^x, \sigma^y, \sigma^z\}/\sqrt{2}$.

For any $A \in \mathcal{B}(\mathcal{H})$ we can then expand $A = \sum_i a_i F_i$, and in particular $\mathcal{L}(F_i) = \sum_j \ell_{ji} F_j$ (note the transposed index order), where ℓ is the matrix representation of $\mathcal{L}$ in the given basis, with the matrix elements given by $\ell_{ij} = \mathrm{Tr}[F_i \mathcal{L}(F_j)]$. The basis $\{F_i\}$ "coordinatizes" both $\mathcal{L}$ and the operators in $\mathcal{B}(\mathcal{H})$.

Assume that the superoperator $\mathcal{L}$ satisfies the Hermiticity-preservation condition

$$[\mathcal{L}(A)]^\dagger = \mathcal{L}(A^\dagger). \tag{C3}$$

This is true, in particular, for the Lindbladian (C1), as is easily checked. Let us show that then the matrix ℓ representing $\mathcal{L}$ in the chosen basis is real:

$$\ell_{ij}^* = \mathrm{Tr}\left([\mathcal{L}(F_i)]^\dagger F_j^\dagger\right) = \mathrm{Tr}\left[\mathcal{L}(F_i) F_j\right] = \ell_{ij}. \tag{C4}$$

Thus, after coordinatization the superoperator $\mathcal{L}$ can be seen as a $d^2 \times d^2$-dimensional matrix $\ell \in \mathbb{R}^{d^4}$.

Now assume that A is a right eigenoperator of $\mathcal{L}$ with eigenvalue λ, i.e., $\mathcal{L}(A) = \lambda A$. Then,

$$\mathcal{L}(A) = \sum_{ij} a_i \ell_{ji} F_j = \lambda A = \lambda \sum_i a_i F_i. \tag{C5}$$

Taking the trace of both sides after multiplying from the right by F_k yields $\sum_i a_i \ell_{ki} = \lambda a_k$, i.e.,

$$\ell \vec{a} = \lambda \vec{a}, \tag{C6}$$

where $\vec{a} = (a_0, \ldots, a_{d^2-1})^t$ is a column vector (the superscript t denotes the transpose). Thus, if R_i is a right eigenoperator of $\mathcal{L}$ then its coordinates-vector $\vec{r}_i$ in the expansion

$$R_i = \sum_j (\vec{r}_i)_j F_j. \tag{C7}$$

is a right eigenvector of ℓ. Conversely, by solving the linear algebra problem of finding the set of right eigenvectors $\{\vec{r}_i\}$ of ℓ, we can construct the right eigenoperators of $\mathcal{L}$ using Eq. (C7).

Now consider the set of left eigenvectors $\vec{l}_i^{\,t}$ of ℓ: $\vec{l}_i^{\,t} \ell = \lambda_i \vec{l}_i^{\,t}$. These are also the right eigenvectors of ℓ^t: $\ell^t \vec{l}_i = \lambda_i \vec{l}_i$. Note that the left and right eigenvalues of ℓ are identical since the determinant of a matrix equals the determinant of its transpose.

We define $\mathcal{L}^\dagger$ as usual via the inner product relation

$$\langle \mathcal{L}^\dagger(A), B \rangle = \langle A, \mathcal{L}(B) \rangle, \tag{C8}$$

where we use the Hilbert-Schmidt inner product

$$\langle A, B \rangle \equiv \mathrm{Tr}(A^\dagger B). \tag{C9}$$

Specifically, for the Lindbladian in Eq. (C1), this implies that

$$\mathcal{L}^\dagger(A) = i[H, A] + \sum_k \gamma_k \left(V_k^\dagger A V_k - \frac{1}{2}\left\{V_k^\dagger V_k, A\right\} \right), \tag{C10}$$

as can easily be verified by direct substitution of this form of $\mathcal{L}^\dagger(A)$ into Eq. (C8).

Let us show that ℓ^t is the matrix representation of $\mathcal{L}^\dagger$. To do so, consider the expansion $\mathcal{L}^\dagger(F_i) = \sum_j \tilde{\ell}_{ji} F_j$; we show that in fact $\tilde{\ell} = \ell^t$. Indeed, on the one hand we have from Eq. (C1):

$$(\ell^t)_{ji} = \ell_{ij} = \mathrm{Tr}[F_i \mathcal{L}(F_j)] = -i\mathrm{Tr}(F_i[H, F_j])$$
$$+ \sum_k \gamma_k \left(\mathrm{Tr}[F_i V_k F_j V_k^\dagger] - \frac{1}{2}\mathrm{Tr}[F_i\{V_k^\dagger V_k, F_j\}] \right), \tag{C11}$$

and on the other hand we have from Eq. (C10):

$$\tilde{\ell}_{ji} = \mathrm{Tr}[F_j \mathcal{L}^\dagger(F_i)] = i\mathrm{Tr}(F_j[H, F_i])$$
$$+ \sum_k \gamma_k \left(\mathrm{Tr}[F_j V_k^\dagger F_i V_k] - \frac{1}{2}\mathrm{Tr}[F_j\{V_k^\dagger V_k, F_i\}] \right), \tag{C12}$$

which is easily checked to be equal to the expression for $(\ell^t)_{ji}$ in Eq. (C11) by cycling operators under the trace. Thus,

$$\mathcal{L}^\dagger(F_i) = \sum_j \ell_{ij} F_j, \tag{C13}$$

and the same reasoning that we used above for the right eigenvectors and eigenoperators now leads to the conclusion that if L_i is a right eigenoperator of $\mathcal{L}^\dagger$, i.e., a left eigenoperator of $\mathcal{L}$, then its coordinates-vector $\vec{l}_i$ in the expansion

$$L_i = \sum_j (\vec{l}_i)_j F_j. \tag{C14}$$

is a right eigenvector of ℓ. Conversely, by solving the linear algebra problem of finding the set of right eigenvectors $\{\vec{l}_i\}$

of ℓ, we can construct the right eigenoperators of $\mathcal{L}$ using Eq. (C14).

Finally, it is well known that each left eigenvector is orthogonal to all right eigenvectors except its corresponding one (the one it shares an eigenvalue with), and vice versa [65]. By choice of normalization, the inner products of corresponding left and right eigenvectors can always be made unity for any matrix with nondegenerate eigenvalues. Assume nondegeneracy and that we have normalized ℓ's inner products of corresponding left and right eigenvectors, i.e., $\vec{l}_i \cdot \vec{r}_j = \delta_{ij}$. Let us show in which sense this property is inherited by the left and right eigenoperators of $\mathcal{L}$:

$$\mathrm{Tr}(L_i R_j) = \sum_{kl}(\vec{l}_i)_k(\vec{r}_j)_l \mathrm{Tr}(F_k F_l) = \sum_k (\vec{l}_i)_k(\vec{r}_j)_k$$
$$= \vec{l}_i \cdot \vec{r}_j = \delta_{ij}. \tag{C15}$$

Note that $\mathrm{Tr}(L_i R_j) \neq \langle L_i, R_j \rangle$ since we do not take the Hermitian conjugate of L_i under the trace [this only becomes possible if ℓ is symmetric, since we would need its eigenvalues to be real in order for $\mathrm{Tr}(L_i R_j) = \mathrm{Tr}(L_i^\dagger R_j)$ to hold].

