

Non-Markovian Quantum Process Tomography

G.A.L. White^{1,*}, F.A. Pollock², L.C.L. Hollenberg¹, K. Modi^{2,†} and C.D. Hill^{1,3,‡}

¹*School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia*

²*School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia*

³*School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria 3010, Australia*



(Received 11 August 2021; revised 22 December 2021; accepted 10 May 2022; published 27 May 2022)

Characterization protocols have so far played a central role in the development of noisy intermediate-scale quantum (NISQ) computers capable of impressive quantum feats. This trajectory is expected to continue in building the next generation of devices—ones that can surpass classical computers for particular tasks—but progress in characterization must keep up with the complexities of intricate device noise. A missing piece in the zoo of characterization procedures is tomography, which can completely describe non-Markovian dynamics over a given time frame. Here, we formally introduce a generalization of quantum process tomography, which we call process tensor tomography. We detail the experimental requirements, construct the necessary postprocessing algorithms for maximum-likelihood estimation, outline the best-practice aspects for accurate results, and make the procedure efficient for low-memory processes. The characterization is a pathway to diagnostics and informed control of correlated noise. As an example application of the hardware-agnostic technique, we show how its predictive control can be used to substantially improve multitime circuit fidelities on superconducting quantum devices. Our methods could form the core for carefully developed software that may help hardware consistently pass the fault-tolerant noise threshold.

DOI: [10.1103/PRXQuantum.3.020344](https://doi.org/10.1103/PRXQuantum.3.020344)

I. INTRODUCTION

Central to the theme of progress in quantum computing has been the development and application of quantum characterization, verification, and validation (QCVV) procedures [1–7]. These techniques model and identify the presence of errors in a quantum-information processor (QIP). These errors may have different origins, such as coherent control noise, decoherence, crosstalk, or state-preparation-and-measurement (SPAM) errors. The operational description of open quantum dynamics has been immensely useful in describing the noise present in QIPs [8,9]. Typified by mappings of density operators, a discrete snapshot of a given noisy process can be characterized through a series of experiments on the QIP. The resulting object is the gold standard for describing two-time

errors on a quantum device: a completely positive trace-preserving (CPTP) map.

CPTP maps, however, are not sufficient to describe all dynamics present on real quantum devices. A generic quantum stochastic process represents many times and carries correlations across multiple time scales as a rule rather than the exception [10]. The emergence of adverse effects from temporal correlations in quantum processes is known as non-Markovian noise and arises from mutual interaction between a system and its complex environment. Standard CP maps—such as those characterized by quantum process tomography (QPT)—cannot describe reduced system-environment (*SE*) dynamics arising from a correlated state. Famously, this leads to CP divisibility of a process as a granular measure of non-Markovianity [11–13] and prohibits their use in the detailed study of multitime quantum correlations.

The discourse on device quality is mostly shaped through conventional benchmarks either at the low level—by emphasis on the fidelity of individual gates—or at the high level—through holistic measures such as quantum volume [14]. However, there is a discontinuity in these abstractions: the former does not consider the effects of intrinsic *temporal context*, by which we mean the tendency of past gate choices to be correlated with future gate outcomes. Meanwhile, the latter coarsely summarizes the

*white.g@unimelb.edu.au

†kavan.modi@monash.edu

‡cdhill@unimelb.edu.au

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

average performance of a QIP in high-width high-depth random circuits. Indeed, recent progress in quantum computing has led to the engineering of very-low-error gates [15–17] but evidence has shown that NISQ devices do not behave like the sum of their parts. Typically, they perform worse than predicted by constituent gate errors alone, which ignore non-Markovian quantum processes present in reality [4,6,18]. A simple example of how this complex noise can impact the outcome of an experiment is that the choice of a gate in the past may influence the action of a gate applied at the present. The two gates do not simply multiply out. Correlated noise can be particularly deleterious not only in its complexity but in its ability to reduce or even completely eliminate the effectiveness of quantum error-correcting codes [19,20]. Importantly, there is ample evidence that Markov models are insufficient to fully capture the dynamics exhibited by current generation quantum devices [7,18,21–23], thus motivating the present work.

The characterization of general quantum stochastic processes is the first step in explicating correlated noise on quantum processors. Contextual noise detection has been rigorously studied under a null-hypothesis standpoint in prior work. That is, a successful approach has been to probe for a change in measurement statistics—or some other invariant property of the data—both with and without the presence of a context variable [24–27]. This includes non-Markovian noise under its umbrella, where the context is limited to circuit-level decisions. However, a clear procedure and a systematic framework for characterizing contextual noise has been notably absent; there are several technical and fundamental challenges that make this task highly nontrivial.

In this work, we formally flesh out a part of this missing puzzle piece in quantum tomography, which we call process tensor tomography (PTT). Specifically, PTT addresses the subclass of non-Markovian noise. Dynamical maps—estimated by QPT—break down when employed for multitime processes with memory. Instead, these dynamics are well described by the process tensor, a recently developed mathematical framework to represent quantum stochastic processes [28]. We first focus on formalizing the tomographic reconstruction of generic multitime quantum processes and discuss the challenges and pitfalls that lie in the way. We then turn our attention to sparingly characterizing processes while overcoming many of these obstacles by integrating PTT with maximum-likelihood estimation (MLE). This approach is then applied efficiently for the case of processes with low memory, described by their Markov order [29]. We emphasize that this work adopts a constructive approach to solve the problem of statistically robust non-Markovian quantum process tomography in full generality. Consequently, we consider relatively small numbers of steps on single-qubit systems. The scalability of the method is intrinsically linked to the complexity of the process under scrutiny.

For environments with relatively few relevant degrees of freedom or processes with quickly decaying memory, we foresee no fundamental obstacles to simplifying and scaling up the techniques introduced here [30–33]

The MLE-based PTT (or MLE PTT) has several advantages over linear-inversion PTT (LI PTT); it is significantly more efficient than standard inversion methods, as well as overcoming the difficulties of positivity and causality conditions. Yet, it grows exponentially in the number of time steps. To overcome this, we accommodate for sparseness in the complexity of the dynamics. Low-memory processes are ubiquitous in nature and our characterization tools leverage this to offer both an experimentally and computationally more efficient description of the process. In particular, this permits us to quantify the degree to which a limited-memory model fails to predict laboratory observations. This significantly enhances previous results that quantify how memoryless models break down [22,34]. Moreover, our methods establish a trade-off between the level of characterization complexity and a desired approximation. Our measures have both simple interpretations and a clear recipe to expand the model should they fail to adequately explain the data.

To show the efficacy of this approach, we report the real device characterization of multitime quantum stochastic processes in a way that is both consistent and fully inclusive of non-Markovian dynamics. There are three key motivators for these implementations: to benchmark quantum devices and diagnose non-Markovian noise, to study the structure of quantum stochastic processes, and for precise enhanced control of non-Markovian systems. As such, we first show MLE PTT to be highly reliable at characterizing multitime dynamics. The reliability and the efficiency of these methods allows us to implement noise-aware control to significantly reduce the noise present across a variety of contexts. We demonstrate not only the utility of our technique here but also its necessity by showing that improvement is contingent on the inclusion of higher-order (i.e., multitime) temporal correlations in the model.

Our paper is structured as follows. In Sec. II, we give a brief background on quantum state tomography (QST) and QPT before continuing into the theory of quantum stochastic processes and tomographic reconstruction of a process tensor. This analyzes our non-Markovian characterization in context. In Sec. III, we derive the main components of a MLE-PTT protocol, including a positive causal projection necessary for reconstructed process tensors to be physical. This procedure is then benchmarked and validated on IBM Quantum devices. Specifically, the number of circuits required for accurate LI PTT in a real setting scales as $\mathcal{O}(N_{\text{oc}}^k)$, where N_{oc} is the number of (overcomplete) inputs per time step and k is the number of time steps. The MLE-PTT procedure reduces this to $\mathcal{O}(N_{\text{MLE}}^k)$, where $N_{\text{oc}} = 24$ and $N_{\text{MLE}} = 10$. Using the tools of Sec. III, in Sec. IV we motivate and advance the theory of quantum Markov order,

in which we determine how to adaptively truncate weaker long-time temporal correlations in the model. This tames the exponential scaling to a linear scaling, i.e., $\mathcal{O}(k \times N_{\text{MLE}}^\ell)$, where ℓ is the fixed Markov order. This is not only practical but it provides an accessible diagnostic to the complexity of device noise. In Sec. V, we demonstrate noise-aware control over a non-Markovian process, which is based on the noise-characterization methods of previous sections. Namely, we characterize noise for Markov orders of $\ell = \{1, 2, 3\}$ and use this information to significantly increase the fidelity of several NISQ devices by using non-Markovian correlations as a resource. We show that the performance of the machine improves when the Markov order is chosen to be higher. This leads to a trade-off relation between characterization complexity and accuracy of the characterization.

A. Summary of quantum characterization

Before we proceed, we briefly analyze where PTT sits among the existing characterization techniques. Frameworks for dynamics in the literature can be broadly classed under dynamical-map or master-equation formalisms. The former captures only two-time correlations that, in the presence of non-Markovian noise, will fail to describe multitime processes when composed together [8]. This applies clearly to QPT but also to any QCVV procedures born out of the umbrella of quantum channels, such as gate-set tomography (GST), randomized benchmarking (RB), and Hamiltonian tomography [1,22,35–37]. The latter is generally a function of at most three-time correlations and be reduced to a family of dynamical maps [38].

In the case where these approaches are insufficient to characterize features of non-Markovian noise, *ad hoc* extension techniques exist to detect a departure from the Markov assumptions. These detect some parts of the non-Markovian character but are not generalizable or predictive. Their application is typically as witnesses or for shallow diagnostics but cannot rigorously measure the memory or be employed systematically to control the system. Common examples include memory kernels for master equations [39] or statistical tests to establish causal connections between environmental factors or gate choices, and system-level dynamics [23–27]. Statistical tests have also been employed to quantify the confidence with which breakdown of Markovianity can be described [22].

In a precise way, PTT is a direct generalization of the QPT framework: instead of estimating a single dynamical map, it estimates a sequence of possibly correlated dynamical maps. These temporal correlations may be arbitrarily strong. Thus it is, to the best of our knowledge, the only procedure demonstrated to fit the criteria of systematically capturing this difficult and important class of noisy dynamics. Moreover, it is fully general, without relying on underlying microscopic models. It carries some of

the same associated baggage (exponential scaling in full generality, assumptions about prior calibrations)—but can also be imbued with the modifications and simplifications that make the QPT framework a rich tool for quantum characterization. A guiding set of simplifying assumptions can reduce the experimental burden of characterization at the expense of either a possible sacrifice in accuracy or information gain; for example, in compressed-sensing or tensor-network models [30,40]. We also foresee that many of the ideas that either arise from or utilize QPT could be applied to PTT, further extending this branch of QCVV and error mitigation [2].

We conclude this section by emphasizing the extent to which the methods and theory in this work offer a departure from those introduced in Ref. [21]. This previous work has shown that non-Markovian characterization is possible on real quantum devices by implementing the process-tensor mapping through linear inversion. In particular, it is a demonstration of how to effectively wield the spatiotemporal version of Born’s rule [Eq. (10) and Fig. 1(c)] on quantum processors—thus permitting characterization and control regardless of the strength of the *SE* interaction. However, the shortcomings are that the process tensor itself (and thus any rigorous non-Markovian measures)

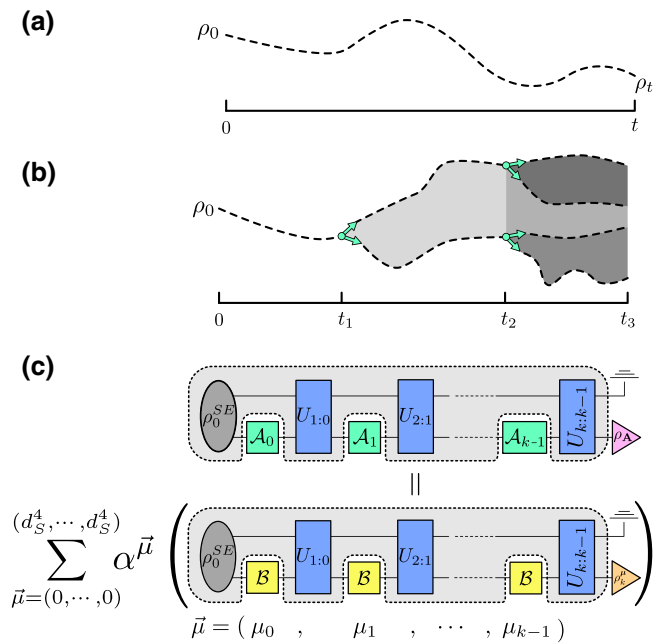


FIG. 1. (a) The conventional model of open quantum systems tracks the state of the system as a function of time, but cannot build up multitime joint statistics. (b) The quantum stochastic process picture considers the response of the system to different sequences of gates; by considering all trajectories of the system, correlations between different times may be exactly characterized and quantified. (c) A pictorial description of how linear expansion in a basis by the process tensor preserves intermediate dynamics and expresses an arbitrary sequence.

is not estimated. Further, the averaging over sampling statistics is made very expensive through an overcomplete basis—and no guarantees are made of a completely positive or causal mapping. Finally, the structure of the process is left opaque, such that all possible operational trajectories are considered. This leaves it impossible to leverage any sparseness in the real process.

By fleshing out PTT in a sophisticated way, the present work provides a clear framework by which a process tensor may be estimated using a minimal complete basis in a manner that is both physical and statistically robust. We also show how decaying correlations in the memory may be truncated to offer an efficient characterization of the non-Markovian process and finally how all of the above may be employed to provide superior control of a quantum device.

II. A HIERARCHY OF QUANTUM TOMOGRAPHY

We start by outlining the fundamentals of QST, QPT, and PTT. These procedures build on each other; for a d -dimensional system, QST requires a set of $\mathcal{O}(d^2)$ experiments, QPT is most easily thought of as d^2 QSTs and requires $\mathcal{O}(d^4)$ experiments, and PTT requires $\mathcal{O}(d^{4k})$ experiments, where $k = 1, 2, \dots$ is the number of time steps [21,41,42]. The familiarity of the first two lays the groundwork for the latter. The treatment and practical concerns of each technique are similar with respect to real data. One key difference lies in the fact that due to the higher-dimensional superoperator basis, especially for PTT, small errors can become magnified and require closer attention. In addition to an overview of tomography, we present the conceptual developments of PTT in this section and scrutinize its proclivities—such as with respect to hardware control restrictions and SPAM error.

Fundamentally, quantum tomography is an exercise in reconstructing linear maps from experimental data. This can be accomplished by measuring the input-output relations on a complete basis for the input space. A disconnect between theory and experiment occurs when, in practice, the input vectors are faulty in some way (such as noisy

preparation) or the measured output frequencies differ from that of the real population (due to a noisy probe or finite sampling error) [41]. As well as producing an object that may disagree with experiment for inputs away from the characterization, the resulting estimate might not even be physical. A variety of different methods may be employed to overcome some of these problems: the collection of more data, the elevation of inputs and outputs to the model [42,43], employing an overcomplete basis in the characterization [44], and the treatment of the measured data to fit a physical model [45]. These techniques are applicable, regardless of the model type, and we discuss their utility in PTT.

Before examining the PTT description of quantum stochastic processes, we emphasize parallels with more conventional tomography in QST and QPT, such that the content can appear more familiar to readers. Along our exposition, we emphasize a “hierarchy” in the sense that the information of each level is strictly contained within the characterization of the next level. That is to say, QPT can describe the reconstructed state of QST and PTT can describe the dynamical map of QPT. We present a summary of each map in Table I, as well as their physical requirements, and continue to flesh out here.

A. Quantum state and quantum process tomography

The foundation of most QCVV procedures is the estimation of quantum states and quantum channels on an experimental device. By quantum state, we mean the density-matrix representation of a system at a given time. A quantum channel—or quantum stochastic matrix or dynamical map—then expresses the evolution of some state between two times and has a freedom in representation. A convenient choice for QPT employs the Choi representation. Here, using the Choi-Jamiolkowski isomorphism (CJI), CP maps may be given by a positive-matrix representation as a quantum state, exploiting the correspondence between $\mathcal{B}(\mathcal{H}_{\text{in}}) \rightarrow \mathcal{B}(\mathcal{H}_{\text{out}})$ and $\mathcal{B}(\mathcal{H}_{\text{out}}) \otimes \mathcal{B}(\mathcal{H}_{\text{in}})$. Here, $\mathcal{B}(\mathcal{H})$ denotes the space of bounded linear

TABLE I. Detail on different levels of the quantum tomography hierarchy, pertinent to experimental reconstruction. QST reconstructs a density operator, ρ , a positive matrix with unit trace representing the quantum state. QPT reconstructs a quantum channel, \mathcal{E} , through its action on different states. This map must be both CP and TP, conditions which, in Choi form, manifest themselves as positivity and affine constraints on the matrix. Finally, PTT reconstructs a process tensor $\mathcal{T}_{k:0}$ through its action on different control operations. This object must have a positive-matrix Choi form and must respect causality. The information of each column is strictly contained in the column to the right.

	Quantum state	Quantum process	Process tensor
Characterization object	Density operator ρ	Quantum channel \mathcal{E}	Process tensor $\mathcal{T}_{k:0}$
Mapping	$\mathcal{H}_S \rightarrow \mathcal{H}_S$	$\mathcal{B}(\mathcal{H}_S) \rightarrow \mathcal{B}(\mathcal{H}_S)$	$\bigotimes_{i=1}^k \mathcal{B}(\mathcal{B}(\mathcal{H}_S)) \rightarrow \mathcal{B}(\mathcal{H}_S)$
Observed probabilities	$p_i = \text{Tr}[\Pi_i \rho]$	$p_{ij} = \text{Tr}[(\Pi_i \otimes \rho_j^T) \hat{\mathcal{E}}]$	$p_{i,\tilde{i}} = \text{Tr}[(\Pi_i \otimes \mathcal{B}_{k-1}^{\mu_{k-1}T} \otimes \dots \otimes \mathcal{B}_0^{\mu_0T}) \Upsilon_{k:0}]$
Positivity constraint	$\rho \succcurlyeq 0$	$\hat{\mathcal{E}} \succcurlyeq 0$	$\Upsilon_{k:0} \succcurlyeq 0$
Affine constraint	$\text{Tr}[\rho] = 1$	$\text{Tr}_{\text{out}}[\hat{\mathcal{E}}] = \mathbb{I}_{\text{in}}$	$\text{Tr}_{\text{out}}[\Upsilon_{k:0}] = \mathbb{I}_{\text{in}} \otimes \Upsilon_{k-1:0} \forall k$

operators on a Hilbert space \mathcal{H} . Explicitly, for some channel \mathcal{E} , its Choi state $\hat{\mathcal{E}}$ is constructed through the action of \mathcal{E} on one half of an unnormalized maximally entangled state $|\Phi^+\rangle = \sum_{i=1}^d |ii\rangle$, with identity map \mathcal{I} on the other half:

$$\hat{\mathcal{E}} := (\mathcal{E} \otimes \mathcal{I}) [|\Phi^+\rangle\langle\Phi^+|] = \sum_{i,j=1}^d \mathcal{E}[|i\rangle\langle j|] \otimes |i\rangle\langle j|. \quad (1)$$

All reconstruction of experimental properties must begin with a probe to read out quantum information. This extraction comes from a known positive-operator-valued measure (POVM) $\mathcal{J} := \{\Pi_i\}_{i=1}^L$ with associated elements called ‘‘effects.’’ To reconstruct any state, \mathcal{J} must span the space of density matrices $\mathcal{B}(\mathcal{H}_S)$, a characteristic known as informational completeness (IC). For some density matrix ρ , a POVM yields observable probabilities for each effect in accordance with Born’s rule:

$$p_i = \text{Tr}[\Pi_i \rho]. \quad (2)$$

For a quantum channel, its action is given in terms of its Choi state by

$$\mathcal{E}[\rho_{\text{in}}] = \text{Tr}_{\text{in}} \left[(\mathbb{I} \otimes \rho_{\text{in}}^T) \hat{\mathcal{E}} \right] = \rho_{\text{out}}. \quad (3)$$

These input-output relations are sufficient to reconstruct both a state ρ and a channel \mathcal{E} . That is, the measurement p_i for each element of \mathcal{J} is sufficient to construct ρ and the measurement $\{\rho'_i\} := \{\mathcal{E}[\rho_i]\}_{i=1}^n$ for a full basis of inputs is sufficient to construct $\hat{\mathcal{E}}$. This is accomplished with the construction of a *dual* set $\mathcal{D} := \{\Delta_i\}_{i=1}^L$ to linearly independent \mathcal{J} , satisfying $\text{Tr}[\Pi_i \Delta_j] = \delta_{ij}$. Similarly, let $\{\omega_j\}$ be the dual set to $\{\rho_i\}$. Note that in practice, linear independence may be relaxed with an overcomplete basis in the case where $L, n > d^2$, where a matrix pseudoinverse is used to find the dual set, rather than an inverse (for further details, see Appendix A). Then, we may (respectively) express ρ and $\hat{\mathcal{E}}$ as

$$\rho = \sum_{i=1}^L p_i \Delta_i \quad \text{and} \quad \hat{\mathcal{E}} = \sum_{i=1}^n \rho'_i \otimes \omega_i^T, \quad (4)$$

which, by design, are consistent with Eqs. (2) and (3), respectively. By combining the two above equations, $\hat{\mathcal{E}}$ may also be decomposed in terms of POVM effects as

$$\hat{\mathcal{E}} = \sum_{j=1}^L \sum_{i=1}^n p'_j \Delta_j \otimes \omega_i^T. \quad (5)$$

Note that the Δ_j and ω_i matrices are not usually positive but the resulting ρ is both positive with unit trace and the resulting $\hat{\mathcal{E}}$ is both positive and trace preserving, in that its marginal input is maximally mixed. These conditions are written explicitly in the final row of Table I.

