

User-specified random sampling of quantum channels and its applicationsJun Yan Sim,^{1,*} Jun Suzuki^{2,†} Berthold-Georg Englert^{1,3,4,‡} and Hui Khoon Ng^{5,1,4,§}¹*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*²*Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan*³*Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551, Singapore*⁴*MajuLab, International Joint Research Unit UMI 3654, CNRS, Université Côte d'Azur, Sorbonne Université, National University of Singapore, Nanyang Technological University, Singapore*⁵*Yale-NUS College, 16 College Avenue West, Singapore 138527, Singapore*

(Received 17 July 2019; revised manuscript received 4 October 2019; accepted 10 January 2020; published 7 February 2020)

Random samples of quantum channels have many applications in quantum information processing tasks. Due to the Choi-Jamiołkowski isomorphism, there is a well-known correspondence between channels and states, and one can imagine adapting *state* sampling methods to sample quantum channels. Here, we discuss such an adaptation, using the Hamiltonian Monte Carlo method, a well-known classical method capable of producing high-quality samples from arbitrary, user-specified distributions. Its implementation requires an exact parametrization of the space of quantum channels, with no superfluous parameters and no constraints. We construct such a parametrization, and demonstrate its use in three common channel sampling applications.

DOI: [10.1103/PhysRevA.101.022307](https://doi.org/10.1103/PhysRevA.101.022307)**I. INTRODUCTION**

Quantum channels, or completely positive (CP) and trace-preserving (TP) maps, are a central concept in describing the dynamics of quantum systems. They form the basic models for imperfect quantum operations used for quantum information processing (QIP). Random—according to some specified distribution—samples of quantum channels are needed in many QIP tasks, including the evaluation of the distributional average of channel-related quantities, the computation of error bars for quantum process tomography, the exploration of typical properties of quantum channels, the numerical optimization of functions of channels over a complicated landscape, and others.

Sampling from specific distributions over the quantum *state* space is a well-studied problem, with many different approaches, including the Monte Carlo (MC) technique for arbitrary distributions, and other methods for sampling from specific distributions [1–8]. Due to the Choi-Jamiołkowski isomorphism [9,10], which gives a correspondence between CP channels and states, these state sampling methods can be adapted to sample quantum channels. Indeed, in the recent work by Tinh *et al.* [11], a Metropolis-Hasting Markov chain (MHMC) MC approach was used to sample channels from arbitrary distributions, by sampling the purification of the Choi-Jamiołkowski state corresponding to the channel. References [12,13] discuss a procedure for generating samples of quantum channels with a specific distribution on the channel

space by making use of the channel-state correspondence. An alternative way of generating the same distribution of channels is to couple the input state and an ancilla initially in an arbitrary pure state by a Haar-random unitary operator and then taking the partial trace over the ancilla. See Ref. [14] for more discussion on the properties of the distribution of channels generated by these two procedures.

As a general method for sampling from arbitrary distributions, MC methods stand out in their wide-ranging applicability and efficiency. The MHMC variety of MC methods used, for example, in Ref. [11], however, suffer from strong correlations between sample points, and one requires large samples for reliable answers not biased by these correlations. This was observed, for instance, in the MHMC state sampling algorithm of Ref. [1]. A significant improvement in the quality of the samples was seen when we switched to the Hamiltonian Monte Carlo (HMC) approach [2], reaffirming the advantage of HMC over MHMC MC also observed in other settings [15–19].

The HMC method requires the availability of a parametrization of the domain space with exactly the right number of parameters, with no superfluous parameters and no constraints. The parametrization of the channel and state space used in Ref. [11], which has superfluous parameters, cannot be used for HMC. The exact parametrization of states used in the HMC algorithm in Ref. [2] gives, through the Choi-Jamiołkowski isomorphism, a parametrization of the set of all CP, but not necessarily TP, maps. The TP property has to be imposed as an explicit constraint, thus rendering the parametrization unsuitable in a HMC algorithm for sampling CPTP channels.

In this work, we construct one exact parametrization of the space of CPTP maps, with no superfluous parameters, and no constraints. This can then be used in a HMC

*e0012429@u.nus.edu

†junsuzuki@uec.ac.jp

‡cqtebg@nus.edu.sg

§huikhooon.ng@yale-nus.edu.sg

procedure for sampling from arbitrary, user-specified, distributions over the channel space. To illustrate the usefulness of our parametrization and the HMC algorithm, we apply our methods to three quantum sampling problems. Our examples are focused on problems in quantum process tomography, reflecting the interests of the authors; our parametrization and the HMC method, however, are just as useful for sampling problems in other areas of QIP. As an aside, our construction exactly parametrizes the space of all bipartite mixed quantum states with the completely mixed state for one of the parties.

Here is the brief outline of our paper. We first review the Choi-Jamiołkowski isomorphism in Sec. II. Section III explains our main contribution: the exact parametrization of the space of CPTP channels. In Sec. IV, we illustrate the use of our parametrization in a HMC sampling algorithm through three examples from quantum process tomography: (a) the construction of error regions in process estimation; (b) marginal likelihood for estimating specific properties of the channel; (c) model selection among candidate channel families. The reader is referred to Ref. [2] or Appendix A for an introduction to the HMC algorithm used here. We conclude in Sec. V.

II. CHANNEL-STATE DUALITY

There are many ways of writing the CPTP map of a quantum channel. Given our desire to make the connection with the sampling of quantum states, we make use of the channel-state duality and describe the quantum channel by a state via the Choi-Jamiołkowski isomorphism. Here, we remind the reader of this isomorphism, and, in the process, define the notation used throughout the article.

We begin with the d -dimensional Hilbert space \mathcal{H} describing the state vectors (pure states) of the system. We define a map $*$: $\mathcal{H} \rightarrow \mathcal{H}$,

$$*(|\psi\rangle) \equiv |\bar{\psi}\rangle \in \mathcal{H} \quad \text{for } |\psi\rangle \in \mathcal{H}, \quad (1)$$

such that

$$\langle \bar{\psi} | \bar{\phi} \rangle = \langle \phi | \psi \rangle \quad \forall |\psi\rangle, |\phi\rangle \in \mathcal{H}, \quad (2)$$

and $*$ is “*-linear,” i.e.,

$$*\left(\sum_i c_i |\psi_i\rangle\right) = \sum_i c_i^* |\bar{\psi}_i\rangle, \quad (3)$$

where c_i^* is the complex conjugate of c_i . Note that Eq. (2) specifies the $*$ map only up to a unitary transformation of no consequence. One specific realization of the $*$ map, and what we use in our numerical examples below, is to first pick a basis $\{|i\rangle\}$ on \mathcal{H} , define $|\bar{i}\rangle \equiv |i\rangle$, and then extend the action of $*$ to arbitrary vectors using the *-linearity property. See also Sec. 3.1 in Ref. [20] for qubit examples of the $*$ map.

We extend the action of the $*$ map to adjoint vectors $*(\langle\psi|) = \langle\bar{\psi}| = (\langle\psi|)^\dagger = [*(|\psi\rangle)]^\dagger$, and further to the set of operators on \mathcal{H} , denoted as $\mathcal{B}(\mathcal{H})$,

$$*\left(\sum_{ij} c_{ij} |\psi_i\rangle\langle\phi_j|\right) \equiv \sum_{ij} c_{ij}^* |\bar{\psi}_i\rangle\langle\bar{\phi}_j|. \quad (4)$$

We write $*(X) \equiv \bar{X}$, for any $X \in \mathcal{B}(\mathcal{H})$. Note that $\bar{\bar{X}} = X$, and we denote $X^T \equiv \bar{X}^\dagger$, a basis-independent transpose operation. If X is non-negative, then so is X^T .

Using the $*$ map, we define the *vectorization map*, a linear map from operators to vectors in a vector space \mathcal{V} , $\text{vec}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{V}$,

$$\text{vec}(|\psi\rangle\langle\phi|) \equiv *(|\phi\rangle) \otimes |\psi\rangle = |\bar{\phi}\rangle \otimes |\psi\rangle = |\bar{\phi}\psi\rangle, \quad (5)$$

for any $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ and extended to all operators by linearity. We write, for any $X \in \mathcal{B}(\mathcal{H})$, $\text{vec}(X) \equiv |X\rangle \in \mathcal{V}$. Note the useful identity

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B). \quad (6)$$

Also, if $\{|i\rangle\}$ is an orthonormal basis for \mathcal{H} , then so is $\{|\bar{i}\rangle\}$. Consequently, the vectorized identity operator $|\mathbb{1}\rangle = \text{vec}(\mathbb{1}) = \sum_{i=1}^d |\bar{i}\rangle|i\rangle$ can be regarded as a bipartite maximally entangled (unnormalized) state on $\mathcal{H} \otimes \mathcal{H}$.