APPENDIX D: ANALYTICAL SOLUTION OF THE PMME

Here we present the analytical solution of the PMME, Eq. (1), for our model. We take the Laplace transform, and the PMME becomes

$$s\tilde{\rho}(s) - \rho(0) = \mathcal{L}_0\tilde{\rho}(s)$$
$$+ \mathcal{L}_1 \mathrm{Lap}\left[k(t)\exp\left(\mathcal{L}_0 + \mathcal{L}_1\right)t\right]\tilde{\rho}(s). \tag{D1}$$

To deal with $e^{\mathcal{L}t} := e^{(\mathcal{L}_0 + \mathcal{L}_1)t}$, it is convenient to work in the damping basis of $\mathcal{L}$, as defined in Appendix C. Recall that the sets of right and left eigenoperators of $\mathcal{L}$, $\{R_i\}$ and $\{L_i\}$, are complete and mutually orthonormal in the sense of Eq. (C15). We therefore expand ρ in the basis of right eigenoperators of $\mathcal{L}$:

$$\rho(t) = \sum_i \mu_i(t)R_i, \tag{D2}$$

where the expansion coefficients are

$$\mu_j(t) = \sum_i \mu_i(t)\mathrm{Tr}\left(L_j R_i\right) = \mathrm{Tr}\left[L_j \rho(t)\right]. \tag{D3}$$

Substituting Eq. (D2) into the PMME Eq. (1), we obtain

$$\sum_i \frac{\partial \mu_i(t)}{\partial t}R_i = \sum_i \mu_i(t)\mathcal{L}_0 R_i$$
$$+ \sum_i \int_0^t dt'k(t')\exp\left(\lambda_i t\right)\mu_i(t - t')\mathcal{L}_1 R_i. \tag{D4}$$

Notice that if we assume that $[\mathcal{L}_0, \mathcal{L}_1] = 0$ (as is the case for us), then $\mathcal{L}_0$ and $\mathcal{L}_1$ both commute with $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1$ and hence share the same set of left and right eigenoperators with it, i.e., $\mathcal{L}_0(R_i) = \lambda_i^0 R_i$, $\mathcal{L}_0^\dagger(L_i) = \lambda_i^0 L_i$, $\mathcal{L}_1(R_i) = \lambda_i^1 R_i$, $\mathcal{L}_1^\dagger(L_i) = \lambda_i^1 L_i$. Multiplying both sides of Eq. (D4) by L_j from left and taking the trace, we obtain, under this assumption:

$$\frac{\partial \mu_i(t)}{\partial t} = \lambda_i^0 \mu_i(t) + \lambda_i^1 \int_0^t dt'k(t')\exp\left[\lambda_i t'\right]\mu_i(t - t'). \tag{D5}$$

Take the Laplace transform of both sides and use the shifting property of the Laplace transform, we have

$$s\tilde{\mu}(s) - \mu_i(0) = \lambda_i^0 \tilde{\mu}_i(s) + \lambda_i^1 \mathrm{Lap}\left[k(t)e^{\lambda_i t}\right]\tilde{\mu}_i(s) \tag{D6a}$$
$$= \lambda_i^0 \tilde{\mu}_i(s) + \lambda_i^1 \tilde{k}(s - \lambda_i)\tilde{\mu}_i(s). \tag{D6b}$$

Therefore,

$$\tilde{\mu}_i(s) = \frac{1}{s - \lambda_i^0 - \lambda_i^1 k(s - \lambda_i)}. \tag{D7}$$

Taking the inverse Laplace transform:

$$\mu_i(t) = \xi_i(t)\mu_i(0), \tag{D8}$$

where

$$\xi_i(t) = \mathrm{Lap}^{-1}\left[\frac{1}{s - \lambda_i^0 - \lambda_i^1 k(s - \lambda_i)}\right], \tag{D9a}$$

$$\mu_i(0) = \mathrm{Tr}\left[L_i \rho(0)\right]. \tag{D9b}$$

1. Solution with the specific Lindbladian and kernels

Choosing the operator basis as $\{F_i\} = \{I, \sigma^x, \sigma^y, \sigma^z\}/\sqrt{2}$, we follow the methodology of Appendix C and find the matrix representation of $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1$ of the specific

PMME model we seek to construct in Sec. II B to be

$$\ell = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{\Gamma_s}{2} - 2\gamma_z & -\omega_z & 0 \\ 0 & \omega_z & -\frac{\Gamma_s}{2} - 2\gamma_z & 0 \\ \gamma_- & -\gamma_+ & 0 & 0 & -\Gamma_s \end{pmatrix}. \quad (D10)$$

The eigenvalues of ℓ are

$$\{\lambda_i\} = \left\{0, -\frac{1}{2}\Gamma_s - 2\gamma_z + i\omega_z, -\frac{1}{2}\Gamma_s - 2\gamma_z - i\omega_z, -\Gamma_s\right\}. \quad (D11)$$

It is straightforward to compute and normalize the corresponding right and left eigenvectors $\{\vec{r}_i, \vec{l}_i\}_{i=1}^4$ such that they are mutually orthonormal, i.e., $\vec{l}_i \cdot \vec{r}_j = \delta_{ij}$. This allows us to find the right and left eigenoperators of $\mathcal{L}$ using Eqs. (C7) and (C14), which yields

$$R_1 = \begin{pmatrix} \frac{1}{\Gamma_r+1} & 0 \\ 0 & \frac{\Gamma_r}{\Gamma_r+1} \end{pmatrix}, R_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, R_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad (D12a)$$

$$R_4 = \begin{pmatrix} -\frac{1}{\Gamma_r+1} & 0 \\ 0 & \frac{1}{\Gamma_r+1} \end{pmatrix} \quad (D12b)$$

$$L_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, L_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, L_3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (D12c)$$

$$L_4 = \begin{pmatrix} -\Gamma_r & 0 \\ 0 & 1 \end{pmatrix}. \quad (D12d)$$

It is simple to verify that this set satisfies Eq. (C15) as required.

Applying $\mathcal{L}_0$ and $\mathcal{L}_1$ to $\{R_i\}$, we find the corresponding sets of eigenvalues:

$$\{\lambda_i^0\} = \left\{0, -\frac{\Gamma_s}{2} + i\omega_z, -\frac{\Gamma_s}{2} - i\omega_z, -\Gamma_s\right\}, \quad (D13a)$$

$$\{\lambda_i^1\} = \{0, -2\gamma_z, -2\gamma_z, 0\}. \quad (D13b)$$

Next, we need to evaluate Eq. (D9a) with the specific forms of the kernels we have chosen. Regardless of the kernel,

$$\tilde{\xi}_1(s) = \frac{1}{s} \iff \xi_1(t) = 1, \quad (D14a)$$

$$\tilde{\xi}_4(s) = \frac{1}{s + \Gamma_s} \iff \xi_4(t) = e^{-\Gamma_s t}, \quad (D14b)$$

while for $i = 2, 3$:

$$f(t) \equiv \xi_2(t) = \xi_3^*(t), \quad (D15a)$$

$$= \text{Lap}^{-1}\left[\frac{1}{s - \lambda_2^0 - \lambda_2^1 \tilde{k}(s - \lambda_2)}\right]. \quad (D15b)$$

For the exponentially decaying kernel in Eq. (7), we make the parameter substitution $x \equiv 2 + b_0/\gamma_z$ and $y \equiv \gamma_s/2\gamma_z - i\omega_z/\gamma_z$ and transform the variables in the Laplace transform correspondingly as $s = \gamma_z z$ and $\tau = \gamma_z t$, to get

$$f(\tau) = \text{Lap}^{-1}\left[\tilde{f}(z)\right], \quad (D16a)$$

$$\tilde{f}(z) = \frac{z + x + y}{(z + y)^2 + x(z + y) + 2}. \quad (D16b)$$

The analytical solution can be found by using the residue theorem [66]:

$$f(\tau) = \text{sum of residues of } e^{z\tau}\tilde{f}(z) \text{ at poles of } \tilde{f}(z). \quad (D17)$$

The rational function $\tilde{f}(z)$ has two poles z_1 and z_2

$$z_{1,2} = \frac{1}{2}\left(-x - 2y \pm \sqrt{D}\right), \quad D \equiv x^2 - 8, \quad (D18)$$

and using Eq. (D17):

$$f(\tau) = e^{z_1\tau}\frac{z_1 + y + x}{z_1 - z_2} + e^{z_2\tau}\frac{z_2 + y + x}{z_2 - z_1}. \quad (D19)$$