B. Process tensor tomography

A typical approach to studying dynamical processes consists of monitoring the state of the system as a function of time, as in Fig. 1(a) [11]. Although effects such as coherent state oscillation can flag non-Markovianity, any interrogation of the system necessarily disrupts its future evolution. For this reason, joint statistics cannot be measured across time and, consequently, multitime correlations cannot be characterized. As a result, quantum non-Markovian effects, which can have a variety of different physical sources, have been historically difficult to theoretically describe, much less experimentally capture. Often, non-Markovian effects are quantified in terms of ‘‘leftover’’ error, inferred by the extent to which Markov models break down [13,22,34].

Recently, the process-tensor framework [28] (and the process-matrix framework [46]) have been proposed as a generalization of classical stochastic processes to the quantum domain. Importantly, this generalization allows for the study of multistep temporal correlations—or non-Markovianity—in quantum systems. An important feature of the PTT formalism is that it maps all possible temporal correlations onto spatial correlations over sequences of CPTP channels, leaving non-Markovian measures as operationally well defined as for any quantum or classical spatial correlation. The application of conventional many-body techniques allows for necessary and sufficient measures of device non-Markovianity, as well as the more fine-grained study of operation-specific context dependence. This fills a gap in the library of QCVV resources [1].

Any continuous-time quantum stochastic process can be discretized in a number of time steps: $\mathbf{T}_k = \{t_0, t_1, \dots, t_k\}$ (for example, in the context of a quantum circuit). A finite-time process tensor is then a marginal of the continuous-time process tensor [47], which represents all possible correlations in \mathbf{T}_k . To capture statistics for each $t_i \in \mathbf{T}_k$, the experimenter applies an IC basis of control operations, each of which changes the trajectory of the state and maps to an output. In completing this procedure for all times, all trajectories consistent with \mathbf{T}_k may be inferred. This is depicted in Fig. 1(b). This is sufficient both to construct all joint statistics, and to predict the output subject to any generic sequence. Figures 1(a) and 1(b) contrast the traditional approach to open quantum dynamics and PTT framework.

To be precise, we consider the situation in which a k -step process is driven by a sequence $\mathbf{A}_{k-1:0}$ of control operations, each represented mathematically by CP maps— $\mathbf{A}_{k-1:0} := \{\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_{k-1}\}$ —after which we obtain a final state $\rho_k(\mathbf{A}_{k-1:0})$ conditioned on this choice of interventions. These controlled dynamics have the form

$$\rho_k(\mathbf{A}_{k-1:0}) = \text{tr}_E[U_{k:k-1} \mathcal{A}_{k-1} \cdots U_{1:0} \mathcal{A}_0(\rho_0^{SE})], \quad (6)$$

where $U_{k:k-1}(\cdot) = u_{k:k-1}(\cdot)u_{k:k-1}^\dagger$. Equation (6) can be used to define a mapping from past controls $\mathbf{A}_{k-1:0}$ to future states $\rho_k(\mathbf{A}_{k-1:0})$, which is the process tensor $\mathcal{T}_{k:0}$:

$$\mathcal{T}_{k:0}[\mathbf{A}_{k-1:0}] = \rho_k(\mathbf{A}_{k-1:0}). \quad (7)$$

The logic of the process tensor is depicted in Fig. 1(c), mirroring the trajectory sketch in Fig. 1(b).

In this sense, the process tensor is designed to account for intermediate control operations and quantifies quantum correlations between past operations and future states. Just as in the case of CPTP maps, it can be shown that the process tensor too has a many-body Choi representation [28]. Both states ρ and channels $\hat{\mathcal{E}}$ have affine conditions ensuring unit probability and trace preservation, respectively. Similarly, the Choi state of the process tensor $\Upsilon_{k:0}$ also has affine conditions: these guarantee causality. That is, any future control operations cannot affect the past statistics. These facets of the process tensor are all possible to reconstruct experimentally, using many of the techniques from QPT and QST. We directly employ and build upon these ideas in this work. We start from the direct linear-inversion construction of the process-tensor Choi state and then proceed with maximum-likelihood estimation and truncated Markov models. Along the way, we make explicit the parallels between QPT and PTT as a generalization in Fig. 2 for pedagogical purposes. In particular, Fig. 2(b) contrasts the generalized CJI of the process tensor with the standard channel CJI, emphasizing the extension made.

We begin by discussing the *linear-inversion construction* of the process tensor direct from experimental data, which has recently been reported in Ref. [21]. The estimate here comes from (pseudo)inverting the feature matrix on observed data. The resulting object need not be physical and though it may be consistent with its measurement basis, it may not even serve as a good indicator for the behavior of other sequences.

The sequence of interventions $\mathbf{A}_{k-1:0}$ is a CP map represented through the CJI as a $2k$ -partite quantum state—i.e., through action on a maximally entangled state at each of the k time steps. When the operations at each time step are chosen independently, the sequence is given by $\mathbf{A}_{k-1:0} = \bigotimes_{j=0}^{k-1} \mathcal{A}_j$. Each time-local operation may be expanded into a basis $\{\mathcal{B}_j^{\mu_j}\}$ such that any CP map can be expressed as $\mathcal{A}_j = \sum_{\mu_j=1}^{d^4} \alpha_j^{\mu_j} \mathcal{B}_j^{\mu_j}$. The subscript j allows for the possibility of a different basis at each time; meanwhile the superscript μ_j denotes the elements of that particular set. The complete spatiotemporal basis is

$$\{\mathbf{B}_{k-1:0}^{\vec{\mu}}\} = \left\{ \bigotimes_{j=0}^{k-1} \mathcal{B}_j^{\mu_j} \right\}_{\vec{\mu}=(1,1,\dots,1)}, \quad (8)$$

with vector of indices $\vec{\mu}$. To construct the process tensor, therefore, it suffices to measure the output $\rho_k^{\vec{\mu}} := \rho_k(\mathbf{B}_{k-1:0}^{\vec{\mu}})$ for each $\vec{\mu}$ [see Fig. 2(f)]. To do so, we make use of the dual set $\{\Delta_j^{\mu_j}\}$ such that $\text{Tr}[\mathcal{B}_j^{\mu_j} \Delta_j^{\nu_j}] = \delta_{\mu_j \nu_j}$. Then, the Choi state $\Upsilon_{k:0}$ of the process tensor $\mathcal{T}_{k:0}$ is given by

$$\Upsilon_{k:0} = \sum_{\vec{\mu}} \rho_k^{\vec{\mu}} \otimes \mathbf{A}_{k-1:0}^{\vec{\mu} T}, \quad (9)$$

where $\{\mathbf{A}_{k-1:0}^{\vec{\mu}}\} = \{\bigotimes_{j=0}^{k-1} \Delta_j^{\mu_j}\}$ satisfies $\text{Tr}[\mathbf{B}_{k-1:0}^{\vec{\mu}} \mathbf{A}_{k-1:0}^{\vec{\nu}}] = \delta_{\vec{\mu} \vec{\nu}}$. We remark here that the Choi form of a process tensor is a $(2k+1)$ -partite state with alternating input and output indices.

We use the notation \mathfrak{o}_j to denote an output leg of the process at time t_j and \mathfrak{i}_j for the input leg of the process at time t_{j-1} . The collection of indices is therefore $\{\mathfrak{o}_k, \mathfrak{i}_k, \dots, \mathfrak{o}_2, \mathfrak{i}_2, \mathfrak{o}_1, \mathfrak{i}_1, \mathfrak{o}_0\}$. These correspond to the marginals of the process, $\{\hat{\mathcal{E}}_{k:k-1}, \dots, \hat{\mathcal{E}}_{2:1}, \hat{\mathcal{E}}_{1:0}, \rho_0\}$ as shown in Fig. 2(b). Note that this ordering is opposite to the arrow of time in quantum circuits, following instead the convention of matrix multiplication.

Once characterized, the action of the process tensor on a sequence of operations is found by projecting the process tensor onto the Choi state of this sequence (up to a transpose). That is,

$$\rho_k(\mathbf{A}_{k-1:0}) = \text{Tr}_{\bar{\mathfrak{o}}_k} \left[\Upsilon_{k:0} (\mathbb{I}_{\mathfrak{o}_k} \otimes \mathcal{A}_{k-1} \otimes \dots \otimes \mathcal{A}_0)^T \right], \quad (10)$$

where $\bar{\mathfrak{o}}_k$ is every index except \mathfrak{o}_k . This equation, reminiscent of the Born rule [48,49], can determine the output of any sequence of operations (that are consistent with \mathbf{T}_k) and is inclusive of all intermediate *SE* dynamics as well as any initial correlations. The prediction of outcomes subject to control sequences naturally makes the process tensor a very useful object for quantum control.

The generalized CJI of the process tensor maps the multitime process onto a many-body state. Consequently, most of the analytical tools used for the description of spatial correlations can also be employed to describe the temporal correlations of the process. This provides a key motivator for carrying out PTT: the reconstructed process-tensor Choi state provides all information about any spatiotemporally correlated behavior in a quantum device. This makes it a very useful diagnostic tool for near-term quantum devices: one can qualitatively and quantitatively describe the complexity of the interaction of the system with its environment.

Conventional measures of non-Markovianity are well motivated but typically only describe a subset of non-Markovian processes. That is to say, they are sufficient

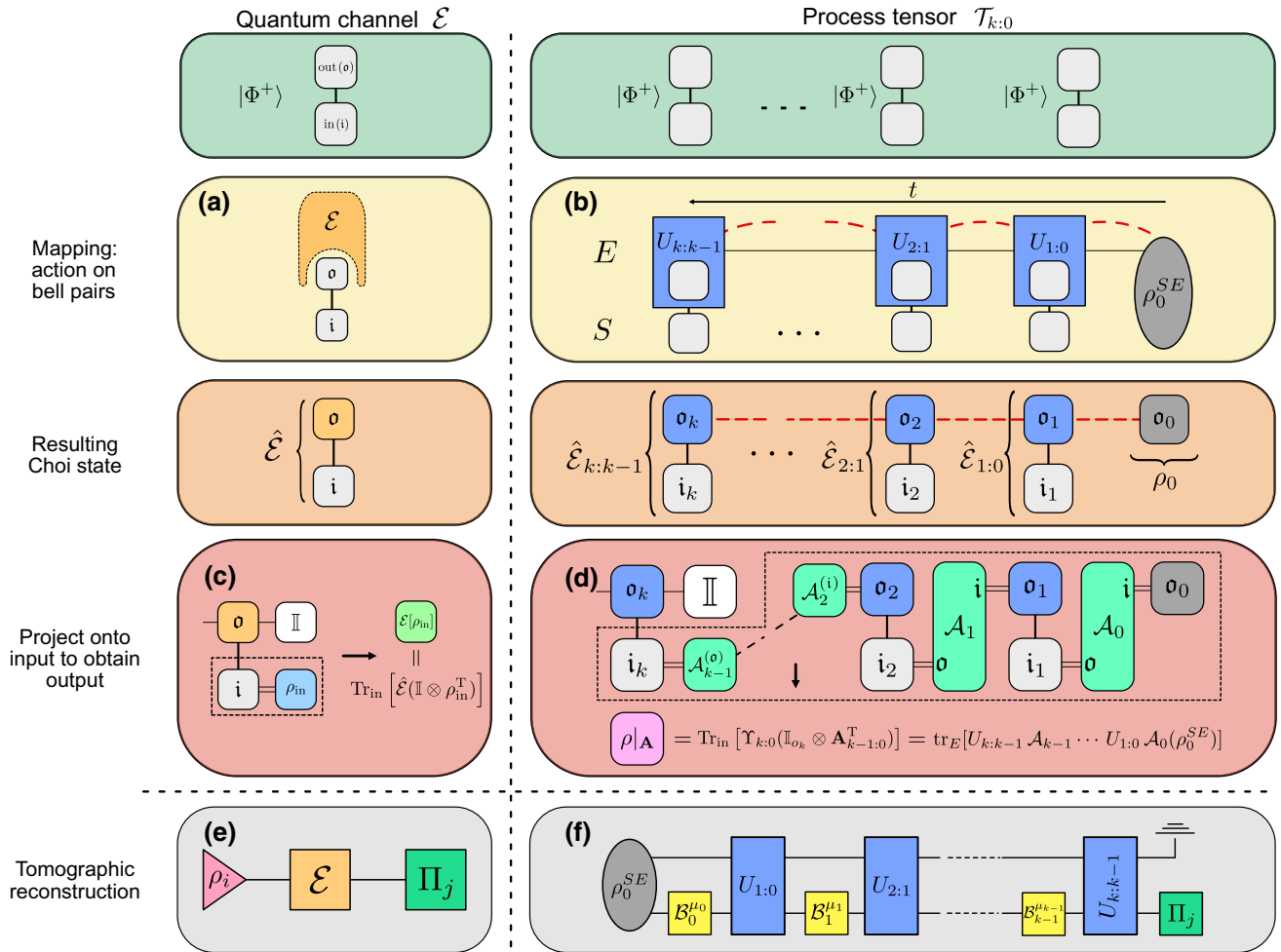


FIG. 2. Operation, manipulation, and characterization analogs between quantum process tomography and process tensor tomography. (a) The Choi-Jamiołkowski isomorphism represents a quantum process \mathcal{E} by the density matrix $\hat{\mathcal{E}}$ using the action of \mathcal{E} on one half of a maximally entangled state. (b) In the generalized CJL, SE unitaries act on one half of a maximally entangled state per time step. Correlations between times (denoted using red dashed lines) are then mapped onto spatial correlations between each output leg of a Bell pair. The result is a collection of (possibly correlated) CPTP maps, as well as the average initial state. (c) The outcome of \mathcal{E} on some ρ_{in} is obtained by projecting the Choi state onto $\mathbb{I} \otimes \rho_{\text{in}}^T$ and tracing over the input space. (d) The outcome of a process conditioned on a sequence of operations $\mathcal{A}_{k-1:0}$ is obtained by projecting the Choi state of $\Upsilon_{k:0}$ onto the Choi state $\bigotimes_{i=0}^{k-1} \mathcal{A}_i$ and tracing over the input. Each \mathcal{A}_i maps the output state of the i th CPTP map to the input leg of the $(i+1)$ th CPTP map. (e) To reconstruct $\hat{\mathcal{E}}$ experimentally, prepare a complete set of states $\{\rho_i\}$, apply \mathcal{E} , and reconstruct the output state with an IC POVM $\{\Pi_j\}$. (f) To reconstruct $\Upsilon_{k:0}$, measure each final state of the system subject to a complete basis of CP maps $\{\mathcal{B}_j^{\mu_j}\}$ at each time. Note that the blue unitaries here are symbolic of any SE interactions and not gates that need to be performed.

but not necessary measures [10–13,50–54]. The process tensor gives rise to both a necessary and sufficient measure of non-Markovianity through all CP-contractive quasidistance measures between $\Upsilon_{k:0}$ and its closest Markov process tensor according to that distance. A Markovian process is one without any correlations in its Choi state—i.e., a product state of some CPTP maps $\hat{\mathcal{E}}_{j+1:j}$ and average initial state ρ_0 . In general, the closest product state is not found with an analytic form; however, there exist convenient choices. One example is the relative entropy, $\mathcal{S}[\rho\|\sigma] = \text{Tr}[\rho(\log \rho - \log \sigma)]$. For this, the closest Markov process

tensor is obtained by discarding the correlations. That is,

$$\Upsilon_{k:0}^{\text{Markov}} = \text{Tr}_{\bar{k}}[\Upsilon_{k:0}] \otimes \text{Tr}_{\bar{k-1}}[\Upsilon_{k:0}] \otimes \cdots \otimes \text{Tr}_{\bar{0}}[\Upsilon_{k:0}], \quad (11)$$

where \bar{j} is the trace over every index except o_j and i_j . These marginals constitute the above average CPTP maps exactly. Once reconstructed with PTT, this informs the user how close their device performs when compared to a temporally uncorrelated (not necessarily noiseless) Markov model.

C. Performing PTT on NISQ devices

To reconstruct the process tensor, a minimal complete basis for the process tensor requires d_S^4 operations spanning the superoperator space $\mathcal{B}(\mathcal{B}(\mathcal{H}_S))$ at each time step. One mathematically convenient basis is an IC POVM, followed by a set of IC preparation states that are independent of the measurement outcomes. However, this procedure requires fast projective control in device hardware. Although some progress has been made on this front [55], fast control, in practice, is often too noisy and leads to a poor reconstruction. An exception to this is the recent work on process characterization with intermediate measurements completed by Xiang *et al.* in Ref. [56]. A different approach to this measure-and-prepare strategy could implement this entire basis set through an interaction between the system qubit and an ancilla, followed by projective measurement in which the outcome of the ancilla is recorded. This too is problematic in practice, as system-ancilla interactions will generate an operation that effectively depends on the circuit.

For a typical NISQ device, all intermediate operations are limited to unitary transformations and a measurement is only allowed at the end—or if midcircuit measurement with feed-forward is possible, it is typically much slower than SE dynamics. Nevertheless, it is possible to work within the experimental limitations and implement an informationally incomplete set of basis operations. This constructs what is known as a *restricted* process tensor [57] and has full predictive power for any operation in a subspace of operations. That is, these objects are well defined as maps over the span of their incomplete basis but do not form positive operators and do not uniquely fix a Choi state of a process tensor.

We expand on the notion of a restricted process tensor here by offering an analogous quantum state perspective. The measurement of ρ_k subject to some sequence of operations $\mathbf{B}_{k-1:0}^{\bar{\mu}}$ is akin to measuring a many-body observable on $\Upsilon_{k:0}$, as per Eq. (10). The reconstruction of $\Upsilon_{k:0}$ then lies on the hyperplane defined by

$$\text{span}(\{\Pi_i, \mathbf{B}_{k-1:0}^{\bar{\mu}}\}). \quad (12)$$

When the set of operations is tomographically incomplete, $\Upsilon_{k:0}$ is nonuniquely fixed and thus termed “restricted.” Consider, for example, in the state case, if only Z and X measurements are performed on each subsystem. Then a consistent nonunique state can be estimated with the correct Z and X expectations. However, the Y expectations will be a free parameter (up to positivity of the state).

In the special case of unitary control, the Choi states representing each intervention are rank-1 projections onto a maximally entangled state of dimension d_S^2 . The expansion of these entangled measurements into a Hermitian basis yields only nonzero coefficients on the nonlocal terms.

That is, in the Pauli basis for example, we have $\text{Tr}(\mathcal{B}_i^{\mu_l} \times \mathbb{I}^{(i_{l+1})} \otimes P_j^{(o_l)}) = \text{Tr}(\mathcal{B}_i^{\mu_l} \times P_i^{(i_{l+1})} \otimes \mathbb{I}^{(o_l)}) = 0 \forall i, j$. Unitaries are consequently fully orthogonal to the span of nonunital (where the maximally mixed state is mapped to something more pure) and trace-decreasing (stochastically applied) maps. Linear-inversion reconstruction of the process tensor, then, omits these local expectation values.

A restricted process tensor is not a model restriction as such, but an observational restriction. Its properties are fully consistent with the discussed facets of full process tensors but the mapping is only valid across the span of observed data. This means, for example, that it will provide a recipe for complete control (under the restriction) that is fully inclusive of non-Markovianity. However, the actual strength of the memory can only be inferred rather than directly measured. For example, in the absence of measurement causal breaks, correlations between past and future measurement statistics cannot be established. Moreover, any measure relying on a full eigendecomposition of the state (such as quantum mutual information, for example) is similarly out of reach. The relevant analogy, then, is to perform QST without measuring in all possible bases. The state will not be fully determined but the information provided through Born’s rule to predict the future will still be valid, so long as predictions are made within linear combinations of the measured bases.