Now, we are ready to state the channel-state duality. Consider a CP map $\mathcal{E}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{H})$, acting as $\mathcal{E}(\cdot) = \sum_a E_a(\cdot)E_a^\dagger$ for a (nonunique) set of Kraus operators $\{E_a\}$. We define

$$\begin{aligned} \rho_{\mathcal{E}} &\equiv \sum_a |E_a\rangle\langle\langle E_a| = \sum_a (\mathbb{1} \otimes E_a)|\mathbb{1}\rangle\langle\langle\mathbb{1}|(\mathbb{1} \otimes E_a^\dagger) \\ &= (\mathbb{1} \otimes \mathcal{E})(|\mathbb{1}\rangle\langle\langle\mathbb{1}|), \end{aligned} \quad (7)$$

where we have used the identity in Eq. (6); the $\mathbb{1}$ in $\mathbb{1} \otimes \mathcal{E}$ denotes the identity map. Thus defined, $\rho_{\mathcal{E}}$ is a non-negative operator on \mathcal{V} ; it can also be regarded as an unnormalized state (density operator) on the bipartite Hilbert space $\mathcal{H} \otimes \mathcal{H} \equiv \mathcal{H}_1 \otimes \mathcal{H}_2$, labeling the two subsystems by 1 and 2. In the latter picture, one regards $|\mathbb{1}\rangle\langle\langle\mathbb{1}|$ as the density operator for a maximally entangled state on $\mathcal{H} \otimes \mathcal{H}$, and $\rho_{\mathcal{E}}$ is the density operator that results from the action of the map $\mathbb{1} \otimes \mathcal{E}$ on it.

That $\rho_{\mathcal{E}}$ is invariant under a change of Kraus representation for the \mathcal{E} is manifest in the last line of Eq. (7). We can turn the logic around: Any bipartite state on $\mathcal{H} \otimes \mathcal{H}$ possesses a spectral decomposition into eigenvectors, and the identification of those eigenvectors, with their corresponding (square root of the) eigenvalues, as vectorized Kraus operators immediately gives an associated CP map on $\mathcal{B}(\mathcal{H})$. Equation (7) hence states a duality between CP maps \mathcal{E} and states $\rho_{\mathcal{E}} \geq 0$. $\rho_{\mathcal{E}}$ is sometimes called the “Choi state” of the CP map \mathcal{E} . Observe that

$$\mathcal{E}(X) = \text{tr}_1\{\rho_{\mathcal{E}}(X^T \otimes \mathbb{1})\}. \quad (8)$$

We are primarily interested in CP maps that are also TP. In this case, the state $\rho_{\mathcal{E}}$ dual to the CP and TP channel satisfies the partial trace condition

$$\text{tr}_2(\rho_{\mathcal{E}}) = \mathbb{1}, \quad (9)$$

i.e., \mathcal{E} is CPTP if and only if $\rho_{\mathcal{E}} \geq 0$ and $\text{tr}_2(\rho_{\mathcal{E}}) = \mathbb{1}$. A simple count verifies that we have just the right number of parameters: A CP \mathcal{E} is represented by d^4 real parameters—a positivity-preserving map that specifies how a d^2 -element basis of operators on \mathcal{H} is mapped back to itself—and this is the same number of real parameters needed to specify an unnormalized non-negative $\rho_{\mathcal{E}}$; the TP condition removes d^2 parameters, leaving $d^2(d^2 - 1)$ real parameters for a CPTP

map, i.e., a quantum channel. Note that the set of $\rho_{\mathcal{E}}$'s corresponding to quantum channels form a convex set of states, each with trace d . We denote the convex set of all $\rho_{\mathcal{E}}$ that satisfy Eq. (9) by \mathcal{S}_{TP} , and refer to $\rho_{\mathcal{E}} \in \mathcal{S}_{\text{TP}}$ as a TP state.

This duality between quantum channels and states enables us to sample quantum channels with algorithms for sampling quantum states (see the next section). Furthermore, the problem of process tomography—the estimation of the full description of a quantum channel acting on a quantum system—can be recast as that of *state* tomography. As the applications of our channel sampling algorithm discussed below are related to estimating quantum channels, we use the remainder of this section to recall this connection between state and process tomography, stemming from the channel-state duality [21].

Quantum process tomography seeks to discover the full description of some unknown quantum channel \mathcal{E} , through N uses of the channel. Standard strategies involve choosing a set of input states $\{\rho^{(i)}\}$, sending $N^{(i)}$ copies of state $\rho^{(i)}$ through the channel \mathcal{E} , and then measuring the output state using a positive-operator-valued measure (POVM) $\Pi^{(i)} \equiv \{\Pi_k^{(i)}\}$. For each i , the tomographic outcome probabilities come from the Born rule

$$p_k^{(i)} = \text{tr}\{\Pi_k^{(i)} \mathcal{E}(\rho^{(i)})\} = \text{tr}\{\rho_{\mathcal{E}} \Lambda_k^{(i)}\}, \quad (10)$$

where $\Lambda_k^{(i)} \equiv (\rho^{(i)})^{\mathcal{T}} \otimes \Pi_k^{(i)}$. Written in this manner, the expression for $p_k^{(i)}$ reminds one of the situation of state tomography of $\rho_{\mathcal{E}}$, where the set $\{\Lambda_k^{(i)}\}$ forms a pseudo-POVM in that $\Lambda_k^{(i)} \geq 0 \forall k, i$, and $\sum_k \Lambda_k^{(i)} = \rho^{(i)} \otimes \mathbb{1}$ for any i . Note that $\sum_k p_k^{(i)} = 1$, as guaranteed by the TP condition in Eq. (9) together with the normalization $\text{tr}(\rho^{(i)}) = 1$.

The likelihood function for the data $D = \{D^{(i)} = (n_1^{(i)}, n_2^{(i)}, \dots)\}$ — $n_k^{(i)}$ denotes the number of clicks in detector $\Pi_k^{(i)}$ when $\rho^{(i)}$ is sent, and $\sum_k n_k^{(i)} = N^{(i)}$ —collected is

$$L(D|\rho_{\mathcal{E}}) = \prod_i L(D^{(i)}|\rho_{\mathcal{E}}) = \prod_i \left[\prod_k (p_k^{(i)})^{n_k^{(i)}} \right], \quad (11)$$

where we omit the combinatorial factors that are needed for proper normalization but are not important here. Disregarding quantum constraints, the likelihood is maximized, over all $\{p_k^{(i)}\}$, by setting $p_k^{(i)} = \frac{n_k^{(i)}}{N^{(i)}}$; with quantum constraints, a constrained maximization of $L(D|\rho_{\mathcal{E}})$ over all permissible probabilities—those $p_k^{(i)}$'s that could have come from a non-negative $\rho_{\mathcal{E}}$ and which satisfy $\sum_k p_k^{(i)} = 1 \forall i$ —yields what is known as the maximum-likelihood estimator (MLE) for $\rho_{\mathcal{E}}$ [21].

III. PARAMETRIZING CHANNELS

A. Arbitrary channels

To obtain a sample of quantum channels according to some specified distribution, we generate Choi states $\rho_{\mathcal{E}}$ with the HMC algorithm. The HMC method demands a parametrization of the state space (in this case the space of $\rho_{\mathcal{E}}$) with no superfluous parameters and no external constraints. In Ref. [2], the ability to sample quantum states with the HMC algorithm was demonstrated using a parametrization of the full quantum state space. Because of the TP condition, sampling of quantum

channels demands a parametrization of, not the full quantum state space as in Ref. [2], but only of the set \mathcal{S}_{TP} of TP states. Here, as our central result, we explain how to accomplish this.

We first choose a product basis $\{|\bar{i}j\rangle\}_{i,j=1}^d$ on $\mathcal{H} \otimes \mathcal{H}$ and represent $\rho_{\mathcal{E}}$ as a $d^2 \times d^2$ matrix—also denoted as $\rho_{\mathcal{E}}$, to simplify notation—with complex entries. Positivity of $\rho_{\mathcal{E}}$ means that we can write $\rho_{\mathcal{E}} = A^\dagger A$, where A is a $d^2 \times d^2$ upper triangular complex matrix with real entries in the last column. The d^2 columns of A are labeled using a double index

$$A = \begin{pmatrix} | & | & \dots & | \\ \varphi_{11} & \varphi_{12} & \dots & \varphi_{1d} \\ | & | & & | \end{pmatrix}, \quad (12)$$

so that $\rho_{\mathcal{E}} = \sum_{ijkl=1}^d \varphi_{ij}^\dagger \varphi_{kl} |\bar{i}j\rangle \langle \bar{k}l|$, as the abstract, basis-independent object. Stacking the columns of A to form columns with d^3 entries,

$$\varphi_i \equiv \begin{pmatrix} \varphi_{i1} \\ \varphi_{i2} \\ \vdots \\ \varphi_{id} \end{pmatrix} \quad \text{for } i = 1, 2, \dots, d, \quad (13)$$

permits writing the TP condition in Eq. (9), that is, $\text{tr}_2(\rho_{\mathcal{E}}) = \sum_{ij} (\sum_k \varphi_{ik}^\dagger \varphi_{jk}) |\bar{i}\rangle \langle \bar{j}| = \mathbb{1} = \sum_{ij} \delta_{ij} |\bar{i}\rangle \langle \bar{j}|$, as an orthonormality condition on the φ_i 's:

$$\varphi_i^\dagger \varphi_j = \delta_{ij} \quad \text{for } i, j = 1, 2, \dots, d. \quad (14)$$

Hence, to sample quantum channels, we simply need to find a parametrization for the orthonormal set $\{\varphi_i\}_{i=1}^d$.