For the other kernel in Eq. (8), we make the parameter substitution $a_0 \equiv \gamma_z^2 x$, $y \equiv \gamma_s/2\gamma_z - i(\omega_z/\gamma_z) + 2$, $a_1 \equiv \gamma_z w$, $b_0 \equiv \gamma_z^2 u$, $b_1 \equiv \gamma_z v$, and transform the variables in the Laplace transform correspondingly as $s = \gamma_z z$ and $\tau = \gamma_z t$, to get

$$f(\tau) = \text{Lap}^{-1}\left[\tilde{f}(z)\right], \quad (D20a)$$

$$\tilde{f}(z) = \frac{p_1(z)}{(z + y - 2)p_1(z) + 2w(z + y) + 2x}, \quad (D20b)$$

where $p_1(z) = (z + y)^2 + v(z + y) + u$. The analytical solution can therefore be found in terms of the roots of the cubic polynomial

$$(z + y - 2)p_1(z) + 2w(z + y) + 2x \quad (D21a)$$

$$= z^3 + c_2 z^2 + c_1 z + c_0 = 0, \quad (D21b)$$

where the coefficients are

$$c_2 = 3y + v - 2, \quad (D22a)$$

$$c_1 = 3y^2 + 2vy - 4y + 3w - 2v, \quad (D22b)$$

$$c_0 = y^3 + vy^2 - 2y^2 + 2wy - 2vy + 2x - 2w. \quad (D22c)$$

The corresponding depressed cubic is found by the substitution $z = z' - c_2/3$,

$$r(z') = z'^3 + pz' - q = 0, \quad (D23)$$

where the coefficients are

$$p = \frac{3c_1 - c_2^2}{3}, \tag{D24}$$

$$q = \frac{9c_1c_2 - 27c_0 - 2c_2^3}{27}, \tag{D25}$$

which yields the cubic discriminant

$$D = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2. \tag{D26}$$

Further denote

$$S = \sqrt[3]{\frac{q}{2} + \sqrt{D}}, \quad T = \sqrt[3]{\frac{q}{2} - \sqrt{D}}. \tag{D27}$$

FIG. 7. The tomography dataset and the corresponding model predictions used to calculate the degree of non-Markovianity in Fig. 6(a),(b) when the spectator qubits are initialized in the ground state. (a0)–(c0) The fitting dataset with the qubit initialized in $|\psi_0(0)\rangle$ that is used to find the best-fit Lindblad model $\mathcal{M}_0$ (orange lines), the PMME model with the type 1 kernel $\mathcal{M}_1$, and the PMME model with the type 2 kernel $\mathcal{M}_2$. (a1)–(c2) The tomography dataset with qubit initialized in $|\psi_{+x}\rangle$, $|\psi_{-x}\rangle$ and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 6 (a),(b). (a3)–(c4) The tomography dataset with the qubit initialized in $|\psi_{+y}\rangle$, $|\psi_{-y}\rangle$, and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 10 (a),(b). (d0)–(d4) The distance between the tomographically constructed state and the state predicted by the best-fit models.
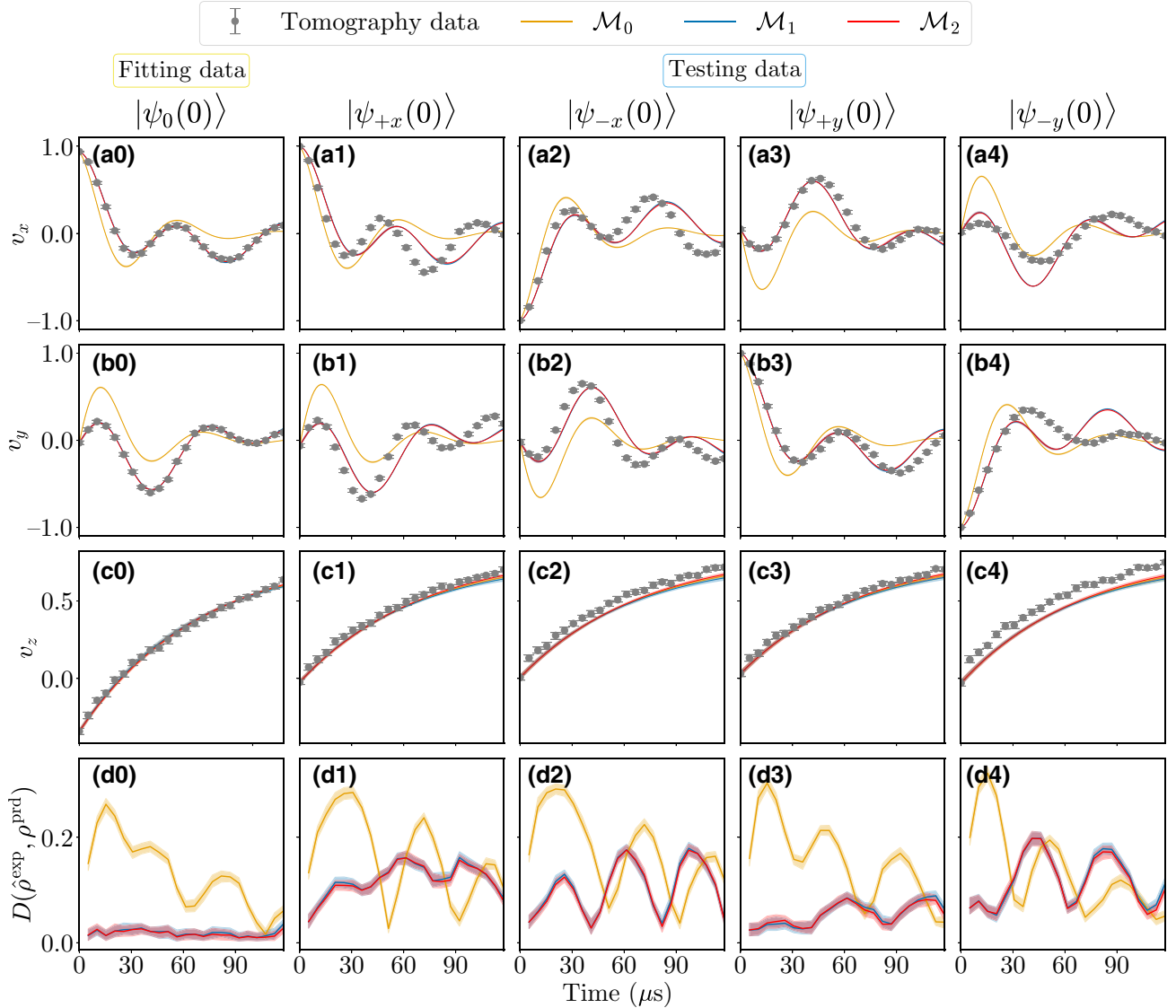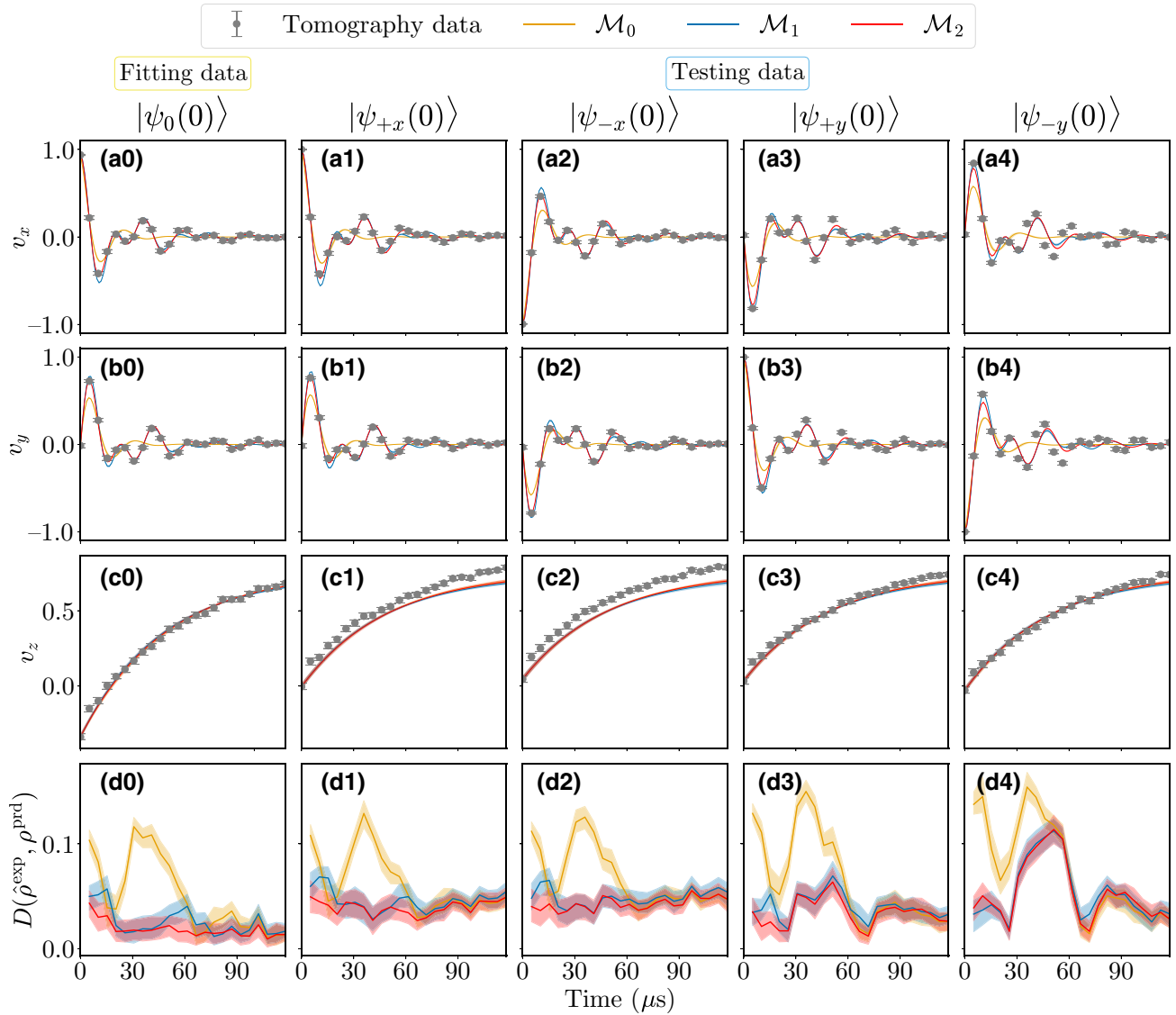
FIG. 8. The tomography dataset and the corresponding model predictions used to calculate the degree of non-Markovianity in Fig. 6(c),(d) when the spectator qubits are initialized in the excited state. (a0)–(c0) The fitting dataset with the qubit initialized in $|\psi_0(0)\rangle$ that is used to find the best-fit Lindblad model $\mathcal{M}_0$ (orange lines), the PMME model with the type 1 kernel $\mathcal{M}_1$, and the PMME model with the type 2 kernel $\mathcal{M}_2$. (a1)–(c2) The tomography dataset with qubit initialized in $|\psi_{+x}\rangle$, $|\psi_{-x}\rangle$ and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 6(c),(d). (a3)–(c4) The tomography dataset with the qubit initialized in $|\psi_{+y}\rangle$, $|\psi_{-y}\rangle$, and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 10(c),(d). (d0)–(d4) The distance between the tomographically constructed state and the state predicted by the best-fit models.

