Quantum process identification: a method for characterizing non-Markovian quantum dynamics

Ryan S Bennink and Pavel Lougovski
Quantum Information Science Group, Oak Ridge National Laboratory, United States of America
E-mail: benninkrs@ornl.gov

Keywords: system identification, quantum process tomography, device characterization, non-Markovian dynamics, open quantum dynamics, Hamiltonian learning, generalized probabilistic theory

Abstract
Established methods for characterizing quantum information processes do not capture non-Markovian (history-dependent) behaviors that occur in real systems. These methods model a quantum process as a fixed map on the state space of a predefined system of interest. Such a map averages over the system’s environment, which may retain some effect of its past interactions with the system and thus have a history-dependent influence on the system. Although the theory of non-Markovian quantum dynamics is currently an active area of research, a systematic characterization method based on a general representation of non-Markovian dynamics has been lacking. In this article we present a systematic method for experimentally characterizing the dynamics of open quantum systems. Our method, which we call quantum process identification (QPI), is based on a general theoretical framework which relates the (non-Markovian) evolution of a system over an extended period of time to a time-local (Markovian) process involving the system and an effective environment. In practical terms, QPI uses time-resolved tomographic measurements of a quantum system to construct a dynamical model with as many dynamical variables as are necessary to reproduce the evolution of the system. Through numerical simulations, we demonstrate that QPI can be used to characterize qubit operations with non-Markovian errors arising from realistic dynamics including control drift, coherent leakage, and coherent interaction with material impurities.

1. Introduction
The experimental characterization of dynamical processes is fundamental to physics. In the burgeoning field of quantum information science, the characterization of quantum processes in a general, systematic way has become particularly important. For example, in quantum key distribution, characterization of the information-preserving properties of the quantum channel between communicating parties is crucial to establishing the security of the generated key [1]. In the circuit model of quantum computing, a computation is represented as a sequence of primitive quantum logic operations or ‘gates’, each of which is implemented via a controlled quantum process. Characterization of these processes essential to assessing and improving their performance [3–5]. The paradigmatic way of characterizing a process involving some quantum system of interest is quantum process tomography (QPT) [6]. In QPT the process is tested many times using a variety of initial states and final measurements that collectively span the state space of the system. The process is then expressed as a quantum channel—a linear, completely positive, trace-preserving (CPTP) map on the system state space—that can be estimated from the experimental results.

A modern application of QPT (and the main context of this work) is to construct models for the behavior of fabricated quantum computing devices. Each qubit operation or ‘gate’ the device is capable of performing is characterized via QPT as a quantum channel on the targeted qubits. The execution of an arbitrary gate sequence is then modeled as a corresponding sequence of quantum channels on the device’s qubits. While newer and more robust characterization methods such as gate set tomography (GST) [3, 5, 7] and randomized...
benchmarking (RB) [4, 8–14] have largely replaced QPT, these methods continue the approach of modeling each gate as a quantum channel involving only the targeted qubits.

This nearly ubiquitous approach to quantum device characterization is not valid, however, when the dynamics of interest are non-Markovian. For the purposes of this work, a process is Markovian if the dynamical map for any given time interval can be expressed as a sequence of maps representing the dynamics over successive subintervals [15]. Informally, this is the property that the (statistical) state of the system at any point in time is sufficient to determine its (statistical) state at any future time; there is no additional dependence on prior states of the system or on any external degrees of freedom. By this definition a sequence of quantum channels constitutes a Markovian process. While the theory of non-Markovian quantum dynamics has received growing attention in recent years [16, 17], experimental strategies for characterizing such dynamics [18–20] are less well developed. Meanwhile, incoherent (Markovian) noise processes in quantum information technologies have been reduced to the point that non-Markovian errors are starting to become significant [3, 5].

In this paper we present a general, systematic method to empirically characterize arbitrary ‘black box’ quantum processes (i.e. processes not involving intermediate measurements), including non-Markovian dynamics of open quantum systems. More specifically, we show how to construct a minimal dynamical model of a quantum system solely from observations of that system at selected times. We call this task quantum process identification (QPI) in analogy with the systems engineering task known as system identification [21, 22]. Operationally, QPI consists of time-resolved process tomography followed by a numerical search for an appropriate model. But unlike QPT and its modern replacements, QPI assumes almost nothing about the system of interest, its environment, or their joint dynamics. In particular it does not restrict analysis to a presupposed system of interest, but identifies and characterizes as many degrees of freedom as are needed to account for the observations. It assumes no particular model, nor even quantum mechanics specifically. QPI assumes only that (1) there exists a finite-dimensional probabilistic model capable of predicting the observed behavior to desired accuracy, and (2) experiments on the system are statistically independent. Furthermore, QPI is fully self-contained; it requires no calibrated measurements or physical references. In the context of quantum computing, QPI can be used to better assess the quality of qubit operations and inform the development of more effective control and error mitigation techniques. Dynamical models produced by QPI may also be used to assess non-Markovianity or quantumness in processes of interest; however, such questions are beyond the scope of this work. We present QPI simply as a very general method—free of the usual assumptions of Markovianity—to obtain an accurate, empirically-derived mathematical model of a dynamical system’s observable behavior.

The development of QPI was motivated by studies addressing the limitations of GST and RB for processes with non-Markovian aspects [5, 23–25] and by the emergence of ad-hoc adaptations of these methods to characterize specific types of non-Markovian errors [13, 26]. QPI provides a general, systematic solution to these challenges. (Unlike GST and RB, however, the version of QPI presented here characterizes only an individual gate, not a set of gates; see section 6 for further discussion.) As will be explained, our method for identifying quantum processes is based on a classic technique for reconstructing classical discrete-time linear systems from time traces, a technique also utilized in [27] for Markovian quantum processes. [19, 20, 28, 29] also present general frameworks for representing and characterizing non-Markovian quantum dynamics, but address slightly different situations than QPI and employ rather different formalisms. [28, 29] consider quantum communication channels with memory, wherein the system is prepared fresh for each use of the channel; in QPI, each occurrence of the process in question begins with the state resulting from the preceding occurrence, a communication channels with memory, wherein the system is prepared afresh for each use of the channel; in slightly different situations than QPI and employ rather different formalisms.

In the remainder of this article we give a detailed description and numerical demonstration of QPI. We begin in section 2 with a thought exercise to develop the intuition underlying QPI and illustrate several important principles. In section 3 we formulate the problem and review a model inference technique from the field of linear systems engineering that underlies QPI. The QPI protocol itself is described in section 4, followed in section 5 by
simulation results which demonstrate the effectiveness of QPI for representative non-Markovian quantum processes. Lastly, in section 6 we summarize our results and identify possible directions of further development. Mathematical and algorithmic details are provided in the Appendices.

2. A thought exercise

The means of detecting and characterizing non-Markovian behavior can be understood through a simple example. Consider a slot machine which produces a 0 or 1 each time its lever is pulled (figure 1). A frequent player notices that each time the machine is powered on, the first two digits produced are equally likely to be 00, 01, 10, or 11. If this were the extent of Alice’s observations, she might reasonably conclude that pulling the lever simply outputs a random digit. That is, she would expect the probability of observing a 1 would be 0.5, regardless of the value of previously observed digits.

But suppose Alice observes longer output sequences and discovers that each time the machine is powered on, one of two distinct behaviors occurs: In half the cases, each pull of the lever repeats whatever digit appeared first. In the remaining cases, the output alternates between 0 and 1. Note that the value of the $(k + 1)$th digit is not predicted by the $k$th digit alone (the conditional probability is 1/2), but it is predicted by the $(k - 1)$th digit (they are always the same). In other words, the next output of the machine depends not only on the most recent output, but on prior outputs; the output is non-Markovian. Alice can explain the observed behavior by positing the existence of a ‘control’ bit inside the machine which determines whether the output is constant or alternating. The control bit is set to a random value when the machine is powered on and does not change.

Let us now imagine that the slot machine contains an internal counter which causes the control bit to change state after 7 pulls, thereby switching the behavior from constant to alternating or vice versa. If Alice only observes sequences of length $\leq T$, she will see nothing that indicates that the behavior of longer sequences is any different; her model will be incomplete.

Although this example is not a perfect analogy for QPI—the slot machine yields an output at every iteration whereas a disruptive measurement is needed to query a quantum process—it illustrates several relevant principles. The first principle is that repeated interactions between a system and non-forgetful ancillary degrees of freedom can yield non-Markovian system behavior. (Note that if the control bit was randomized at each time step, the target bit would be randomized at each time step independent of its previous value. In this case, there would be no need to track the state of the control bit; a stochastic model of the target bit alone would suffice.) The second principle is that the pattern of system behavior over an extended period of time can be used to infer the existence of, and determine the dynamics of, hidden degrees of freedom. In this example the inference could be made simply by inspection, but we will show how to do it in a systematic way. The third principle is that since experiments can have only finite duration, it is not possible to guarantee that one has observed enough behavior to obtain a complete model, i.e. with enough degrees of freedom to accurately predict all future behavior. But as we will show, it is possible to guarantee that an experimental characterization yields a complete model if one has an upper bound for the effective dimension of the process.

Figure 1. A fictional slot machine illustrates how repeated observations of a system over time can be used to infer the existence of, and to model, latent degrees of freedom responsible for non-Markovian behavior. (left) The slot machine produces a digit (0 or 1) each time its lever is pulled. A frequent player observes that each time the slot machine is powered on it exhibits one of two distinct behavior patterns: the output either alternates with each pull of the lever or remains constant. This behavior cannot be explained by any stochastic model involving the only the most recently output digit. (right) The behavior can be explained by positing the existence of an ancillary bit that determines whether the output is constant or alternating.
3. Theory of QPI

3.1. Problem formulation

The problem we address is to how to experimentally characterize the behavior of a quantum system of interest under some dynamical process, without assuming anything about the system’s size or composition or the nature of the process. The process in question may be a repeatable procedure \( \mathcal{P} \) with definitive start and end, or it may be a continuous, ongoing dynamic. In the latter case we let \( \mathcal{P} \) denote evolution for some fixed time \( \Delta \tau \), where \( \Delta \tau \) becomes the time resolution at which the dynamics are resolved.

Ultimately, all that can be known about a system comes from one’s interactions with it. The kind of characterization we develop is relative to, and in terms of, the available ways of interacting with the system. In this work we consider only interactions in the form of ‘experiments’ in which the system is prepared, made to evolve under \( \mathcal{P} \), and subsequently measured. (In section 6 we will speculate how QPI may be extended to more complicated interactions.) Let \( \mathcal{I} \) be the set of ways the system can be prepared (initialized) and let \( \mathcal{M} \) be the set of ways it can be measured. While in principle there exists a continuum of possible initializations and possible measurements, in practice one can utilize only finitely many of these; thus we assume \( \mathcal{I} \) and \( \mathcal{M} \) are finite sets.