Let us count the number of parameters needed. Since A is upper triangular, φ_{ik} has $(ik) + 1$ generically nonzero entries, where $(ik) \equiv (i - 1)d + (k - 1)$ is a d -nary number. Each φ_i thus has $K_i \equiv \sum_k [(ik) + 1] = id^2 - \frac{1}{2}d(d - 1)$ nonzero entries. These nonzero entries are all complex, except for the d^2 of them in φ_{dd} , which are real. The orthonormality conditions on the φ_i 's remove d^2 real parameters. Altogether then, the φ_i 's are described by $2 \sum_i K_i - d^2 - d^2 = d^2(d^2 - 1)$ real parameters, exactly the number needed to describe a quantum channel.

To specify an appropriate parametrization of the φ_i set, it is convenient to reshuffle the rows of φ_i so that all the identically zero entries of each φ_i are collected together. We first define the matrix

$$\Phi \equiv \begin{pmatrix} | & | & \dots & | \\ \varphi_1 & \varphi_2 & \dots & \varphi_d \\ | & | & & | \end{pmatrix}. \quad (15)$$

Observe that the orthonormality conditions on the φ_i 's translate into the requirement that $\Phi^\dagger \Phi = \mathbb{1}$. Let P be a $d^3 \times d^3$ permutation matrix such that

$$\Psi \equiv P\Phi = \begin{pmatrix} | & | & \dots & | \\ \psi_1 & \psi_2 & \dots & \psi_d \\ | & | & & | \end{pmatrix} \quad (16)$$

has columns ψ_i 's, each of which is a reshuffled φ_i with all identically zero entries located below the generically nonzero ones, i.e., the k th entry of ψ_i , which we denote as ψ_{ik} , is generally nonzero for $k = 1, \dots, K_i$, and zero for $k = K_i + 1, \dots, d^3$. Such a P matrix exists because A is upper

triangular. Requiring $\Phi^\dagger \Phi = \mathbb{1}$ is equivalent to demanding $\Psi^\dagger \Psi = \Phi^\dagger P^{-1} P \Phi = \mathbb{1}$.

We are now ready to state the parametrization for the ψ_i 's, thereby giving a parametrization for \mathcal{S}_{TP} . We begin with ψ_d , parametrizing it with spherical coordinates so that it is normalized,

$$\psi_{dk} \equiv \begin{cases} e^{i\phi_k} (\cos \theta_{k-1}) S_k & \text{for } k = 1, \dots, K_d, \\ 0 & \text{for } k = K_d + 1, \dots, d^3, \end{cases} \quad (17)$$

where $\theta_0 \equiv 0$ fixed, and the S_k 's are recursively defined as $S_k = (\sin \theta_k) S_{k+1}$, with $S_{K_d} = 1$. Here, the ϕ_k 's for the ψ_{dk} 's that come from the real entries of φ_{dd} are understood to be set to zero (which ones they are, depends on the choice of P). ψ_d is hence parametrized by real parameters $\theta_1, \dots, \theta_{K_d-1}$, and $K_d - d^2 \phi_k$ (real) parameters, giving $2K_d - 1 - d^2$ real parameters in all. Note the identity

$$\sum_{k=1}^m |\psi_{dk}|^2 = S_m^2 \quad \text{for any } m = 1, 2, \dots, K_d, \quad (18)$$

so that the norm square of ψ_d is simply $\psi_d^\dagger \psi_d = \sum_{k=1}^{K_d} |\psi_{dk}|^2 = S_{K_d}^2 = 1$, i.e., ψ_d has length 1.

Next, let v_n , for $n = 1, \dots, K_d - 1$, be the d^3 -long column vector with the k th entry defined as

$$\begin{aligned} v_{nk} &\equiv \frac{1}{S_{n+1}} \begin{cases} \psi_{dk} |_{\theta_n \rightarrow \theta_n + \frac{\pi}{2}} & \text{for } k = 1, \dots, n+1, \\ 0 & \text{for } k = n+2, \dots, d^3 \end{cases} \\ &= \frac{1}{S_{n+1}} \begin{cases} \psi_{dk} \frac{\cos \theta_n}{\sin \theta_n} & \text{for } k = 1, \dots, n, \\ \psi_{d(n+1)} \frac{-\sin \theta_n}{\cos \theta_n} & \text{for } k = n+1, \\ 0 & \text{for } k = n+2, \dots, d^3. \end{cases} \end{aligned} \quad (19)$$

Observe that v_n is orthogonal to ψ_d , for every n , since

$$\begin{aligned} S_{n+1} v_n^\dagger \psi_d &= \frac{\cos \theta_n}{\sin \theta_n} \sum_{k=1}^n |\psi_{dk}|^2 - \frac{\sin \theta_n}{\cos \theta_n} |\psi_{d(n+1)}|^2 \\ &= \cos \theta_n \sin \theta_n S_{n+1}^2 - \sin \theta_n \cos \theta_n S_{n+1}^2 = 0. \end{aligned} \quad (20)$$

One can check, in a similar manner, that the v_n column vectors form an orthonormal set.

The span of $\{v_n\}_{n=1}^{K_d-1}$ lies in the orthogonal subspace of ψ_d . $\psi_1, \psi_2, \dots, \psi_{d-1}$ are to be orthogonal to ψ_d , so we can set them to be in the linear span of $\{v_n\}$. Note that both ψ_{d-1} and v_{K_d-1} have the same number ($= K_d - 1$) of nonzero entries, the largest among the ψ_i 's ($i = 1, \dots, d - 1$) and v_n 's. Specifically, we define

$$\begin{pmatrix} | & | & \dots & | \\ \psi_1 & \psi_2 & \dots & \psi_{d-1} \\ | & | & & | \end{pmatrix} \equiv V \tilde{\Psi}, \quad (21)$$

where V is the (nonsquare) matrix with columns $v_1, v_2, \dots, v_{K_d-1}$. $\tilde{\Psi}$ is defined such that its columns are the coefficients of the ψ_i 's when expressed as a linear combination of the v_n 's, i.e., $\psi_i = V \tilde{\psi}_i = \sum_n \tilde{\psi}_{in} v_n$, where $\tilde{\psi}_i$ is the i th column of $\tilde{\Psi}$, and $\tilde{\psi}_{in}$ are its entries. Note that V is a $d^3 \times (K_d - 1)$ matrix with the last $d^3 - K_d - 1$ rows completely zero, while $\tilde{\Psi}$ is a $(K_d - 1) \times (d - 1)$ matrix.

Observe that the orthonormality of the ψ_i 's, for $i = 1, \dots, d - 1$, is equivalent to the orthonormality of the columns of $\tilde{\Psi}$, i.e., $\tilde{\Psi}^\dagger \tilde{\Psi} = \mathbb{1}$. This is then the same problem as before, for $\tilde{\Psi}$, with now one fewer column. We hence repeat the procedure above, parametrizing $\tilde{\psi}_{d-1}$ using a new set of spherical coordinates (θ 's and ϕ 's; note that none of the ϕ are set to zero as the $\psi_{i \neq d}$'s are generally complex), defining new v vectors orthogonal to it, getting a new $\tilde{\Psi}$, and so forth. We do this recursively until all ψ_i 's are parametrized.

Let us check that the recursive procedure yields the right number of parameters for the full set of orthonormal ψ_i 's. As mentioned earlier, in the first round, ψ_d (and the V there) is parametrized by $2K_d - 1 - d^2$ parameters, that subtraction of d^2 coming from the d^2 zero ϕ_k 's done for ψ_d only. In the next round, ψ_{d-1} is parametrized by an additional (on top of the ones that go into V) $2(K_{d-1} - 1) - 1$ real parameter; in yet the next round, ψ_{d-2} is parametrized by an additional $2(K_{d-2} - 2) - 1$ real parameter; and so forth. Altogether then, we have $-d^2 + \sum_{i=0}^{d-1} [2(K_{d-i} - i) - 1] = d^4 - d^2$ real parameters, exactly the right number needed for parametrizing d -dimensional quantum channels.