Working within the constraints of NISQ devices, we need to account for $d_S^4 - 2d_S^2 + 2$ unitary operations at each time step. For a qubit, this amounts to $N = 10$ unitary gates per time step. However, any estimation procedure will come with sampling error, leading to both an incorrect and unphysical representation of the map. In a practical setting with finite sampling error, it is best to set up tomographic protocols without bias in the basis vectors [58]. This is especially true in high-dimensional spaces, where even small errors may become significantly magnified. In Ref. [21], we have found that a minimal single-qubit ($N_{\min} = 10$) unitary basis incurs substantial error in reconstructing the process. We have thus resorted to an overcomplete basis of $N_{\text{oc}} = 24$, leading to very-high-fidelity reconstruction for the process tensor, within shot-noise precision. However, this is very resource demanding, since

$$\text{number of experiments} \sim \mathcal{O}(N_{\text{oc}}^k) \quad (13)$$

for k time steps.

Our focus here is to reduce the requirements of PTT reconstruction while obtaining accurate estimates. To do so, in Sec. III we first integrate the maximum-likelihood optimization for PTT. This has the advantage that we no longer need an overcomplete basis and thus we can reduce the base from 24 to 10. In order to do this, we ensure that the MLE accommodates the affine conditions of the process tensor; devise a numerical method to generate

an approximately unbiased basis, which minimizes reconstruction errors; and develop a projection method to ensure the physicality of the process tensor. Of course, the MLE alone is not enough if we also cannot reduce the exponent k . In Sec. IV, we integrate the MLE tools with a more generic truncation method of Markov order. The idea here is to truncate small long-time memory to exponentially reduce the number of reconstruction circuits, while retaining high fidelities for the reconstruction. Along the way, we demonstrate that both the MLE and the Markov order methods are practically implementable by applying these ideas to superconducting quantum devices.

D. SPAM errors in process tensor tomography

It is key to know the limitations of any QCVV procedures, in particular how the estimate is affected by faulty control—or SPAM errors. In the context of PTT, the usual notions of SPAM need to be broadened. For example, quantum channels are assumed to act on “known” input states; when this assumption breaks, it can be problematic for QPT. In contrast, the initial state is a marginal to the process tensor and any error is naturally estimated. A measurement probe, on the other hand, is required to read out information. Errors that are insensitive to the POVM effect will absorb into the process. If they are common to all bases, the predictive capabilities will be unaffected. However, the estimated process tensor itself will then look slightly noisier. Accounting for either this or basis-specific noise can be straightforwardly achieved by using an estimate of the device POVM in the model. Estimates may be obtained with consistent detector-tomography outputs from procedures such as GST [22].

For PTT, it is the input quantum operations that are assumed to be known, replacing “state preparation” in QPT. In order of increasing consequence, violation of this assumption can occur in three ways: (i) with gate-independent error (such as decoherence), (ii) with gate-dependent coherent error, and, finally, (iii) with significant SE interaction during the finite-time gate. Similar to the measurement case, PTT is insensitive to independent error for the purposes of control, since it does not change the linear relation between basis elements. The second consideration is more problematic because it can lead to an inconsistent characterization. This can be resolved in two ways: by *a priori* characterizing the gates themselves through GST or by using an overcomplete basis to average over the coherent error. Lastly, if a non-Markovian interaction occurs with coupling $\mathcal{O}(1/\tau_p)$ for control width τ_p , then the process-tensor model will break down. However, we expect this final possibility to be extremely rare for any functioning device—but, indeed, it could be circumvented with virtual gates, such as those described in Ref. [59].

Because the input control must be high fidelity, we view PTT predominantly as a useful tool for quantum devices

clean enough to be sensitive to non-Markovian dynamics. In this work, the demonstrated results focus on single-qubit unitary gates, for which the error is $\mathcal{O}(10^{-4})$. A future extension to PTT that one might consider is a self-calibrating simultaneous estimation of both the process tensor and the input interventions, as in GST. Though incorrect characterizations through gate errors are not implausible, single-qubit gate errors for a typical NISQ device are already smaller than the expected $1/\sqrt{N_{\text{shots}}}$ sensitivity and most of this error is represented by decoherence during the small finite pulse width. Moreover, the control aspect may be self-consistently checked by comparing predictions made from estimates of the process tensor with random gate sequences on the real device, offering certification to the characterization. We explore this concept further in Sec. III B.

III. MAXIMUM-LIKELIHOOD PROCESS TENSOR TOMOGRAPHY

A major gap in the process tensor tomography toolkit is its lack of integration with standard tomography estimation tools such as MLE, the underlying principle of which is to find a physical model estimate that maximizes the probability with the observed data. Due to the intricate affine conditions of causality, this integration is nontrivial in general. The complexity of the procedure grows further when applied to restricted process tensors, e.g., when control operations are restricted and/or when a finite Markov order model is imposed. Our integration naturally accommodates all of these variations. We now close this gap and present a MLE construction for PTT, which helps to put this on the same footing as other tomographic techniques. The MLE procedure estimates the physical quantum map most consistent with the data, according to some desired measure, while respecting the constraints listed in Table I. The circuits required for tomography, depicted in Fig. 3(a), are the same as for linear inversion. That is, the scaling is the same. However, linear inversion typically requires an overcomplete basis to naturally average over inconsistencies. Meanwhile MLE treats the data such that a minimal tomographically complete basis suffices for accurate results.

An estimate for the map is coupled with a metric of goodness (the likelihood) that quantifies how consistent the map is with the data. The cost function is then minimized while enforcing the physicality of the map. The stored data vector in PTT is the object $n_{i,\bar{\mu}}$, which contains the observed measurement probabilities for the i th effect of an IC POVM, subject to a sequence of k operations $\bigotimes_{j=0}^{k-1} \mathcal{B}_j^{\mu_j}$. As is typical in MLE tomography, these data are fitted to a model for the process, $\Upsilon_{k;0}$, such that

$$p_{i,\bar{\mu}} = \text{Tr} \left[(\Pi_i \otimes \mathcal{B}_k^{\mu_{k-1}T} \otimes \cdots \otimes \mathcal{B}_0^{\mu_0T}) \Upsilon_{k;0} \right]. \quad (14)$$

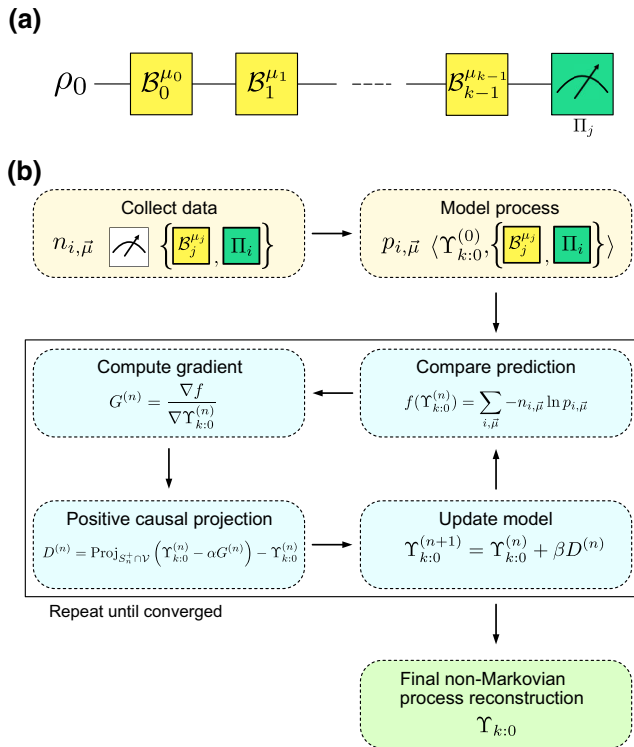


FIG. 3. (a) The circuit structure for PTT. An arbitrary state is fed in, the experimenter acts with all combinations of different basis elements at different times, and a final measurement is recorded. (b) The logical flow of PGDB in the context of our MLE-PTT procedure. We maximize the likelihood of the model through iterative gradient descent and the physical projection of Sec. III A until some convergence condition is achieved. Full details of the algorithm are shown in Appendix A.

These predictions are then compared to the observed frequencies, $n_{i,\bar{\mu}}$. The “likelihood” of $\Upsilon_{k:0}$ subject to the data is given by $\mathcal{L} = \prod_{i,\bar{\mu}} (p_{i,\bar{\mu}})^{n_{i,\bar{\mu}}}$. The cost function of MLE algorithms is then the log likelihood, i.e.,

$$f(\Upsilon_{k:0}) = -\ln \mathcal{L} = \sum_{i,\bar{\mu}} -n_{i,\bar{\mu}} \ln p_{i,\bar{\mu}}, \quad (15)$$

the minimization of which is the maximizer of the likelihood. A key part of the appeal to MLE is that the cost function given in Eq. (15) is convex.

An extensive selection of different semidefinite program packages exists in the literature for the log-likelihood minimization in QST and QPT under the appropriate constraints. In our construction of the MLE-PTT procedure, we employ and adapt the algorithm from Ref. [60], used for QPT. This algorithm is termed the “projected gradient descent with backtracking” (PGDB). We select this both for its simplicity and because it has been benchmarked as both faster and more accurate than other MLE-QPT algorithms. In this approach, the log likelihood is minimized using conventional gradient descent but at each iteration,

a projection is made on the step direction to keep the map physical. The main steps are summarized in Fig. 3(b).

The relevant projection—onto the intersection of the cone of CP maps with the affine space of TP channels—is performed using a procedure known as Dykstra’s alternating-projection algorithm [61]. We offer two key advancements here for PTT. First, we determine the exact affine space generated by causality conditions on process tensors, such that the physical constraints are mathematically elucidated. Then, we adapt and introduce a conic projection technique from the optimization literature in order to project onto the space of completely positive causal processes [62]. We find this projection method to far outperform Dykstra’s alternating-projection algorithm in the problem instances, a fact that may be of independent interest for QPT. We detail each of these aspects in the following subsection. Finally, we benchmark the performance of MLE PTT on superconducting quantum devices. These devices, as mentioned above, are limited to unitary control in the middle and a measurement at the end. This is insufficient to uniquely reconstruct the complete Choi state of a process. As such, our MLE procedure yields an operationally well-defined restricted process tensor that has been completed into a full process tensor. One might consider a “family” of process tensors generated by the intersection of positive causal matrices with the affine space of observed experimental data. This yields all possible process tensors the restriction of which to unitary operations is consistent with the observed data. Therefore, further information about the full dynamics may be inferred even with limited data. We focus on the (nonunique) properties of the restricted process-tensor family in Ref. [63] and the performance of the restricted process tensor in the present work.

A. Projecting onto the space of physical process tensors

Here, we describe in detail the physical conditions imposed on process tensors, as well as the approach used for projections onto the space of physical process tensors. Generally, this projection can be described as a problem of conic optimization: finding the closest point lying on the intersection of a cone with an affine subspace. The affine constraints differ in each category: unit trace for state tomography, trace preservation for process tomography, and causality for the process tensor. Fundamentally, however, these techniques are applicable to all forms of quantum tomography, as shown in Table I.

Let $\Upsilon_{k:0}$ be the Choi form for a k -step process tensor (we occasionally switch to Υ for brevity if the number of steps is not pertinent) and let $|\Upsilon\rangle\rangle := \text{vec}(\Upsilon)$, where we employ the row-vectorized convention [64]. We discuss the mathematical demands of positivity and causality first,

their individual projections, and then their simultaneous realization.

Similar to a quantum channel, complete positivity of a process tensor is guaranteed by positivity of its Choi representation,

$$\Upsilon_{k:0} \in \mathcal{S}_n^+, \quad (16)$$

where \mathcal{S}_n^+ is the cone of $n \times n$ positive-semidefinite (PSD) matrices with complex entries. For k time steps, $n = 2^{2k+1}$. The Euclidean projection is computed with a single eigen-decomposition. Diagonalization of Υ gives $\Upsilon = UDU^\dagger$, where $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ is real. Then the projection onto \mathcal{S}_n^+ is

$$\text{Proj}_{\mathcal{S}_n^+}(\Upsilon) = U \text{diag}(\lambda_0^+, \dots, \lambda_n^+) U^\dagger \quad (17)$$

with $\lambda_j^+ := \max\{\lambda_j, 0\}$.

The Choi state must also obey causality, a generalization of trace preservation. This is nontrivial to enforce and ensures that future events do not influence past statistics. In the CJI picture of Fig. 2(b), there should be no correlations between the final input leg and the rest of the process when the final output leg is traced out. This is also a statement of containment of the process tensor: that the process over a subset of the total period is contained within the larger process tensor:

$$\text{Tr}_{\sigma_k} [\Upsilon_{k':0}] = \mathbb{I}_{i_k} \otimes \Upsilon_{k'-1:0}, \quad (18)$$

iterated for all values of k' from 1 to k . This statement is equivalent to causality in that the past stochastic process is unaffected by averaging over all future operations.

We approach the problem of causality enforcement in the Pauli basis. Examining the Choi state, this condition places constraints on the values of these expectations. Let $\mathbf{P} := \{\mathbb{I}, X, Y, Z\}$ denote the single-qubit Pauli basis, \mathbf{P}^n its n -qubit generalization, and $\tilde{\mathbf{P}} := \{X, Y, Z\}$. Focusing on the $i_{k'}$ subsystem, Eq. (18) can be enforced if all Pauli strings connecting the identity on the left subsystems with $\tilde{\mathbf{P}}$ on the $i_{k'}$ subsystem have coefficients of zero. If this condition is imposed iteratively for all input legs on the process tensor, then Eq. (18) will hold for all k' . For example, in a two-step single-qubit process (represented by a five-partite system), we have

$$\begin{aligned} \langle \mathbb{I}_{\sigma_2} P_{i_2} P_{\sigma_1} P_{i_0} P_{\sigma_0} \rangle &= 0 \forall P_{i_2} \in \tilde{\mathbf{P}}; P_{\sigma_1}, P_{i_0}, P_{\sigma_0} \in \mathbf{P} \\ \langle \mathbb{I}_{\sigma_2} \mathbb{I}_{i_2} \mathbb{I}_{\sigma_1} P_{i_1} P_{\sigma_0} \rangle &= 0 \forall P_{i_0} \in \tilde{\mathbf{P}}; P_{\sigma_0} \in \mathbf{P}. \end{aligned} \quad (19)$$

A simple way to enforce this condition is with the projection of Pauli coefficients. In particular, let \mathcal{P} be the elements of \mathbf{P}^{2k+1} the expectations of which must be zero from Eqs. (18) and (19). We can write this as a single affine

constraint in the matrix equation:

$$\begin{pmatrix} \langle \mathcal{P}_0 | \\ \langle \mathcal{P}_1 | \\ \vdots \\ \langle \mathcal{P}_{m-2} | \\ \langle \mathbb{I} | \end{pmatrix} \cdot |\Upsilon\rangle\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ d \end{pmatrix}, \quad (20)$$

where d is the normalization chosen for the Choi matrix (in this work, $d = 1$).

Letting this set the context for our discussion of the projection routine, consider a full-rank constraint matrix $A \in \mathbb{C}^{m \times n^2}$, variable vector v , and fixed right-hand side coefficient vector b . Let \mathcal{V} be the affine space:

$$\mathcal{V} = \{v \in \mathbb{C}^{n^2} | Av = b\}. \quad (21)$$

The projection onto \mathcal{V} is given by

$$\text{Proj}_{\mathcal{V}}(v_0) = [\mathbb{I} - A^\dagger(AA^\dagger)^{-1}A] v_0 + A^\dagger(AA^\dagger)^{-1}b. \quad (22)$$

In general, however, a projection onto \mathcal{S}_n^+ and a projection onto \mathcal{V} is not a projection onto $\mathcal{S}_n^+ \cap \mathcal{V}$. The conic and affine constraints are difficult to simultaneously realize. One approach to this is to use Dykstra's alternating-projection algorithm, as performed in Ref. [60] for quantum process tomography. This applies a select iterative sequence of Eqs. (22) and (17). Although this method is straightforward and has guaranteed convergence, we find it unsuitable for larger-scale problems. For large gradient steps, the convergence can take unreasonably many steps. More importantly, however, each step of the gradient descent requires many thousands of applications of Eq. (22). Although much of this expression can be pre-computed, the complexity grows strictly with n , rather than the number of constraints. Moreover, the matrix inverse requirement can reduce much of the advantage of having a sparse A .

In our MLE PTT, instead of Dykstra's alternating-projection algorithm, we integrate a variant of the technique introduced in Ref. [62] and discussed further in Ref. [65]. This method regularizes the projection into a single unconstrained minimization, such that only eigen-decompositions and matrix-vector multiplications by A are necessary, avoiding the need for Eq. (22). Note that to guarantee uniqueness of the projection as well as convergence of projected gradient descent in general, the closest physical process tensor at each step is found in terms of the Euclidean distance. For further detail on this, see Refs. [62,66]. Because the projection is the only component of PGDB that we change with respect to Ref. [60], we explicitly walk through the steps in the following paragraph. We also benchmark this direct conic projection routine on normally distributed random matrices with

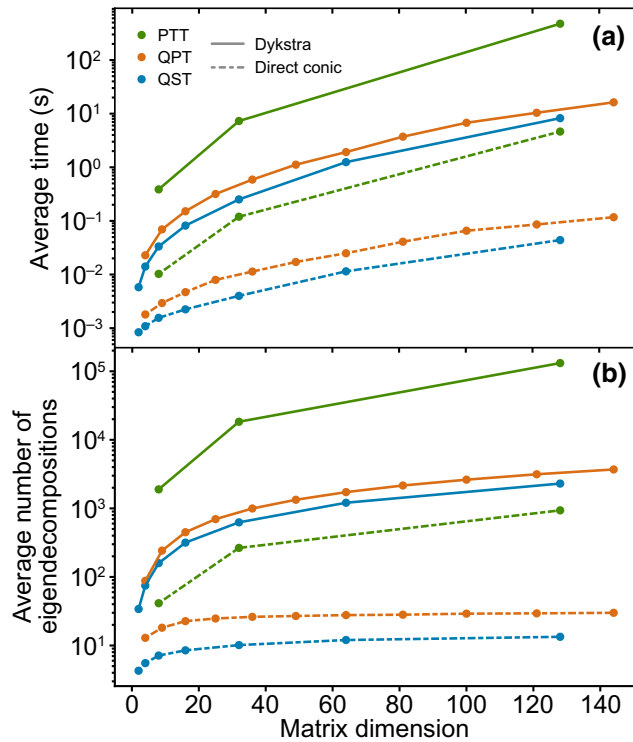


FIG. 4. A comparison between projection methods imposing physical conditions on 500 normally distributed random matrices. (a) The average time taken for a single projection for both Dykstra’s alternating-projection algorithm and the direct conic projection. We compare conditions set by QST, QPT, and PTT. (b) The average number of eigendecompositions for each of the above. This dominates the run time of each method.

respect to Dykstra’s alternating-projection algorithm for the case of QST, QPT, and PTT in Fig. 4. The scaling for each method is similar (dominated by the cost of eigendecompositions) but the absolute savings are of 2 orders of magnitude.

In each respective regime of tomography, the increased number of constraints increases the amount of time, on average, for the projection to complete. However, we find substantial improvements in both the run time and in the number of eigendecomposition calls between the direct conic projection in comparison to Dykstra’s. This is especially necessary for the fitting of process tensors, where the difference between the two can be the difference between a run time of days or of fractions of a minute. We include QST here for completeness; however, note that the fixed projection of eigenvalues onto the canonical simplex with a single diagonalization is more appropriate [67].

We now explicitly step through the direct conic projection method. For a given v_0 , we wish to find the closest (in Euclidean terms) $v \in \mathcal{S}_n^+ \cap \mathcal{V}$; that is, to compute

$$\arg \min_{v \in \mathcal{S}_n^+ \cap \mathcal{V}} \|v - v_0\|^2. \quad (23)$$

Note that when we talk about the vector v being PSD, we mean that its matrix reshape is PSD. The dual approach introduces the Lagrangian, which is a function of the primal variable $v \in \mathcal{S}_n^+$ and dual variable $\lambda \in \mathbb{R}^m$ (for m affine constraints):

$$\mathcal{L}(v; \lambda) = \|v - v_0\|^2 - \lambda^\dagger (Av - b). \quad (24)$$

Since v is PSD, it is Hermitian, meaning that the matrix-vector product Av is always real. This avoids the need for recasting the complex problem into real and imaginary pairs.