An experiment is a specified by a triple \( (i, t, m) \) which specifies the initialization \( i \in \mathcal{I} \), followed by \( t \) repetitions of \( \mathcal{P} \), followed by the measurement \( m \in \mathcal{M} \). For notational convenience we suppose that each measurement has only two possible outcomes, YES or NO. (A measurement with \( k \) possible outcomes can be cast as \( k \) distinct YES/NO measurements.) Under these premises, the observable behavior of the system is the set of YES outcome probabilities for all possible experiments. The YES probability for experiment \( (i, t, m) \) will be denoted by \( F^{(i)}_{t,m} \).

It is worth reiterating that the characterization of \( \mathcal{P} \) produced by QPI is completely relative to the set of initializations \( \mathcal{I} \) and set of measurements \( \mathcal{M} \) considered. \( \mathcal{I} \) and \( \mathcal{M} \) need not be tomographically complete for the system, but if they are not then the resulting characterization may not cover the full state space of the system. Furthermore, if an ‘absolute’ characterization is desired, \( \mathcal{I} \) and \( \mathcal{M} \) should include the procedures that define the reference frame. As a side result, QPI also yields a characterization of the initializations and measurements relative to one another.

For our purposes, characterization amounts to constructing a mathematical model of the process \( \mathcal{P} \), initializations \( \mathcal{I} \), and measurements \( \mathcal{M} \). We opt not to use the traditional Hilbert-space formulation of quantum mechanics, but instead the more general framework known as generalized probabilistic theory (GPT) \([59–61]\). GPT is a framework for a class of physical theories that includes quantum mechanics and classical mechanics as special cases. For us GPT has two appealing features: first, physical phenomena are described in operational terms. Rather than expressing phenomena in terms of complex operators that are not directly observable, GPT expresses phenomena in terms of the outcome probabilities of measurements that could be performed. This matches the task of experimental characterization extremely well. Secondly, the GPT formalism does not distinguish between quantum and classical behaviors; it treats all behaviors in a uniform way. This is useful because although the system of interest and the environment are nominally quantum, their behavior may have some purely classical aspects. GPT allows one to construct the most economical model of a process without choosing it to be quantum or classical. On the other hand, some symmetries required by quantum mechanics (such as complete positivity) are less conveniently expressed in GPT. We emphasize that QPI is not tied to the GPT formalism; if desired, QPI could be formulated entirely in terms of the traditional Hilbert space formalism.

One approach to representing non-Markovian dynamics is to construct dynamical maps that are not linear and/or not CPTP \([36]\). Another strategy is to go beyond time-local maps and use structures that explicitly encode temporal correlations, such as spectral densities \([62, 63]\), autoregression parameters \([19]\), or process tensors \([20, 64]\). QPI takes a different approach: the system’s temporal evolution (whether Markovian not) is modeled by a fixed Markovian process involving both the system of interest and a sufficiently large abstract environment \([65]\). (This is analogous to Stinespring dilation of a CPTP channel \([66, 67]\).) That is, an extended Markovian model is formulated to predict the system’s observable (possibly non-Markovian) behavior. In technical terms, the central premise of QPI is that the system’s observable behavior can be modeled to desired accuracy by a sufficiently large finite-dimensional, time-independent, time-local dynamical model \([2]\). Significantly, such a model can be inferred from measurements that nominally involve the system alone. This is important because the experimentalist often does not have access to, or even knowledge of, all the pertinent parts of the

---

1 There are two ways in which a characterization involving finitely many different initializations \( \mathcal{I} \) and measurements \( \mathcal{M} \) can be extrapolated to the continuum of initializations and measurements. Firstly, the linearity of probability means that a characterization involving elements of \( \mathcal{I} \) and \( \mathcal{M} \) automatically applies to any probabilistic mixture of such operations. More generally, suppose one has a continuously-parameterized initialization procedure \( I \). If it can be assumed that \( I(\theta) \) describes a continuous family of states, then by using a sufficiently dense set \( \mathcal{I} = \{ I(\theta_1), \ldots, I(\theta_M) \} \) one could interpolate the state produced by \( I(\theta) \) for \( \theta_1 \leq \theta \leq \theta_M \). An analogous argument holds for measurements.

2 Time-independent means the dynamical equations have no explicit dependence on time. Time-local means that all quantities in the dynamical equation are evaluated at the same time.
environment. By inferring a model on an extended state space, QPI implicitly characterizes the involvement of pertinent environmental degrees of freedom (though these degrees of freedom are abstract and might not be readily associated with particular physical degrees of freedom).

The only other assumption of QPI is that each experiment is statistically independent. One way of satisfying this second assumption is to use strong initializations that erase any correlations between the system and the environment due to previous experiments. For example, the initialization procedure may include waiting a time much longer than the plausible memory lifetime of the environment. Note, however, that we allow the initialization procedure itself to correlate the system and environment. A weaker way of satisfying the independence assumption is to randomize the order of the experiments so that any residual correlations between the system and the environment at the start of each experiment are sampled fairly. In this case, QPI infers a distribution of correlated initial states.

In QPI, a model of dimension \( d \) consists of a state vector \( s_i \in \mathbb{R}^d \) for each initialization procedure \( i \in I \), a property vector \( p_m \in \mathbb{R}_m^d \) for each measurement procedure \( m \in M \), and a transfer matrix \( T \in \mathbb{R}^{d \times d} \) for the process \( P \). (The relation between these model elements and conventional quantities is shown in table 1.) Then the YES probability of experiment \((i, t, m)\) is

\[
F_{i,m}^{(t)} = s_i T^t p_m.
\]

(We have adopted the convention in which states are row vectors and operators are applied on the right.) In terms of the matrices \( S = [s_1; s_2; \ldots] \) and \( P = [p_1^t; p_2^t; \ldots] \), we have

\[
F^{(t)} = ST^t P.
\]

The goal of QPI is to find matrices \( S, P, \) and \( T \) that accurately predict \( F^{(t)} \) for all \( t = 0, 1, 2, \ldots \), i.e. the observable behavior of the system as defined above. Here \( S \) and \( P \) are a characterization of the state preparations and measurements, respectively, while \( T \) characterizes the process \( P \) itself. We note that these matrices can be determined only up to a similarity transformation, since for any invertible matrix \( G \) the mapping \( S \rightarrow GS \), \( P \rightarrow G^{-1} P \), \( T \rightarrow G^{-1} T G \) leaves all probabilities unchanged. This indeterminacy of representation is known as ‘gauge indeterminacy’. While gauge indeterminacy does lead to a minor theoretical conundrum regarding the assessment of process fidelity [68], it has (by definition) no experimental consequence and thus is not a limitation of QPI.

In the case of a continuous-time process, QPI can be used to obtain an effective master equation. Let \( \rho(t) \in \mathbb{R}^d \) denote the extended system-environment state at time \( t \) with \( \rho(t) \equiv \rho(t) P \). From (2) we have

\[
\rho(t + \Delta t) = ST^t \rho((t - 1) + \Delta t) T
\]

where \( \Delta t \) is the chosen time resolution. For sufficiently small \( \Delta t \), \( T \approx 1 + L \Delta t \) for some matrix \( L \). Then \( \rho \) approximately obeys the master equation

\[
\frac{d\rho(t)}{dt} = \rho(t)L.
\]

### 3.2. The Ho–Kalman method

A key to solving the problem of characterizing non-Markovian quantum processes is a technique that was devised half a century ago in the context of discrete-time linear systems and is now a standard topic in introductory texts on systems engineering [21, 22]. Consider a classical input-output system whose output \( x \in \mathbb{R}^n \) at each time \( t \in \mathbb{Z} \) is a linear function of its input \( u \in \mathbb{R}^m \) at the previous \( d \) times. Ho and Kalman [69] showed that the system can be described by a \( d \)-dimensional state model of the form

\[
\begin{align*}
\dot{s}(t + 1) &= u(t)A + s(t)B \\
x(t) &= s(t)C
\end{align*}
\]

where \( s \in \mathbb{R}^d, A \in \mathbb{R}^{d \times d}, B \in \mathbb{R}^{d \times d}, \) and \( C \in \mathbb{R}^{d \times n} \). Furthermore, they showed how to obtain the coefficients \( A, B, C \) from the observable quantities \( X(1), \ldots, X(2d) \) where \( X_i(t) \) is the value of the output \( x_i(t) \) in response to a unit impulse on the \( i \)th input \((u_i)\) at time 0. In this case \( X(t + 1) = AB'C \). Let

\footnote{A measurement with a YES/NO outcome space may be regarded as a determination of whether or not a system has a particular property.}
Mathematically, the construction of a \( d \)-dimensional model from \( H \) is possible because \( H \) has rank \( d \). What if \( A \) and \( C \) are already tomographically complete? In that case it ought to be sufficient to use just \( X(1) = AC \) for \( H \) and \( X(2) = ABC \) for \( H' \), as in QPT. In this case, the textbook Ho–Kalman method asks for more data than is actually necessary. The following Lemma, proved in appendix A, states that the number of time steps needed to ensure a complete characterization is determined by the number of latent (unmeasured) degrees of freedom only, not the total number of degrees of freedom:

**Lemma 1.** Let \( X(t) \) be the impulse response function for an arbitrary linear, discrete-time, time-shift invariant system of dimension \( d \). Let \( r = \max(\text{rank} X(t)) \) denote the number of directly accessible dimensions and let \( l = d - r \) denote the number of latent dimensions. Let

\[
H^{(t;k)} = \begin{bmatrix}
X(t) & X(t+1) & \cdots & X(t+k) \\
X(t+1) & \vdots & & \\
\vdots & & \ddots & \\
X(t+k) & \cdots & X(t+2k-1) & X(t+2k)
\end{bmatrix}.
\]

For \( k \geq l \), rank \( H^{(t;k)} = d \), i.e. \( H^{(t;k)} \) captures the full dimensionality of the system. Furthermore a minimal (\( d \)-dimensional) model of the system can be constructed from \( X^{(1)} \), \( X^{(2)} \), \( X^{(2i+1)} \). In particular, if \( LR \) is any rank factorization of \( H^{(t)} \), then \( X(t+1) = AB'C \) where \( A \) is the first \( m \) rows of \( L \), \( C \) is the first \( n \) columns of \( R \), and \( B = L^{-1}H'R^+ \). Some \( d \)-dimensional systems can be reconstructed from fewer observations and/or observations at non-consecutive times, but some cannot. Thus \( k \geq l \) is necessary to guarantee complete characterization of an unknown \( d \)-dimensional system.