To illustrate how one applies the above parametrization, the case of qutrit channels is discussed in Appendix B. In the following sections, we make use of our parametrization in a HMC algorithm to sample quantum channels according to specified distributions, and demonstrate the usefulness of these samples in different applications. Before we get to that, however, let us mention a parametrization designed specifically for unital qubit channels, useful for one of our examples below.

B. Unital qubit channels

A useful class of quantum channels is the set of unital channels, those that preserve the identity operator $\mathcal{E}(\mathbb{1}) = \mathbb{1}$. The unitality condition can be stated in terms of the Choi state as the requirement

$$\text{tr}_i(\rho_{\mathcal{E}}) = \mathbb{1}. \quad (22)$$

A unital quantum channel thus has $\rho_{\mathcal{E}}$ such that $\text{tr}_i(\rho_{\mathcal{E}}) = \mathbb{1}$ for $i = 1, 2$, stating both the TP and unitality conditions. This is generally a difficult pair of conditions to impose, for a parametrization of unital channels with exactly the right number of parameters, as needed for HMC.

For unital qubit channels, however, this can be done in a straightforward manner, as we describe here [22]. The Choi state of a qubit channel is a two-qubit state. Any two-qubit state (normalized to trace 2) can be written as

$$\rho = \frac{1}{2} (\mathbb{1} + \boldsymbol{\sigma} \cdot \boldsymbol{s} + \boldsymbol{t} \cdot \boldsymbol{\tau} + \boldsymbol{\sigma} \cdot \boldsymbol{C} \cdot \boldsymbol{\tau}), \quad (23)$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of Pauli operators for the first qubit and $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)$ is the vector of Pauli operators for the second qubit. (Here, the word ‘‘vector’’ is used in the physicist's sense of a three-dimensional spatial vector.) \boldsymbol{s} and \boldsymbol{t} are the Bloch vectors for qubits 1 and 2, respectively; \boldsymbol{C} is a dyadic, representable by a 3×3 matrix of real numbers corresponding to the coefficients of $\sigma_i \tau_j$, for $i, j = x, y, z$. The TP condition requires $\boldsymbol{s} = 0$; the unitality condition demands $\boldsymbol{t} = 0$. The Choi state of a unital qubit channel thus takes the

form

$$\rho_{\mathcal{E}} = \frac{1}{2}(\mathbb{1} + \boldsymbol{\sigma} \cdot \mathbf{C} \cdot \boldsymbol{\tau}). \quad (24)$$

Up to local unitary transformation, the dyadic \mathbf{C} can always be chosen to be diagonal \mathbf{C}_{diag} . For $\rho_{\mathcal{E}}$ to be positive semidefinite, the three diagonal entries of \mathbf{C}_{diag} must lie within a tetrahedron with the vertices

$$\begin{aligned} \mathbf{v}_1 &= (-1, -1, -1), \\ \mathbf{v}_2 &= (-1, 1, 1), \\ \mathbf{v}_3 &= (1, -1, 1), \\ \text{and } \mathbf{v}_4 &= (1, 1, -1), \end{aligned} \quad (25)$$

where each vertex corresponds to one of four pairwise orthogonal maximally entangled two-qubit states. We parametrize the three entries of \mathbf{C}_{diag} by the convex combination of the four vertices

$$(c_1, c_2, c_3) = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 + \alpha_4 \mathbf{v}_4, \quad (26)$$

where

$$\begin{aligned} \alpha_1 &= \cos^2 \theta_1, \\ \alpha_2 &= \sin^2 \theta_1 \cos^2 \theta_2, \\ \alpha_3 &= \sin^2 \theta_1 \sin^2 \theta_2 \cos^2 \theta_3, \\ \alpha_4 &= \sin^2 \theta_1 \sin^2 \theta_2 \sin^2 \theta_3. \end{aligned} \quad (27)$$

Generally, the dyadic \mathbf{C} can be written as

$$\mathbf{C} = \mathbf{R}_1 \mathbf{C}_{\text{diag}} \mathbf{R}_2^T, \quad (28)$$

where \mathbf{R}_1 and \mathbf{R}_2 are the rotation matrices representing the local unitary transformations (equivalently, spatial rotations in the Bloch-ball picture) of qubits 1 and 2, respectively. \mathbf{R}_1 and \mathbf{R}_2 can each be parametrized by three rotation angles. Altogether, we have a parametrization of the set of all unital qubit channels, specified by nine angle parameters.

IV. APPLICATIONS

The HMC algorithm is a method for generating random samples from any target distribution by making use of pseudo-Hamiltonian dynamics in a mock phase space. Upon identifying the parameters in the parametrizations given in Sec. III—which satisfy the requirements of not having superfluous parameters and no constraints—as the position variables in the mock phase space, we can employ the HMC algorithm. We give a brief review of the HMC algorithm in Appendix A. A more detailed discussion can be found in Ref. [2]. In this section, we demonstrate the use of random samples of channels in three applications related to process tomography. That the examples are related to tomography simply reflects the authors' original motivation and source of interest in the matter of channel sampling. The channel parametrization invented here and the resulting ability to sample according to a user-specified distribution using a HMC algorithm are applicable beyond tomography tasks.

A. Error regions for process estimation

Whether one chooses to use the MLE or some other estimator for $\rho_{\mathcal{E}}$, the *point* estimator will not coincide exactly with the true $\rho_{\mathcal{E}}$ with finite data. It is important then to endow the point estimators with error regions expressing the uncertainty in our knowledge of the identity of the channel. Here, we adopt as error regions the notion of smallest credible regions (SCRs) proposed in Ref. [23]. SCRs were originally proposed for the estimation of quantum states, whether they are TP states or not, but completely analogous notions can be defined for \mathcal{S}_{TP} . Here, we examine the construction of SCRs for the task of quantum process estimation, as an application of our channel sampling algorithm. We first recall a few key points about SCRs pertinent to our discussion here; the reader is referred to [23] for further details.

The SCR is the region—a set of states—in \mathcal{S}_{TP} with the smallest size for a chosen credibility. Size is the prior content of a region in \mathcal{S}_{TP} , i.e., the prior (before any data are taken) probability that the true state is in the region; credibility is the posterior (after incorporating the data) content of that region. The SCRs are bounded-likelihood regions (BLRs), i.e., regions \mathcal{R}_λ comprising all states with likelihood no smaller than a threshold fraction $\lambda \in [0, 1]$ of the maximum likelihood $L_{\text{max}}(D)$,

$$\mathcal{R}_\lambda(D) = \{\rho \in \mathcal{S}_{\text{TP}} : L(D|\rho) \geq \lambda L_{\text{max}}(D)\}, \quad (29)$$

with $\mathcal{R}_0 = \mathcal{S}_{\text{TP}}$. The size s_λ of the BLR \mathcal{R}_λ is its prior content, and its credibility c_λ is its posterior content

$$s_\lambda(D) = \int_{\mathcal{R}_\lambda(D)} (d\rho) \quad \text{and} \quad c_\lambda(D) = \int_{\mathcal{R}_\lambda(D)} (d\rho) \frac{L(D|\rho)}{L(D)}, \quad (30)$$

with $s_0 = c_0 = 1$ when $\lambda = 0$. The volume element $(d\rho)$ expresses the prior distribution; $(d\rho) \frac{L(D|\rho)}{L(D)}$ is the posterior distribution. $L(D) \equiv \int_{\mathcal{R}_0} (d\rho) L(D|\rho)$, a normalizing factor, is the likelihood of obtaining the data D for the chosen prior. For tomography problems, it is often natural to state the prior distribution in terms of the POVM-induced probabilities [see Eq. (10)]

$$(d\rho) = (dp) w_0(p), \quad (31)$$

where $w_0(p)$ is the prior density, nonzero only for $p \equiv (p_1^{(1)}, p_1^{(2)}, \dots, p_2^{(1)}, \dots)$ that corresponds to a $\rho \in \mathcal{S}_{\text{TP}}$, and $(dp) \equiv dp_1^{(1)} dp_1^{(2)} \dots$.

To report the error region for an experiment with data D , following the scheme of Ref. [23], s_λ and c_λ are calculated for all values of λ . The error regions are reported by plotting s_λ and c_λ as functions of λ . For a desired level of credibility, the λ value is read off, and the error region is the \mathcal{R}_λ for that value of λ . The size and credibility of a BLR [see Eq. (30)] cannot, in general, be computed analytically, due to the complicated integration region. Instead, we make use of MC integration: We generate random samples using HMC according to the prior and posterior distributions; the size and credibility are then the fractions of points contained in the BLR for the two distributions.

A related concept is the plausible region [24]. This is the set of all points in \mathcal{S}_{TP} , for which the data provide evidence in favor of— $L(D|\rho) > L(D)$. The plausible region is in fact a

TABLE I. Simulated data for the first example in Sec. IV A. Each entry $n_k^{(i)}$ represents the number of clicks in detector Π_k when the input state $\rho^{(i)}$ is sent.