The vector v that minimizes \mathcal{L} for a given λ provides a lower bound to the solution to the primal problem. We introduce the dual concave function

$$\theta(\lambda) := \min_{v \in \mathcal{S}_n^+} \mathcal{L}(v; \lambda), \quad (25)$$

the maximum of which is exactly the solution to Eq. (23). It is shown in Ref. [68] that the minimum given in Eq. (25) is uniquely attained by $v(\lambda) = \text{Proj}_{\mathcal{S}_n^+}(v_0 + A^\dagger \lambda)$ [68] and can hence be recast (up to a constant) as

$$\theta(\lambda) = -\|v(\lambda)\|^2 + b^\dagger \lambda. \quad (26)$$

It can further be shown that Eq. (26) is differentiable on \mathbb{R}^m with gradient

$$\nabla \theta(\lambda) = -Av(\lambda) + b. \quad (27)$$

Thus, the solution to the projection problem given in Eq. (23) becomes an unconstrained minimization problem of Eq. (26) with respect to λ , opening the door to the application of a wealth of tested optimization packages. The solution to the projection is then $\text{Proj}_{\mathcal{S}_n^+}(v_0 + A^\dagger \lambda_{\min})$. Specifically, in this work we select the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm to perform this minimization, as we find it to give the fastest and most reliable solution [69]. Although we do not implement it here, it is also possible to compute the Clarke-generalized Jacobian of Eq. (27), allowing for exact second-order optimization techniques to be used [70,71]. Note also that the difficulty of this minimization is sensitive to the condition number of A . Thus, we find the best approach to be to always frame affine constraints in the Pauli basis to ensure a uniform spectrum.

Another favorable reason to apply this conic projection method is in the arbitrary application of affine constraints. In Ref. [63], we show how this can be used for searching (and thus bounding quantities of) manifolds of states consistent with an incomplete set of data. The introduction of a feature matrix as part of the affine constraint with observed probabilities permits this exploration. Without the faster method, we find that this is infeasible to perform.

Using this modification of PGDB with an updated projection routine, we are able to implement PTT both in simulation and in real data. We thus formalize MLE PTT and are now in a position to benchmark the performance of MLE-PTT tomography against the linear-inversion method and look at applications. The results, summarized in the following sections, suggest full characterization of quantum non-Markovian dynamics in an object that is both mathematically minimal and that obeys all of the physical constraints of a quantum stochastic process.

B. Reconstruction fidelity

In the linear-inversion regime, the action of the process tensor on basis sequences will result in the experimentally observed density matrices by construction. Note that by “experimentally observed,” we mean the density matrices as reconstructed by QST. Since a process tensor is a linear operator, its action on linear combinations of basis sequences should be exactly the linear combinations of observed basis actions. This idea is expressed in Fig. 1(c): the SE evolution between each operation is the same for all intermediate operations. In a linear combination, these arbitrarily strong dynamics are entirely accounted for. By tracing over the input space, we have the following relationship between the state conditioned on an arbitrary sequence of operations $\mathbf{A}_{k-1:0}$ and the states after each measured basis sequence:

$$\rho_k(\mathbf{A}_{k-1:0}) = \sum_{\bar{\mu}} \alpha^{\bar{\mu}} \rho_k(\mathbf{B}_{k-1:0}^{\bar{\mu}}), \quad (28)$$

which is equivalent to Eq. (10). Hence, we may (in principle) determine the exact response of the system to any sequence of operations in the presence of non-Markovian interaction. We use this as the figure of merit for the quality of characterization. That is, we compute the fidelity over random sequences of operation between the state predicted by the process tensor—Eq. (10)—and what is realized on the device. In Ref. [21], we have introduced this as the concept of reconstruction fidelity.

Formally, let the fidelity F between two process tensors $\mathcal{T}_{k:0}^{(1)}$ and $\mathcal{T}_{k:0}^{(2)}$ for a given sequence of interventions $\mathbf{A}_{k-1:0}$ be given by

$$F_{(1,2)}[\mathbf{A}_{k-1:0}] = F\left(\mathcal{T}_{k:0}^{(1)}[\mathbf{A}_{k-1:0}], \mathcal{T}_{k:0}^{(2)}[\mathbf{A}_{k-1:0}]\right),$$

where $F(\rho, \sigma) = \text{Tr} \left[\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right]^2$. (29)

Here, $\mathcal{T}^{(1)}$ is taken to be the reconstructed process and $\mathcal{T}^{(2)}$ to be the real process, i.e., the experimental outputs. Then,

the average reconstruction fidelity is an estimate of

$$\mathcal{F} := \int d\mathbf{A}_{k-1:0} F_{(1,2)}[\mathbf{A}_{k-1:0}]. \quad (30)$$

We can use this to estimate the quality of our reconstruction. The outputs of a real process are simply the state reconstructions conditioned on a sequence of gates $\{\mathcal{A}_0, \dots, \mathcal{A}_{k-1}\}$. This integral can be estimated by performing sequences of randomly chosen operations and comparing the fidelity of the predictions made by $\mathcal{T}_{k:0}$ with the actual outcomes measured. We use this as a metric for the accuracy with which a process has been characterized.

As well as MLE PTT, we also improve upon LI PTT through particular choice of a basis. The sampling error typically averages out to zero on a circuit-by-circuit basis; however, this noise can be both biased and amplified if certain regions of superoperator space are overrepresented—i.e., if some elements of the basis overlap more than others. For this reason, we build upon the idea of a mutually unbiased basis (MUB) in conventional tomography. Unfortunately, mutually unbiased *unitary* bases (MUUBs) do not exist in ten dimensions [72]. However, in Appendix C, we numerically find the best approximation to a MUUB.

We probe each of these characteristics by looking at three-step PTT on IBM Quantum devices. Using combinations of basis choice and LI-MLE postprocessing, we compute the reconstruction fidelities for random sequences of unitaries. The box plots showing these distributions are shown in Fig. 5. Each data point constitutes a different sequence of random unitary gates. The number is then the fidelity between the density matrix predicted through action of the reconstruction process tensor on these unitary mappings and the actual density matrix reconstructed through QST on the device after executing that specific unitary sequence. The use of MUUB alone finds substantial improvement in characterizing the process. We continue to use these optimal parameter values as our unitary control basis for the remainder of this work. Reconstruction is improved further by MLE PTT, in which we not only see an increase in median reconstruction fidelity but there are far fewer outliers in the distribution. Compared to the random basis linear-inversion case, reconstruction fidelity increases greatly to within shot noise. This is essential for both validation of the process characterization and optimal control of the system.

In Ref. [21] (from which some of this device data is taken), much of the linear-inversion characterization noise is overcome with the use of an overcomplete basis—up to 24 unitaries. Since PTT is exponential in the size of the basis, the employment of maximum-likelihood methods as shown here can offer a significant reduction in experimental requirements. Thus, in addition to offering an algorithm that imposes physicality constraints, we show how to make

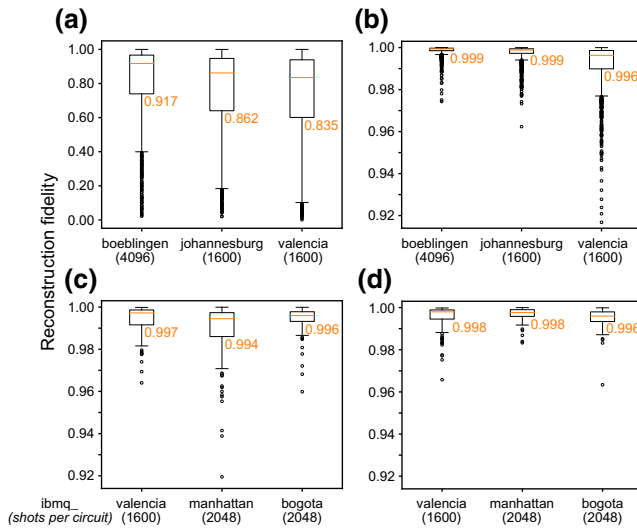


FIG. 5. The reconstruction fidelity of various three-step process-tensor procedures when using a minimal complete basis. Each data point represents a different randomly generated unitary sequence. The top and bottom of the box plots are 75 and 25 percentiles, the orange line is the median (figure also printed), the whiskers are 1.5 times the interquartile range, and any remaining data points are outliers beyond this. We compare reconstructions with (a) a randomly generated basis and linear inversion, (b) a randomly generated basis processed by maximum likelihood, (c) a mutually unbiased unitary basis (MUUB) processed with linear inversion, and (d) a MUUB processed with maximum likelihood. The results showcase high-fidelity physical process tensors with minimal resources, in contrast to Ref. [21] where the random-basis data are taken.

the technique more practical to implement. The scaling of MLE is therefore

$$\text{number of experiments} \sim \mathcal{O}(N_{\text{MLE}}^k), \quad (31)$$

where N_{MLE} can now be 10 regardless of the specific basis choice.

As well as comparing processing methods, we also juxtapose our approximate MUUB with the minimal randomly chosen unitary basis, where $N_{\text{MUUB}} = 10$. This, too, sees a drastic improvement of the method: though it is not guaranteed to produce a physical process tensor, we see that much-higher-quality predictions are possible without any additional effort in the linear-inversion approach.

IV. CONDITIONAL MARKOV ORDER

Even in the classical case, the price of characterizing the joint statistics of a stochastic process in full generality is exponentially high. Often, however, this is unnecessary in practice as physical processes are sparse. This is because the memory must be carried by another physical system, the size of which then bounds then the size of the memory. Often in practice, the *necessary* complexity of a process

characterization only grows modestly with the size of its memory if, after a certain amount of time, the history and the future are independent from one another.

In such cases, the joint statistics are no longer required between those points in time. This motivates the idea of Markov order [38,73]: the number of previous time steps in the process that are relevant to the present. Concretely, in classical theory, a stochastic process is described by the joint probability distribution of a sequence of events $\mathbb{P}(x_k, x_{k-1}, \dots, x_0)$, occurring at times $\{t_k, t_{k-1}, \dots, t_0\}$. A process with Markov order ℓ then conditionally separates the future $F_j = \{t_{j+\ell+1}, \dots, t_k\}$ from the past $P_j = \{t_0, \dots, t_{j-1}\}$, given the knowledge of the state in the memory block $M_j = \{t_j, \dots, t_{j+\ell}\}$. That is, the above distribution takes the form

$$\mathbb{P}(x_k, \dots, x_0) = \sum_{M_j} \mathbb{P}(P_j | M_j) \mathbb{P}(M_j) \mathbb{P}(F_j | M_j). \quad (32)$$

That is, in order know the probability of an event at a given time, we only need to look at the past ℓ events. Anything beyond that will not affect the future. In the case where memory is indeed infinite but decays in time, we can turn the last equation into an approximate statement. Importantly, the complexity of the whole process goes as d^k , where k is not bounded, while the complexity of a process with Markov order ℓ goes as d^ℓ with a fixed ℓ ; that is, the distribution in Eq. (32) is fully determined by knowing $\mathbb{P}(M_j)$.

It is possible, roughly speaking, to extend the notion of Markov order to quantum stochastic processes by replacing \mathbb{P} with its quantum counterpart Υ . We now apply this idea to MLE PTT and derive a resource-efficient way to characterize even processes with very large numbers of steps. Concretely, integrating MLE PTT with a Markov order of ℓ would reduce the exponential scaling in Eq. (31) to

$$\text{number of experiments} \sim \mathcal{O}(k \times N_{\text{MLE}}^\ell). \quad (33)$$

To achieve this, we build upon the ideas established in Ref. [29] and subsequently realized in Ref. [74].

A summary of our approach is to divide a k -step process up into a number of smaller overlapping process tensors. These smaller process tensors are designed to account for a truncated number of past-time correlations. We then use MLE estimation to fit each of the memory process tensors according to a Markov order model chosen by the experimenter. The finite Markov order process-tensor fitting, adaptive memory blocking, and action across sequences that we introduce here are all novel features of quantum Markov order and offer a method by which non-Markovian behavior on NISQ devices can be feasibly characterized and controlled. We note in passing that without MLE PTT, a Markov order integration would not be possible when

working with restricted process tensors. This is because in this regime the partial traces of the process tensor are not well defined, which makes it difficult to split the process into parts.

A. Structure of quantum Markov order

In the quantum realm, the matter of Markov order is more nuanced than for classical processes. A quantum stochastic process has either Markov order one (the output at any leg is affected only by the previous input) or infinite Markov order (the memory persists indefinitely). That is to say, a Choi state may only be written as a product state or there will exist correlations between all points in time (though saying nothing about the strength of these correlations). For practical purposes, however—especially in the context of quantum computing—there exists the useful concept of *conditional* Markov order [29].

To properly explain this statement about conditioning, we first reemphasize that a process tensor represents a quantum stochastic process. As a consequence, if we gain extra information about the past—for example, what operation was applied by the experimenter—then we may update our description of the process when conditioned on that choice of operation. This is akin to the quantum state picture: if a measurement is made on one qubit as part of a many-body system, then the remaining state can be updated based on the outcome of the measurement. For a dynamical process, this intervention can, in full generality, be a quantum instrument. A quantum instrument is a set of completely positive trace-nonincreasing maps, the sum of which is a CPTP map. We employ this terminology but for readers unfamiliar with the object, it suffices to interpret this as any way an experimenter might manipulate a system, including unitaries, measurements, and reparations. Interested readers may consult Ref. [75]. The conditional state of the process may then exhibit past-future independence.

We now introduce Markov order for a quantum stochastic process, as well as the related notion of instrument-specific conditional Markov order. We then make clear that processes conditioned on generic operations may only exhibit *approximate* conditional Markov order. Finally, we explicitly walk through our calculations of tomographically reconstructing processes with an approximate conditional Markov order ansatz.

To begin, we describe a k -step process with Markov order ℓ . When $\ell = 1$, the only information relevant to the CPTP map $\hat{\mathcal{E}}_i$ is the state mapped at the output of the $\hat{\mathcal{E}}_{i-1}$ step, described by the σ_{i-1} leg of the process tensor. Consequently, there is no context to the previous gates. The choice of operation \mathcal{A}_{i-1} is only relevant insofar as it determines the output state for time $(i-1)$. This constitutes a Markovian process and the dynamics are CP divisible.

Otherwise, it is non-Markovian with $\ell = \infty$ [29]. Intuitively, one may think of this as the statement that there is no way to consistently write a generic quantum state with strictly limited correlations—for example, where each subsystem might have nearest-neighbor correlations but zero correlations with any subsystem outside of this.

Although finite $\ell > 1$ Markov order is well defined for classical stochastic processes, where there is only one basis, there is no generic way to write a quantum state with correlations persisting to the last ℓ subsystems. However, future and past statistics may be independent of one another for quantum stochastic processes when conditioned on the choice of an intermediate instrument. This notion of conditional Markov order may be described as follows. Consider a process tensor $\Upsilon_{k;0}$, which we denote as Υ_{FMP} with the groupings for the past, the memory, and the future, respectively:

$$\begin{aligned} P_j &= \{t_0, \dots, t_{j-1}\}, \\ M_j &= \{t_j, \dots, t_{j+\ell}\}, \\ F_j &= \{t_{j+\ell+1}, \dots, t_k\}. \end{aligned} \quad (34)$$

Let a sequence of operations $\mathbf{C}_{j+\ell;j}$, with $k > j + \ell$ act on the memory block of the process. For the moment, while discussing the basic properties of Markov order in quantum processes, we omit the j and ℓ . However, these become important when propagating processes with a Markov order assumption. Thus, henceforth, $\mathbf{C}_{j+\ell;j}$ is expressed as \mathbf{C}_M . Let $\{\mathbf{B}_M^{\bar{\mu}}\}$ be a minimal IC basis for these times, which includes \mathbf{C}_M , and let $\{\Delta_M^{\bar{\mu}}\}$ be its dual set. The conditional process is given by

$$\Upsilon_{FP}^{(C_M)} = \text{Tr}_M [\Upsilon_{FMP} \mathbf{C}_M^T], \quad (35)$$

where Υ_X is the process tensor across the legs given by the set(s) X . If the past and the future are independent in this conditional process, then it can be written as

$$\Upsilon_{FP}^{(C_M)} = \Upsilon_F^{(C_M)} \otimes \Upsilon_P^{(C_M)}, \quad (36)$$

where

$$\Upsilon_X^{(C_M)} = \text{Tr}_{M\bar{X}} [\Upsilon_{FMP} \mathbf{C}_M^T] \quad X \in \{F, P\}. \quad (37)$$

Note that the condensed language used here is identical to the description used in Eq. (10). If Eq. (36) holds for all elements of $\{\mathbf{B}_M^{\bar{\mu}}\}$, then the process, by construction, can be written as

$$\Upsilon_{FMP} = \sum_{\bar{\mu}} \Upsilon_F^{(\mathbf{B}_M^{\bar{\mu}})} \otimes \Delta_M^{\bar{\mu}} \otimes \Upsilon_P^{(\mathbf{B}_M^{\bar{\mu}})}. \quad (38)$$

A fact of practical importance is that for all sequences of operations $\mathbf{A}_M \notin \{\mathbf{B}_M^{\bar{\mu}}\}$, Eq. (36) cannot hold. This is

because $\{\mathbf{B}_M^{\vec{v}}\}$ is informationally complete, meaning that some operation sequence from outside the set can be expressed as a linear combination:

$$\mathbf{A}_M = \sum_{\vec{v}} \alpha_{\vec{v}} \mathbf{B}_M^{\vec{v}}. \quad (39)$$

Contraction of this operation into the process then yields

$$\begin{aligned} \Upsilon_{FMP}^{(\mathbf{A}_M)} &= \text{Tr}_M [\Upsilon_{FMP} \mathbf{A}_M^T] \\ &= \text{Tr}_M \left[\Upsilon_{FMP} \left(\sum_{\vec{v}} \alpha_{\vec{v}} \mathbf{B}_M^{\vec{v}T} \right) \right] \\ &= \sum_{\vec{v}} \alpha_{\vec{v}} \Upsilon_F^{(\mathbf{B}_M^{\vec{v}})} \otimes \Upsilon_P^{(\mathbf{B}_M^{\vec{v}})}, \end{aligned} \quad (40)$$

which is no longer a product state, and thus the future and the past are *separable* but not completely uncorrelated.

The complexity of characterizing a process grows exponentially in ℓ ; we would prefer to drop the instrument-specific component and employ a generic conditional Markov order model. Explicitly, in this model, we truncate all conditional future-past correlations, treating the conditional state as a product, i.e.,

$$\Upsilon_{FMP}^{(\mathbf{A}_M)} \approx \Upsilon_F^{(\mathbf{A}_M)} \otimes \Upsilon_P^{(\mathbf{A}_M)}. \quad (41)$$

The cost, or approximation, in doing so is determined by the actual memory strength of the process over different times. One meaningful measure of this is the quantum mutual information (QMI) of the conditional state, the left-hand side of Eq. (41). Of course, this information is inaccessible in our truncated characterization. Instead, we continue to use the reconstruction fidelity and experimentally estimate this model error in the ability of each Markov order to predict the behavior of actual sequences of random unitaries [76].

B. Stitching together finite Markov order processes

We turn now to our work in extending the concept of conditional quantum Markov order to a quantum circuit context and to tomographic characterization. Here, we are interested not only in dividing up the process into a single past, memory, and future but doing this for all times in the process. Then, for each time, the previous $\ell - 1$ operations are taken into account. This is performed by iterating through the above computation: at each time step, dividing the circuit up into past, memory, and future. Correlations due to the memory are taken care of via contraction of the relevant operations, leaving the future and the past conditionally independent. This requires a tomographically reconstructed process tensor for the relevant memory steps. These memory process tensors are then stitched together

by the overlapping operation's map of the earlier output state. Our goal is to predict $\rho_\ell(\mathbf{A}_{\ell-1:0})$ by making use of the Markov order structure. At the first step, we have no past, $P_0 = \{\emptyset\}$, and the memory is given by the first ℓ operations. That is, we have to construct the full process tensor $\Upsilon_{M_0} = \Upsilon_{\ell:0}$, which contains all of the conditional states $\rho_j(\mathbf{A}_{j-1:0})$ in M_0 , i.e., $j \leq \ell$. To go beyond time t_ℓ , we need the conditional state $\rho_\ell(\mathbf{A}_{\ell-1:0})$ given by Eq. (10) [see also the first line of Eq. (42)]: contracting Υ_{M_0} with $\mathbf{A}_{\ell-1:0}$. Importantly, this is the state propagated along with the sequences of operations. For this reason, the first memory process tensor is the only one for which the output state is not traced over, since we are not tracing over any alternative pasts. This means that it contracts one more local operation than the remainder.