The Ho–Kalman method is the basis of our approach to QPI. However, there are two important differences between a (classical) linear input-output system and a quantum process that require consideration:

- In the Ho–Kalman problem \( X_j(t) \) is (in principle) a directly observable quantity, whereas in QPI the measurable quantity \( F_{i;m}^{(t)} \) is a probability. One consequence is that in QPI, many repetitions of an experiment are needed to obtain each value \( F_{i;m}^{(t)} \). Even then, one only obtains an estimate \( \hat{F}_{i;m}^{(t)} \) whose precision is limited by the number of repetitions of the experiment. There is a large body of work on the extension of the Ho–Kalman approach to noisy data, but it considers additive or multiplicative Gaussian noise and is not applicable to QPI, where \( F_{i;m}^{(t)} \) has a binomial distribution whose width depends on the (unknown) value of \( F_{i;m}^{(t)} \).

- The observables in the Ho–Kalman method are all mutually compatible. In principle, the values \( X_j(t) \) for all \( j \) and all \( t \geq 1 \) can be obtained in a single experiment with initial state \( i \). In contrast, measuring a quantum system generally disturbs it (an aspect of the uncertainty principle). Thus in QPI, not all measurable properties of the system can be measured at the same time, nor is the evolution of the system after a measurement the same as if the system had not been measured. Each \( F_{i;m}^{(t)} \) must be estimated from a distinct set of experiments, in which the system evolves unobserved for \( t \) process repetitions and is then measured.
In the next section we present a method for QPI based on an extension of the Ho–Kalman approach to the quantum domain.

4. The QPI protocol

4.1. The experimental protocol

QPI is accomplished by performing a series of experiments, each repeated many times (figure 2). Each experiment is designated by a triple \((i, t, m)\) which designates the \(i\)th way of initializing the system, followed by \(t\) repetitions of the process to be characterized, followed by the \(m\)th way of measuring the system. For each experiment \((i, t, m)\) one records \(N_{i,m}^{(t)}\), the number of times the experiment was performed, and \(Y_{i,m}^{(t)}\), the number of times the YES outcome was observed. The frequency \(\hat{F}_{i,m}^{(t)} \equiv Y_{i,m}^{(t)}/N_{i,m}^{(t)}\) is the experimental estimate of \(F_{i,m}^{(t)}\). (Throughout this work, an experimental estimate of a quantity \(Q\) is denoted by \(\hat{Q}\).)

The set of experiments performed must be chosen with some care in order for QPI to be effective. If the experimental probabilities could be determined perfectly, then by lemma 1 \(F^{(0)}, \ldots, F^{(2^t+1)}\) would be sufficient to characterize any process of dimension \(d \leq I + \text{rank } F^{(0)}\). (Note that rank \(F^{(0)}\) is just the dimension of the space spanned by the initial states and measurements. If these are tomographically complete for the system of interest, rank \(F^{(0)}\) is just the system dimension.) However, some of the degrees of freedom may be slow to manifest in the sense that it takes many repetitions of the process for their impact on observable properties to become significant. In such cases the matrix \(H^{(0,\bar{T})}\) is ill-conditioned and even a small amount of statistical error in the data yields an extremely poor characterization. For this reason it is advantageous to perform experiments covering a wide range of \(t\) values. As a general rule, an experiment with \(2t\) process repetitions is twice as sensitive to the process parameters \(T\) as an experiment with \(t\) process repetitions [7].

The other main consideration is that the Ho–Kalman reconstruction technique requires the matrix \(H\) to have a generalized Hankel block structure:

\[
H = \begin{bmatrix}
F^{(t_i+j_1)} & F^{(t_i+j_2)} & \ldots & F^{(t_i+j_r)} \\
F^{(t_i+j_1)} & F^{(t_i+j_2)} & \ldots & F^{(t_i+j_r)} \\
\vdots & \vdots & \ddots & \vdots \\
F^{(t_i+j_1)} & F^{(t_i+j_2)} & \ldots & F^{(t_i+j_r)} 
\end{bmatrix}
\]  

(8)

for some choice of integers \(i_1, \ldots, i_m\) and \(j_1, \ldots, j_m\).

Our approach to satisfying these criteria is to use multiple ‘flights’ of experiments, where a flight is a set of experiments \(\{(i, t, m)\}\) with \(2I + 1\) consecutive \(t\) values and all values of \(i, m\) for each \(t\) (figure 3(a)). By lemma 1, each flight contains enough information to characterize a process of dimension at least \(\bar{T} + \text{rank } F^{(0)}\). To efficiently cover a wide range of \(t\) values we use flights whose base \(t\) values increase exponentially, \(\{0, 1, 2, 4, \ldots, 2^s\}\).

We then add flights as needed to support the Ho–Kalman structure, equation (8). The result is that the experiments performed are those with \(t\) values in the set

\[
T = T_R + T_C,
\]  

(9)

where

\[
T_R = \{0, 1, 2, 4, \ldots, 2^s\} + \{0, 1, 2, \ldots, \bar{T}\}
\]  

(10)

\[\text{Figure 2. The experimental procedure for quantum process identification. An experiment consists of an initialization } i \text{, } t \text{ repetitions of the process, and a final YES/NO measurement } m. \text{ Initializations and measurements nominally concern the system but are considered to involve the environment as well. Each experiment is repeated many times, yielding a frequency of YES outcomes. For the most complete and accurate results, all available initializations and measurements should be used, and the set of } t \text{ values should consist of widely spaced ‘flights’ of consecutive values (see section 4.1).} \]
Here + is the sumset operation and $a \oplus b$, $\tilde{I}$ are chosen integers. We found that such a pattern of experiments is generally necessary to obtain accurate models, especially for systems in which the non-Markovian aspects manifest only over long timescales. Notably, the characterization accuracy grows as $\max T$, whereas the number of different experiments to be performed grows asymptotically as only $\log_{\max T} \hat{\mathcal{R}}(2)$. 

4.2. Model inference procedure

Once the experiments are performed, the data is analyzed to infer a model of the process $\mathcal{P}$, the initial states, and the measurements. Our basic approach is to seek the model with the highest likelihood for the given observations. This is a challenging task for two reasons: (1) The size of the model (number of dimensions) is unknown. (2) The likelihood is extremely nonlinear in $T$, $T$ appears with powers up to $\max T$, which in practice can be on the order of $10^3$ or more. This creates an extremely unforgiving optimization landscape with many local (and very suboptimal) maxima.

Regarding the first challenge, a number of techniques have been devised to compare the likelihood of models with different numbers of parameters (e.g. the Akaike information criterion [70]) or to construct models whose complexity automatically scales with the complexity of the data (such as Dirichlet processes [71]). We use a simple but effective technique to first estimate the dimensionality of the process in question, which then determines the set of parameters to be estimated. The second challenge is essentially the same as that faced in gate set tomography. Building upon the progressive fitting approach described in [7], we devised a 4-stage model inference procedure, carefully honed over the course of our studies to reliably find concise, accurate process models. An overview of our procedure is given below; additional details are given in appendices B–D.

1. Estimate the Model Dimension. First, the data is arranged into a matrix $\hat{H}$ which is the experimental estimate of a Ho–Kalman matrix $H$. $\hat{H}$ has the structure given in equation (8) with rows indexed by values from $T_R$ and columns indexed by values from $\mathcal{T}_C$. This may be viewed as an arrangement of the data from all the flights as shown in figure 3. The effective dimension $d$ is then estimated as the number of statistically significant non-zero singular values of $\hat{H}$ [72], where significance is assessed by a chi-square test. In brief, the uncertainties in the experimental data are estimated and propagated to yield uncertainties in the singular values of $\hat{H}$. This yields for each $k = 1, 2, \ldots$ a threshold for the sum of squares of the $k$ smallest singular

$\mathcal{T}_C = \{0, 1, 2, 4, \ldots, 2^b\} + \{0, 1, 2, \ldots, \hat{I}\}$.  

Here $\oplus$ is the sumset operation and $a, b, \hat{I}$ are chosen integers. We found that such a pattern of experiments is generally necessary to obtain accurate models, especially for systems in which the non-Markovian aspects manifest only over long timescales. Notably, the characterization accuracy grows as $\max T$, whereas the number of different experiments to be performed grows asymptotically as only $\log_{\max T} \hat{\mathcal{R}}(2)$.
values. If for some $k$ the sum falls below the threshold, those singular values are discounted (deemed insignificant). The dimension $d$ is taken to be the number of singular values that remain after the largest subset of singular values has been discounted. We note that this procedure does not determine the inherent dimension of a process so much as the dimension that is revealed by the data.

2. Obtain an Initial Model. This stage is a variation of the Ho–Kalman method, modified to account for uncertainty in the experimental data. Let $\hat{H}$ be the ‘time shifted’ version of $\hat{H}$, i.e. the matrix obtained by increasing the $t$ value of each element of $\hat{H}$ by 1. First, the variance of each experimental quantity $\hat{F}_{it}$ is estimated. Then a rank-$d$ factorization $\hat{H} \approx LR$ is found by minimizing $\sum_{i,m} W_{i,m}(LR - \hat{H})_{i,m}^2$, where $W_{i,m}$ is the inverse variance of $\hat{H}_{i,m}$. The process matrix is estimated as the matrix $T$ that minimizes $\sum_{i,m} W_{i,m}'(LTR - \hat{H}')_{i,m}$ where $W_{i,m}'$ is the inverse variance of $\hat{H}'_{i,m}$. The first few rows of $L$ and columns of $R$ may be taken as initial estimates of $S$ and $P$, respectively.

3. Obtain an Improved Model. Although the previous stage uses all the data, it uses only the fact that $\hat{H}$ and $\hat{H}'$ are related by a (single) factor of $T$. It does not use the fact that data from different flights should be related by higher powers of $T$. To incorporate this relationship into the model estimate, the data is organized into groups of flights or ‘blocks’ with exponentially increasing $t$ values (figure 3(c)). Starting with the matrices $L$, $T$, $R$ obtained in stage 2, a search is performed to find matrices $A$, $T$, $B$ that best fit the data in all the blocks. The first few rows of $A$ and columns of $B$ may be taken as improved estimates of $S$ and $P$, respectively. To ensure that the search finds a good solution and does not get stuck in a local extremum, the search is performed progressively, starting with the lowest-$t$ blocks and gradually incorporating data from blocks with higher $t$ values. The low-$t$ blocks yield coarse estimates of $T$ which are gradually refined using the data in higher-$t$ blocks.

4. Obtain a physically valid model estimate. The model produced by stage 3 is usually fairly accurate, but tends to make slightly unphysical predictions (i.e. probabilities outside the interval [0, 1]). In this last stage, the model likelihood is maximized with the addition of penalty terms to encourage the validity of predicted probabilities. As in stage 3, the log-likelihood is approximated by a weighted sum of squared errors; however, this time the weights are not based on the (fixed) experimental frequencies but are based on the predicted probabilities. Also, in this step the data is not arranged into Ho–Kalman matrices; the model $F(t) = ST^TP$ (equation (2)) is fit directly to each experiment.

In our simulation studies, the model inference procedure described above proved to be both fast and reliable. In over 99% of the thousands of simulations we performed, it found a high-likelihood model without any intervention. In the few cases in which it did not, making minor adjustments to optimization parameters in stage 3 or 4 (e.g. penalty weights, convergence thresholds) invariably led to success. We found that all four stages were usually necessary to obtain accurate models. On a standard desktop computer, the time to complete all four stages ranged from just a few seconds for 7 dimensional processes to a few minutes for a 19-dimensional process.