The zeros of the cubic are

$$z_1 = -\frac{2x}{3} + (S+T), \tag{D28}$$

$$z_2 = -\frac{2x}{3} - \frac{1}{2}(S+T) + \frac{i}{2}\sqrt{3}(S-T), \tag{D29}$$

$$z_3 = -\frac{2x}{3} - \frac{1}{2}(S+T) - \frac{i}{2}\sqrt{3}(S-T). \tag{D30}$$

This completes the exact solution of the PMME.

APPENDIX E: COMPLETE POSITIVITY OF THE PMME

The complete positivity of the PMME is not guaranteed because of the freedom in choosing the kernel $k(t)$. A complete positivity test for the PMME was provided in Ref. [26]. Below we apply this test to the kernels specified in the model $\mathcal{M}_i$.

The solution of the PMME can be viewed as a map Φ acting on the operators represented by $d \times d$
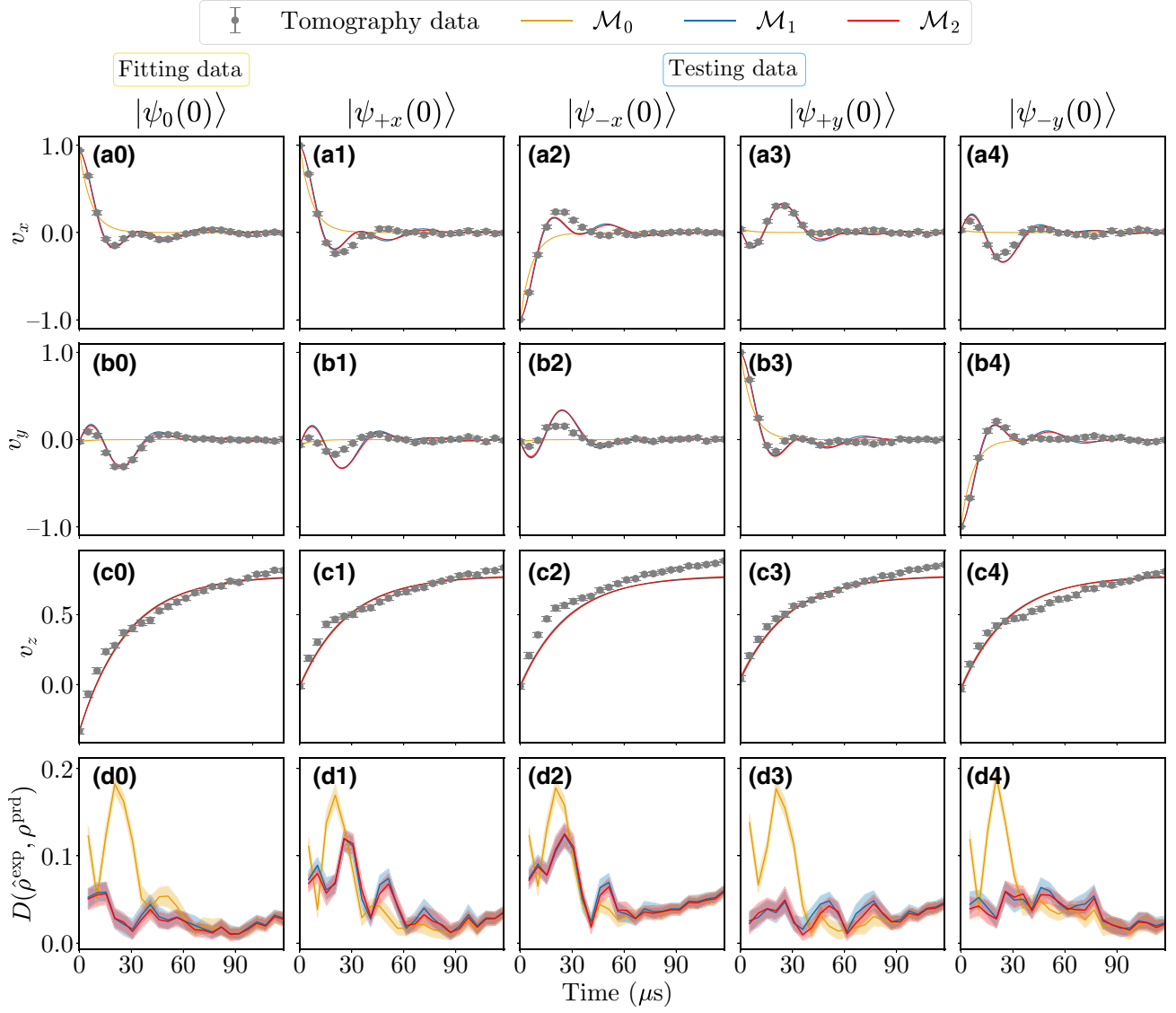
FIG. 9. The tomography dataset and the corresponding model predictions used to calculate the degree of non-Markovianity in Fig. 6(e),(f) when the spectator qubits are initialized in the $|+\rangle$ state. (a0)–(c0) The fitting dataset with the qubit initialized in $|\psi_0(0)\rangle$ that is used to find the best-fit Lindblad model $\mathcal{M}_0$ (orange lines), the PMME model with the type 1 kernel $\mathcal{M}_1$, and the PMME model with the type 2 kernel $\mathcal{M}_2$. (a1)–(c2) The tomography dataset with qubit initialized in $|\psi_{+x}\rangle$, $|\psi_{-x}\rangle$ and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 6(e),(f). (a3)–(c4) The tomography dataset with the qubit initialized in $|\psi_{+y}\rangle$, $|\psi_{-y}\rangle$, and the prediction from the best-fit models from the fitting dataset in (a0)–(c0). The datasets are used to evaluate the degree of non-Markovianity in Fig. 10(e),(f). (d0)–(d4) The distance between the tomographically constructed state and the state predicted by the best-fit models.