To get state $\rho_{\ell+1}(\mathbf{A}_{\ell:0})$, we move one step forward with P_1 , M_1 , F_1 . The relevant information is stored in Υ_{M_0} and $\Upsilon_{M_1} = \Upsilon_{\ell+1:1}$. For this process (and for all intermediate blocks in the process), there are three considerations: first, we must account for the memory through its action on the sequence $\mathbf{A}_{\ell-1:1}$ on Υ_{M_1} and trace over its output index at time t_ℓ [see the second line of Eq. (42)], since this state corresponds to a different fixed past. Finally, the operation \mathcal{A}_ℓ connects Υ_{M_0} and Υ_{M_1} by mapping the state $\rho_\ell(\mathbf{A}_{\ell-1:0})$ to time $t_{\ell+1}$ since, as per our conditional Markov order assumption, once M_1 is accounted for, F_1 and P_1 are independent, i.e., their dynamics can be treated as a tensor product. See Fig. 6 for a graphical tensor-network depiction. Note the distinction between here and Fig. 2(d). For a full four-step process tensor, estimation of a single expectation value involves contraction of a tensor of matrix size 512×512 . With $\ell = 2$ conditional Markov order, however, this requires only three tensors with matrix size 32×32 .

Following this recipe, we proceed forward in single steps, generating blocks of P_j , M_j , F_j until we reach time t_k , at which point the final state may be read out. For clarity, the sequence of conditional memory process tensor states is given by

$$\begin{aligned} \rho_\ell(\mathbf{A}_{\ell-1:0}) &= \text{Tr}_{\vec{o}_\ell} \left[\Upsilon_{M_0} \bigotimes_{i=0}^{\ell-1} \mathcal{A}_i^T \right], \\ \Upsilon_j^{(\mathbf{A}_{j-2:j-\ell})} &:= \text{Tr}_j \left[\Upsilon_{M_{j-\ell}} \bigotimes_{i=j-\ell}^{j-2} \mathcal{A}_i^T \right]. \end{aligned} \quad (42)$$

The conditional state $\rho_\ell(\mathbf{A}_{\ell-1:0})$ has the free index \vec{o}_ℓ , corresponding to its output state. All others $\Upsilon_j^{(\mathbf{A}_{j-2:j-\ell})}$ have free indices \vec{i}_j and \vec{o}_j , which, respectively, are contracted with the output and input legs of the operation \mathcal{A}_j and \mathcal{A}_{j+1} , respectively. These are the operations that stitch together the different conditional memory process tensors, where the conditional independence means that the state

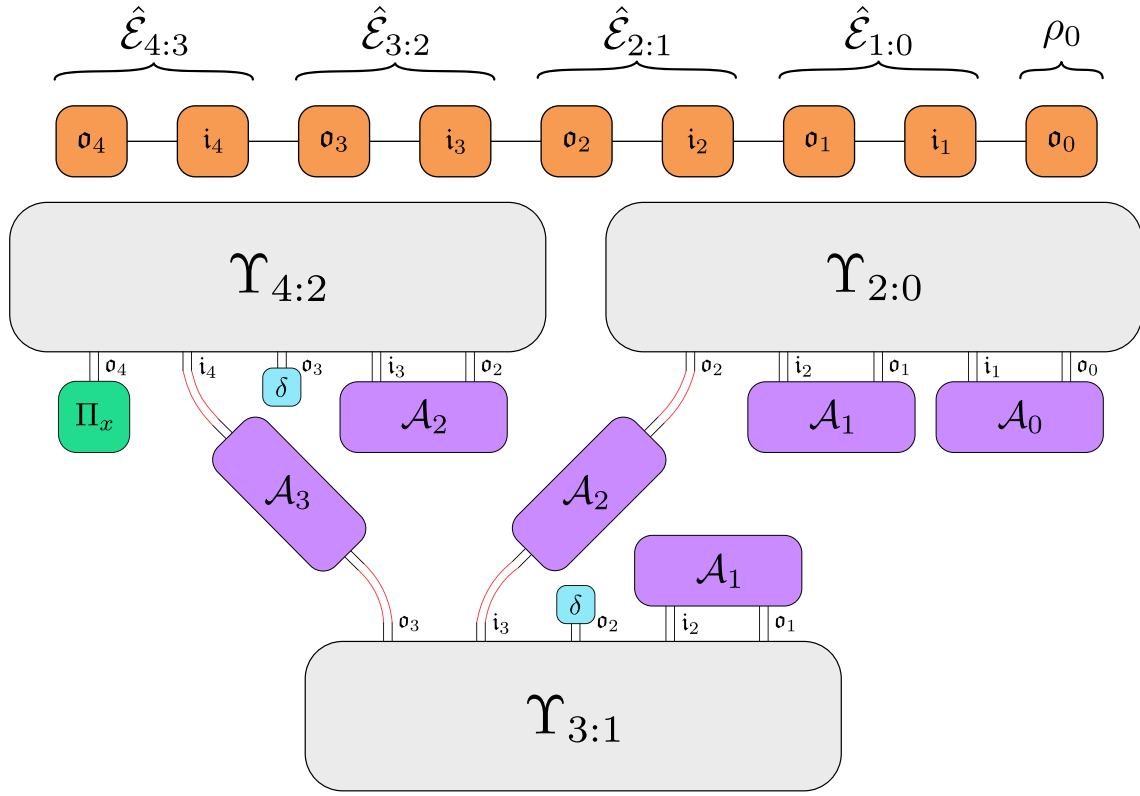


FIG. 6. A contraction strategy for mapping multitime gate sequences using a conditional Markov order ansatz. Here, we show how a four-step process can be modeled by two-step memory process tensors. The memory process tensors are stitched together by first contracting the relevant operations to their times to account for correlations. After tracing over the state output (denoted by δ), the conditionally independent parts can be treated as tensor products and thus stitched together with the latest common operation, where the output of the earlier process tensor is mapped to the input of the later one.

can be mapped as though it is a tensor product. Finally, the last output leg σ_k is read out by some POVM.

We condense this k -step Markov order ℓ process in the object $\Upsilon_{k:0}^\ell := \{\Upsilon_{M_{k-\ell}}, \Upsilon_{M_{k-\ell-1}}, \dots, \Upsilon_{M_0}\}$. That is, the final state is defined by the collective action of each Υ_{M_j} as

$$\begin{aligned} \rho_k(\mathbf{A}_{k-1:0}) &\approx \Upsilon_{k:0}^\ell * \mathbf{A}_{k-1:0} \\ &:= \text{Tr}_{\bar{\sigma}_k} \left[\Upsilon_k^{(\mathbf{A}_{k-2:k-\ell})} \left(\bigotimes_{j=\ell}^{k-1} \Upsilon_j^{(\mathbf{A}_{j-2j-\ell})} \mathcal{A}_j^T \right) \right]. \end{aligned} \quad (43)$$

Note that since the same control operation may contract into multiple different memory process tensors, this action is no longer linear in $\mathbf{A}_{k-1:0}$.

Recalling our earlier depiction of the process tensor in Fig. 2, the dynamics can be described as a collection of correlated CPTP maps $\{\hat{\mathcal{E}}_{j:j-1}\}$. In the CJI picture, past operations are equivalently seen as measurements on these earlier states. Thus, the $\Upsilon_j^{(\mathbf{A}_{j-2j-\ell})}$ are exactly the conditional memory states $\hat{\mathcal{E}}_{j:j-1}^{(\mathbf{A}_{j-2j-\ell})}$. With correlations accounted for, they can be treated locally in time. Any

process may be written exactly as a sequence of conditional CPTP maps, but in full generality they depend on the whole past. Here, they only depend on the memory. The difference in complexity of characterization is $\mathcal{O}(N^k)$ vs $\mathcal{O}(N^\ell)$. Putting it all together, we have an equivalent form of Eq. (43):

$$\begin{aligned} \rho_k(\mathbf{A}_{k-1:0}) &\approx \mathcal{E}_{k:k-1}^{(\mathbf{A}_{k-2:k-\ell})} \circ \mathcal{A}_{k-1} \circ \dots \circ \mathcal{E}_{\ell+2:\ell+1}^{(\mathbf{A}_{\ell:2})} \\ &\quad \circ \mathcal{A}_{\ell+1} \circ \mathcal{E}_{\ell+1:\ell}^{(\mathbf{A}_{\ell-1:1})} \circ \mathcal{A}_\ell[\rho_\ell(\mathbf{A}_{\ell-1:0})] \\ &\text{with } \rho_\ell(\mathbf{A}_{\ell-1:0}) = \mathcal{T}_{M_0}[\mathbf{A}_{\ell-1:0}]. \end{aligned} \quad (44)$$

To summarize, the process with a conditional Markov order ℓ ansatz $\Upsilon_{k:0}^\ell$ is represented by a collection of memory process tensors $\{\Upsilon_{k:k-\ell}, \Upsilon_{k-1:k-\ell-1}, \dots, \Upsilon_{\ell+1:1}, \Upsilon_{\ell:0}\}$. As discussed, it cannot be represented generically by a quantum state but this collection of memory process tensors defines its action on a sequence of k operations. The contraction strategy for a series of control operations (with $k = 4$ and $\ell = 2$) is shown in Fig. 6. In short:

- (1) Contract the first ℓ operations into $\Upsilon_{\ell:0}$, producing the output state at time t_ℓ .

- (2) Contract operations 2 to ℓ into $\Upsilon_{\ell+1:1}$.
- (3) Trace over the output index of $\Upsilon_{\ell+1:1}$ at time t_ℓ (since this is not representative of the actual state of the system subject to all operations).
- (4) Taking the $(\ell + 1)$ th operation to be conditionally independent of the first, this can be applied across the tensor product of the two process tensors into the indices for the output state at time t_ℓ and the input for time $t_{\ell+1}$.
- (5) Repeat this pattern for the next $k - \ell - 2$ process tensors.
- (6) Read out the final state at time t_k .

The full details of this computation can be found in Appendix B, where we fully describe an efficient tensor-network contraction for the action of $\Upsilon_{k:0}^\ell$.

C. Circuits for Υ_{M_j}

With a framework established for constructing and operating a finite Markov order ansatz, we now explicitly detail how to tomographically reconstruct this model on a real device. This procedure does not deviate significantly from Sec. III. In order to estimate $\Upsilon_{k:0}^\ell$, we must estimate each of the memory-block process tensors. Recall that each $\Upsilon_{M_j} := \Upsilon_{\ell+j:j}$ is equivalent to $\Upsilon_{\ell+j:0}$ with the first j times projected out onto some series of interventions. In order to experimentally reconstruct each Υ_{M_j} , then, it suffices to fix the first j operations in the circuit and then perform a complete basis of operations in each position from t_j to $t_{\ell+j}$ and estimate the associated ℓ -step process tensor. The circuits required for each of these are illustrated in Fig. 7, with the fixed operation labeled \mathcal{B}_f . As well as sufficiently describing a Markov order ℓ model, these circuits contain all of the information required for any lower-order Markov model if it is a full process tensor. Under the unitary-only restriction, there will be a small number of extra experiments required for any smaller memory blocks terminating earlier than t_ℓ . We also note here that the maximum-likelihood

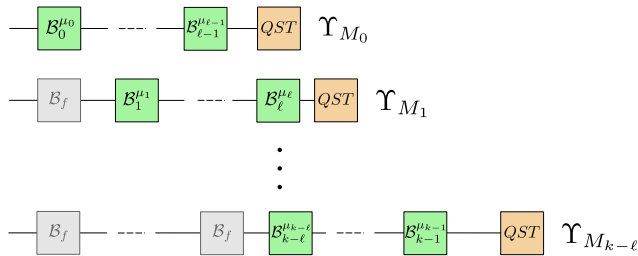


FIG. 7. Circuits to construct a process tensor with conditional Markov order ℓ . For a k -step process, $\Upsilon_{k:0}^\ell$, there are $k - \ell + 1$ memory process tensors that need to be constructed—each circuit represents the estimation of each of these, by varying all combinations of all indices from i_0 to $i_{\ell-1}$. In the other gate positions, a fixed operation \mathcal{B}_f is applied.

procedure of Sec. III is necessary for conditional Markov order models if the set of instruments is restricted to the unitaries. This is because Eq. (43) requires local partial traces but unitary gates are equivalent to entangled measurements in the Choi picture. Hence, if a linear-inversion restricted process tensor is constructed, the partial traces will not be well defined.

In the action of $\Upsilon_{k:0}^\ell$, the state generated by each Υ_{M_j} is traced over for all $j > 0$. Consequently, the fixed operations that precede M_j in the reconstruction circuits should, in principle, not affect the final outcome ρ_k . Since, however, the conditional Markov order (CMO) ansatz is an approximation to the true dynamics, then the fixed past operations will, in practice, affect this approximation. In the generic case, there is no reason to suspect that any operation will put forth a better or worse approximation, hence we arbitrarily set this operation to be the first element of the basis set each time. This choice may require closer attention in practical situations.

With this, we describe how to adaptively characterize a process with quantum and classical requirements only as large as the complexity of the noise (or up to the error that the experimenter is willing to tolerate). Since a process characterization may be verified through the reconstruction fidelity, the best approach to this is to progressively build up and verify a more complex model until the desired precision is reached. This will depend on the intended applications of the characterization. We believe that this is the first procedure to methodically characterize finite Markov order quantum processes. We expect this to be extremely useful in stemming the effects of both correlated and uncorrelated noise on NISQ devices, where the open dynamics are already clean enough so as to be mostly—but not strongly—non-Markovian.

D. Estimating memory build-up

We can combine some of the ideas introduced to give a more fine-grained measure of non-Markovian memory on quantum devices, both in terms of its length and its strength. Specifically, these simplified models of the process can either be employed to streamline control of the quantum system or they may be used as a diagnostic tool by observing how well different restrictions describe the dynamics. This also validates our method of reconstructing processes with conditional Markov order in an efficient way. To estimate non-Markovianity, we construct increasingly complex models by accounting for increasing memory and quantify how well they describe observed device dynamics under the measure of reconstruction fidelity. This measure is computationally convenient, scaling linearly in time steps; has an immediately available interpretation; and may be performed up to acceptable approximation or where costs become prohibitive. The breakdown of conditional Markov order is bounded by

the maximum conditional quantum mutual information (CQMI) as described in Ref. [76]. For a three-step process tensor, the CQMI is taken to be

$$\max_{\mathcal{A}_1} S[\Upsilon_{3:0}^{(\mathcal{A}_1)} || \hat{\mathcal{E}}_{3:2}^{(\mathcal{A}_1)} \otimes \Upsilon_{2:0}^{(\mathcal{A}_1)}], \quad (45)$$

where $S[\rho||\sigma] := \text{Tr}[\rho(\log \rho - \log \sigma)]$ is the von Neumann relative entropy. The conditional Markov order approximation is illustrated in Fig. 8(a).

We perform this procedure as an example on *ibmq_guadalupe* to observe non-Markovianity as a function of time. For a four-step process, we construct process tensors $\Upsilon_{4:0}^\ell$ for $\ell \in \{1, 2, 3\}$. With a $|+\rangle$ state neighbor, the duration of each step is varied across ten different times, ranging from 180 ns up to 1800 ns. For each value of t , we execute 100 sequences of random unitaries $\{\mathbf{U}_{3:0}^i\}$ followed by state reconstruction. Then the action of $\Upsilon_{4:0}^\ell$ on $\mathbf{U}_{3:0}^i$ is used to predict the resulting states. The fidelity between the predicted state and the actual state is computed and the distribution for each data point is shown in Fig. 8. We see here the build-up of memory effects; the

time scales constitute relatively short-depth effective circuits, meaning that these temporal correlations are likely to accumulate across practical circuits. Interestingly, the $\ell = 3$ model performs significantly better than the other two, whereas $\ell = 2$ is only marginally better at predicting the dynamics than $\ell = 1$ [77]. This suggests that most of the memory effects in these devices are higher order—they persist across multiple times. This observation is substantiated by computing the QMI from Eq. (11), as well as finding the operation that maximizes the conditional QMI for the three-step process tensors. The ability to compute these measures comes from our compressed-sensing approach to PTT, detailed in Ref. [63]. The increase of both of these non-Markovian measures is shown in Fig. 8(c). Because of the cumulative build-up, for long-time dynamical processes in real situations, mitigating the effects of these correlations would require either decoupling early or fine graining the process into many more time steps. These memory effects manifest themselves over a time frame of only a few controlled-NOT (CNOT) gates, indicating that non-Markovian dynamics are likely a significant class of noise in regular circuits.

Markovianity breakdown has been previously quantified in terms of model violation in gate sets or in the loss of CP divisibility [4,7,53]. However, these approaches only coarsely diagnose temporal correlations and are not generic to the process. We present a systematic framework by which different levels of finite conditional Markov order may be tested on quantum devices with both a rigorous foundation and practical interpretation.

V. APPLICATIONS OF MULTITIME CHARACTERIZATION

The characterization given in PTT can be useful for qualitatively different applications. Broadly speaking, these applications fall into two different camps: non-Markovian diagnostics and non-Markovian optimal control. In the former, conventional many-body tools are applied to the Choi state to probe the characteristics of the temporal correlations via correlations between the CPTP marginals $\hat{\mathcal{E}}_{j:j-1}$. These characteristics can reveal a great deal about the noise: its complexity, the probability of Markov-model confusion, the size of the environment, as well as its quantum or classical nature, to give some examples [63,78,79]. Since the Choi state is, in general, nonuniquely defined for a restricted process tensor, we do not comment on these aspects here. In Ref. [63], we focus on applying our methods to extract and bound information about these quantities on NISQ devices, with examples shown on IBM Quantum devices.

Here, we focus on control. We present some examples of how a process-tensor characterization can straightforwardly yield superior circuit fidelities on real QIPs and the extent to which a conditional Markov model can be

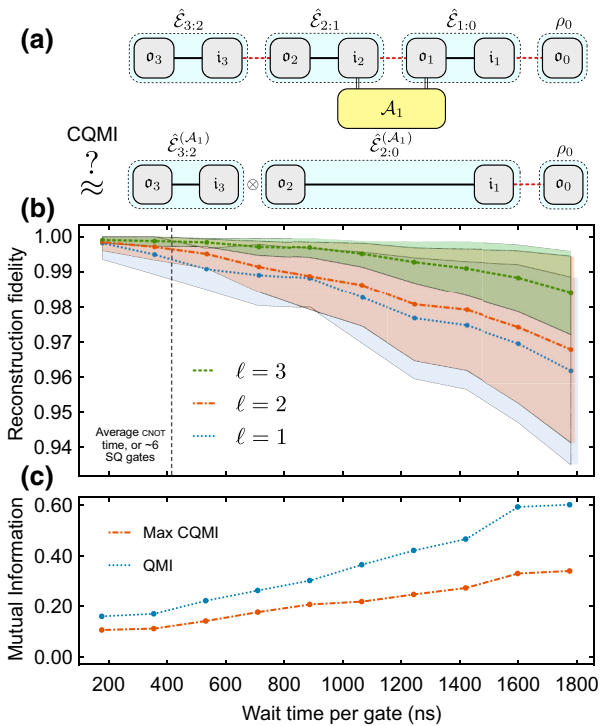


FIG. 8. The build-up of temporal correlations with increasing interaction time. (a) The CQMI quantifies the error in truncating correlations beyond $\ell = 2$ when conditioned on intermediate operations. (b) A four-step process is considered, with an increasing wait time after each gate. Using Markov orders $\ell = 1$, $\ell = 2$, and $\ell = 3$, we quantify how each model reconstructs 100 random unitary sequences. The points indicate the average fidelity and the shaded regions indicate the standard deviation. (c) The QMI and CQMI both increase with the interaction time, bounding the breakdown of $\ell = 1$ and $\ell = 2$ models, respectively.

used for this. Reconstruction fidelity validates the ability of the process tensor to accurately map a given sequence to its final state. This is especially applicable to near-term quantum devices the control operations of which are high in fidelity but the dynamics of which (non-Markovian or otherwise) are not under control. In the same way that a mathematical description of a quantum channel may be used to predict its behavior on any input state, the process tensor can predict the output state of a process, subject to any sequence of input operations. A mapping from unitary gates to outcomes allows an experimenter to ask “What is the optimal sequence of gates that best achieves this outcome?” Two key features distinguishing this from regular quantum optimal control are, first, that after characterization, all optimization can be performed classically with confidence and, second, that the process is fully inclusive of non-Markovian dynamics, allowing for the suppression of correlated errors. Simply choose an objective function \mathcal{L} of the final state. Then $\mathcal{L}(\mathcal{T}_{k:0}[\mathbf{A}_{k-1:0}])$ classically evaluates \mathcal{L} conditioned on some operation sequence $\mathbf{A}_{k-1:0}$ using the process characterization. This can be cast as a classical optimization problem to find the sequence of gates that best results in the desired value of \mathcal{L} . This idea has been preliminarily explored in Ref. [21] and applied to pulse shaping in Ref. [80].