We note that the exact arrangement of the data in steps 1–3 does not appear critical. Beyond satisfying the Ho–Kalman structure, it appears to be important only that each row and column of $\hat{H}$ cover both a wide range of $t$ values and multiple runs of roughly $l$ consecutive $t$ values.

4.3. Example
To illustrate the QPI procedure let us consider how Alice would use it to characterize a ‘black box’ version of the slot machine described in section 2. In this version the slot machine has a display screen, a lever, and buttons labeled I0, I1, and M. Pushing I0 or I1 turns the machine on. When the lever is pulled the machine makes a momentary noise but otherwise has no observable effect. Pressing M causes the screen to briefly display the number 0 or 1, after which the machine turns off. In terms of the QPI formalism the initialization procedures are $I = \{0, 1\}$. Pulling the lever performs an unknown process $P$. Measurement is accomplished by pressing M. This is treated as two procedures, M0 and M1, which test whether the outcome is 0 or 1, respectively. Thus $M = \{M0, M1\}$.

The behavior of the machine, which Alice will attempt to discover, is as follows: If the machine is turned on using I0, the output after an even number of lever pulls is always 0; after an odd number of pulls it is 0 with probability 0.5 and 1 with probability 0.5. Likewise, if the machine is turned on using I1, the output after an even number of lever pulls is always 1; after an odd number of pulls it is 0 with probability 0.5 and 1 with probability 0.5. This behavior is expressed by the series of matrices $F^{(0)}, F^{(1)}, F^{(2)}$, ... with
Alice seeks a model \((S, T, P)\) such that \(ST^tP = F^{(t)}\) for all \(t\). The rows of \(S \in \mathbb{R}^{2 \times d}\) will express the states produced by \(I_0\) and \(I_1\) respectively in some \(d\)-dimensional state space. \(T \in \mathbb{R}^{d \times d}\) will express the action of the lever in this state space; and the columns of \(P \in \mathbb{R}^{d \times 2}\) will express the properties measured by \(M_0\) and \(M_1\) respectively.

To perform QPI Alice first estimates the probabilities \(F^{(t)}_{i,m}\) for all \(i \in \{0, 1\}\), all \(m \in \{M_0, M_1\}\), and for selected values of \(t\). Each experiment involves initializing the machine using either \(I_0\) or \(I_1\) and pulling the lever some fixed number of times \(t\) before performing readout \(M\). She repeats each experiment 100 times and records the fraction of times outcome \(m = \{0, 1\}\) occurs as \(F_{i,m}^{(t)} \approx F^{(t)}_{i,m}\). Alice first obtains

\[
F^{(0)} = \begin{bmatrix} 0.48 & 0.52 \\ 0.53 & 0.47 \end{bmatrix}, \quad F^{(1)} = \begin{bmatrix} 0.49 & 0.51 \\ 0.46 & 0.54 \end{bmatrix}, \quad F^{(2)} = \begin{bmatrix} 0.52 & 0.48 \\ 0.57 & 0.43 \end{bmatrix}.
\]

Since \(F^{(0)} \approx F^{(0)}\) she surmises that rank \(F^{(0)} = 2\), i.e. the initializations and readout span a 2 dimensional space. Alice decides that she is willing to construct a model that can capture at least one latent dimension, \(l = 1\). (In practice she should probably consider a larger value, but here we use a small value for simplicity of presentation.) Each flight of experiments will involve \(2^l + 1 = 3\) consecutive \(t\) values. She chooses the flights to have base values from 0 up to \(2^1 + 2^1 = 4\). The complete set of \(t\) values used in Alice’s characterization is \(T = T_R + T_C = \{0, 1, 2, 3, 4, 5, 6\}\) where \(T_R = T_C = \{0, 2^0, 2^1\} + \{0, 1\}\). For the sake of discussion, we suppose Alice obtains

\[
F^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad F^{(1)} = \begin{bmatrix} 0.48 & 0.52 \\ 0.53 & 0.47 \end{bmatrix}, \quad F^{(2)} = \begin{bmatrix} 0.49 & 0.51 \\ 0.46 & 0.54 \end{bmatrix}, \quad F^{(3)} = \begin{bmatrix} 0.52 & 0.48 \\ 0.57 & 0.43 \end{bmatrix}.
\]

Having obtained experimental estimates of \(F^{(t)}\) for a number of \(t\) values, Alice proceeds to perform the four steps of data analysis:

1. Estimate the model dimension. Alice arranges nearly all the data into a Ho–Kalman matrix

\[
\hat{H} = \begin{bmatrix} F^{(0)} & F^{(1)} & F^{(2)} \\ F^{(3)} & F^{(4)} & F^{(5)} \end{bmatrix}.
\]

(The column that would contain \(F^{(6)}\) is excluded because step 2 will require all the \(t\) values to be increased by one, and no data was taken for \(t = 7\).) Using the procedure detailed in appendix B, Alice computes the singular values of \(\hat{H}\), the cumulative sum of squares of the singular values from smallest to largest, and the corresponding chi-square thresholds:

\[
\begin{array}{ccccccc}
\text{dimension} & 1 & 2 & 3 & 4 & 5 & 6 \\
\text{singular value} & 3.46 & 2.00 & 1.42 & 0.130 & 0.034 & 0.000 \\
\text{cum. sum. of squares} & 18.03 & 6.02 & 2.02 & 0.018 & 0.001 & 0.000 \\
\chi^2 \text{ threshold} & 0.086 & 0.066 & 0.049 & 0.036 & 0.019 & 0.010
\end{array}
\]

Since the sum of squares falls below the threshold for \(d \geq 4\) but above the threshold for \(d \leq 3\), Alice concludes \(d = 3\).

2. To obtain an initial estimate, Alice shifts the entries in \(\hat{H}\) by one lever pull to obtain

\[
\tilde{H}' = \begin{bmatrix} F^{(1)} & F^{(2)} & F^{(3)} \\ F^{(4)} & F^{(5)} & F^{(6)} \end{bmatrix}.
\]

She numerically obtains a rank-3 factorization \(LR \approx \tilde{H}'\) that minimizes the total weighted error. The first 2 rows of \(L\) yield an initial estimate of \(S\) while the first two columns of \(R\) yield an initial estimate of \(P\). She
obtains an initial estimate of $T$ by minimizing total weighted error of the approximation $LTR \approx \tilde{H}'$, keeping $I$ and $R$ fixed. Her initial model estimate is

$$S = \begin{bmatrix} \begin{array}{ccc} -0.659 & 0.705 & -0.010 \\ -0.657 & -0.709 & -0.016 \end{array} \end{bmatrix}, \quad T = \begin{bmatrix} \begin{array}{ccc} 1.00 & 0.015 & 0.013 \\ 0.005 & -0.064 & -0.839 \\ 0.006 & -1.19 & 0.057 \end{array} \end{bmatrix}, \quad \tilde{P} = \begin{bmatrix} \begin{array}{ccc} -0.761 & -0.758 \\ 0.706 & -0.708 \\ -0.016 & 0.009 \end{array} \end{bmatrix}. \quad (19)$$

3. To obtain a better estimate, Alice arranges her data into three blocks:

$$D_0 = \begin{bmatrix} \bar{F}^{(0)} \\ \bar{F}^{(1)} \\ \bar{F}^{(2)} \\ \bar{F}^{(3)} \end{bmatrix}, \quad D_1 = \begin{bmatrix} \bar{F}^{(1)} \\ \bar{F}^{(2)} \\ \bar{F}^{(3)} \\ \bar{F}^{(4)} \end{bmatrix}, \quad D_2 = \begin{bmatrix} \bar{F}^{(2)} \\ \bar{F}^{(3)} \\ \bar{F}^{(4)} \\ \bar{F}^{(5)} \end{bmatrix}$$

Alice then uses the numerical search procedure described in appendix C to find $A \in \mathbb{R}^{5 \times 3}$, $B \in \mathbb{R}^{3 \times 4}$, and $T \in \mathbb{R}^{3 \times 3}$ such that $AB \approx D_0$, $ATB \approx D_1$, and $AT^2B \approx D_2$. For the starting values of $A$, $T$, $B$ she uses $L$, $T$, and the first four columns of $P$ from the previous step. This yields

$$S = \begin{bmatrix} \begin{array}{ccc} -0.659 & 0.705 & -0.014 \\ -0.657 & -0.709 & 0.003 \end{array} \end{bmatrix}, \quad T = \begin{bmatrix} \begin{array}{ccc} 1.00 & 0.015 & 0.013 \\ 0.005 & -0.064 & -0.834 \\ 0.006 & -1.19 & 0.057 \end{array} \end{bmatrix}, \quad P = \begin{bmatrix} \begin{array}{ccc} -0.761 & -0.758 \\ 0.706 & -0.708 \\ -0.017 & 0.011 \end{array} \end{bmatrix}. \quad (21)$$

4. Alice improves her estimate even further using the procedure described in appendix D to directly minimize the total error of the model $ST^tP = F^{(t)}$ over all $t \in T$, using the values of $S$, $T$, $P$ obtained from step 3. This yields her final estimated model

$$S = \begin{bmatrix} \begin{array}{ccc} -0.659 & 0.701 & -0.025 \\ -0.657 & -0.712 & 0.000 \end{array} \end{bmatrix}, \quad T = \begin{bmatrix} \begin{array}{ccc} 1.00 & 0.010 & 0.009 \\ 0.005 & -0.061 & -0.840 \\ 0.006 & -1.19 & 0.055 \end{array} \end{bmatrix}, \quad P = \begin{bmatrix} \begin{array}{ccc} -0.764 & -0.755 \\ 0.706 & -0.708 \\ -0.020 & 0.013 \end{array} \end{bmatrix}. \quad (22)$$

For comparison, an exact minimal model $(S', T, P)$ for this system is

$$S' = \begin{bmatrix} \begin{array}{ccc} -0.5 & -0.5 & -0.707 \\ 0.5 & 0.5 & -0.707 \end{array} \end{bmatrix}, \quad T' = \begin{bmatrix} \begin{array}{ccc} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \end{bmatrix}, \quad P' = \begin{bmatrix} \begin{array}{ccc} -0.5 & 0.5 \\ -0.5 & -0.5 \\ -0.707 & -0.707 \end{array} \end{bmatrix}. \quad (23)$$

Alice’s final model does not appear to be close to this. However, recall that models are meaningful only up to similarity transformations. A suitable similarity transformation on Alice’s final model yields

$$S = \begin{bmatrix} \begin{array}{ccc} -0.486 & -0.498 & -0.707 \\ 0.511 & 0.493 & -0.707 \end{array} \end{bmatrix}, \quad T = \begin{bmatrix} \begin{array}{ccc} -1.003 & 0.004 & 0.000 \\ -0.004 & 0.997 & 0.000 \\ 0.006 & 0.000 & 1.000 \end{array} \end{bmatrix}, \quad P = \begin{bmatrix} \begin{array}{ccc} -0.503 & 0.503 \\ -0.497 & 0.497 \\ -0.710 & -0.704 \end{array} \end{bmatrix}. \quad (24)$$

which clearly approximates equation (23). As it turns out, this model is accurate only for $t \leq 10$, a consequence of the fact that Alice tested only small values of $t$ and repeated each experiment relatively few times.