	Π_k			
	1	2	3	4
$\rho^{(i)}$	7	5	6	6
1	7	5	6	6
2	3	13	4	4
3	2	7	8	7
4	3	7	8	6

BLR, with a critical value of λ ,

$$\lambda_{\text{crit}}(D) = \frac{L(D)}{L_{\text{max}}(D)}. \quad (32)$$

Once we have computed the size and credibility curves, we can also identify the plausible region for the data.

As a first example, we look at single-qubit channels. The input states $\rho^{(i)}$ for process tomography are taken to be the tetrahedron states

$$\rho^{(i)} = \frac{1}{2}(\mathbb{1} + \mathbf{a}_i \cdot \boldsymbol{\sigma}), \quad i = 1, 2, 3, \text{ and } 4 \quad (33)$$

where $\mathbf{a}_i = 3^{-1/2}\mathbf{v}_i$ with the vertex vectors of Eq. (25). For every i , we use the same POVM, the four-outcome tetrahedron measurement, with outcomes

$$\Pi_k = \frac{1}{4}(\mathbb{1} + \mathbf{a}_k \cdot \boldsymbol{\sigma}), \quad k = 1, 2, 3, \text{ and } 4. \quad (34)$$

We simulate data using an amplitude-damping channel described by the Kraus operators

$$E_0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \quad \text{and} \quad E_1 \equiv \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \quad (35)$$

where γ , the damping parameter, is set to 0.4. The matrices above refer to the computational basis. For the matrices appearing in the rest of this paper, it should be assumed that they refer to the computational basis as well. 24 copies of each input state $\rho^{(i)}$ are measured (simulated), giving a total of 96 counts over the four input states. The simulated data are reported in Table I.

For the prior distribution, we choose the conjugate prior

$$(dp) w_0(p) \propto (dp) \prod_{i,k=1}^4 (p_k^{(i)})^{48\bar{p}_k^{(i)}}, \quad (36)$$

where $\bar{p} = \{\bar{p}_k^{(i)}\}$ corresponds to the Born probabilities [see Eq. (10)] for an amplitude-damping channel with $\gamma = 0.5$, expressing our prior belief that that is the actual channel. Figure 1(a) shows the size and credibility curves, obtained from MC integration using 500 000 sample points generated from HMC with the channel parametrization of Sec. III. The critical λ value for the plausible region is indicated with a red dashed line, with size value $s = 0.2102$ and credibility value $c = 0.8586$. The true channel is contained in all BLRs with $\lambda < 0.0302$ and $c_\lambda > 0.5511$, and is thus in the plausible region.

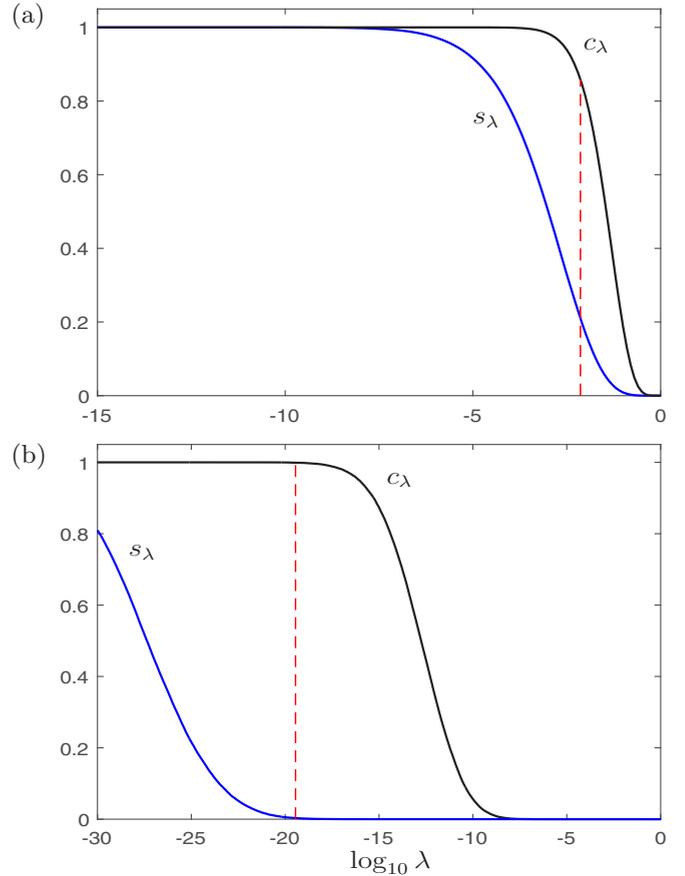


FIG. 1. Size s_λ and credibility c_λ of the BLRs \mathcal{R}_λ , plotted against $\log_{10} \lambda$, for (a) the qubit amplitude-damping channel, and (b) the qutrit amplitude-damping channel. The red vertical dashed lines mark the respective critical λ values $\lambda_{\text{crit}} = 0.0073$ for (a), and $\lambda_{\text{crit}} = 3.5598 \times 10^{-20}$ for (b). These identify the plausible regions.

Now, qubit channels are simple to characterize and there are many ways of sampling from the space of qubit channels. It is hence useful to see how our sampling algorithm works for examples beyond the qubit situation, for which proper sampling is more challenging. As a second example, we consider an amplitude-damping qutrit (three-dimensional quantum system) channel with the Kraus operators

$$E_0 \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-\gamma_1} & 0 \\ 0 & 0 & \sqrt{1-\gamma_2} \end{pmatrix}, \quad (37)$$

$$E_1 \equiv \begin{pmatrix} 0 & \sqrt{\gamma_1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad E_2 \equiv \begin{pmatrix} 0 & 0 & \sqrt{\gamma_2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

for $\gamma_1 = 0.1$ and $\gamma_2 = 0.5$.

The POVM used is one of the symmetric, informationally complete POVM (SIC-POVM) from the one-parameter family of qutrit SIC-POVMs. It can be described by a set of states $\{|\mu_i\rangle\}$; when written in the computational basis, they are given

TABLE II. Simulated data for the second example in Sec. IV A. Each entry $n_k^{(i)}$ represents the number of clicks in detector Π_k when the input state $\rho^{(i)}$ is sent.

	Π_k									
	1	2	3	4	5	6	7	8	9	
$\rho^{(1)}$	1	8	1	4	1	8	2	1	2	0
	2	3	10	3	1	1	2	3	3	1
	3	1	1	9	1	2	2	3	5	3
	4	8	3	3	4	0	1	4	3	1
	5	2	3	3	1	10	2	4	2	0
	6	3	4	4	2	0	8	2	2	2
	7	3	2	4	2	0	2	9	4	1
	8	2	0	4	0	0	2	6	8	5
	9	3	3	3	1	5	0	1	3	8

explicitly by

$$(|\mu_1\rangle|\mu_2\rangle \dots |\mu_9\rangle) \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \omega & \omega^* & 1 \\ \omega & \omega^* & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega & \omega^* & 1 & 1 & 1 & 1 \end{pmatrix}, \quad (38)$$

where $\omega = e^{i2\pi/3}$, $\omega^* = \omega^2$, and $1 + \omega + \omega^2 = 0$. The POVM elements are

$$\Pi_i = \frac{1}{3} |\mu_i\rangle\langle\mu_i|, \quad i = 1, 2, \dots, 9. \quad (39)$$

The input states are

$$\rho^{(i)} = |\mu_i\rangle\langle\mu_i|, \quad i = 1, 2, \dots, 9. \quad (40)$$

For each of the input states, the number of copies measured is 27, giving a total of 243 counts. The simulated data are reported in Table II. The prior is the primitive prior, i.e., $w_0(p)$ is a constant wherever it is nonzero. Figure 1(b) shows the size and credibility curves, obtained from MC integration with 100 000 sample points using HMC and our channel parametrization. As before, the critical λ value for the plausible region is indicated by the vertical dashed line. The size and credibility of the plausible region are $s = 0.0032$ and $c = 0.9990$, respectively. The true channel is contained in all BLRs with $\lambda < 1.1560 \times 10^{-6}$, and is thus in the plausible region.

B. Marginal likelihood for channel properties

Often, one is only interested in certain properties of a channel, like the fidelity between the output of the channel and its input, rather than a full channel description in the form of its process matrix. If one could directly measure that one quantity of interest, one expects to accomplish the estimation task with significantly fewer uses of the channel than needed for full tomography. However, a direct measurement of the quantity of interest may be difficult to design and implement, while the process tomography measurement is often standard procedure. Even in the latter case, one should still estimate the quantity of interest directly from the tomography data, rather than first estimating the full process matrix and then computing the quantity of interest from that estimate [25].