matrices, where d is the dimension of the Hilbert space $\mathcal{H} = \text{span}\{|i\rangle\}_{i=1}^{d}$. Using Eq. (D2),

$$\rho(t) = \sum_i \mu_i(t) R_i = \sum_i \xi_i(t)\mu_i(0) R_i \tag{E1a}$$

$$= \sum_i \xi_i(t)\text{Tr}[L_i\rho(0)] R_i = \Phi[\rho(0)], \tag{E1b}$$

where

$$\Phi[X] \equiv \sum_i \xi_i(t)\text{Tr}[L_iX] R_i. \tag{E2}$$

Let $|\phi\rangle = \sum_i |i\rangle \otimes |i\rangle$ be a maximally entangled state in $\mathcal{H} \otimes \mathcal{H}$. According to Choi's theorem [67], Φ is CP if and only if the Choi matrix $C \geq 0$, where

$$C = (\mathbb{I} \otimes \Phi)|\phi\rangle\langle\phi| = \sum_{ij} |i\rangle\langle j| \otimes \Phi[|i\rangle\langle j|]. \tag{E3}$$

We construct the Choi matrix for the PMME. Let us pick the basis state $|i\rangle$ to be a column vector of zeros, except for a 1 in position i; we have,

$$C = \sum_{ij} |i\rangle\langle j| \otimes \sum_k \xi_k(t) \text{Tr}\left[L_k|i\rangle\langle j|\right] R_k, \quad \text{(E4a)}$$

$$= \sum_k \xi_k(t) \sum_{ij} |i\rangle\langle j| \otimes \langle j|L_k|i\rangle R_k, \quad \text{(E4b)}$$

$$= \sum_k \xi_k(t) \sum_{ij} |i\rangle\langle j| \left(L_k^T\right)_{ij} \otimes R_k. \quad \text{(E4c)}$$

Hence,

$$C = \sum_k \xi_k(t) L_k^T \otimes R_k > 0 \quad \text{(E5)}$$

is the complete positivity condition for the kernel, under a given Lindbladian $\mathcal{L}$ and its set of left and right eigenvectors.

For the Lindbladian $\mathcal{L}$ in Eq. (3a) and its set of left and right eigenvectors in Eq. (D12), the Choi matrix is

$$C = \begin{pmatrix} \frac{1+\Gamma_r\xi_4}{1+\Gamma_r} & 0 & 0 & \xi_2 \\ 0 & \frac{\Gamma_r(1-\xi_4)}{1+\Gamma_r} & 0 & 0 \\ 0 & 0 & \frac{1-\xi_4}{1+\Gamma_r} & 0 \\ \xi_3 & 0 & 0 & \frac{\Gamma_r+\xi_4}{1+\Gamma_r} \end{pmatrix}. \quad \text{(E6)}$$

Its eigenvalues are found to be

$$\lambda_1^c = \frac{1-\xi_4}{1+\Gamma_r}, \quad \text{(E7a)}$$

$$\lambda_2^c = \frac{\Gamma_r(1-\xi_4)}{1+\Gamma_r}, \quad \text{(E7b)}$$

$$\lambda_{3,4}^c = \frac{1+\xi_4}{2} \quad \text{(E7c)}$$

$$\pm \sqrt{\left(\frac{\xi_4+1}{2}\right)^2 - \frac{\Gamma_r+\xi_4+\Gamma_r^2\xi_4+\Gamma_r\xi_4^2-|\xi_2|^2}{(\Gamma_r+1)^2}}. \quad \text{(E7d)}$$

Therefore, the PMME in this case corresponds to a CP map if and only if

$$|\xi_4| < 1, \quad \text{(E8a)}$$

$$|\xi_2| = |\xi_3| < \frac{\sqrt{(\Gamma_r+\xi_4)(1+\Gamma_r\xi_4)}}{1+\Gamma_r} < 1, \quad \text{(E8b)}$$

which is a condition on the problem parameters γ_z, γ_+, γ_- and the kernel parameters $\vec{a}$ and $\vec{b}$.

APPENDIX F: PMME MODEL CONSTRUCTION RESULT FOR DIFFERENT SPECTATOR QUBIT STATES

In this section we present data supplementing the results of the PMME model construction reported in Fig. 4, for the following three initial states of the spectator qubits: ground state $|0\rangle$, the excited state $|1\rangle$, and the $|+\rangle$ state. This data is also used to calculate the degree of non-Markovianity reported in Fig. 6, which uses different initial states of the main qubit. Figures 7, 8, and 9 show the fitting data used to construct the model and the testing data to validate it. The initial states in the testing datasets are $\{|\psi_{+x}\rangle = |+\rangle, |\psi_{-x}\rangle = |-\rangle, |\psi_{+y}\rangle = |+i\rangle, |\psi_{-y}\rangle = |-i\rangle\}$ and they are used to evaluate the degree of non-Markovianity as in Eq. (14), plotted in Figs. 6 and 10. The initial state pairs $\rho_1(0) = |+\rangle\langle+|$ and $\rho_2(0) = |-\rangle\langle-|$ (or $\rho_1(0) = |+i\rangle\langle+i|$ and $\rho_2(0) = |-i\rangle\langle-i|$) are optimal pairs such that they feature a maximal flow of information from the environment back to the system [68].

The yellow, blue, and red solid lines in Figs. 7, 8, and 9 represent the constructed models $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$, respectively, with their best-fit model parameters summarized in Table V. On the fitting dataset, we find that the Lindblad model $\mathcal{M}_0$ does not adequately describe the data, while the PMME models $\mathcal{M}_1$ and $\mathcal{M}_2$ describe the data accurately. The kernels of the constructed PMME models are plotted in Fig. 11. Although $\mathcal{M}_2$ has a more elaborate kernel with more free parameters, it does not blue necessarily provide a better fit to the data; $\mathcal{M}_1$ and $\mathcal{M}_2$ result in similar fits, which suggests that we do not need a very complicated kernel to significantly improve upon the standard Lindblad model. On the testing dataset, the constructed models provide a qualitatively adequate prediction for different initial states, but deviations do arise relative to the experimentally constructed states (see Figs. 7–9 columns 1, 2, and 4). Due to system fluctuations, which the PMME models cannot capture, the model constructed from the fitting dataset may lose some of its predictive power for the testing dataset. Examples of such system fluctuations include fluctuations of qubit T_1 relaxation time and $1/f$ noise in the qubit frequency. As further evidence of such fluctuations, the data shown in Fig. 4(a0)–(d0) and Fig. 7(a0)–(d0) are taken under nominally identical conditions but nonetheless show different behavior. These data are taken days apart, during which time the system parameters drifted.

Nonetheless, compared with the Lindblad model, the PMME models provide higher levels of agreement for the fitting datasets and a more accurate prediction for the testing datasets in all spectator qubit configurations. This can be seen in the bottom row of Figs. 7–9, and is summarized in Fig. 12, which compares the models using the Akaike information criterion (AIC) and the trace-distance metric discussed below.
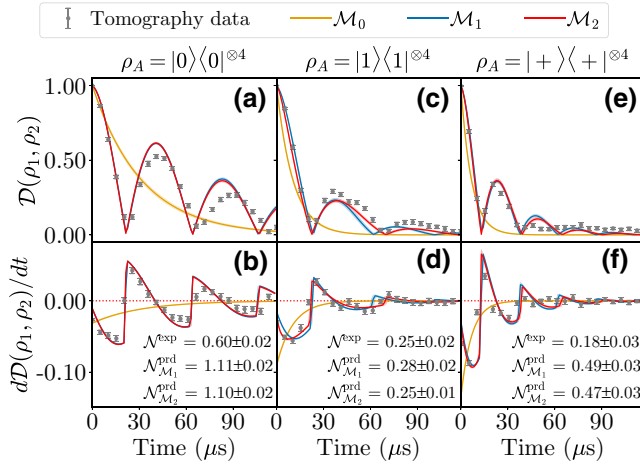
FIG. 10. Non-Markovianity of qubit free-evolution dynamics for spectator qubits in the ground state (a),(b), the excited state (c),(d), and the $|+\rangle$ state (e),(f). (a),(c),(e) The trace-norm distance $\mathcal{D}[\rho_1(t),\rho_2(t)]$ predicted by the best-fit models (solid lines) and experimentally measured by performing free-evolution tomography with a pair of initial states $\rho_1(0) = |+i\rangle\langle+i|$ and $\rho_2(0) = |-i\rangle\langle-i|$ (gray circles). (b),(d),(f) The derivative $\sigma(t)$, defined in Eq. (14), predicted by the best-fit models (solid lines), and approximated experimentally using forward differencing based on the tomography data in (a),(c),(d) (gray circles).