5. Simulations

To demonstrate the effectiveness of QPI, we simulated the characterization of several different non-Markovian processes relevant to quantum computing. Each process consists of a simple unitary operation on a qubit with a weak, physically-motivated non-Markovian error. In each case, the system of interest is the qubit and the non-Markovian behavior of the qubit is generated from a Markovian model of the qubit interacting with some external degree(s) of freedom. For these studies the error processes were chosen to be not only weak but also slow to reveal their non-Markovian nature, in order to present a strong characterization challenge.

For each process, we first simulated the evolution of the quantum state for several different initial states and predicted the time-dependent outcome probabilities for the standard tomographic measurements. The ‘standard tomographic measurements’ are the measurements of the dimensionless Pauli operators $\sigma_x, \sigma_y, \sigma_z$. These operators have eigenvalues $+1$ (‘YES’) and $-1$ (‘NO’) with corresponding eigenstates denoted $|\pm\rangle_x, |\pm\rangle_y, |\pm\rangle_z$, respectively. For simplicity, state preparation and measurement were assumed to be ideal. Then, we simulated the collection of experimental data by choosing a random number of YES outcomes for each measurement from the computed probability distribution. The simulated data for each experiment was then
passed to our data analysis code, which estimated the dimension of the process and found the most likely model using the procedure outlined above. The error of the characterization was calculated by using the inferred model to predict the time-dependent qubit state and comparing the predicted state to the true simulated state. The error at each time step was averaged over all initial states and over 500 independent simulated characterizations (200 for the system described in section 5.4).

In each plot below, black dots show the times of the (simulated) measurements and their average error, where error is defined as the trace distance \[ \frac{1}{2} \text{Tr} |\rho(t) - \rho_{\text{est}}(t)| \] between the true state \( \rho(t) \) and the state \( \rho_{\text{est}}(t) \) estimated from the measurements at time \( t \) only. The blue line is the QPI error, defined as the average trace distance between the true state \( \rho(t) \) and the state \( \rho_{\text{QPI}}(t) \) predicted by QPI, which is based on all the simulated measurements. For comparison, the dotted black line is the corresponding error for perfect QPT, obtained using exact measurement outcome probabilities at \( t = 0 \) and \( t = 1 \) only. As will be evident, the intrinsically Markovian models produced by process tomography (even with no statistical measurement error) cannot accurately represent the non-Markovian processes considered below.

5.1. Qubit rotation with slowly varying systematic error
In quantum computing technologies, operations on qubits are typically accomplished by applying control pulses that temporarily modify the Hamiltonian to induce a unitary rotation of one or more qubits. In many implementations the amount of rotation is proportional to the pulse energy. A miscalibration of the pulse intensity or duration results in over- or under-rotation of the qubits, which generally contributes to computational error \[73\].

In this example the process of interest is an intended \( \pi \) rotation of a qubit about the \( y \) axis, i.e. a bit flip. We suppose however that the intensity of the control pulse drifts sinusoidally over time, such that the actual angle of rotation produced by the pulse at time \( t \) (where each pulse takes one unit of time) is

\[ \theta_t = \pi + \epsilon \sin \Omega t. \]  

We choose \( \epsilon = 0.01 \) and \( \Omega = 0.02 \), which yields a small, slow oscillation of the qubit about the intended state in the \( x-z \) plane. This oscillation cannot be described by any fixed Markovian process involving only the qubit. In fact, the error can be described as phase modulation, a process which has an infinite number of sidebands. In other words, this process technically has an infinite dimension. However, relatively few dimensions are needed to accurately reproduce typical experimental data.

The evolution of the qubit under this process was simulated for two different initial states, \( |+\rangle \) and \( |+\rangle_x \). For each initial state, tomographic measurements of the qubit were simulated at 304 times ranging from \( t = 0 \) to \( t = 1035 \), forming 57 flights of length 12. Each measurement was repeated \( 10^4 \) times. In nearly all realizations, model inference yielded an 11-dimensional model. Figure 4 shows the average characterization error as a function of \( t \).

5.2. Coherent qubit leakage
Leakage refers to the loss of a qubit, or more precisely, the transition of a qubit to a non-computational state. This process is usually modeled as irreversible and incoherent, in which case the qubit has no effect on any subsequent operation. More realistically, qubit operations create and manipulate superpositions of computational and non-computational states \[26\]. In such cases a qubit can effectively leave and later return to the computational subspace, interfering with the computational evolution. The particular details of the qubit’s behavior will depend strongly on the physics of the physical implementation of the qubit and the control

![Figure 4. Average characterization error (solid line) for a qubit undergoing \( \pi \) rotations with small, sinusoidally varying error in the rotation angle. Also shown for comparison are the raw measurement error (dots) and error of perfect process tomography (dotted line).](image)
mechanisms employed. For this example we use the 3-level anharmonic oscillator model which is applicable to popular qubit technologies including superconducting flux qubits and trapped ions \[74\]. In this model, the control pulse drives not only the intended \( |0\rangle \rightarrow |1\rangle \) transition, but also off-resonant transitions between \( |1\rangle \) and the next excited state \( |2\rangle \). The interaction Hamiltonian is

\[
H = \Omega(t) |0\rangle \langle 1| + \sqrt{2} |1\rangle \langle 2| + \text{h.c.} + \Delta |2\rangle \langle 2|,
\]

(26)

where \( \Omega(t) \) is the control amplitude and \( \Delta \) is the detuning of the non-computational state. For simulations we used Gaussian control pulses of width 0.25 time steps (truncated to have 1 time step duration) and chose an excited state detuning \( \Delta = 20 \) yielding a system in which the leakage in and out of the computational space is small but not negligible. For these parameter values the total probability of state \( |2\rangle \) oscillates with a period of nearly 30 time steps, with maximal value of a few percent. Meanwhile, the computational state slowly precesses about the intended state.

The evolution of this 3-level system was simulated for four different initial qubit states \( |+\rangle_2 = |0\rangle, |-\rangle_2 = |1\rangle, |+\rangle_x, \) and \( |+\rangle_y \). For each initial state, tomographic measurements of the qubit were simulated at 196 times ranging from \( t = 0 \) to \( t = 1029 \), forming 57 flights of length 6. Each measurement was repeated \( 10^4 \) times. In nearly all cases, the model inference yielded a 7-dimensional model. The average characterization error for this process is shown in figure 5.

### 5.3. A qubit coupled to a single impurity

Another potential source of error in qubit devices is unintended coupling between qubits, or between a qubit and a nearby impurity. A simple model for such coupling is the isotropic spin exchange Hamiltonian

\[
H = \gamma (\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z).
\]

(27)

Taking \( \gamma = 0.01 \) to model weak coupling, the evolution of a qubit and impurity spin was simulated for three initial qubit states \( |+\rangle_x, |+\rangle_y, \) and \( |+\rangle_2 \). The impurity is always initially in the state \( |+\rangle_2 \). For each initial state, tomographic measurements were simulated at 64 times ranging from \( t = 0 \) to \( t = 1030 \), forming 12 flights of length 7. Each measurement was repeated \( 10^4 \) times. In nearly all cases, model inference yielded a 7-dimensional
model. This may be compared with the nominal dimension $4^2 = 16$ of a two qubit system. The average characterization error for this process is shown in figure 6.

5.4. $^{31}$P qubit in silicon

For our final case study, we considered a $^{31}$P donor in $^{28}$Si with $^{29}$Si impurities. The nuclear spin of the $^{31}$P defines a qubit. The qubit is manipulated by a time-varying magnetic field which also affects the impurities. Meanwhile, the $^{31}$P and $^{29}$Si impurities interact via hyperfine coupling to the donor electron. A detailed model for this system was formulated in [75]. We simulated the behavior of the qubit and impurities in response to a sequence of pulses that each nominally produce a $\pi$ rotation of the qubit (bit flip). For these simulations, the donor was imagined to be at the center of a $(5 \text{ nm})^3$ cube with a $^{29}$Si concentration of 800 ppm, which translates to 6 impurities in the qubit volume. The impurities were distributed randomly throughout the volume; in all, 200 random configurations were simulated. The background magnetic field $B_z$ was 1.5 T, the amplitude $B_x$ of the driving magnetic field was 0.15 T, and the ambient temperature was 250 mK. The total length of each pulse sequence was 2.9 $\mu$s. (For further details see [75].)

For each configuration of the impurities, the evolution of the donor and impurities was simulated for 4 different initial qubit states $\{\langle \uparrow \rangle_x, \langle \uparrow \rangle_y, \langle \downarrow \rangle_x, \langle \downarrow \rangle_y\}$ and with the impurities initially in thermal equilibrium. For each initial state, tomographic measurements of the qubit were simulated at 272 different time steps ranging from $t = 0$ to $t = 1033$, forming 57 flights of length 10. Each measurement was repeated $10^3$ times. In most realizations the inferred model dimension was 19 or 20, which is much less than the nominal dimension of the $^{31}$P + $^{29}$Si system (7 nuclear spins + 1 electron spin = 8 spins with a nominal dimension of $(2^8)^2 = 65536$). Thus QPI reveals that the effective interaction between the qubit and the impurities involves only a small subspace of the available Hilbert space, an insight that is not readily apparent from the underlying model. The average characterization error for this system is shown in figure 7.

6. Discussion

In all of the simulation studies above, our QPI method was able to accurately capture and reproduce the observable behavior of the qubit over very long time scales, with trace distance errors on the order of $10^{-2}$. While this level of error would not be impressive for tomography of the state at any one time, the fact that this accuracy is maintained over more than a thousand time steps implies that the process itself is actually characterized to an accuracy on the order of $10^{-5}$. Notably, the inferred models remained accurate well beyond the times at which measurements were taken (although the errors did tend to slowly grow the farther in time the models were extrapolated).

In contrast, standard quantum process tomography (which uses measurements only at $t = 0$ and $t = 1$) yielded inaccurate predictions of qubit behavior past a few tens of time steps, even under the assumption of zero statistical measurement error. For some of the examples the error of process tomography is oscillatory. We attribute this to an oscillatory aspect of the non-Markovian evolution, such that every so often it happens to (nearly) coincide with the prediction of a Markovian model. Arguably, it would be fairer to compare QPI with a version of QPT that incorporates the data at all times but assumes the process is Markovian. This can be achieved in our framework by restricting the model dimension to 4 for characterization of a single qubit. In fact we attempted to perform such a comparison, but found that dimension-limited (Markovian) models were such a poor fit to the data that the model inference would not converge.