The key ingredient in making inferences about a property F of a channel from tomographic data D is the marginal likelihood, obtained by integrating the full likelihood $L(D|p)$ over the irrelevant parameters

$$L(D|F) = \frac{\int (dp) w_r(p) \delta(F - f(p)) L(D|p)}{\int (dp) w_r(p) \delta(F - f(p))} \equiv \frac{W_{r,D}(F)}{W_{r,0}(F)}, \quad (41)$$

where $W_{r,D(0)}(F)$ is the integral in the numerator (denominator). $f(p)$ is the function that expresses F in terms of the tomographic probabilities p , and $w_r(p)$ is the prior density on p , which induces a prior density on F . $\delta(F - f(p))$ is the Dirac delta function that enforces $f(p) = F$. Once we have the marginal likelihood, we can proceed in an analogous way as in Sec. IV A to construct the smallest credible interval (SCI) and the plausible interval for F , as well as perform other statistical inference tasks based on the marginal likelihood.

We thus need a general procedure for computing the marginal likelihood $L(D|F)$. In Ref. [25], an iterative algorithm was developed for that purpose, requiring the use of random samples according to specified distributions. The reader is referred to Ref. [25] for the full description of the iterative algorithm, and to Appendix C for the details relevant for our examples below. Here, we give only a brief account of the basic ideas. The delta functions in the defining equation (41) are difficult to handle in a numerical evaluation of the integrals. Instead, we evaluate the antiderivatives $P_{r,i}(F)$, with respect to F , of $W_{r,i}(F)$,

$$P_{r,i}(F) \equiv \int dF W_{r,i}(F), \quad i = D, 0 \quad (42)$$

with step functions in place of the delta functions. $P_{r,i}$ can be computed by MC integration. The results are closely fitted with several-parameter functions, and then differentiated to give $W_{r,i}$, and hence the marginal likelihood. This procedure works, in principle; in practice, one runs into numerical accuracy problems. If $w_r(p)$ has little weight over some range of F , a rather generic situation, $P_{r,0}$ will be very flat there, and its derivative cannot be reliably estimated. To overcome this problem, the crux is to note that, because of the delta functions, the marginal likelihood is invariant under the replacement $w_r(p) \rightarrow w_r(p)g(f(p))$ for any function $g(F)$ positive over the entire range of F . We thus have the freedom to choose the $w_r(p)$ used to evaluate $L(D|F)$. This freedom of choice is exploited in the iterative procedure described in Ref. [25], where the estimate of $W_{r,0}$ is successively improved by using an $w_r(p)$ modified by the previous (possibly inaccurate) estimate of $W_{r,0}$, until the desired convergence level is reached. Each iterative step requires the ability to sample according to the new $w_r(p)$; that is where the HMC algorithm, permitting sampling in accordance to a user-specified distribution, comes in.

Below, we carry out the iterative algorithm and compute the marginal likelihood for two common channel properties, average fidelity F_{avg} and minimum fidelity F_{min} . We make use of the HMC algorithm made possible by our channel parametrization of Sec. III. Both examples are for qubit channels, and use the same (simulated) tomographic data

TABLE III. Simulated data for the examples in Sec. IV B. Each entry $n_k^{(i)}$ represents the number of clicks in detector Π_k when the input state $\rho^{(i)}$ is sent.

	Π_k				
	1	2	3	4	
$\rho^{(i)}$	1	9	4	4	7
	2	6	6	3	9
	3	3	5	10	6
	4	8	4	5	7

obtained from tetrahedron input states [see Eq. (33)] and the tetrahedron POVM [see Eq. (34)] for the true channel

$$\mathcal{E}_{\text{Pauli}}(\cdot) \equiv \left(1 - \sum_{i=x,y,z} p_i\right)(\cdot) + \sum_{i=x,y,z} p_i \sigma_i(\cdot) \sigma_i, \quad (43)$$

a Pauli channel. Here, the σ_i 's are the standard Pauli operators, and $(p_x, p_y, p_z) = (0.05, 0.15, 0.2)$. The data are generated from 96 uses of the channel. The simulated data are reported in Table III. We regard the Pauli channel as noise acting on our quantum system. We are interested in the fidelity measures F_{avg} and F_{min} quantifying the effect of this noise channel on our system.

1. Average fidelity

The average fidelity F_{avg} is defined here as the (squared) fidelity between the input and output of the channel \mathcal{E} , averaged over all input pure states according to the Haar measure. We write $F(\psi, \rho) \equiv \langle \psi | \rho | \psi \rangle$ for the square of the fidelity between a pure state $\psi \equiv |\psi\rangle\langle\psi|$ and an arbitrary state ρ . Then, the average fidelity for the channel \mathcal{E} is

$$\begin{aligned} F_{\text{avg}}(\mathcal{E}) &\equiv \int d\psi \langle \psi | \mathcal{E}(\psi) | \psi \rangle \\ &= \left\langle \psi_0 \left[\int dU U^\dagger \mathcal{E}(U \psi_0 U^\dagger) U \right] \psi_0 \right\rangle \\ &= \frac{1}{d} [1 + (d-1)q]. \end{aligned} \quad (44)$$

Here, dU is the Haar measure for the space of unitary operators, and ψ_0 is some fiducial pure state. In arriving at the last line, we have used a standard result of the twirling operation [26] (namely, the expression in the brackets in the second-to-last line), with q given by

$$q \equiv \frac{1}{d^2 - 1} \sum_i \text{tr}[\rho_{\mathcal{E}}(O_i^T \otimes O_i)], \quad (45)$$

where O_i 's are all the traceless elements of an orthonormal (according to the Hilbert-Schmidt inner product) operator basis, containing an element proportional to the identity operator, for the d -dimensional \mathcal{H} . In the qubit case, q has the explicit formula

$$q = \frac{1}{3} \text{tr}[\rho_{\mathcal{E}}(\sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z)], \quad (46)$$

where we have chosen the $*$ map such that $|\bar{i}\rangle = |i\rangle$ for $\{|i\rangle\}_{i=0}^1$, the σ_z basis for the qubit (see comment about this choice in the second paragraph of Sec. II).

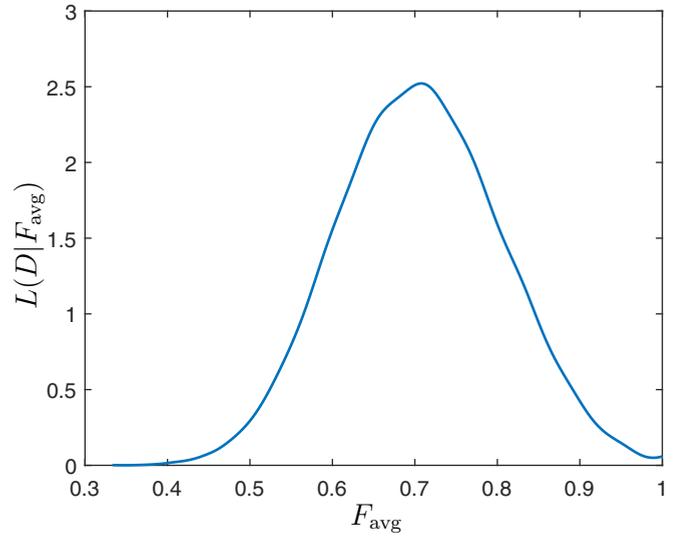


FIG. 2. The marginal likelihood $L(D|F_{\text{avg}})$, computed using the iterative procedure of Ref. [25] and HMC with our channel parametrization.

We use the iterative procedure of Ref. [25] to compute the marginal likelihood $L(D|F_{\text{avg}})$, for $F \equiv F_{\text{avg}}$. The final result is shown in Fig. 2; the intermediate steps of the iterative algorithm are described in Appendix D 1. With the marginal likelihood at hand, as an example of its usefulness, we can construct, as in Sec. IV A, the SCI for our estimate of F_{avg} . Figure 3(a) gives the size and credibility curves, as well as the critical λ value for the plausible region. Figure 3(b) shows the SCI for F_{avg} for different credibility values. The horizontal black line specifies the plausible interval, which includes the true value of $F_{\text{avg}} = 0.7333$ (indicated with an arrow).