Legend: Tomography data — $\mathcal{M}_0$ — $\mathcal{M}_1$ — $\mathcal{M}_2$

Panel (a) $\rho_A = |0\rangle\langle0|^{\otimes4}$; Panel (c) $\rho_A = |1\rangle\langle1|^{\otimes4}$; Panel (e) $\rho_A = |+\rangle\langle+|^{\otimes4}$

Panel (b): $\mathcal{N}^{\text{exp}} = 0.60\pm0.02$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_1} = 1.11\pm0.02$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_2} = 1.10\pm0.02$

Panel (d): $\mathcal{N}^{\text{exp}} = 0.25\pm0.02$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_1} = 0.28\pm0.02$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_2} = 0.25\pm0.01$

Panel (f): $\mathcal{N}^{\text{exp}} = 0.18\pm0.03$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_1} = 0.49\pm0.03$, $\mathcal{N}^{\text{prd}}_{\mathcal{M}_2} = 0.47\pm0.03$

To compare the goodness of the fit of the models and the simplest model that best describes the fitting dataset, we use the AIC [35], which is defined as

$$\mathcal{A} = -2\ln\left(\hat{L}(\hat{\theta}|D)\right) + 2p, \tag{F1}$$

where $\hat{L}$ denotes the likelihood function and p is the number of free model parameters. The second term in the AIC ($2p$) is called the bias term and it penalizes the models with higher complexity. The AIC can be interpreted as the "information loss" (in the Kullback-Leibler divergence sense) of using some candidate model to approximate the "true" model. Akaike showed that the maximized log-likelihood in the first term of Eq. (F1) is a biased estimator, and that under certain assumptions, the bias correction approximately equals p. The calculation of the first term in AIC depends strongly on the sample size used, and the bias-correction term becomes exact when the sample size diverges. For these reasons, the numerical value of AIC has no intrinsic significance as such. With these caveats in mind, we can use the AIC metric to find a model that accurately describes the data and at the same time avoids overfitting, as lower (more negative) values indicate a better model.

The likelihood function $\hat{L}$ is defined over the observed dataset, and quantifies the likelihood of observing the dataset as a function of the model parameters $\vec{\theta}$. It measures how well the data supports that particular choice of

TABLE V. Best-fit parameters of the models $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_2$ constructed using the fitting datasets in Fig. 4, Fig. 7, Fig. 8 and Fig. 9. The values in the parentheses correspond to the 2σ error bars (95% confidence intervals) estimated using the bootstrap method. Parameters with best-fit values smaller than 1e-5 are expressed as zero as their effects on the dynamical predictions are negligible and well below the measurement precision of the tomography experiment. For the same reason, we report the non-zero best-fit parameters with two significant figures and their confidence intervals with one significant figures.

|	ω_z (MHz)	γ_z (MHz)	γ_- (MHz)	γ_+ (MHz)	b_0	a_0	b_1
				Fig. 4			
$\mathcal{M}_0$	$9.6\times10^{-2}\,(3\times10^{-4})$	$1.2\times10^{-2}\,(8\times10^{-4})$	$1.9\times10^{-2}\,(8\times10^{-4})$	$1.9\times10^{-3}\,(3\times10^{-4})$			
$\mathcal{M}_1$	$5.6\times10^{-2}\,(2\times10^{-4})$	$3.3\times10^{-3}\,(2\times10^{-5})$	$1.7\times10^{-2}\,(3\times10^{-4})$	$1.1\times10^{-3}\,(1\times10^{-4})$	0		
$\mathcal{M}_2$	$5.6\times10^{-2}\,(2\times10^{-4})$	$3.3\times10^{-3}\,(2\times10^{-5})$	$1.7\times10^{-2}\,(3\times10^{-4})$	$1.1\times10^{-3}\,(1\times10^{-4})$	$9.6\times10^{-5}\,(3\times10^{-6})$	$5.5\times10^{-5}\,(9\times10^{-5})$	0
				Fig. 7			
$\mathcal{M}_0$	$-5.3\times10^{-2}\,(3\times10^{-4})$	$1.2\times10^{-2}\,(7\times10^{-4})$	$1.3\times10^{-2}\,(8\times10^{-4})$	$1.5\times10^{-3}\,(4\times10^{-4})$			
$\mathcal{M}_1$	$-1.7\times10^{-2}\,(2\times10^{-4})$	$2.8\times10^{-3}\,(3\times10^{-5})$	$1.4\times10^{-2}\,(8\times10^{-4})$	$1.8\times10^{-3}\,(4\times10^{-4})$	$2.3\times10^{-3}\,(1\times10^{-3})$		
$\mathcal{M}_2$	$-1.7\times10^{-2}\,(2\times10^{-4})$	$3.0\times10^{-3}\,(9\times10^{-5})$	$1.3\times10^{-2}\,(8\times10^{-4})$	$1.3\times10^{-3}\,(4\times10^{-4})$	$8.6\times10^{-2}\,(2\times10^{-2})$	$9.7\times10^{-2}\,(2\times10^{-2})$	0
				Fig. 8			
$\mathcal{M}_0$	$-1.2\times10^{-1}\,(5\times10^{-3})$	$4.3\times10^{-2}\,(2\times10^{-3})$	$1.8\times10^{-2}\,(9\times10^{-4})$	$2.4\times10^{-3}\,(4\times10^{-4})$			
$\mathcal{M}_1$	$-1.3\times10^{-1}\,(7\times10^{-4})$	$3.7\times10^{-3}\,(2\times10^{-4})$	$1.9\times10^{-2}\,(1\times10^{-3})$	$2.8\times10^{-3}\,(4\times10^{-4})$	$4.7\times10^{-2}\,(4\times10^{-3})$		
$\mathcal{M}_2$	$-1.3\times10^{-1}\,(7\times10^{-4})$	$1.0\times10^{-2}\,(1\times10^{-3})$	$1.8\times10^{-2}\,(9\times10^{-4})$	$2.4\times10^{-3}\,(4\times10^{-4})$	$1.0\times10^{-1}\,(2\times10^{-2})$	$5.2\times10^{-1}\,(1\times10^{-1})$	0
				Fig. 9			
$\mathcal{M}_0$	0	$6.5\times10^{-2}\,(2\times10^{-3})$	$3.2\times10^{-2}\,(1\times10^{-3})$	$4.0\times10^{-3}\,(3\times10^{-4})$			
$\mathcal{M}_1$	$-2.5\times10^{-2}\,(9\times10^{-4})$	$8.6\times10^{-3}\,(3\times10^{-4})$	$3.2\times10^{-2}\,(1\times10^{-3})$	$4.1\times10^{-3}\,(3\times10^{-4})$	$3.1\times10^{-2}\,(3\times10^{-3})$		
$\mathcal{M}_2$	$-2.5\times10^{-2}\,(9\times10^{-4})$	$1.1\times10^{-2}\,(9\times10^{-4})$	$3.2\times10^{-2}\,(1\times10^{-3})$	$4.0\times10^{-3}\,(3\times10^{-4})$	$1.1\times10^{-1}\,(4\times10^{-2})$	$2.0\times10^{-1}\,(7\times10^{-2})$	0
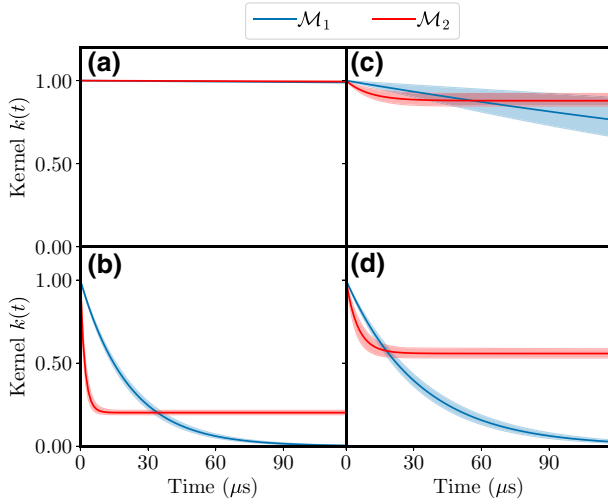
FIG. 11. The kernels in the constructed models $\mathcal{M}_1$ and $\mathcal{M}_2$ using the fitting datasets in Fig. 4(a), Fig. 7(b), Fig. 8(c), and Fig. 9(d). The shaded regions denote the 95% confidence region of the kernel function due to the uncertainty in the best-fit kernel parameters in Table V.

parameters. Since each tomography sample $\hat{\rho}$ is independently drawn, the likelihood $\hat{L}$ is a product of conditional probabilities:

$$\hat{L}(\vec{\theta}|\mathcal{D}) = \prod_i \prod_k p_k(t_i; \vec{\theta}), \quad \text{(F2)}$$

where p_k is the probability of observing the data point in measurement basis k at time t_i.