![Figure 7. Average characterization error (solid line) for a $^{31}$P donor qubit with 6 nearby $^{29}$Si impurities. Also shown for comparison are the raw measurement error (dots) and error of perfect process tomography (dotted line). Each time step is 2.9 $\mu$s.](image-url)
For the finite dimensional processes studied (leakage and spin exchange), QPI inferred a process dimension that was equal to or close to the true dimension. For the high-dimensional processes studied (control drift and \( ^{31}P + ^{29}Si \) systems), QPI was able to reproduce the observed behavior using relatively low-dimensional models. Fortunately, it was not necessary to use flights as long as suggested by the worst-case bound of lemma 1.

Although these simulation studies addressed processes nominally involving a single qubit, QPI is in principle just as applicable to the characterization of qudit or multiqubit operations. However, the state space size (and hence also the cost of data collection and data analysis) grows exponentially with the number of qubits. Thus like all forms of tomography, QPI is really only feasible for characterizing small quantum systems. Indeed, a possible concern is that QPI appears to require much more experimental data than QPT, which already requires a moderately large amount of data. To some extent this is to be expected: whereas QPT needs only enough information to produce a 4-dimensional model for a qubit process, QPI needs enough information to first of all determine the effective size of the environment’s memory and then determine all the dynamical parameters involving the qubit, the pertinent part of the environment, and their interaction. We expect, however, that QPI can be made much more efficient than the studies above would suggest. In these studies, all measurements were arbitrarily chosen to be performed the same number of times. Almost certainly some measurements are more informative than others. We expect that an adaptive testing strategy, in which the current model estimate is used to choose the most informative experiment to perform next, will significantly reduce the cost of QPI [44, 45, 76].

As presented here, QPI has two main limitations. The first is that it characterizes only one process. In the context of a quantum computer, one would like a self-consistent characterization of all qubit operations (‘gates’) that may be performed; under the assumption of Markovianity, GST provides just such a characterization. As formulated here, QPI would yield a set of separate characterizations, one for each gate, that cannot be related. More specifically, the relationship between latent variables in the models for different gates would be indeterminate, i.e. one would not know whether different gates interact with the same or different environmental degrees of freedom. We conjecture that QPI can be extended to jointly characterize a set of processes: The principles of the Ho–Kalman method extend in a straightforward way to this case. What is unclear at present is how to extend lemma 1, i.e. how to identify a relatively small set of operation sequences that are likely to yield a complete model. This is an obvious direction for future work.

In principle, extending QPI to multiple processes would allow cross talk and other spatial correlations to be characterized by treating different combinations of simultaneous gates as different processes. However, due to the very large number of different combinations this would constitute (even in a small device), this does not seem to be a practical approach to characterizing spatial correlations.

The other main limitation of QPI as presented here is that it assumes all measurements are final; no attempt is made to model the effect of the measurement on the state of the system or environment. But intermediate measurements are a key component of quantum error correction protocols, thus the ability to characterize their impact on the remainder of a computation is important. We suspect the ability to characterize intermediate measurements would follow from the ability to characterize multiple processes, since each measurement outcome can be treated as a distinct, randomly selected process.

As a final remark, it may be acknowledged that QPI models are abstract and often not manifestly insightful. However, capturing empirical behavior in a concise and accurate mathematical model, as QPI does, is an essential first step toward understanding. Methods of extracting further insight from QPI models would be an interesting topic for future work.

7. Conclusion

In conclusion, we have developed a new method for the characterization of quantum dynamical processes, called QPI. QPI goes beyond state-of-the-art methods by providing the means to systematically characterize non-Markovian dynamics as well as Markovian dynamics. We presented a detailed experimental protocol and data analysis procedure and demonstrated their effectiveness using numerical simulations of realistic non-Markovian error processes affecting qubits. Directions for future work include extending QPI to enable the characterization of multiple processes and non-final measurements, and using adaptive strategies to reduce its experimental cost. Apart from experimental applications, the theory underlying QPI elucidates the relationship between the temporal evolution of an open quantum system and its effective environment, providing a useful alternative to existing representations of non-Markovian quantum dynamics.

Acknowledgments

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for
publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy. The authors would like to express appreciation to Nicholas Peters for helpful comments on the preparation of this manuscript.

Appendix A. Proof of lemma 1

In this section we show that the minimum number of time observations needed to guarantee complete characterization of a discrete-time linear system depends on the number of latent (unmeasured) degrees of freedom, not the total number of degrees of freedom. Let \( d \) be the true dimension of the system and let \( a \) be the dimension of the subspace spanned by the initial states and measured quantities. We show that \( \text{rank}(H_{k+1}) = d \) where \( l = d - a \) and \( H \) is the block-Hankel matrix formed from \( X(1), \ldots, X(2l - 1) \). It follows from the Ho–Kalman theory that \( X(1), \ldots, X(2l + 1) \) are sufficient to obtain a state model of the system.

The goal is to find matrices \( L, M, R \) such that \( X(k) = LM^{-1}R \). In the remainder of this section, \( X(k) \) will be denoted more compactly as \( X_k \). For convenience, we assume that redundant initial states and redundant measurements have been omitted, so that \( X(k) = a \times a \). Since rank \( X_i \leq a \) for every \( X_i \), we may choose a representation in which \( L \) has the form \( L = \begin{bmatrix} 1_{a \times a} & 0_{a \times l} \end{bmatrix} \). It follows that \( R = \begin{bmatrix} X_0^T \end{bmatrix} \) where \( Y_0 \) is some unknown \( l \times a \) matrix. We write \( M \) as

\[
M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{A1}
\]

where \( A \) is \( a \times a \), \( D \) is \( l \times l \), and \( B \) and \( C \) are sized correspondingly. We have

\[
\begin{bmatrix} X_n \\ Y_n \end{bmatrix} = M^n \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X_{n-1} \\ Y_{n-1} \end{bmatrix} \quad \tag{A2}
\]

Then

\[
X_{n+1} = AX_n + BY_n \tag{A4}
\]

\[
= AX_n + B(CX_{n-1} + DY_{n-1}) \tag{A5}
\]

\[
\begin{array}{c}
\vdots \\
\end{array}
\]

\[
= AX_n + \sum_{i=1}^{l} BD^{i-1}CX_{n-i} + BD^{l}Y_{n-l} \tag{A7}
\]

or

\[
X_{n+k+1} = AX_{n+k} + \sum_{i=1}^{l} BD^{i-1}CX_{n+i-1} + BD^{l}Y_{n} \quad \tag{A8}
\]

for \( k = 0, 1, 2, \ldots, n \). Let \( k = l \). Our goal is to eliminate \( Y_n \) and obtain a recurrence relation involving only the \( X_j \)’s. By the Cayley–Hamilton theorem there are coefficients \( a_0, \ldots, a_{l-1} \) such that \( D^l = \sum_{j=0}^{l-1} a_j D^j \). Thus

\[
X_{n+l+1} = AX_{n+l} + \sum_{i=1}^{l} BD^{i-1}CX_{n+i-1} + \sum_{j=0}^{l-1} a_j BD^jY_n. \quad \tag{A9}
\]

Setting \( k = 0 \) in (A8) yields

\[
X_{n+j+1} = AX_{n+j} + \sum_{i=1}^{j} BD^{i-1}CX_{n+j-i} + BD^jY_n. \quad \tag{A10}
\]

Solving for \( BD^jY_n \) and substituting into (A9) gives

\[
X_{n+j+1} = AX_{n+j} + \sum_{i=1}^{j} BD^{i-1}CX_{n+j-i} + \sum_{j=0}^{l-1} a_j \left( X_{n+j+1} - AX_{n+j} - \sum_{i=1}^{l} BD^{i-1}CX_{n+i-1} \right) \quad \tag{A11}
\]
\[ AX_{n+1} + \sum_{i=1}^{l} \left( 1 - \sum_{j=i}^{l-1} \alpha_j \right) BD^{i-1} CX_{n+1-i} + \sum_{j=0}^{l-1} \alpha_j (X_{n+j+1} - AX_{n+j}). \] (A12)

The important thing to note about this last equation is that \( X_{n+1} \) is written in terms of \( X_n, \ldots, X_{n+l} \). More precisely, there are matrices \( Q_l, \ldots, Q_1 \) such that

\[ X_{n+l+1} = \sum_{i=0}^{l} Q_i X_{n+i}. \] (A13)

In other words, the rows of \( X_{n+l+1} \) are linear combinations of the rows of \( X_n, \ldots, X_{n+l} \). (Note that if the \( Q_i \)'s were scalars instead of matrices, the effective dimension of the system would be at most \( l + 1 \).) By extension,

\[ T_{n+l+1} = \sum_{i=0}^{l} Q_i T_{n+i} \] (A14)

where \( T_k \equiv [ X_k \ X_{k+1} \ \cdots ] \) is the \( k \)th block of rows of \( H_\infty \). Thus \( T_0, \ldots, T_{l+1} \) span \( T_{l+1} \) and, by induction, all rows of \( H_\infty \).

Now, we could just as well have chosen a representation in which \( Y_n \) is replaced by \( Z_n \). Here we derive a reliable criterion for determining which singular values are small enough to be considered spurious.

A naive approach would be to set a threshold for each singular value—say, based on its estimated variance—and count the number of singular values that exceed their threshold. There are problems with this approach, however, the most serious being that the number of singular values that exceed their threshold simply by chance grows with the size of \( \tilde{H} \), i.e. grows with the number of experiments performed. A way to overcome this problem is to test singular values collectively rather than individually: Given a threshold \( \sigma \) for statistical significance of the \( k \) smallest singular values, one finds the largest \( k \) for which the threshold is not met and takes \( d \) to be the number of singular values that remain.

Appendix B. Dimension estimation

In this section we consider the problem of estimating the dimension \( d \) of a process from an experimental estimate \( \tilde{H} \) of a Ho–Kalman matrix \( H \). Ho and Kalman showed if \( H \) is sufficiently large, then \( d \) is just the rank of \( H \). But \( \tilde{H} \), which is a random perturbation of \( H \), has a rank that is generally much higher than \( d \). The singular values of \( \tilde{H} \) are much more informative than the rank, as they reveal the magnitude of each dimension’s contribution to the data. If the statistical errors in the data are small, all but \( d \) singular values of \( \tilde{H} \) will be small. Here we derive a reliable criterion for determining which singular values are small enough to be considered spurious.

A naive approach would be to set a threshold for each singular value—say, based on its estimated variance—and count the number of singular values that exceed their threshold. There are problems with this approach, however, the most serious being that the number of singular values that exceed their threshold simply by chance grows with the size of \( \tilde{H} \), i.e. grows with the number of experiments performed. A way to overcome this problem is to test singular values collectively rather than individually: Given a threshold \( \sigma \) for statistical significance of the \( k \) smallest singular values, one finds the largest \( k \) for which the threshold is not met and takes \( d \) to be the number of singular values that remain.