2. Minimum fidelity of unital qubit channels

As a second example, also to illustrate the use of the parametrization of the unital qubit channels of Sec. III B, we look at the minimum, or worst-case, (squared) fidelity of a unital channel. The minimum fidelity for a channel \mathcal{E} is the fidelity of the output of \mathcal{E} with its (pure) input, minimized over all input states, i.e.,

$$F_{\text{min}} \equiv \min_{|\psi\rangle} F(\psi, \mathcal{E}(\psi)). \quad (47)$$

In the qubit case, F_{min} can be written explicitly using the Bloch-ball representation as

$$F_{\text{min}} = \min_{s:|s|=1} \frac{1}{2} (1 + s \cdot s_{\mathcal{E}}), \quad (48)$$

where s is the Bloch vector of the input state ψ , and $s_{\mathcal{E}}$ is that of the output $\mathcal{E}(\psi)$. For a unital qubit channel, $s_{\mathcal{E}}$ is the image of a linear map on the Bloch vector: $s_{\mathcal{E}} = \mathbf{M}s$. The minimum fidelity can thus be written simply as

$$F_{\text{min}} = \min_{s:|s|=1} \frac{1}{2} (1 + s^T \mathbf{M} s) = \frac{1}{2} (1 + \mu_{\text{min}}), \quad (49)$$

where μ_{min} is the smallest eigenvalue of $\frac{1}{2}(\mathbf{M} + \mathbf{M}^\dagger)$. This provides the direct connection between the unital qubit channel and F_{min} , and, in particular, allows us to express F_{min}

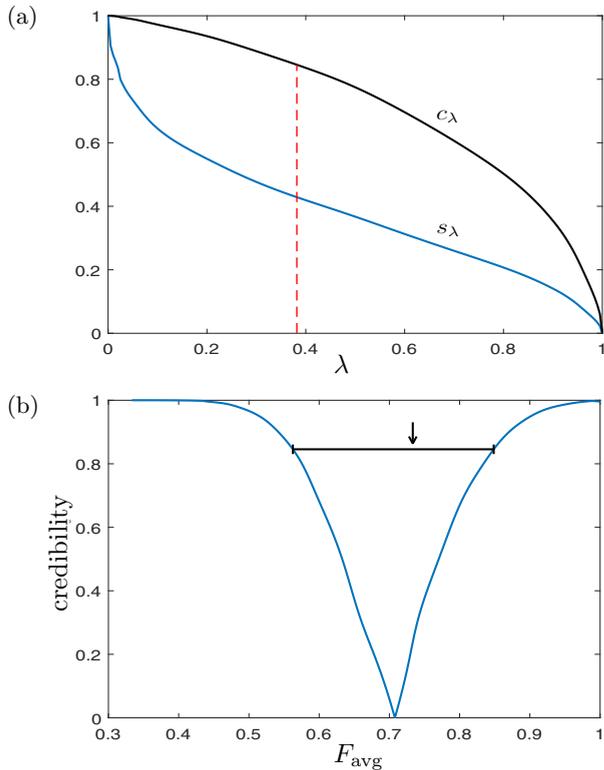


FIG. 3. (a) Size (blue) and credibility (green) curves for the bounded likelihood intervals for F_{avg} . The red vertical dashed line marks the critical value of λ , at $\lambda_{\text{crit}} = 0.3819$. (b) SCI for F_{avg} . The blue curve indicates the boundaries of the SCIs for different credibility values. The black horizontal line marks the plausible interval and the arrow indicates the true value of $F_{\text{avg}} = 0.7333$.

in terms of the tomographic probabilities associated with a channel \mathcal{E} .

Here, we assume the promise that the unknown channel is a unital one; the Pauli channel used to simulate the data is indeed unital. In effect, this unitality assumption restricts the relevant space of Choi states dual to the channels, to a strict subset of \mathcal{S}_{TP} , namely, to those that also satisfy Eq. (22). Any channel sampling is thus done only from this subset. Using the parametrization of Sec. III B, we employ HMC integration to compute the marginal likelihood $L(D|F_{\text{min}})$. The result is given in Fig. 4; the intermediate steps are provided in Appendix D 2. With this marginal likelihood, one can construct the corresponding SCIs and the plausible region, as well as perform other statistical inferences about the unital qubit channel.

C. Model selection

Often, one may not need the full generality of a CPTP channel to describe the dynamics of a quantum system. Instead, a simpler model with fewer parameters may suffice. Simpler models are computationally easier to work with, are likely more easily motivated from a physical standpoint, and may already describe the tomographic data well. One can phrase this problem as one of model selection in statistics, where the best model, among a few candidate models, is chosen, given the available data. Here, we discuss the quantum

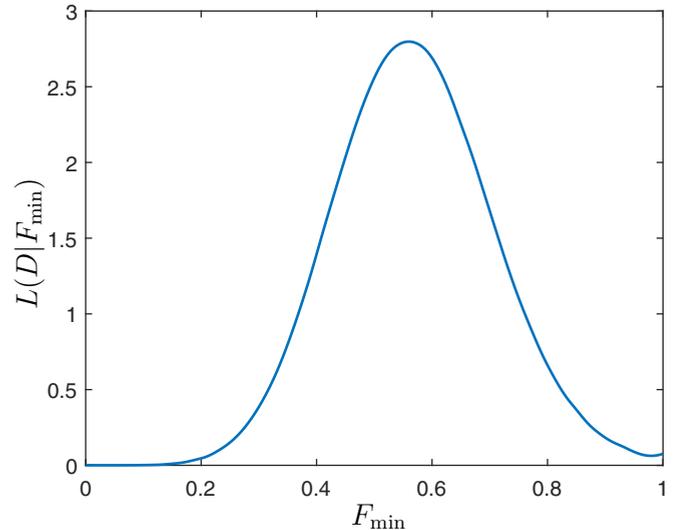


FIG. 4. The marginal likelihood $L(D|F_{\text{min}})$, computed using the iterative procedure of Ref. [25] and HMC with our channel parametrization.

problem of model selection for channel families. Our sampling algorithm is used for two purposes here: (1) to evaluate a criterion—based on the notion of relative belief—for the “best” model; (2) to assess and compare the performance of different model selection criteria by testing them on many randomly chosen true channels.

Two criteria for model selection commonly used in classical problems are the Akaike information criterion (AIC) [27] and the Bayesian information criterion (BIC) [28]. The AIC is based on the quantity (which we denote also as “AIC”)

$$\text{AIC} = 2k - 2 \ln(L_{\text{max}}), \quad (50)$$

k is the number of parameters in the model and L_{max} is the maximum value of the likelihood of the model for the data. The best model is the one with the smallest AIC value. The BIC is defined in a similar manner, but uses the value of N , the number of copies measured,

$$\text{BIC} = k \ln(N) - 2 \ln(L_{\text{max}}). \quad (51)$$

The best model according to this criterion is again the one with the smallest BIC value.

Another approach to model selection is based on the relative belief ratio (RBR) of Ref. [24]. The RBR of a model M is the ratio of its posterior to prior probabilities,

$$\text{RBR}(M|D) = \frac{P(M|D)}{P(M)}, \quad (52)$$

where

$$P(M|D) = \int_M (d\rho_{\mathcal{E}}) \frac{L(D|\rho_{\mathcal{E}})}{L(D)} \quad \text{and} \quad P(M) = \int_M (d\rho_{\mathcal{E}}). \quad (53)$$

If the posterior probability for a model M increases after the data, i.e., $\text{RBR}(M|D) > 1$, the data provide evidence in favor of the model; the data provide evidence against the model if $\text{RBR}(M|D) < 1$. It is also useful to have a measure of strength of evidence since the data might provide evidence

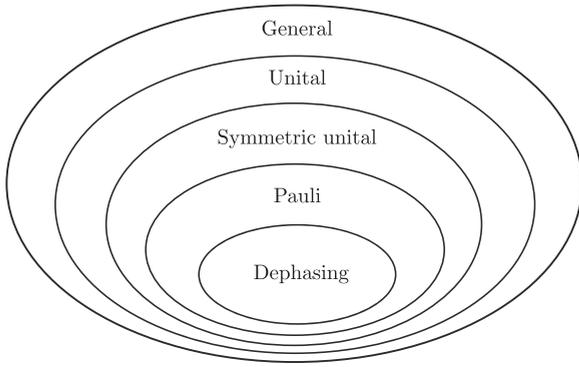


FIG. 5. The hierarchy of the five candidate models for the example in Sec. IV C.

in favor of more than one model from our candidate set, and one would like some basis of choosing among those models. The RBR value by itself is not a measure of the strength of evidence (see Ref. [24] for a discussion of various aspects, and also Ref. [29]). We supplement it with the posterior probability

$$P_{M_0} \equiv P[\text{RBR}(M|D) = \text{RBR}(M_0|D) | D] \quad (54)$$

for the model M_0 in question, and M ranges over the set of candidate models. If $\text{RBR}(M_0|D) > 1$ and P_{M_0} is large, then there is strong evidence in favor of M_0 . The best model, according to the RBR criterion of relative belief ratio, is the one with the largest posterior probability P_M , among all candidate models with $\text{RBR}(M|D) > 1$.