We report the AIC values computed in this manner for the fitting data using the three models in Fig. 12 (purple squares), for three different initial states of the ancilla qubits. However, as mentioned above, the numerical value of the AIC is not intrinsically meaningful. In practice, it is convenient to scale AICs with respect to the minimum AIC value among all models:

$$\Delta_i = \mathcal{A}_i - \min_i \mathcal{A}_i, \quad \text{(F3)}$$

where $\min_i \mathcal{A}_i$ is the AIC value of the best model in the set. As seen in Fig. 12, we find that the minimum is achieved for the $\mathcal{M}_2$ model. The AIC difference Δ_i estimates the information loss when using model i rather than the estimated best model. Hence, the larger Δ_i, the less plausible is model i. Some guidelines for the interpretation of AIC difference in the case of nested models are given in Ref. [69], as summarized in Table VI. Given the much larger $\Delta_0 = \mathcal{A}_0 - \mathcal{A}_1$ value, we find that the data in Fig. 4 is considerably less in favor of the Lindblad model $\mathcal{M}_0$, despite the fact that it has the smallest number of free parameters.

Another metric we use here for comparison is the trace-norm distance between the experimentally constructed
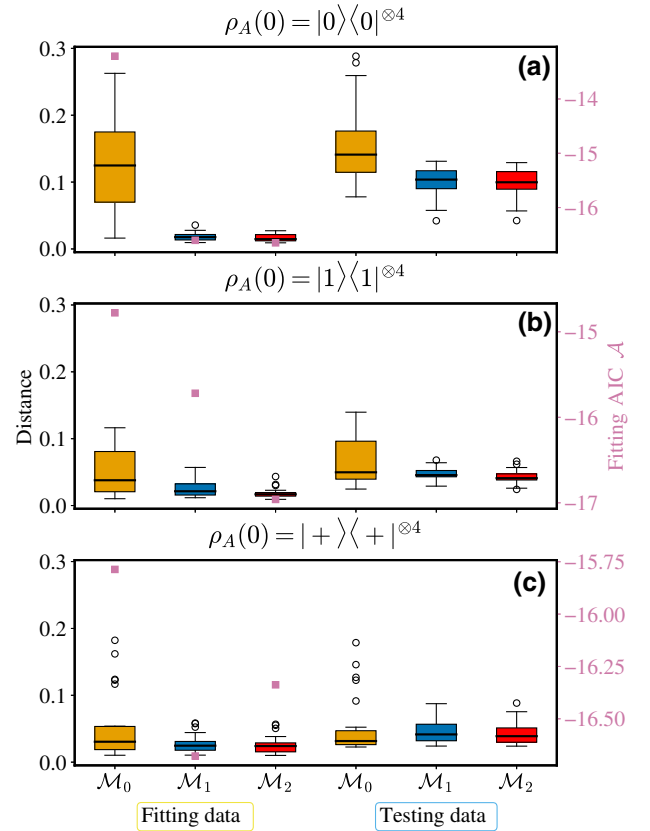

FIG. 12. Summary of the results of running the PMME tomography protocol on the IBMQE processor ibmq_athens, showing how well the model candidates describe the fitting dataset and the testing dataset in Fig. 7 for the spectator qubits in the ground state (a), for those in Fig. 8 for the spectator qubits in the excited state (b) and for those in Fig. 9 for the spectator qubits in the $|+\rangle$ state (c). The box plots show the trace-norm distance (see text) and the median is reported as the middle value of $\{\mathcal{D}_j\}$. The lower line of the box corresponds to the lower quartile of the data (25th percentile, Q1), and the upper line of the box corresponds to the upper quartile of the data (75th percentile, Q3). Let IQR denote the interquartile range: IQR = Q3–Q1. The outliers, plotted in circles, are the data outside the range (Q1–1.5 × IQR, Q3 + 1.5 × IQR).

state and the model predicted state during its evolution. Let D_j^k denote the distance between the experimentally constructed state $\hat{\rho}_k^{\text{exp}}$ at time t_j with initial state $\rho_k(0) = |\psi_k(0)\rangle\langle\psi_k(0)|$ and that predicted by the model $\rho_k^{\text{prd}}(t_j)$, i.e., $\mathcal{D}_k^j = \mathcal{D}[\hat{\rho}_k^{\text{exp}}, \rho_k^{\text{prd}}(t_j)]$, the values reported in the box plot in Fig. 12 are averaged over the test dataset with different initial states: $\mathcal{D}_j = 1/4 \sum_{k \in \{1,2,3,4\}} \mathcal{D}_j^k$. After doing this averaging, we arrive at an array $\{\mathcal{D}_j\}_{j=1}^{j=24}$, corresponding to the trace distance, again, averaged over four different initial states in the testing dataset, between the experimentally constructed states and the model's predicted states at 24 sampled time points during the evolution.

TABLE VI. A heuristic interpretation of AIC differences Δ_i reported in Fig. 5. The larger Δ_i is, the less plausible it is that the model $\mathcal{M}_i$ is the best model.

Δ_i	Level of empirical support for model i	Models	$\Delta_i = \mathcal{A}_i - \mathcal{A}_2$
0–2	Substantial	$\mathcal{M}_1$	0
4–7	Considerably less	$\mathcal{M}_2$	0.20
>10	Essentially none	$\mathcal{M}_0$	9.11

Finally, using the tomography data with the qubit initialized in $\rho_0(0) = |+i\rangle\langle+i|$ and $\rho_2(0) = |-i\rangle\langle-i|$, we re-evaluate the degree of non-Markovianity for different spectator qubit states in Fig. 10 according to Eqs. (14) and (15). The nonmonotonic decay in the trace distance $\mathcal{D}$ and the estimated non-Markovian measure $\mathcal{N}$ in Fig. 10 agrees well with those in Fig. 6, serves as a supplementary quantitative demonstration of non-Markovian effects present in the device. The degree of non-Markovianity $\mathcal{N}_{\mathcal{M}_1}^{\mathrm{prd}}$ and $\mathcal{N}_{\mathcal{M}_2}^{\mathrm{prd}}$ calculated from the constructed PMME models agrees well with that from the experimental data $\mathcal{N}^{\mathrm{exp}}$, showing that the PMME models have the ability to quantitatively describe and predict the degree of non-Markovianity of the dynamics during the qubit free evolution.

---

[1] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of N-level systems, J. Math. Phys. **17**, 821 (1976).

[2] G. Lindblad, On the generators of quantum dynamical semigroups, Comm. Math. Phys. **48**, 119 (1976).

[3] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).

[4] A. Rivas and S. F. Huelga, *Open Quantum Systems: An Introduction, SpringerBriefs in Physics* (Springer Berlin Heidelberg, Heidelberg, 2011).

[5] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, *Colloquium*: Non-markovian dynamics in open quantum systems, Rev. Mod. Phys. **88**, 021002 (2016).

[6] G. Burkard, Non-markovian qubit dynamics in the presence of $1/f$ noise, Phys. Rev. B **79**, 125317 (2009).

[7] M. Malekakhlagh, A. Petrescu, and H. E. Türeci, Non-markovian dynamics of a superconducting qubit in an open multimode resonator, Phys. Rev. A **94**, 063848 (2016).

[8] IBM Quantum (2021).

[9] J. Morris, F. A. Pollock, and K. Modi, Non-Markovian memory in IBMQX4, (2019), ArXiv:1902.07980.

[10] A. J. Landahl, J. T. Anderson, and P. R. Rice, Fault-tolerant quantum computing with color codes, (2011), ArXiv:1108.5738.

[11] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[12] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Rep. Prog. Phys. **76**, 076001 (2013).

[13] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, Quantum **4**, 321 (2020).

[14] R. Alicki, D. A. Lidar, and P. Zanardi, Internal consistency of fault-tolerant quantum error correction in light of rigorous derivations of the quantum markovian limit, Phys. Rev. A **73**, 052311 (2006).

[15] D. Aharonov, A. Kitaev, and J. Preskill, Fault-Tolerant Quantum Computation with Long-Range Correlated Noise, Phys. Rev. Lett. **96**, 050504 (2006).