An appropriate threshold can be derived from the uncertainty in the raw data. Each quantity \( F_{i,m}^{(t)} \) that appears in \( \tilde{H} \) is an independent random variable, distributed binomially with mean \( F_{i,m}^{(t)} \) and variance \( F_{i,m}^{(t)^2} / X_{i,m}^{(t)} \). Let \( \epsilon_{i,t,m} = F_{i,m}^{(t)} - F_{i,m}^{(t)} \) denote the statistical deviation of \( F_{i,m}^{(t)} \) and let \( \partial_{i,t,m} H = \partial H / \partial F_{i,m}^{(t)} \). Then
\[ \hat{H} = H + \sum_x (\partial_x H) \epsilon_x, \]  

(B1)

where \( x \) ranges over all experiments \((i, t, m)\). Let \( h_1 \times h_2 \) be the size of \( H \) and let \( h = \min(h_1, h_2) \). Let \( H = U S V^\dagger \) be a 'full' singular value decomposition of \( H \), i.e. \( U \) is \( h_1 \times h_1 \) and \( V \) is \( h_2 \times h_2 \), and \( S \) is \( h_1 \times h_2 \) with diagonal elements \( s_1 \geq s_2 \geq \cdots \geq s_h \). Let \( U_k \) be the matrix consisting of columns \( k \) through \( h_1 \) of \( U \), and \( V_k \) the matrix consisting of columns \( k \) through \( h_2 \) of \( V \).

Suppose \( H \) has rank \( r \). Then \( s_1, \ldots, s_r = 0 \) and \( s_{r+1} = \cdots = s_h = 0 \). According to perturbation theory, the corresponding singular values \( \tilde{s}_{r+1}, \ldots, \tilde{s}_h \) of \( \hat{H} \) are the singular values of

\[ A \equiv U_{r+1}^\dagger \hat{H} V_{r+1} \]

(B2)

where

\[ A_x \equiv U_{r+1}^\dagger (\partial_x H) V_{r+1}. \]

(B5)

Let \( \chi_r \) denote the residual 'energy' of the \( h - r \) smallest singular values,

\[ \chi_r = \sum_{i>r} \tilde{s}_i^2 = \text{Tr}(A^2) \]

(B6)

\[ = \sum_{i,j} \text{Tr}(A^2) \epsilon_i \epsilon_j = \sum_{i,j} \text{Tr}(A_i^T A_j) \epsilon_i \epsilon_j. \]

(B7)

Owing to the independence of experiments, we have \( \langle \epsilon_x \epsilon_y \rangle = 0 \) for \( x \neq y \), yielding

\[ \langle \chi_r \rangle = \sum_{i,j>r} \sum_x B_{ij}^{(x)} \]

where \( B_{ij}^{(x)} \equiv A_{ij} \sqrt{\langle \epsilon_x^2 \epsilon_y^2 \rangle} \).

If the threshold for rejecting \( \tilde{s}_r, \ldots, \tilde{s}_h \) were set at \( \langle \chi_r \rangle \), then approximately half of the time the subset \( \tilde{s}_{r+1}, \ldots, \tilde{s}_h \) would be incorrectly accepted as statistically significant, resulting in an overestimate of \( d \). A higher threshold reduces the probability of this occurring, but increases the probability of rejecting \( \tilde{s}_1, \ldots, \tilde{s}_h \) and thereby underestimating \( d \). Thus the threshold should be just high enough to reject \( \tilde{s}_{r+1}, \ldots, \tilde{s}_h \) a majority of the time. To that end we calculate the variance of \( \chi_r \). We have

\[ \chi_r^2 = \text{Tr}(A^2) = \sum_{x} \sum_{y \neq x} \text{Tr}(A_x^T A_y) \epsilon_x \epsilon_y. \]

(B8)

Since the deviations are independent and zero-mean, \( \langle \epsilon_x \epsilon_y \epsilon_f \epsilon_q \rangle \) vanishes unless it is of the form \( \langle \epsilon_x^2 \epsilon_y^2 \rangle \) or \( \langle \epsilon_x^2 \epsilon_y^2 \rangle \). The second form simplifies to \( \langle \epsilon_x^2 \rangle \langle \epsilon_y^2 \rangle \). Provided the statistical errors are small compared to the quantities being estimated, \( \epsilon_x \) has an approximately normal distribution, for which \( \langle \epsilon_x^2 \rangle = 3 \langle \epsilon_x^2 \rangle \). Then

\[ \langle \chi_r \rangle = 3 \sum_x \text{Tr}(A_x^2) \langle \epsilon_x^2 \rangle = 2 \sum_{x \neq y} \text{Tr}(A_x^T A_y) \langle \epsilon_x \epsilon_y \rangle + 2 \text{Tr}(A_x^T A_x \langle \epsilon_x^2 \rangle) \langle \epsilon_x^2 \rangle \]

\[ = \left( \sum_{x} \text{Tr}(A_x^2) \langle \epsilon_x^2 \rangle \right)^2 = 2 \sum_{x \neq y} \text{Tr}(A_x^T A_y) \langle \epsilon_x^2 \rangle \langle \epsilon_y^2 \rangle \]

(B9)

where

\[ D_{i,j,k,l} \equiv \sum_x B_{ij}^{(x)} B_{kl}^{(x)}. \]

Our criterion for accepting the hypothesis \( s_{r+1} = s_{r+2} = \cdots = 0 \) is

\[ \chi_r \leq \langle \chi_r \rangle + \sqrt{\text{Var} \chi_r}, \]

(B11)
\[
\begin{align*}
= \sum_{i,j>r} \sum_{x} B_{ij}^{(x)} + \sqrt{2 \sum_{i,j>r} \sum_{k,l>r} \left( \sum_{x} B_{ij}^{(x)} B_{kl}^{(x)} \right)^2}.
\end{align*}
\]

We take \( d \) be the smallest \( r \) for which this criterion is met. In theory the probability of overestimating \( d \) is then at most 0.16.

In practice this criterion cannot be applied exactly: The quantities on the right side of equation (B12) are not known, but can only be estimated. The singular values may not even be in quite the right order due to statistical variations. Furthermore, the approximations made in the derivation might not always be accurate. Nevertheless, in our simulations we observed that the criterion (B12) is effective: It usually yields an estimate of \( d \) quite close to the true value, provided one has taken enough data.

**Appendix C. Progressive fitting of data blocks**

For large times \( t \), the experimental quantity \( F^{(i)} = S^TP \) is an extremely nonlinear function of \( T \), presenting a considerable challenge for estimation of the parameters \( S, T, P \) from the data. Building on the approach described in [7], we devised a progressive fitting method which first uses low-\( t \) data to obtain a coarse estimate of \( T \), then gradually incorporates data with higher \( t \) to increase the accuracy of the model. Our implementation involves multiple passes over the data, alternating between optimization of \( T \) given the current estimate of \( S, P \) and optimization of \( S, P \) given the current estimate of \( T \). We found that if too much or too little data was used at any given step, either \( S, P \) or \( T \) could ‘lock in’ to the wrong region of parameter space from which the search could not recover. The procedure below was carefully honed to ensure that at each step, only reliable intermediate estimates were used. In our studies, 5–15 passes were usually sufficient to obtain good fits to the data.

Let \( H^{(b)} \) denote the \( b \)th block of data (figure 3(c)), \( W^{(b)} \) the matrix of corresponding statistical weights, and \( N_b \) the number of elements in blocks 0 through \( b \). (The weights \( W^{(b)} \) are adjusted to account for the different multiplicity of different experiments, so that each experiment has the same total weight.) Let

\[
\Theta_k = \begin{cases} 0 & k = 0, \\ 2^{k-1} & k > 0. \end{cases}
\]

The function to be minimized is

\[
\Phi_b(A, T, B) \equiv \frac{1}{N_b} \sum_{b'=0}^{b} \sum_{i \in T} \sum_{m \in M} W^{(b')}_{im} (A^T \varphi^b B - H^{(b')} y^b_m)^2
\]

which is the average weighted error of the model over blocks 0 through \( b \). Under the true model, the distribution of \( \Phi_b \) has a strong peak at 1. A value significantly larger than this (say, 1.5) indicates a poor fit. Model fitting with the correct dimension usually yields \( \Phi_b \) values around 0.8; overfitting leads to \( \Phi_b \) around 0.5.

Let \( b_{\text{max}} \) denote the index of the last block. Let \( b_{\text{highest}} \) denote the index of the highest block for which a fit has been attempted. The fitting procedure is as follows:

3.1 Let \( (L, T, R) \) be the outputs of Ho–Kalman estimation. Set \( A \) equal to \( L \) and set \( B = \) equal to the first \( n(l + 1) \) columns of \( R \). Find the smallest \( b \) such that \( \Phi_b > 1.5 \) and set \( b_{\text{highest}} \) to this value (or to \( b_{\text{max}} \) if there is no such \( b \)).

3.2 Keeping \( T \) fixed, optimize \( A, B \). This is accomplished by minimizing a modified form of equation (C2) with \( b = b_{\text{max}} \). Whereas the current estimate of \( T \) is used for blocks \( b' < b_{\text{highest}} \) for blocks \( b' > b_{\text{highest}} \), we replace \( T \varphi^b \) by the matrix \( Y^{(b)} \) that is optimal with respect to \( A, B \). In this way, \( A, B \) are made to fit all the data but are not constrained by high powers of \( T \) that have not yet been determined to be accurate. Minimization can be accomplished via iterative linear algebraic methods.

3.3 Keeping \( A, B \) as fixed, optimize \( T \).

3.3.1 Set \( b = 2 \).

3.3.2 If \( b < b_{\text{highest}} \) and \( \Phi_b < 1.5 \), continue to 3.3.3. Otherwise, find \( T \) that minimizes \( \Phi_b \), keeping \( A, B \) fixed. This can be accomplished by various nonlinear optimization methods; we found the Gauss–Newton method with line search to work well.

3.3.3 If \( b < b_{\text{max}} \) and \( \Phi_b < 1.5 \), increase \( b \) by 1, set \( b_{\text{highest}} = \max(b, b_{\text{highest}}) \), and go back to 3.3.2.

3.4 If \( b = b_{\text{max}} \) and \( \Phi_{b_{\text{max}}} \) did not improve significantly (say, by more than 0.001) compared to the previous pass, optimization has converged; continue to step 3.5. Otherwise, go back to step 3.2 (perform another pass).
3.5 If $\Phi_{b_{new}} \leq 1.5$, the fit is deemed a success; in this case return $T$, return the first $m$ rows of $A$ for $S$, and return the first $n$ columns of $B$ for $P$. Otherwise, increase $d$ by one, perform Ho–Kalman estimation again (stage 2, described in section 4.2), and start again at step 3.1.