A. Optimization for states using a full process tensor

To demonstrate the utility of this idea, we use the process tensor to improve the fidelity of IBM Quantum devices over multitime processes. Note that this characterization overcomes both Markovian and non-Markovian errors. In particular, we apply many sequences of random unitaries to a single qubit and measure the final state. Interleaved between each operation is a delay time roughly equivalent to the implementation duration of a CNOT gate. We then compare the fidelity of this output state to the ideal output subject to those unitaries. That is, we generate a set of ideal outputs,

$$\rho_{\text{ideal}}^{ijk} = \mathcal{A}_3^k \circ \mathcal{A}_2^j \circ \mathcal{A}_1^i[|0\rangle\langle 0|], \quad (46)$$

with a set of values $\mathcal{F}(\rho_{\text{ideal}}, \rho_{\text{actual}})$. Mirroring these dynamics, we construct a process tensor the basis of the inputs of which is at the same time as the target unitaries. We supplement the reconstruction by using GST to estimate the noisy-device POVM. The estimated POVM is then used in the MLE processing, rather than the ideal projective measurements. The purpose of this is to avoid inflating any circuit improvement. For example, a relaxation process during the measurement operation would be absorbed into the process-tensor estimate and could be artificially overcome by increasing the $|1\rangle$ population. By accounting for measurement errors in PTT and QST, we are considering only dynamics during the circuit as

a more representative depiction of generic PTT capabilities. Finally, using the PTT characterization, we determine which set of unitaries \mathcal{V} should be used (instead of the native ones, \mathcal{A}) in order to achieve the ideal output state. Let each unitary gate V corresponding to the map \mathcal{V} be parametrized in terms of θ , ϕ , and λ as

$$V(\theta, \phi, \lambda) = \begin{pmatrix} \cos(\theta/2) & -e^{i\lambda} \sin(\theta/2) \\ e^{i\phi} \sin(\theta/2) & e^{i\lambda+i\phi} \cos(\theta/2) \end{pmatrix}. \quad (47)$$

The process prediction is then given with

$$\rho_{\text{predicted}}^{ijk}(\vec{\theta}, \vec{\phi}, \vec{\lambda}) = \mathcal{T}_{3:0}[\mathcal{V}_3^k, \mathcal{V}_2^j, \mathcal{V}_1^i]. \quad (48)$$

Then, for each combination ijk , we find

$$\arg \max_{\vec{\theta}, \vec{\phi}, \vec{\lambda}} \mathcal{F}(\rho_{\text{ideal}}^{ijk}, \rho_{\text{predicted}}^{ijk}) \quad (49)$$

and use these optimal values in sequences on the device. The results across 216 random sequences are summarized in Fig. 9(a). The average observed improvement is 0.045, with a maximum of 0.10. In addition, the distribution of fidelities is much tighter, with the standard deviation of device-implemented unitaries at 0.0241, compared with our computed values at 0.00578. We also repeat similar runs on *ibmq_bogota*, intended to drive some characteristics of crosstalk: starting the neighbor in a $|+\rangle$ state followed by four sequential CNOT gates to its other-side nearest-neighbor between each unitary. Initializing the neighbor in a $|+\rangle$ state is intended to generate a passive entangling interaction between the two qubits due to the always-on ZZ interaction found in superconducting transmons. We find that these native fidelities are much worse than on the *ibmq_manhattan*, despite possessing similar error rates—implicating the effects of crosstalk. We emphasize that the noise encountered in all our results is naturally occurring from device fabrication, rather than a contrived environment. Nevertheless, the process-tensor-optimal fidelities in Fig. 9(b) are nearly as high. This suggests a path forward whereby quantum devices may be characterized using PTT and circuits compiled according to the correlated noise of that device. An obvious drawback of this is the characterization requirements. We now investigate carrying out a similar task with our conditional Markov order model.

B. Optimization for arbitrary circuits with finite Markov order process tensors

The use of a complete process-tensor model to optimize circuit sections may be feasible for a small number of gates and, indeed, may be necessary for highly correlated noise. However, it is not desirable in a generic sense to characterize redundant information. Moreover, it is impractical

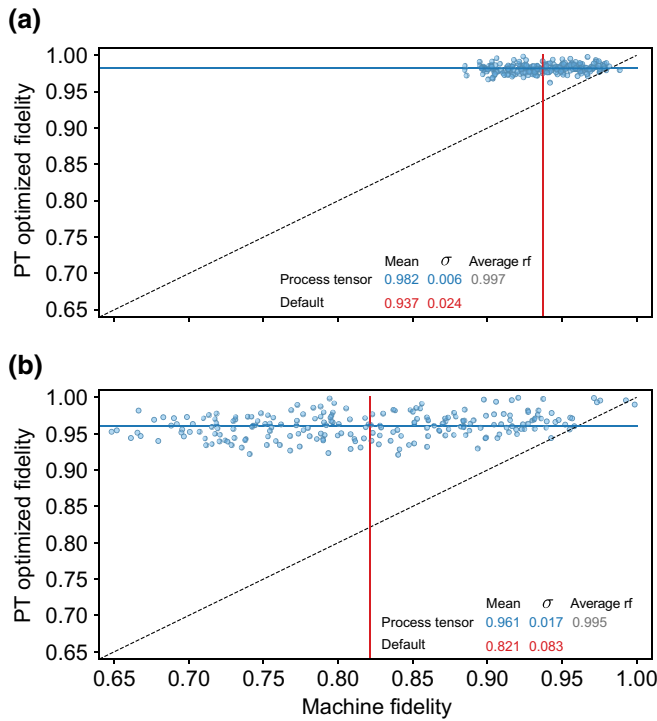


FIG. 9. The results of using the process tensor (PT) to optimize multitime circuits with different random single-qubit unitaries. The x axis indicates the fidelity compared to ideal when the sequences are run on the device. The y axis is the fidelity of the sequence when the process tensor is used to optimize to the ideal case. (a) Here, we have three sequential unitaries with a wait time interleaved similar to that of two CNOT gates ($0.71 \mu\text{s}$) on the *ibmq_manhattan* (no background). (b) A similar setup is considered on *ibmq_bogota* but with the neighboring qubit in a $|+\rangle$ state and subject to four CNOT gates per time step ($2.5 \mu\text{s}$) ($|+\rangle$ and CNOT background).

to optimize over specific circuits in a state-dependent way when inputs may be reduced subsystems of a larger register. Here, we address both of these points. We target longer circuits with larger values of k by using a truncated Markov model. In doing so, we both validate our conditional Markov order methodology and demonstrate the need to approach the problem of NISQ noise with temporal correlations in mind. Further, we also change our optimization approach: instead of trying to create a specific state on a circuit-by-circuit basis, we numerically find the sequence of gates that most closely takes the effective process to be the identity channel. This allows for arbitrary addressing of non-Markovian noise without *a priori* knowing the input state.

A five-step process is considered with delays of approximately $1.2 \mu\text{s}$ after each gate. We characterize this process using conditional Markov order models of $\ell = 1$, $\ell = 2$, and $\ell = 3$ under three different cases: no operations on the background qubits, one nearest neighbor to the system initialized in a $|+\rangle$ state, and, finally, two nearest neighbors

and one next-to-nearest neighbor in a $|+\rangle$ state. The first job takes place on *ibmq_montreal* and the second and third on *ibmq_guadalupe*. The purpose of the latter two analyses is to encourage any (predominantly ZZ) interaction that realistically might occur between qubits in an algorithm. We then generate 100 sequences of five random unitary gates, followed by QST. These sequences serve two purposes: first, we evaluate the reconstruction fidelity for the different Markov order models and, second, we use these as our benchmark for adaptively improving the native fidelity of the device. We select an IC set of unitary gates $\{\mathcal{A}^i\}$ to be applied in the first circuit position, generating a set of ideal states $\{\rho_{\text{ideal}}^i\} := \{\mathcal{A}[|0\rangle\langle 0|]\}$. That is, i indexes a set of random states. We then parametrize the next four gates, again, in terms of $\vec{\theta}$, $\vec{\phi}$, and $\vec{\lambda}$. However, this time, the gates are the same for each input. Finally, using each $\Upsilon_{5,0}^\ell$, we compute

$$\arg \max_{\vec{\theta}, \vec{\phi}, \vec{\lambda}} \sum_i \left[\mathcal{F}(\rho_{\text{ideal}}^i, \rho_{\text{predicted}}^i) \right]^2. \quad (50)$$

In plain words, we are finding the four gates that simultaneously best preserve all of our random input states. After running this optimization for each Markov order model and each background, we then aim to create the ideal output from the 100 random sequences, first, by creating the state with the first gate and then, applying the four gates found from Eq. (50). Following QST at the end, we compute the fidelity of each final state with respect to the ideal. The purpose of this routine is twofold: to determine whether active circuit improvements (akin to dynamical decoupling) can be systematically found, even in the presence of non-Markovian noise, and to ascertain how the inclusion of higher-order temporal correlations in the model can help achieve this task. Without randomizing over the inputs, we find that the $\ell = 1$ model will hide each state in a decoherence-free subspace until the last gate, which is not a generalizable strategy. The results of these runs are shown in Fig. 10 for each sequence and each Markov model, with both the mean circuit fidelities and reconstruction fidelities printed. With no activity on neighboring qubits, we find a moderate amount of non-Markovian noise at this time scale. Interestingly, $\ell = 3$ predicts the dynamics moderately better than $\ell = 2$, signaling the presence of higher-order correlations in the dynamics. However, the optimal interventions improve the average circuit fidelity to a similar level for each. The generic correctability for given circuit structures may saturate, regardless of the completeness of characterization. For the second and third situations, the dynamics are more complex and we see a clear separation between the different Markov orders. By accounting for these higher-order temporal correlations, we are able to more substantially increase the circuit fidelities, both in terms of the mean value and in terms of the tightness of the distribution. Only

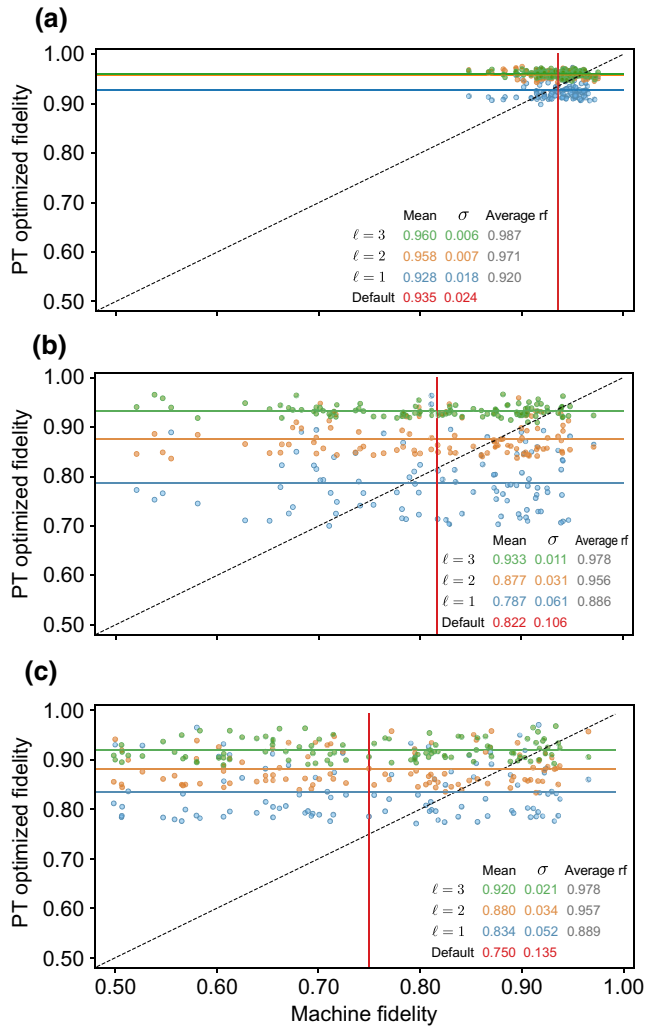


FIG. 10. The results of using conditional Markov models to improve the fidelities of a five-step circuit on *ibmq_montreal* and *ibmq_guadalupe*. We construct conditional Markov order models for $\ell = 1$, $\ell = 2$, and $\ell = 3$ across five steps under a variety of background conditions. For randomized inputs, these models are used to optimize the next four operations. The four optimal operations are then applied to each input and the results compared to the machine fidelity. The mean and standard deviation for each set of circuits is listed, as well as the average reconstruction fidelity of the models. (a) No background operations are applied. (b) The nearest neighbor to the system is initialized in a $|+\rangle$ state. (c) Two nearest neighbors and one next-to-nearest neighbor from the system are initialized in a $|+\rangle$ state.

in the last case do we find that a Markov model $\ell = 1$ is able to achieve an improvement, further highlighting the need for our multitime process characterization on NISQ devices.

VI. DISCUSSION AND CONCLUSIONS

In NISQ devices, circuit performance is not solely determined by the simple composition of high-fidelity gates and measurements but exhibits complex non-Markovian

effects. It is therefore unavoidable to pivot the focus of characterization techniques onto the emergent holistic behavior of multitime processes. In this paper, we formally introduce a multitime generalization of quantum process tomography, in the form of the estimation of process-tensor models. We present several key advancements that we believe will be valuable contributions to the community: we show how to obtain reliable high-fidelity minimal-resource estimations of quantum non-Markovian processes through our fast MLE procedure; we derive a method by which low-memory ansätze can be implemented; and we show how to use our tools to improve the performance of NISQ devices. This is, to the best of our knowledge, the first development of a maximum-likelihood technique for reconstructing multitime processes. Moreover, it permits a modular description of non-Markovian memory. These facets are important not just for quantum-information processing but in the study of multitime correlations that naturally occur in out-of-equilibrium quantum stochastic phenomena, such as in cold atoms, condensed-matter physics, and quantum biology [81–86].

Our technique requires computational resources that scale only linearly in time, while being exponential in the Markov order. The required Markov order plays two important roles: it limits the computational requirements and it benchmarks the degree of non-Markovianity in the device by answering “How many previous time steps are relevant to our current description of the dynamics?” The next step is to further compress the process description by employing the myriad techniques for efficient tomography that are well described in the context of QST and QPT [30,40,87–89]. Alternatively, if one was interested in an informative snapshot of the temporal correlations, shadow tomography could efficiently estimate linear functions of the Choi state, supposing that an informationally complete basis of operations was available [90].

Applications of the process tensor have been promising for improving circuit fidelities. However, much more work needs to be accomplished to render this practical. What is the minimal characterization required to realize this superior control? When can it be reused on other qubits? Can the characterization be used to improve generic (possibly unknown) circuits? The practical and theoretical tools developed here pave the way for answering many of these questions.

One important aspect of the results so far is that they allow for optimizations that mitigate highly correlated noise. These suggest that current devices may be closer to fault tolerance than presently realized. For example, the native error per gate suggested in Fig. 10(c) is approximately 0.02 in the presence of crosstalk, whereas the application of an optimization based on non-Markovian error characterization reduces this to approximately 0.005. But this is not the only consideration; not only do our methods allow for the device to be cleaned up in an absolute sense

but also the reduction of correlated errors has the effect of lowering the stringency of quantum error-correcting thresholds (when compared to a correlated error model). Thus, a non-Markovian characterization can both raise the performance while lowering the bar. We strongly believe that characterization techniques and software will play a large role in the eventual realization of a fault-tolerant quantum computer.

ACKNOWLEDGMENTS

This work was supported by the University of Melbourne through the establishment of an IBM Quantum Network Hub at the University. G.A.L.W. is supported by an Australian Government Research Training Program Scholarship. C.D.H. is supported through a Laby Foundation grant at The University of Melbourne. K.M. is supported through the Australian Research Council Future Fellowship under Grant No. FT160100073. K.M. and C.D.H. acknowledge the support of the Australian Research Council's Discovery Project under Grant No. DP210100597. K.M. and C.D.H. were recipients of the International Quantum U Tech Accelerator award by the U.S. Air Force Research Laboratory.

APPENDIX A: PROCESS-TENSOR CONTRACTION AND MAXIMUM LIKELIHOOD

Throughout the main text, we omit some of the lengthier computations and descriptions relevant to both linear-inversion and maximum-likelihood PTT. We include these as follows for completeness, as well as an outline of our algorithmic implementation of PTT.

1. Construction of a dual set

The procedure to construct the dual operators is as follows: we compile an IC operation set $\{\mathcal{B}_i\}$ into a single matrix \mathfrak{B} . Write each $\mathcal{B}_i = \sum_j b_{ij} \Gamma_j$, where the $\{\Gamma_j\}$ form a Hermitian self-dual linearly independent basis satisfying $\text{tr}[\Gamma_j \Gamma_k] = \delta_{jk}$. In our case, we select $\{\Gamma_j\}$ to be the standard basis, meaning that the k th column of the matrix $\mathfrak{B} = \sum_{ij} b_{ij} |i\rangle\langle j|$ is \mathcal{B}_k flattened into a 1D vector. Because the $\{\mathcal{B}_i\}$ are linearly independent, \mathfrak{B} is invertible. Let the matrix $\mathfrak{F}^\dagger = \mathfrak{B}^{-1}$ such that $\mathfrak{B} \times \mathfrak{F}^\dagger = \mathbb{I}$. This means that the rows of \mathfrak{F}^\dagger are orthogonal to the rows of \mathfrak{B} . The dual matrices can then be defined as $\Delta_i = \sum_j f_{ij} \Gamma_j$, ensuring that $\text{tr}[\mathcal{B}_i \Delta_j] = \delta_{ij}$. Note that in this work, our basis is restricted to the submanifold of unitary matrices. This means that the dimension d of the space is less than the order n of the matrices. Therefore we construct \mathfrak{F}^\dagger as the Moore-Penrose or the right inverse of \mathfrak{B} . If a set of duals is with respect to an overcomplete basis, the same strategy may also be used. Here, we relax the duality condition $\text{tr}[\mathcal{B}_i \Delta_j] = \delta_{ij}$ but retain $\sum_i \Delta_i = \mathbb{I}$ to ensure that the expansion of any operation within the basis is complete.

2. Action of Choi states

When written in terms of its dual construction, it becomes apparent that the action of a quantum channel through its Choi representation is a linear expansion in terms of its action on an IC basis of inputs. Here, we step through this computation and then through the same computation for the process tensor in order to emphasize their parallels. Explicitly, consider $\sigma = \sum_{i=1}^n \alpha_i \rho_i$. The action of \mathcal{E} on σ is given by

$$\begin{aligned} & \text{Tr}_{\text{in}} \left[(\mathbb{I}_{\text{out}} \otimes \sigma^T) \hat{\mathcal{E}} \right] \\ &= \text{Tr}_{\text{in}} \left[(\mathbb{I}_{\text{out}} \otimes \sigma^T) \sum_{i=1}^n \rho'_i \otimes \omega_i^T \right] \\ &= \text{Tr}_{\text{in}} \left[(\mathbb{I}_{\text{out}} \otimes \sum_{j=1}^n \alpha_j \rho_j^T) \sum_{i=1}^n \rho'_i \otimes \omega_i^T \right] \\ &= \text{Tr}_{\text{in}} \left[\sum_{i=1}^n \sum_{j=1}^n \rho'_i \otimes \alpha_j \rho_j^T \omega_i^T \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \rho'_i \alpha_j \text{Tr}[\rho_j \omega_i] \\ &= \sum_{i=1}^n \alpha_i \rho'_i. \end{aligned} \tag{A1}$$

Similarly, for the action of a process tensor on a generic sequence of operations $\mathbf{A}_{k-1:0}$,

$$\begin{aligned} \mathcal{T}_{k:0} [\mathbf{A}_{k-1:0}] &= \text{tr}_{\text{in}} \left[\left(\hat{\mathbf{A}}_{k-1:0} \otimes \mathbb{I}_{\text{out}} \right)^T \Upsilon_{k:0} \right] \\ &= \text{tr}_{\text{in}} \left[\left(\bigotimes_{i=0}^{k-1} \hat{\mathcal{A}}_i^T \otimes \mathbb{I} \right) \sum_{\vec{v}} (\Delta_{k-1:0}^{\vec{v}})^T \otimes \rho_k^{\vec{v}} \right] \\ &= \text{tr}_{\text{in}} \left[\sum_{\vec{\mu}} \alpha^{\vec{\mu}} \bigotimes_{i=0}^{k-1} \hat{\mathcal{B}}_i^{\mu_i T} \sum_{\vec{v}} \bigotimes_{j=0}^{k-1} \Delta_j^{v_j T} \otimes \rho_k^{\vec{v}} \right] \\ &= \text{tr}_{\text{in}} \left[\sum_{\vec{\mu}, \vec{v}} \alpha^{\vec{\mu}} \bigotimes_{i,j=0}^{k-1} \{\hat{\mathcal{B}}_i^{\mu_i T} \Delta_j^{v_j T}\} \otimes \rho_k^{\vec{v}} \right] \\ &= \sum_{\vec{\mu}, \vec{v}} \alpha^{\vec{\mu}} \prod_{i,j=0}^{k-1} \text{tr} \left[\hat{\mathcal{B}}_i^{\mu_i} \Delta_j^{v_j} \right] \rho_k^{\vec{v}} \\ &= \sum_{\vec{\mu}, \vec{v}} \alpha^{\vec{\mu}} \prod_{i=0}^{k-1} \delta_{\vec{\mu} \vec{v}} \rho_k^{\vec{v}} \\ &= \sum_{\vec{\mu}} \alpha^{\vec{\mu}} \rho_k^{\vec{\mu}} \\ &= \rho_k (\mathbf{A}_{k-1:0}). \end{aligned} \tag{A2}$$

The direct calculation of each expansion coefficient is therefore given by

$$\begin{aligned} \alpha^{\bar{\mu}} &= \text{tr} \left[\hat{\mathbf{A}}_{k-1:0} \Delta_{k-1:0}^{\bar{\mu}} \right] \\ &= \text{tr} \left[\bigotimes_{i=0}^{k-1} \hat{\mathbf{A}}_i \Delta_i^{\mu_i} \right] \\ &= \prod_{i=0}^{k-1} \text{tr} \left[\hat{\mathbf{A}}_i \Delta_i^{\mu_i} \right] = \prod_{i=0}^{k-1} \alpha_i^{\mu_i}. \end{aligned} \quad (\text{A3})$$

3. Maximum likelihood, cost evaluation, and gradient

Full details and benchmarking of the PGDB algorithm for QPT can be found in Ref. [60]. Here, we provide the pseudocode in this context, which forms the basis for our implementation of MLE PTT.