[16] H. K. Ng and J. Preskill, Fault-tolerant quantum computation versus gaussian noise, Phys. Rev. A **79**, 032318 (2009).

[17] H. K. Ng, D. A. Lidar, and J. Preskill, Combining dynamical decoupling with fault-tolerant quantum computation, Phys. Rev. A **84**, 012305 (2011).

[18] S. Mavadia, C. L. Edmunds, C. Hempel, H. Ball, F. Roy, T. M. Stace, and M. J. Biercuk, Experimental quantum verification in the presence of temporally correlated noise, npj Quantum Inf. **4**, 7 (2018).

[19] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, S. M. Clark, R. Blume-Kohout, I. Siddiqi, and K. C. Young, Experimental characterization of crosstalk errors with simultaneous gate set tomography, PRX Quantum **2**, 040338 (2021).

[20] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar, Demonstration of Fidelity Improvement Using Dynamical Decoupling with Superconducting Qubits, Phys. Rev. Lett. **121**, 220502 (2018).

[21] V. Tripathi, H. Chen, M. Khezri, K.-W. Yip, E. M. Levenson-Falk, and D. A. Lidar, Suppression of crosstalk in superconducting qubits using dynamical decoupling (2021), ArXiv:2108.04530.

[22] E. Paladino, Y. M. Galperin, G. Falci, and B. L. Altshuler, $1/f$ noise: Implications for solid-state quantum information, Rev. Mod. Phys. **86**, 361 (2014).

[23] R. Zwanzig, Ensemble method in the theory of irreversibility, J. Chem. Phys. **33**, 1338 (1960).

[24] L. Ferialdi, Exact Closed Master Equation for Gaussian Non-Markovian Dynamics, Phys. Rev. Lett. **116**, 120402 (2016).

[25] A. A. Budini, Embedding non-markovian quantum collisional models into bipartite markovian dynamics, Phys. Rev. A **88**, 032115 (2013).

[26] A. Shabani and D. A. Lidar, Completely positive post-markovian master equation via a measurement approach, Phys. Rev. A **71**, 020101(R) (2005).

[27] D. A. Lidar, Lecture notes on the theory of open quantum systems, (2019), arXiv preprint ArXiv:1902.00967.

[28] K. Kraus, General state changes in quantum theory, Ann. Phys. **64**, 311 (1971).

[29] D. Chruściński, in *Classical and Quantum Physics*, Springer Proceedings in Physics, edited by G. Marmo, D. Martín de Diego, and M. Muñoz Lecanda (Springer International Publishing, Cham, 2019), p. 147.

[30] C. Sutherland, T. A. Brun, and D. A. Lidar, Non-markovianity of the post-markovian master equation, Phys. Rev. A **98**, 042119 (2018).

[31] C. Fisher, Quantum computing software and programming tools.

[32] J. Wolberg, *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments* (Springer Berlin Heidelberg, Heidelberg, 2006).

[33] J. M. Dominy and D. A. Lidar, Beyond complete positivity, Quant. Inf. Proc. **15,** 1349 (2016).

[34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[35] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. **19,** 716 (1974).

[36] M. Ganzhorn, G. Salis, D. J. Egger, A. Fuhrer, M. Mergenthaler, C. Müller, P. Müller, S. Paredes, M. Pechal, M. Werninghaus, and S. Filipp, Benchmarking the noise sensitivity of different parametric two-qubit gates in a single superconducting quantum computing platform, Phys. Rev. Res. **2,** 033447 (2020).

[37] D. A. Lidar, A. Shabani, and R. Alicki, Conditions for strictly purity-decreasing quantum markovian dynamics, Chem. Phys. **322,** 82 (2006).

[38] W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, Rev. Mod. Phys. **75,** 715 (2003).

[39] M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, Dynamics of superconducting qubit relaxation times, (2021), ArXiv:2105.15201.

[40] H.-P. Breuer, E.-M. Laine, and J. Piilo, Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems, Phys. Rev. Lett. **103,** 210401 (2009).

[41] S. Dasgupta and T. S. Humble, Stability of noisy quantum computing devices, (2021), ArXiv:2105.09472.

[42] 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).

[43] N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory, Robust method for estimating the lindblad operators of a dissipative quantum process from measurements of the density operator at multiple time points, Phys. Rev. A **67,** 042322 (2003).

[44] M. Howard, J. Twamley, C. Wittmann, T. Gaebel, F. Jelezko, and J. Wrachtrup, Quantum process tomography and Linblad estimation of a solid-state qubit, New J. Phys. **8,** 33 (2006).

[45] M. Mohseni and D. A. Lidar, Direct Characterization of Quantum Dynamics, Phys. Rev. Lett. **97,** 170501 (2006).

[46] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes, Integrated Tool set for Control, Calibration, and Characterization of Quantum Devices Applied to Superconducting Qubits, Phys. Rev. Appl. **15,** 034080 (2021).

[47] L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-markovian quantum processes with recurrent neural networks, New J. Phys. **20,** 123030 (2018).

[48] S. Krastanov, K. Head-Marsden, S. Zhou, S. T. Flammia, L. Jiang, and P. Narang, Unboxing Quantum Black Box Models: Learning Non-Markovian Dynamics, (2020), ArXiv:2009.03902.

[49] E. Flurin, L. S. Martin, S. Hacohen-Gourgy, and I. Siddiqi, Using a Recurrent Neural Network to Reconstruct Quantum Dynamics of a Superconducting Qubit from Physical Observations, Phys. Rev. X **10,** 011006 (2020).

[50] I. A. Luchnikov, S. V. Vintskevich, D. A. Grigoriev, and S. N. Filippov, Machine Learning Non-Markovian Quantum Dynamics, Phys. Rev. Lett. **124,** 140502 (2020).

[51] I. A. Luchnikov, E. O. Kiktenko, M. A. Gavreev, H. Ouerdane, S. N. Filippov, and A. K. Fedorov, Probing non-markovian quantum dynamics with data-driven analysis: Beyond "black-box" machine learning models (2021), ArXiv:2103.14490.

[52] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018), p. 325.

[53] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16,** 1050 (2020).

[54] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers (2021), ArXiv:2110.02965.

[55] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77,** 032322 (2008).

[56] G. O. Samach, A. Greene, J. Borregaard, M. Christandl, D. K. Kim, C. M. McNally, A. Melville, B. M. Niedzielski, Y. Sung, D. Rosenberg, M. E. Schwartz, J. L. Yoder, T. P. Orlando, J. I.-J. Wang, S. Gustavsson, M. Kjaergaard, and W. D. Oliver, Lindblad Tomography of a Superconducting Quantum Processor, (2021), ArXiv:2105.02338.

[57] S. Xue, L. Tan, R. Wu, M. Jiang, and I. R. Petersen, Inverse-system method for identification of damping-rate functions in non-markovian quantum systems, Phys. Rev. A **102,** 042227 (2020).

[58] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Non-markovian quantum processes: Complete framework and efficient characterization, Phys. Rev. A **97,** 012127 (2018).

[59] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, and K. Modi, Demonstration of non-markovian process characterisation and control on a quantum processor, Nat. Commun. **11,** 6301 (2020).

[60] L. Xiang, Z. Zong, Z. Zhan, Y. Fei, C. Run, Y. Wu, W. Jin, C. Xiao, Z. Jia, P. Duan, J. Wu, Y. Yin, and G. Guo, Quantify the Non-Markovian Process with Intermediate Projections in a Superconducting Processor, (2021), ArXiv:2105.03333.

[61] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, *Advances in Neural Information Processing Systems* (2018), Vol. 31.

[62] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119,** 180509 (2017).

[63] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, Unfolding quantum computer readout noise, npj Quantum Inf. **6,** 84 (2020).

[64] H.-J. Briegel and B.-G. Englert, Quantum optical master equations: The use of damping bases, Phys. Rev. A **47,** 3311 (1993).

[65] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, Numerical Recipes in C (1992).

[66] J. H. Mathews and R. W. Howell, *Complex Analysis: for Mathematics and Engineering,* (Jones and Bartlett Pub. Inc., Sudbury, MA, 2012), sixth ed.

[67] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).

[68] S. Wissmann, A. Karlsson, E.-M. Laine, J. Piilo, and H.-P. Breuer, Optimal state pairs for non-markovian quantum dynamics, Phys. Rev. A **86**, 062108 (2012).

[69] K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach,* (Springer, New York, 2002), 2nd ed.