Appendix D. Final model optimization

The fourth and final stage of our model inference procedure is to fit the model $F^{(t)} = STP$ directly to the data. This last stage yields the most accurate models but requires a very good starting estimate $S$, $T$, $P$. As in the previous stage, we minimize a sum of weighted squared errors. But unlike in previous stages, the weights are determined by the model itself rather than estimated from the data, making the total squared error a good proxy for the negative log likelihood. Besides being simpler to work with than the log likelihood, weighted errors allow out-of-range probabilities (less than 0 or greater than 1) to be handled much more gracefully. In this stage we also impose soft constraints on the eigenvalues $\{\lambda_i\}$ of $T$: Since the process in question is presumed to occur in a finite state space, the eigenvalues of $T$ should not have magnitude greater than 1, as otherwise repeated application of the process would eventually take the state out of the state space.

The function to be minimized is

$$
\Psi(S, T, P) \equiv \frac{1}{|T||A||T^T|} \sum_{i \in I} \sum_{t \in T} \sum_{m \in M} W_{i,m}^{(t)}(S, T, P)(F^{(t)} - F^{(t)})^2 + E(T). \tag{D1}
$$

where $E(T) = \sum_{i,m} \max(0, |\lambda_i| - 1)^2$ is the eigenvalue constraint. Nominally, $W_{i,m}^{(t)}$ is the inverse variance of $F_{i,m}^{(t)}$. But since the inverse variance becomes negative or infinite if the model ever predicts $F_{i,m}^{(t)} \leq 0$ or $F_{i,m}^{(t)} \geq 1$, we take

$$
W_{i,m}^{(t)} = \frac{1}{V_{i,m}^{(t)} + \frac{1}{N_{i,m}^{(t)} + 2\beta_{i,m}^{(t)^2}}},
$$

where

$$
V_{i,m}^{(t)} = \frac{F_{i,m}^{(t)} - F_{i,m}^{(t)}}{N_{i,m}^{(t)}},
$$

is the theoretical variance of $F_{i,m}^{(t)}$ and $\beta_{i,m}^{(t)}$ is a small buffer term that ensures $W_{i,m}^{(t)}$ remains finite, continuous, and positive. We start with $\beta_{i,m}^{(t)} = 1/N_{i,m}^{(t)}$. We then proceed to minimize $D1$ using the Gauss–Newton method with line search. After each search step the validity of the model predictions are assessed. If any $V_{i,m}^{(t)}$ is not in the interval [0, 1], $\beta_{i,m}^{(t)}$ is reduced by 5%. This increases $W_{i,m}^{(t)}$ outside [0,1] and near its endpoints, thereby more strongly encouraging $F_{i,m}^{(t)}$ to become valid. If after a search step all the predicted probabilities are valid and $\Psi$ was not significantly improved, the search concludes.

References

[1] Scarani V, Bechmann-Pasquinucci H, Cerf N J, Dušek M, Lütkenhaus N and Peev M 2009 Rev. Mod. Phys. 81 1301
[2] Kelly J et al 2014 Phys. Rev. Lett. 112 240504
[3] Dehollain J P et al 2016 New J. Phys. 18 103018
[4] Sheldon S, Bishop L S, Magesan E, Filipp S, Chow J M and Gambetta J M 2016 Phys. Rev. A 93 012301
[5] Blume-Kohout R, Gamble J K, Nielsen E, Rudinger K, Mizrahi J, Fortier K and Maunz P 2017 Nat. Commun. 8 14485
[6] Nielsen M A and Chuang I L 2011 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press)
[7] Blume-Kohout R J, Gamble J K, Nielsen E, Maunz P L W, Scholten T L and Rudinger K M 2015 Turbocharging quantum tomography Technical Report Sandia National Laboratories
[8] Magesan E, Gambetta J M and Emerson J 2012 Phys. Rev. A 85 042311
[9] Magesan E et al 2012 Phys. Rev. Lett. 109 080505
[10] Gambetta J M et al 2012 Phys. Rev. Lett. 109 240504
[11] Gaebler J P et al 2012 Phys. Rev. Lett. 108 260503
[12] Kimmel S, da Silva M P, Ryan C A, Johnson B R and Ohki T 2014 Phys. Rev. X 4 011050
[13] Chasseur T and Wilhelm F K 2015 Phys. Rev. A 92 042333
[14] Cross A W, Magesan E, Bishop L S, Smolin J A and Gambetta J M 2016 NPJ Quantum Inf. 2 16012
[15] Buscemi F and Datta N 2016 Phys. Rev. A 93 012101
[16] Breuer H-P, Laine F-M, Piló J and Vacchini B 2016 Rev. Mod. Phys. 88 021002
[17] de Vega I and Alonso D 2017 Rev. Mod. Phys. 89 015001
[18] Bellomo B, Pasquale A D, Gualdi G and Marzoño U 2010 J. Phys. A: Math. Theor. 43 395303
[19] Cerrillo J and Cao J 2014 Phys. Rev. Lett. 112 110401
[20] Pollock F A, Rodriguez-Rosario C, Frauenheim T, Paternostro M and Modi K 2018 Phys. Rev. A 97 012127
[21] De Schutter B 2000 J. Comput. Appl. Math. 121 331
[22] Heij C, Ran A C M and Schagen F V 2007 Introduction to Mathematical Systems Theory (Boston: Birkhäuser) (https://doi.org/10.1007/978-3-7643-7549-2)
[23] Wallman J and Flammini S T 2014 New J. Phys. 16 103032
[24] Epstein M, Cross A W, Magesan E and Gambetta J M 2014 Phys. Rev. A 89 062321
[25] Ball H, Stace T M, Flammini S T and Biercuk M J 2016 Phys. Rev. A 93 022303
[26] Wood C J and Gambetta J M 2018 Phys. Rev. A 97 032306
[27] Zhang J and Sarovar M 2015 Phys. Rev. A 91 052121
[28] Kretschmann D and Werner R F 2005 Phys. Rev. A 72 062323
[29] Rybár T and Žiman M 2015 Phys. Rev. A 92 042315
[30] Monras A, Beige A and Wiesner K 2010 arXiv:1002.2337v2
[31] O’Neill B, Barlow T M, Šafářík D and Beige A 2012 AIP Conf. Proc. 1479 667
[32] Monras A and Winter A 2016 J. Math. Phys. 57 015219
[33] Cholewa M, Gawron P, Glomb P and Kurzyk D 2017 Quantum Inf. Process. 16 101
[34] Burgarth D and Yuasa K 2012 Phys. Rev. Lett. 108 080502
[35] Chitambar E, Abu-Nada A, Cabellos-R and Byrd M 2015 Phys. Rev. A 92 052110
[36] Dominy J M and Lidar D A 2016 Quantum Inf. Process. 15 1349
[37] Pechukas P 1994 Phys. Rev. Lett. 73 1060
[38] Dive B, Mintert F and Burgarth D 2015 Phys. Rev. A 92 032111
[39] Luchnikov I, Vintskevich S, Ouerdane H and Filippov S 2019 Phys. Rev. Lett. 122 160401
[40] Mohseni M and Lidar D A 2006 Phys. Rev. Lett. 97 170501
[41] Aaronson S 2007 Proc. R. Soc. A 463 3089
[42] Bisio A, Chiribella G, D’Ariano G M, Facchini S and Perinotti P 2010 Phys. Rev. A 81 032324
[43] da Silva M P, Landon-Cardinal O and Poulin D 2011 Phys. Rev. Lett. 107 210404
[44] Huszár F and Houlsby N M T 2012 Phys. Rev. A 85 052120
[45] Mahler D H, Rozema L A, Darabí A, Ferrie C, Blume-Kohout R and Steinberg A M 2013 Phys. Rev. Lett. 111 183601
[46] Lee J Y and Landon-Cardinal O 2015 Phys. Rev. A 91 062128
[47] Howland G A, Knarr S H, Schneebock J, Cum D J and Howell J C 2016 Phys. Rev. X 6 021018
[48] Dumitrescu E and Humble T S 2016 Phys. Rev. A 94 042107
[49] Haah J, Harrow A W, Ji Z, Wu X and Yu N 2017 IEEE Trans. Inf. Theory 63 5628–41
[50] Carleo G and Troyer M 2017 Science 355 602
[51] Shrapnel S, Costa F and Milburn G 2018 Int. J. Quantum Inf. 16 1840010
[52] Luchnikov I A, Vintskevich S V, Grigoriev D A and Filippov S N 2019 arXiv:1902.07019v2
[53] Wohrer S, Christandl M and Doherty A C 2008 Phys. Rev. A 78 062112
[54] Gallego R, Brunner N, Hadley C and Acín A 2010 Phys. Rev. Lett. 105 230501
[55] Ahrens J, Rudzigg P, Cabello A and Bourennane M 2012 Nat. Phys. 8 592
[56] Brunner N, Navascues M and Vértesi T 2013 Phys. Rev. Lett. 110 150501
[57] D’Ambrosio V, Bisio F, Scarinno F, Barra J F, Lima G and Cabello A 2014 Phys. Rev. Lett. 112 140503
[58] Costa F and Shrapnel S 2016 New J. Phys. 18 063032
[59] Hardy L 2001 arXiv:quant-ph/0101012v4
[60] Barrett J 2007 Phys. Rev. A 75 032304
[61] Masanes L and Müller M P 2011 New J. Phys. 13 063001
[62] Addis C, Bylicka B, Chruściński D and Maniscalco S 2014 Phys. Rev. A 90 052103
[63] Guarnieri G, Smirne A and Vacchini B 2014 Phys. Rev. A 90 022310
[64] Taranto P, Pollock F A, Mile S, Tomamichel M and Modi K 2019 Phys. Rev. Lett. 122 140401
[65] Budini A A 2013 Phys. Rev. A 88 032115
[66] Stinespring W F 1955 Proc. Am. Math. Soc. 6 211
[67] Wood C J, Biamonte J D and Cory D G 2015 Quantum Info. Comput. 15 759
[68] Proctor T, Rudinger K, Young K, Sarovar M and Blume-Kohout R 2017 Phys. Rev. Lett. 119 130502
[69] Ho B L and Kalman R E 1966—“Automatisierungs technik” 14 545–8
[70] Akae H 1974 IEEE Trans. Autom. Control 19 716
[71] Buntine W and Hutter M 2010 arXiv:1007.0296v2
[72] Zeiger H and McEwen A 1974 IEEE Trans. Autom. Control 19 153
[73] Barnes J P, Trout C J, Lucarelli D and Cramer B D 2017 Phys. Rev. A 95 062338
[74] Gambetta J M, Motzoi F, Merkel S T and Wilhelm F K 2011 Phys. Rev. A 83 012308
[75] Lougovski P and Peters N A 2018 Phys. Lett. A 382 2113
[76] Grenade C, Ferrie C and Cory D G 2015 New J. Phys. 17 013042