$\text{Proj}_{S_n^+ \cap \mathcal{V}}(\cdot)$ here is the projection subroutine described in Sec. III. Although we fix the gradient step size here to be the same as in Ref. [60], we find this to be slightly problem dependent in terms of its performance. The reason is that the larger the step, the less physical $\Upsilon_{k:0}^{(j)} - \mu \nabla f(\Upsilon_{k:0}^{(j)})$ tends to be, increasing the run time of the projection subroutine. In general, we find that decreasing μ to favor the run time of the projection is favorable overall to the performance of the algorithm.

The process-tensor action described in Eq. (A2) is pedagogically useful; however, in practice, we compute the action of some process tensor $\mathcal{T}_{k:0}$ on a sequence of control operations $\mathbf{A}_{k-1:0}$ by projecting its Choi state onto the sequence of operations, as in Eq. (14). Because the input operations are always tensor product (omitting the case of correlated instruments), this can be performed fast as

```

1:  $j = 0, n = d_S^{2k+1}$ 
2: Initial estimate:  $\Upsilon_{k:0}^{(0)} = \mathbb{I}_{n \times n} / n$ 
3: Set metaparameters:  $\alpha = 2n^2/3, \gamma = 0.3$ 
4: while  $f(\Upsilon_{k:0}^{(j)}) - f(\Upsilon_{k:0}^{(n+1)}) > 1 \times 10^{-6}$  do
5:    $D^{(j)} = \text{Proj}_{S_n^+ \cap \mathcal{V}} \left( \Upsilon_{k:0}^{(j)} - \mu \nabla f(\Upsilon_{k:0}^{(j)}) \right) - \Upsilon_{k:0}^{(j)}$ 
6:    $\beta = 1$ 
7:   while  $f(\Upsilon_{k:0}^{(j)} + \beta D^{(j)}) > f(\Upsilon_{k:0}^{(j)}) + \gamma \beta \langle D^{(j)}, \nabla f(\Upsilon_{k:0}^{(j)}) \rangle$  do
8:      $\beta = 0.5 \beta$ 
9:   end while
10:   $\Upsilon_{k:0}^{(j+1)} = \Upsilon_{k:0}^{(j)} + \beta D^{(j)}$ 
11:   $j = j + 1$ 
12: end while
13: return  $\Upsilon_{k:0}^{(\text{est})} = \Upsilon_{k:0}^{(j+1)}$ 

```

Algorithm 1 PGDB

a tensor-network contraction. In this form, computation of the cost and the gradient is significantly sped up in comparison to multiplying out the full matrices.

Writing the Choi state of a process tensor $\Upsilon_{k:0}$ explicitly with its indices as a rank $2(2k+1)$ tensor, we have $2k+1$ subsystems alternating with outputs from the j th step (o_j) and inputs to the $(j+1)$ th step (i_j), i.e.,

$$\Upsilon_{k:0} \equiv (\Upsilon_{k:0})_{k_{o_k}, k_{i_k}, \dots, k_{o_0}}, \quad (\text{A4})$$

where b is shorthand for bra and k is shorthand for ket. The basis operation at time step j has indices (we write its transpose) $(\mathcal{B}_j^{\mu_j})_{b_{j+1}, b_{o_j}}^{k_{j+1}, k_{o_j}}$; meanwhile the POVM element Π_i is written $(\Pi_i)_{k_{o_k}}^{b_{o_k}}$. Consequently, the full tensor of predicted probabilities for all basis elements is given by

$$p_{i, \bar{\mu}} = \sum_{\substack{k_{o_k}, k_{i_k}, \dots, k_{i_1}, k_{o_0} \\ b_{o_k}, b_{i_k}, \dots, b_{i_1}, b_{o_0}}} (\Upsilon_{k:0})_{k_{o_k}, k_{i_k}, \dots, k_{o_0}}^{b_{o_k}, b_{i_k}, \dots, b_{o_0}} (\Pi_i)_{k_{o_k}}^{b_{o_k}} (\mathcal{B}_{k-1}^{\mu_{k-1}})_{b_{i_k}, b_{o_{k-1}}}^{k_{i_k}, k_{o_{k-1}}} (\mathcal{B}_{k-2}^{\mu_{k-2}})_{b_{i_{k-1}}, b_{o_{k-2}}}^{k_{i_{k-1}}, k_{o_{k-2}}} \dots (\mathcal{B}_{k-1}^{\mu_{k-1}})_{b_{i_1}, b_{o_0}}^{k_{i_1}, k_{o_0}}. \quad (\text{A5})$$

We use the quantum-information PYTHON library QUIMB [91] to perform this and all future tensor contractions straightforwardly. The cost function is then evaluation as in Eq. (15) in the same way: through an element-wise logarithm of $p_{i, \bar{\mu}}$ followed by contraction with the data tensor $n_{i, \bar{\mu}}$. Since the cost function is linear in $\Upsilon_{k:0}$, computation of the gradient $\nabla f / \nabla \Upsilon_{k:0}$ is simply $\nabla p_{i, \bar{\mu}} / \nabla \Upsilon_{k:0} : (n/p)^{i, \bar{\mu}}$, which expands to

$$\frac{\nabla f}{\nabla \Upsilon_{k:0}} = \sum_{i, \bar{\mu}} \left[(\mathcal{B}_{k-1}^{\mu_{k-1}})_{b_{i_k}, b_{o_{k-1}}}^{k_{i_k}, k_{o_{k-1}}} (\mathcal{B}_{k-2}^{\mu_{k-2}})_{b_{i_{k-1}}, b_{o_{k-2}}}^{k_{i_{k-1}}, k_{o_{k-2}}} \dots (\mathcal{B}_0^{\mu_0})_{b_{i_1}, b_{o_0}}^{k_{i_1}, k_{o_0}} \right] \frac{n_{i, \bar{\mu}}}{p_{i, \bar{\mu}}} \quad (\text{A6})$$

i.e., Eq. (A5) without the inclusion of $\Upsilon_{k:0}$.

APPENDIX B: APPROXIMATE CONDITIONAL MARKOV ORDER

In order to estimate a CMO process tomographically, we employ MLE PTT as described by the circuits in Fig. 7. The action of the reduced process tensors under a finite conditional Markov order model is best posed as a tensor-network contraction, so that tasks such as the optimization in Sec. V can be performed quickly.

$$\begin{aligned}
\left(\Upsilon_{\ell:0}^{(\mathbf{B}_{\ell-1:0}^{\vec{\mu}})}\right)_{k_{o\ell}}^{b_{o\ell}} &= \sum_{\substack{k_{o0}, k_{i1}, k_{o1}, \dots, k_{i\ell-1} \\ b_{o0}, b_{i1}, b_{o1}, \dots, b_{i\ell-1}}} (\Upsilon_{\ell:0})_{k_{o\ell}, k_{i\ell}, \dots, k_{i1}, k_{o0}}^{b_{o\ell}, b_{i\ell}, \dots, b_{i1}, b_{o0}} (\mathcal{B}_{\ell-1}^{\mu_{\ell-1}})_{k_{i\ell}, k_{o\ell-1}}^{b_{i\ell}, b_{o\ell-1}} \cdots (\mathcal{B}_0^{\mu_0})_{k_{i1}, k_{o0}}^{b_{i1}, b_{o0}}, \\
\left(\Upsilon_{j:j-\ell}^{(\mathbf{B}_{j-2:j-\ell}^{\vec{\mu}})}\right)_{k_{oj}, k_{ij}}^{b_{oj}, b_{ij}} &= \sum_{\substack{k_{oj-\ell}, k_{ij-\ell+1}, k_{oj-\ell+1}, \dots, k_{ij-1} \\ b_{oj-\ell}, b_{ij-\ell+1}, b_{oj-\ell+1}, \dots, b_{ij-1}}} (\Upsilon_{j:j-\ell})_{k_{oj}, k_{ij}, \dots, k_{ij-\ell+1}, k_{oj-\ell}}^{b_{oj}, b_{ij}, \dots, b_{ij-\ell+1}, b_{oj-\ell}} \delta_{k_{oj-1}}^{b_{oj-1}} (\mathcal{B}_{j-2}^{\mu_{j-2}})_{k_{i\ell-1}, k_{o\ell-2}}^{b_{i\ell-1}, b_{o\ell-2}} \cdots (\mathcal{B}_{j-\ell}^{\mu_{j-\ell}})_{k_{ij-\ell+1}, k_{oj-\ell}}^{b_{ij-\ell+1}, b_{oj-\ell}}, \\
\left(\Upsilon_{k:k-\ell}^{(\mathbf{B}_{k-2:k-\ell}^{\vec{\mu}, \Pi_i})}\right)_{k_{ik}}^{b_{ik}} &= \sum_{\substack{k_{ok-\ell}, k_{ik-\ell+1}, k_{ok-\ell+1}, \dots, k_{ik-1}, k_{ok} \\ b_{ok-\ell}, b_{ik-\ell+1}, b_{ok-\ell+1}, \dots, b_{ik-1}, b_{ok}}} (\Upsilon_{k:k-\ell})_{k_{ok}, k_{ik}, \dots, k_{ik-\ell+1}, k_{ok-\ell}}^{b_{oj}, b_{ij}, \dots, b_{ik-\ell+1}, b_{ok-\ell}} (\Pi_i)_{k_{ok}}^{b_{ok}} \delta_{k_{ok-1}}^{b_{ok-1}} (\mathcal{B}_{k-2}^{\mu_{k-2}})_{k_{i\ell-1}, k_{o\ell-2}}^{b_{i\ell-1}, b_{o\ell-2}} \cdots (\mathcal{B}_{k-\ell}^{\mu_{k-\ell}})_{k_{ik-\ell+1}, k_{ok-\ell}}^{b_{ik-\ell+1}, b_{ok-\ell}}. \tag{B1}
\end{aligned}$$

Then, the tensor of predicted probabilities $p_{i,\vec{\mu}}$ is obtained by stitching each conditional process tensor together with the overlapping control operations; that is,

$$\begin{aligned}
&\sum_{\substack{k_{o\ell}, k_{i\ell+1}, \dots, k_{ik} \\ b_{o\ell}, b_{i\ell+1}, \dots, b_{ik}}} \left(\Upsilon_{k:k-\ell}^{(\mathbf{B}_{k-2:k-\ell}^{\vec{\mu}, \Pi_i})}\right)_{k_{ik}}^{b_{ik}} (\mathcal{B}_{k-1}^{\mu_{k-1}})_{k_{ik}, k_{ok-1}}^{b_{ik}, b_{ok-1}} \\
&\left(\prod_{j=\ell+1}^{k-1} \left(\Upsilon_{j:j-\ell}^{(\mathbf{B}_{j-2:j-\ell}^{\vec{\mu}})}\right)_{k_{oj}, k_{ij}}^{b_{oj}, b_{ij}} (\mathcal{B}_{j-1}^{\mu_{j-1}})_{k_{ij}, k_{oj-1}}^{b_{ij}, b_{oj-1}}\right) \\
&\times (\mathcal{B}_{\ell}^{\mu_{\ell}})_{k_{i\ell+1}, k_{o\ell}}^{b_{i\ell+1}, b_{o\ell}} \left(\Upsilon_{\ell:0}^{(\mathbf{B}_{\ell-1:0}^{\vec{\mu}})}\right)_{k_{o\ell}}^{b_{o\ell}}, \tag{B2}
\end{aligned}$$

which is precisely the generalization of the strategy presented in Fig. 6. When evaluated from left to right, this can be performed efficiently since every contraction is a rank-2 tensor with a rank-4 tensor. Note that in this instance, the index vector $\vec{\mu}$ does not run from $(0, 0, \dots, 0)$ to $(d_S^A, d_S^A, \dots, d_S^A)$ but, rather, for each block of memory, it contains all d_S^A combinations of basis elements,

We show this here. Note that the following extravagant working is equivalent to Fig. 6 but we write it out in full generality in order to make the indices explicit and replication more straightforward. First, the conditional reduced states of each of the process tensors must be taken by contracting the relevant control operations (including final measurement) into the process tensors. These are as follows:

with all other operations fixed at μ_0 . There are therefore $(k - \ell + 1) \cdot d_S^A$ values taken by $\vec{\mu}$.

APPENDIX C: IDENTIFYING MINIMAL UNITARY BASIS OVERLAP

The effects of basis overlap in quantum tomography on the reconstruction have been discussed both with respect to conventional QST and QPT and, more recently, with respect to the process tensor. In particular, the process tensor has shown itself to be highly sensitive to any overlap in its control basis. With access to all 16 dimensions of superoperator space, a mutually unbiased basis can be constructed in the form of a symmetric informationally complete (SIC) POVM followed by an update [92]. However, in the limited case of a unitary-only basis, the ideal method is less straightforward. A randomly chosen unitary basis has been shown to adversely affect the reconstruction fidelity by as much as 30%. Selecting a basis with mutual overlap here would be ideal, akin to the notion of a SIC POVM in conventional quantum state tomography. However, it has been shown that MUUBs do not exist in dimension 10 (the dimension for single-qubit channels).

Because of this limitation, we numerically search for a basis that minimizes its average overlap with the remainder of the set. This procedure is performed as follows.

We parametrize these gates using the standard QISKIT unitary parametrization

$$u(\theta, \phi, \lambda) = \begin{pmatrix} \cos(\theta/2) & -e^{i\lambda} \sin(\theta/2) \\ e^{i\phi} \sin(\theta/2) & e^{i\lambda+i\phi} \cos(\theta/2) \end{pmatrix}. \quad (\text{C1})$$

For two unitaries u and v , let \mathcal{U} and \mathcal{V} be their superoperator equivalent, according to some representation. The overlap between the two channels is given by the Hilbert-Schmidt inner product:

$$\langle A, B \rangle_{\text{HS}} := \text{Tr}[A^\dagger B]. \quad (\text{C2})$$

Importantly, this quantity is independent of representation, allowing us to select a form most desirable for computation. To this effect, we use the row-vectorized convention for states. Here, operations are given by $\mathcal{U} = u \otimes u^*$. The inner product between two unitaries parametrized as in Eq. (C1) is then

$$\begin{aligned} \langle \mathcal{U}, \mathcal{V} \rangle_{\text{HS}} &= \text{Tr}[\mathcal{U}^\dagger \mathcal{V}] \\ &= \text{Tr}[(u \otimes u^*)^\dagger \cdot (v \otimes v^*)] \\ &= \text{Tr}[u^\dagger \cdot v] \otimes (u^T \cdot v^*) \\ &= \text{Tr}[u^\dagger v] \cdot \text{Tr}[u^\dagger v]^* \\ &= |\text{Tr}[u^\dagger v]|^2. \end{aligned} \quad (\text{C3})$$

If we write $u = u(\theta_1, \phi_1, \lambda_1)$ and $v = v(\theta_2, \phi_2, \lambda_2)$, then Eq. (C3) can be straightforwardly written (after some simplification) as

$$\begin{aligned} \text{Tr}[u^\dagger v] &= \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} + e^{i(\phi_2 - \phi_1)} \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \\ &\quad + e^{i(\lambda_2 - \lambda_1)} \sin \frac{\theta_1}{2} \sin \frac{\theta_2}{2} \\ &\quad + e^{i(\lambda_2 + \phi_2 - \lambda_1 - \phi_1)} \cos \frac{\theta_1}{2} \cos \frac{\theta_2}{2} \\ \Rightarrow \langle \mathcal{U}, \mathcal{V} \rangle_{\text{HS}} &= 4 \cos^2 \left(\frac{1}{2} (\lambda_1 - \lambda_2 + \phi_1 - \phi_2) \right) \\ &\quad \times \cos^2 (\theta_1 - \theta_2). \end{aligned} \quad (\text{C4})$$

This simple expression for the inner product of any two single-qubit unitaries allows us to construct an objective function for the straightforward mutual minimization of overlap between all ten elements of the basis set. Let

$$\mathcal{U}(\vec{\theta}, \vec{\phi}, \vec{\lambda}) = \{\mathcal{U}_i\}_{i=1}^{10} \equiv \{(\theta_i, \phi_i, \lambda_i)\}_{i=1}^{10} \quad (\text{C5})$$

be our parametrized basis set. A basis set with the least mutual overlap can then be found by minimizing the sum

TABLE II. A set of 30 parameter values that constitute a set of ten unitary gates with minimal average mutual overlap.

	θ	ϕ	λ
\mathcal{U}_1	1.1148	1.5606	0.8160
\mathcal{U}_2	-2.1993	-2.0552	-0.3564
\mathcal{U}_3	0.9616	-0.8573	1.2333
\mathcal{U}_4	2.2655	-2.7083	0.3154
\mathcal{U}_5	-0.1013	-0.5548	-1.1472
\mathcal{U}_6	1.8434	0.8074	-1.1772
\mathcal{U}_7	-2.2036	1.9589	2.4002
\mathcal{U}_8	-1.2038	-0.2023	1.2355
\mathcal{U}_9	2.1791	3.2836	2.3524
\mathcal{U}_{10}	-1.3116	2.3082	0.2882

of the squares of each unitary with the remainder of the set. This minimizes both the average overlap and the variance of overlaps with the remainder of the set; that is, by computing

$$\begin{aligned} \arg \min_{(\vec{\theta}, \vec{\phi}, \vec{\lambda})} &\sum_{i=1}^{10} \sum_{j>i} (\langle \mathcal{U}_i, \mathcal{U}_j \rangle_{\text{HS}})^2 \\ &= \sum_{i=1}^{10} \sum_{j>i} 16 \cos^4 \left(\frac{1}{2} (\lambda_i - \lambda_j + \phi_i - \phi_j) \right) \\ &\quad \times \cos^4 (\theta_i - \theta_j). \end{aligned} \quad (\text{C6})$$

One such ideal set can be found in Table II. This is the basis set used for the data obtained in the main text. Its overlaps with respect to the Hilbert-Schmidt inner product are listed in Table III.

APPENDIX D: IBM QUANTUM ANALYSES

The quantum device procedures in this work are carried out on IBM Quantum cloud devices: *ibmq_boeblingen*, *ibmq_johannesburg*, *ibmq_valencia*, *ibmq_bogota*, *ibmq_manhattan*, *ibmq_montreal*, and *ibmq_guadalupe*. Data from the first three are the same as the data featured in Ref. [21], with the exception of the MUUB jobs on *ibmq_valencia*, which are newly taken for this work. The detail of these runs can be found within that reference. All other data taken from the remainder of the devices are newly collected for this work. All devices are fixed-frequency superconducting-transmon quantum computers. Below, we step through the circuits conducted in Secs. III B and V on these devices.

1. Reconstruction fidelities

Reconstruction fidelity experimentally validates the quality of a model by comparing predictions made by the model with data generated by the device, where the data is not used to create the model. The process tensor establishes a mapping from a sequence of control operations

TABLE III. The Hilbert-Schmidt overlap between each element of the numerically constructed (approximate) MUUB. We find this to be the most uniformly overlapping unitary basis possible and thus optimal for PTT.

$\text{Tr}[\mathcal{U}_i^T \mathcal{U}_j]$	\mathcal{U}_1	\mathcal{U}_2	\mathcal{U}_3	\mathcal{U}_4	\mathcal{U}_5	\mathcal{U}_6	\mathcal{U}_7	\mathcal{U}_8	\mathcal{U}_9	\mathcal{U}_{10}
\mathcal{U}_1	1									
\mathcal{U}_2	0.196 88	1								
\mathcal{U}_3	0.196 88	0.111 11	1							
\mathcal{U}_4	0.167 58	0.196 88	0.196 88	1						
\mathcal{U}_5	0.167 58	0.196 88	0.196 88	0.167 58	1					
\mathcal{U}_6	0.196 88	0.111 11	0.111 11	0.196 88	0.196 88	1				
\mathcal{U}_7	0.032 86	0.196 88	0.196 88	0.167 58	0.167 58	0.196 88	1			
\mathcal{U}_8	0.167 58	0.196 88	0.196 88	0.167 58	0.032 86	0.196 88	0.167 58	1		
\mathcal{U}_9	0.196 88	0.111 11	0.111 11	0.196 88	0.196 88	0.111 11	0.196 88	0.196 88	1	
\mathcal{U}_{10}	0.167 58	0.196 88	0.196 88	0.032 86	0.167 58	0.196 88	0.167 58	0.167 58	0.196 88	1
Average	0.249 07	0.251 46	0.251 46	0.249 07	0.249 07	0.251 46	0.249 07	0.249 07	0.251 46	0.249 07

to a final state. Therefore, in this context, the comparison is between the final state predicted by a process-tensor model subject to a sequence of operations (from outside the basis set) with the actual state reconstructed when the same sequence of operations is run on the device. The results of Fig. 5 are reconstruction-fidelity distributions for a number of sequences of three random unitaries.

The procedure for constructing a single-qubit three-step process tensor is as follows:

- (1) Initialize the system.
- (2) Wait some time T_1 .
- (3) Apply basis element \mathcal{U}_{μ_0} .
- (4) Wait some time T_2 .
- (5) Apply basis element \mathcal{U}_{μ_1} .
- (6) Wait some time T_3 .
- (7) Apply basis element \mathcal{U}_{μ_2} .
- (8) Wait some time T_4 .
- (9) Measure in X , Y , and Z bases.

The total number of circuits here is $10 \times 10 \times 10 \times 3 = 3000$. Finally, the data are then processed according to the LI-MLE processing methods stipulated in the main text. For LI, this means constructing the density matrix corresponding to each basis sequence. For MLE, this means shaping the data into a (6, 10, 10, 10) array, where the first dimension corresponds to each of the six effects in the POVM $\{|+\rangle\langle+|, |i+\rangle\langle i+|, |0\rangle\langle 0|, |-\rangle\langle-|, |i-\rangle\langle i-|, |1\rangle\langle 1|\}$. Note that, in general, the times T_i can be chosen to be different and to include whatever background dynamics and circuit structure the experimenter is interested in. This is simply a choice of the quantum stochastic process being studied. In the cases of *ibmq_manhattan* and *ibmq_bogota*, the circuit structures are as described in Sec. V.

Once the process-tensor data are collected and an LI-MLE model constructed, a number of random unitary circuits are generated, with each gate chosen by the `SCIPY.STATS.UNITARY_GROUP.RVS()` function. These gates

are then run on the real devices, following the same circuit structure. QST is then performed at the end of each sequence. The conditional state is computed by contracting the random sequence into the process-tensor model. Finally, the state fidelity is computed between the predicted and the actual states. This then forms the data sets shown in Fig. 5. With all things equal, the average reconstruction fidelity will necessarily increase with an increased number of shots per circuit. It will also decrease if the final states are noisier or more mixed, since this will add to the variance of sampling statistics.

Once the model is validated to the desired level of accuracy, it becomes a useful tool for optimal control of the non-Markovian system. In Sec. VA, we use the same process tensors from *ibmq_manhattan* and *ibmq_bogota* to show a circuit-by-circuit improvement of the fidelities of states generated by IBM Quantum devices. So as to avoid readout error obfuscating any results, or overstating any improvements, we first perform GST using the `pyGSTi` software package [93] in order to estimate the actual POVM giving X , Y , and Z projections on the device. This POVM is then used both in reconstructing the states and in the PTT estimate.

2. Conditional Markov order circuit improvement

For each of the conditional Markov order tests, a five-step process with $\ell = 1$, $\ell = 2$, and $\ell = 3$ is considered. This amounts to reconstructing, respectively, five-, four-, and three-memory-block process tensors. The structure of the circuits is similar to the three-step process tensor; however, not all circuit elements are varied. For example, with $\ell = 3$, this means reconstructing the three process tensors corresponding to circuit structure $\mathcal{U}_{\mu_0} - \mathcal{U}_{\mu_1} - \mathcal{U}_{\mu_2} - \Pi_i$, $\mathcal{U}_0 - \mathcal{U}_{\mu_0} - \mathcal{U}_{\mu_1} - \mathcal{U}_{\mu_2} - \Pi_i$, and $\mathcal{U}_0 - \mathcal{U}_0 - \mathcal{U}_{\mu_0} - \mathcal{U}_{\mu_1} - \mathcal{U}_{\mu_2} - \Pi_i$. For $\ell = 2$, a subset of the same data can be reused: fixing $\mu_0 = 0$ and varying μ_1 and μ_2 , for example. The only additional information required is that a projective measurement needs to be made in position 2

of the circuit in order to determine the state at the end of the first $\ell = 2$ memory block. A similar process follows for determination of $\Upsilon_{5,0}^1$, with an extra memory-block process tensor constructed with a projective measurement at position 1. This totals $3 \times (10 \times 10 \times 10 \times 3) = 9000$ circuits for $\Upsilon_{5,0}^3$, an extra $10 \times 10 \times 3 = 300$ circuits for $\Upsilon_{5,0}^2$, and an extra $10 \times 3 = 30$ circuits for $\Upsilon_{5,0}^1$.

Since the state is being propagated along in our finite Markov order stitching procedure, it is important that it is well characterized without measurement error. To this effect, we use GST again to estimate our POVM. This is more essential than before, since now our PTT construction is contingent on inputting the correct form of the operation. This is also true of the unitary gates that we apply; however, the single-qubit error rates are $\mathcal{O}(10^{-4})$, compared with measurement errors of $\mathcal{O}(10^{-2})$, and so a far smaller assumption.

-
- [1] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
- [2] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, *Phys. Rev. X* **8**, 31027 (2018).
- [3] S. Ferracin, T. Kapourniotis, and A. Datta, Accrediting outputs of noisy intermediate-scale quantum computing devices, *New J. Phys.* **21**, 113038 (2019).
- [4] G. A. L. White, C. D. Hill, and L. C. L. Hollenberg, Performance Optimization for Drift-Robust Fidelity Improvement of Two-Qubit Gates, *Phys. Rev. Appl.* **15**, 014023 (2021).
- [5] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, *Nat. Phys.* **16**, 1184 (2020).
- [6] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, and A. Kandala *et al.*, Demonstration of quantum volume 64 on a superconducting quantum computing system, *Quantum Sci. Technol.* **6**, 025020 (2021).
- [7] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2017).
- [8] S. Milz, F. A. Pollock, and K. Modi, An introduction to operational quantum dynamics, *Open Syst. Inf. Dyn.* **24**, 1740016 (2017).
- [9] S. Milz and K. Modi, Quantum Stochastic Processes and Quantum Non-Markovian Phenomena, *PRX Quantum* **2**, 030201 (2021).
- [10] L. Li, M. J. W. Hall, and H. M. Wiseman, Concepts of quantum non-Markovianity: A hierarchy, *Phys. Rep.* **759**, 1 (2018).
- [11] 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).
- [12] I. de Vega and D. Alonso, Dynamics of non-Markovian open quantum systems, *Rev. Mod. Phys.* **89**, 015001 (2017).
- [13] Á. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-Markovianity: Characterization, quantification and detection, *Rep. Prog. Phys.* **77**, 094001 (2014).
- [14] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
- [15] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, and D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [16] I. Pogorelov, T. Feldker, C. D. Marciniak, L. Postler, G. Jacob, O. Kriegelsteiner, V. Podlesnic, M. Meth, V. Negnevitsky, and M. Stadler *et al.*, A compact ion-trap quantum computing demonstrator, [arXiv:2101.11390](https://arxiv.org/abs/2101.11390) (2021).
- [17] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, and D. Kim *et al.*, Realization of High-Fidelity CZ and ZZ-Free iSWAP Gates with a Tunable Coupler, *Phys. Rev. X* **11**, 021058 (2021).
- [18] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Measuring the Capabilities of Quantum Computers, [arXiv:2008.11294](https://arxiv.org/abs/2008.11294) (2020).
- [19] B. D. Clader, C. J. Trout, J. P. Barnes, K. Schultz, G. Quiroz, and P. Titum, Impact of correlations and heavy tails on quantum error correction, *Phys. Rev. A* **103**, 052428 (2021).
- [20] N. H. Nickerson and B. J. Brown, Analysing correlated noise on the surface code using adaptive decoding algorithms, *Quantum* **3**, 131 (2019).
- [21] 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).
- [22] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, [arXiv:2009.07301](https://arxiv.org/abs/2009.07301) (2020).
- [23] 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).
- [24] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing context-dependent errors in quantum processors, *Phys. Rev. X* **9**, 021045 (2019).
- [25] A. Veitia, M. P. da Silva, R. Blume-Kohout, and S. J. van Enk, Macroscopic instructions vs microscopic operations in quantum circuits, *Phys. Lett. A* **384**, 126131 (2020).
- [26] A. Veitia and S. J. van Enk, Testing the context-independence of quantum gates, [arXiv:1810.05945](https://arxiv.org/abs/1810.05945) (2018).
- [27] J. Helsen, F. Battistel, and B. M. Terhal, Spectral quantum tomography, *Npj Quantum Inf.* **5**, 74 (2019).
- [28] 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).
- [29] P. Taranto, F. A. Pollock, S. Milz, M. Tomamichel, and K. Modi, Quantum Markov Order, *Phys. Rev. Lett.* **122**, 140401 (2019).
- [30] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin,

- and Y.-K. Liu, Efficient quantum state tomography, *Nat. Commun.* **1**, 149 (2010).
- [31] M. Cygorek, M. Cosacchi, A. Vagov, V. M. Axt, B. W. Lovett, J. Keeling, and E. M. Gauger, Simulation of open quantum systems by automated compression of arbitrary environments, *Nat. Phys.* (2022).
- [32] A. Dang, G. A. L. White, L. C. L. Hollenberg, and C. D. Hill, Process tomography on a 7-qubit quantum processor via tensor network contraction path finding, [arXiv:2112.06364](https://arxiv.org/abs/2112.06364) (2021).
- [33] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, Scalable Reconstruction of Density Matrices, *Phys. Rev. Lett.* **111**, 020401 (2013).
- [34] R. Blume-Kohout, K. Rudinger, E. Nielsen, T. Proctor, and K. Young, Wildcard error: Quantifying unmodeled errors in quantum processors, [arxiv:2012.12231](https://arxiv.org/abs/2012.12231) (2020).
- [35] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [36] J. Zhang and M. Sarovar, Quantum Hamiltonian Identification from Measurement Time Traces, *Phys. Rev. Lett.* **113**, 080401 (2014).
- [37] S.-T. Wang, D.-L. Deng, and L.-M. Duan, Hamiltonian tomography for quantum many-body systems with arbitrary couplings, *New J. Phys.* **17**, 093017 (2015).
- [38] F. A. Pollock and K. Modi, Tomographically reconstructed master equations for any open quantum dynamics, *Quantum* **2**, 76 (2018).
- [39] S. Lorenzo, F. Ciccarello, and G. M. Palma, Class of exact memory-kernel master equations, *Phys. Rev. A* **93**, 052111 (2016).
- [40] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, *New J. Phys.* **14**, 095022 (2012).
- [41] R. Blume-Kohout, Optimal, reliable estimation of quantum states, *New J. Phys.* **12**, 043034 (2010).
- [42] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, *Phys. Rev. A* **87**, 62119 (2013).
- [43] R. Blume-Kohout, J. Gamble, E. Nielsen, J. Mizrahi, J. Sterk, and P. Maunz, Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, [arXiv:1310.4492](https://arxiv.org/abs/1310.4492) (2013).
- [44] D. Greenbaum, Introduction to Quantum Gate Set Tomography, [arXiv:1509.02921](https://arxiv.org/abs/1509.02921) (2015).
- [45] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, in *Quantum State Estimation*, edited by M. Paris and J. Řeháček (Springer, Berlin, 2004), p. 59.
- [46] F. Costa and S. Shrapnel, Quantum causal modelling, *New J. Phys.* **18**, 063032 (2016).
- [47] S. Milz, F. Sakuldee, F. A. Pollock, and K. Modi, Kolmogorov extension theorem for (quantum) causal modelling and general probabilistic theories, *Quantum* **4**, 255 (2020).
- [48] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory Effects in Quantum Channel Discrimination, *Phys. Rev. Lett.* **101**, 180501 (2008).
- [49] S. Shrapnel, F. Costa, and G. Milburn, Updating the Born rule, *New J. Phys.* **20**, 053010 (2018).
- [50] M. M. Wolf, J. Eisert, T. S. Cubitt, and J. I. Cirac, Assessing Non-Markovian Quantum Dynamics, *Phys. Rev. Lett.* **101**, 150402 (2008).
- [51] 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).
- [52] Á. Rivas, S. F. Huelga, and M. B. Plenio, Entanglement and Non-Markovianity of Quantum Evolutions, *Phys. Rev. Lett.* **105**, 050403 (2010).
- [53] D. Chruściński, A. Kossakowski, and Á. Rivas, Measures of non-Markovianity: Divisibility versus backflow of information, *Phys. Rev. A* **83**, 052128 (2011).
- [54] B. Vacchini, Non-Markovian master equations from piecewise dynamics, *Phys. Rev. A* **87**, 030101 (2013).
- [55] A. D. Corcoles, M. Takita, K. Inoue, S. Lekuch, Z. K. Mineev, J. M. Chow, and J. M. Gambetta, Exploiting dynamic quantum circuits in a quantum algorithm with superconducting qubits, [arxiv:2102.01682](https://arxiv.org/abs/2102.01682) (2021).
- [56] 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, [arXiv:2105.03333](https://arxiv.org/abs/2105.03333) (2021).
- [57] S. Milz, F. A. Pollock, and K. Modi, Reconstructing non-Markovian quantum dynamics with limited control, *Phys. Rev. A* **98**, 012108 (2018).
- [58] R. B. A. Adamson and A. M. Steinberg, Improving Quantum State Estimation with Mutually Unbiased Bases, *Phys. Rev. Lett.* **105**, 030406 (2010).
- [59] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient ZZ gates for quantum computing, *Phys. Rev. A* **96**, 22330 (2017).
- [60] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger, Quantum process tomography via completely positive and trace-preserving projection, *Phys. Rev. A* **98**, 062336 (2018).
- [61] E. Birgin and M. Raydan, Robust Stopping Criteria for Dykstra's Algorithm, *SIAM J. Sci. Comput.* **26**, 1405 (2005).
- [62] D. Henrion and J. Malick, Projection methods for conic feasibility problems: Applications to polynomial sum-of-squares decompositions, *Optim. Methods Softw.* **26**, 23 (2011).
- [63] G. A. L. White, F. A. Pollock, L. C. L. Hollenberg, C. D. Hill, and K. Modi, Diagnosing temporal quantum correlations: Compressed non-Markovian calipers, [arXiv:2107.13934](https://arxiv.org/abs/2107.13934) (2021).
- [64] A. Gilchrist, D. R. Terno, and C. J. Wood, Vectorization of quantum operations and its use, [arXiv:0911.2539](https://arxiv.org/abs/0911.2539) (2009).
- [65] M. F. Anjos and J. B. Lasserre, *International Series in Operations Research and Management Science*, Vol. 166 (Springer U.S., 2012), Chap.20, p. XI, 960.
- [66] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, 2016), p. 225.
- [67] C. Michelot, A finite algorithm for finding the projection of a point onto the canonical simplex of α_n , *J. Optim. Theory Appl.* **50**, 195 (1986).

- [68] J. Malick, A Dual Approach to Semidefinite Least-Squares Problems, *SIAM J. Matrix Anal. Appl.* **26**, 272 (2004).
- [69] D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Math. Program.* **45**, 503 (1989).
- [70] J. Malick and H. S. Sendov, Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices, *Set-Valued Anal.* **14**, 273 (2006).
- [71] H. S. Sendov, Generalized Hadamard product and the derivatives of spectral functions, *SIAM J. Matrix Anal. Appl.* **28**, 667 (2006).
- [72] R. N. M. Nasir, J. S. Shaari, and S. Mancini, Mutually unbiased unitary bases of operators on d -dimensional Hilbert space, *Int. J. Quantum Inf.* **18**, 1941026 (2020).
- [73] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, Memory in network flows and its effects on spreading dynamics and community detection, *Nat. Commun.* **5**, 4630 (2014).
- [74] Y. Guo, P. Taranto, B.-H. Liu, X.-M. Hu, Y.-F. Huang, C.-F. Li, and G.-C. Guo, Experimental Demonstration of Instrument-Specific Quantum Memory Effects and Non-Markovian Process Recovery for Common-Cause Processes, *Phys. Rev. Lett.* **126**, 230401 (2021).
- [75] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).
- [76] P. Taranto, F. A. Pollock, and K. Modi, Memory Strength and Recoverability of Non-Markovian Quantum Stochastic Processes, [arXiv:1907.12583](https://arxiv.org/abs/1907.12583) (2019).
- [77] $\ell = 2$ and $\ell = 1$ models can be constructed from subsets of the $\ell = 3$ data; however, on their own they minimally require $3 \times 300 = 900$ and $4 \times 30 = 120$ circuits per time, respectively.
- [78] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Operational Markov Condition for Quantum Processes, *Phys. Rev. Lett.* **120**, 040405 (2018).
- [79] C. Giarmatzi and F. Costa, A quantum causal discovery algorithm, *Npj Quantum Inf.* **4**, 17 (2018).
- [80] G. E. Fux, E. P. Butler, P. R. Eastham, B. W. Lovett, and J. Keeling, Efficient Exploration of Hamiltonian Parameter Space for Optimal Control of Non-Markovian Open Quantum Systems, *Phys. Rev. Lett.* **126**, 200401 (2021).
- [81] K. Sinha, P. Meystre, E. A. Goldschmidt, F. K. Fatemi, S. L. Rolston, and P. Solano, Non-Markovian Collective Emission from Macroscopically Separated Emitters, *Phys. Rev. Lett.* **124**, 043603 (2020).
- [82] D. Nagy and P. Domokos, Nonequilibrium Quantum Criticality and Non-Markovian Environment: Critical Exponent of a Quantum Phase Transition, *Phys. Rev. Lett.* **115**, 043601 (2015).
- [83] P. Haikka, S. McEndoo, G. De Chiara, G. M. Palma, and S. Maniscalco, Quantifying, characterizing, and controlling information flow in ultracold atomic gases, *Phys. Rev. A* **84**, 031602 (2011).
- [84] C. A. Mujica-Martinez, P. Nalbach, and M. Thorwart, Quantification of non-Markovian effects in the Fenna-Matthews-Olson complex, *Phys. Rev. E* **88**, 062719 (2013).
- [85] A. Nitzan and M. A. Ratner, Electron transport in molecular wire junctions, *Science* **300**, 1384 (2003).
- [86] N. Lambert, Y.-N. Chen, Y.-C. Cheng, C.-M. Li, G.-Y. Chen, and F. Nori, Quantum biology, *Nat. Phys.* **9**, 10 (2013).
- [87] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, Quantum process tomography with unsupervised learning and tensor networks, [arXiv:2006.02424](https://arxiv.org/abs/2006.02424) (2020).
- [88] C. Guo, K. Modi, and D. Poletti, Tensor-network-based machine learning of non-Markovian quantum processes, *Phys. Rev. A* **102**, 062414 (2020).
- [89] M. Rambach, M. Qaryan, M. Kewming, C. Ferrie, A. G. White, and J. Romero, Robust and Efficient High-Dimensional Quantum State Tomography, *Phys. Rev. Lett.* **126**, 100402 (2021).
- [90] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [91] J. Gray, QUIMB: A PYTHON package for quantum information and many-body calculations, *J. Open Source Softw.* **3**, 819 (2018).
- [92] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Minimal qubit tomography, *Phys. Rev. A* **70**, 052321 (2004).
- [93] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, Probing quantum processor performance with pyGSTi, *Quantum Sci. Technol.* **5**, 044002 (2020).