As an example, we consider as candidate models five nested qubit channel families: dephasing channels \subset Pauli channels \subset symmetric unital channels \subset unital channels \subset general CPTP channels; see Fig. 5. The smallest set is the one-parameter family of dephasing channels

$$\{\mathcal{D}_p(\cdot) \equiv (1-p)(\cdot) + p\sigma_z(\cdot)\sigma_z, p \in [0, 1]\}, \quad (55)$$

and its Choi state is given by

$$\rho_{\mathcal{E}} = \begin{pmatrix} 1 & 0 & 0 & 1-2p \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-2p & 0 & 0 & 1 \end{pmatrix}. \quad (56)$$

The set of Pauli channels is a three-parameter family

$$\left\{ \text{Pauli}_{\mathbf{p}}(\cdot) \equiv \left(1 - \sum_i p_i\right)(\cdot) + \sum_i p_i \sigma_i(\cdot) \sigma_i \right\}, \quad (57)$$

for $\mathbf{p} \equiv (p_x, p_y, p_z)$, $p_i \geq 0$, and $\sum_i p_i \leq 1$. The Choi state of a Pauli channel is of the form

$$\rho_{\mathcal{E}} = \begin{pmatrix} p_l + p_z & 0 & 0 & p_l - p_z \\ 0 & p_x + p_y & p_x - p_y & 0 \\ 0 & p_x - p_y & p_x + p_y & 0 \\ p_l - p_z & 0 & 0 & p_l + p_z \end{pmatrix}, \quad (58)$$

where $p_l = 1 - p_x - p_y - p_z$. The six-parameter family of symmetric unital channels refers to the subset of unital qubit channels such that $\mathbf{R}_1 = \mathbf{R}_2$ in Eq. (28). We then have the

9-parameter family of unital qubit channels and, lastly, the 12-parameter set of all CPTP qubit channels.

A natural prior on the model space is one that puts equal weights on each family. This is easily defined by the sampling procedure: the prior sample is constructed by generating 500 000 sample points with the primitive prior for each family. For the dephasing channel, the sample is generated by sampling p uniformly from $[0, 1]$. For the Pauli channel, we obtain the sample by generating (p_x, p_y, p_z) uniformly from the 3-simplex. For the symmetric unital, the unital, and the general channels, we make use of HMC and the parametrizations in Sec. III to generate the sample points. Note that in the numerical procedure that generates the samples for, say, the set of Pauli channels, we will never come across a sample point that is exactly a dephasing channel with $p_x = 0 = p_y$. Thus, even though the channel families are nested sets, one can consider each family to have prior probability of $\frac{1}{5}$. We use this prior to compute the RBR criterion for simulated data of different sizes. For our choice of prior, $P(M) = \frac{1}{5}$ for all models and $P(M|D)$ is calculated by taking the average of $\frac{L(D|\rho)}{L(D)}$ over the sample points for each model. P_{M_0} is computed by summing the posterior probabilities $P(M|D)$ of all the models M with the same RBR as model M_0 . Typically, $P_{M_0} = P(M_0|D)$.

To assess the performance of the three model-selection criteria, for each family of channels, we randomly (according to the primitive prior, as described above) draw 1000 channels. For each channel, we simulate data—with tetrahedron input states and a tetrahedron measurement [see Eqs. (33) and (34)]—for $N = 20, 50, 100, 1\,000, 10\,000$, and $100\,000$ copies measured, and evaluate the AIC, BIC, and RBR criteria for that data. Table IV shows the conclusions when the three criteria are applied to the simulated data. When the number of measured copies is very small, i.e., $N = 20$, the results based on AIC and BIC show a strong bias toward simpler (i.e., fewer-parameters) models. In particular, both criteria rarely identify the right model when the true channel comes from the unital or general families. Results based on RBR, however, show significantly more instances where the correct model is identified for the more complex (i.e., more parameters) models. For a moderate number of measured copies, i.e., $N = 1\,000$, AIC and RBR give equally good results, whereas BIC shows a slight bias toward the simpler models. When the number of measured copies is very large, i.e., $N = 100\,000$, results based on BIC are most accurate whereas results based on AIC have a slight bias to the more complex models. RBR also performs well in this regime.

Another aspect that we can check easily with our sampling procedure is the bias in the prior. This is particularly important for model selection based on the RBR criterion, to be sure that the probability of drawing a wrong conclusion is low. For example, for data that are typical for a unital channel, if we were to conclude regularly that there is evidence in favor of the general CPTP model and evidence against the unital model, there is bias in favor of the general CPTP model and bias against the unital model. To check for the bias, we draw 1000 random channels from each of the channel families and simulate data based on these true channels. The number of instances where the simulated data provide evidence against

TABLE IV. Comparison of results based on AIC, BIC, and RBR criteria, with different number of measured copies N . Candidate models are nested channel families (see main text). Each row below collects the counts for each family of true channels; each column collects the counts for the model that a criterion selects as the best fit for the data.

		True family	No. of cases where the best-fit model is				
			Dephasing	Pauli	SUnital	Unital	General
$N = 20$	AIC	Dephasing	947	43	10	0	0
		Pauli	583	408	9	0	0
		SUnital	629	319	52	0	0
		Unital	562	405	31	2	0
		General	596	372	30	2	0
	BIC	Dephasing	983	17	0	0	0
		Pauli	721	279	0	0	0
		SUnital	795	200	5	0	0
		Unital	741	256	3	0	0
		General	762	235	3	0	0
	RBR	Dephasing	712	62	93	56	77
		Pauli	224	271	173	140	192
		SUnital	239	151	315	117	178
		Unital	191	164	215	234	196
		General	166	156	180	164	334
$N = 50$	AIC	Dephasing	935	42	21	2	0
		Pauli	305	644	45	6	0
		SUnital	343	429	213	13	2
		Unital	257	534	148	57	4
		General	278	531	115	33	43
	BIC	Dephasing	995	4	1	0	0
		Pauli	539	461	0	0	0
		SUnital	646	328	26	0	0
		Unital	601	393	6	0	0
		General	586	404	10	0	0
	RBR	Dephasing	823	52	68	22	35
		Pauli	158	404	139	129	170
		SUnital	155	197	376	140	132
		Unital	91	185	222	343	159
		General	75	178	139	176	432
$N = 100$	AIC	Dephasing	938	41	18	1	2
		Pauli	173	733	72	16	6
		SUnital	141	368	455	30	6
		Unital	87	409	264	215	25
		General	78	442	147	93	240
	BIC	Dephasing	998	2	0	0	0
		Pauli	367	631	2	0	0
		SUnital	427	471	102	0	0
		Unital	357	593	42	8	0
		General	363	609	26	2	0
	RBR	Dephasing	905	43	33	11	8
		Pauli	129	578	105	94	94
		SUnital	76	210	506	126	82
		Unital	55	196	223	394	132
		General	34	132	118	190	526
$N = 1000$	AIC	Dephasing	933	40	18	7	2
		Pauli	22	866	79	26	7
		SUnital	0	64	818	87	31
		Unital	0	10	85	811	94
		General	0	2	3	47	948
	BIC	Dephasing	1000	0	0	0	0
		Pauli	77	923	0	0	0
		SUnital	27	231	742	0	0
		Unital	3	179	250	568	0
		General	2	113	105	112	668
	RBR	Dephasing	987	12	1	0	0
		Pauli	36	936	20	8	0
		SUnital	1	92	864	35	8
		Unital	0	18	122	837	23
		General	0	4	7	106	883
$N = 10000$	AIC	Dephasing	911	49	29	6	5
		Pauli	2	846	99	34	19
		SUnital	0	1	868	94	37
		Unital	0	0	3	889	108
		General	0	0	0	1	999
	BIC	Dephasing	1000	0	0	0	0
		Pauli	11	989	0	0	0
		SUnital	0	11	989	0	0
		Unital	0	0	37	963	0
		General	0	0	0	17	983
	RBR	Dephasing	999	1	0	0	0
		Pauli	6	993	1	0	0
		SUnital	0	7	985	8	0
		Unital	0	2	60	919	19
		General	0	0	7	86	907
$N = 100000$	AIC	Dephasing	921	44	27	6	2
		Pauli	1	848	97	37	17
		SUnital	0	0	865	102	33
		Unital	0	0	0	898	102
		General	0	0	0	0	1000
	BIC	Dephasing	1000	0	0	0	0
		Pauli	2	998	0	0	0
		SUnital	0	1	999	0	0
		Unital	0	0	1	999	0
		General	0	0	0	1	999
	RBR	Dephasing	1000	0	0	0	0
		Pauli	1	999	0	0	0
		SUnital	0	4	994	2	0
		Unital	0	0	68	924	8
		General	0	0	7	80	913

each of the four candidate models are calculated. The results are shown in Table V. As can be seen from the table, there is no significant bias in the prior when $N \geq 100$, and the bias decreases as the number of measured copies increases.

V. CONCLUSIONS

In this work, we constructed one exact parametrization for the space of CPTP channels. This parametrization has no superfluous parameters, and requires no imposition of

TABLE V. A check for bias in the prior. 1000 random channels from each of the channel families are drawn, and data for different number of measured copies N are simulated for each true channel. The table shows the fraction of instances with evidence against each of the channel families.

	True family	Fraction with evidence against				
		Dephasing	Pauli	SUnital	Unital	General
$N = 20$	Dephasing	0.233	0.812	0.746	0.798	0.810
	Pauli	0.724	0.413	0.531	0.401	0.495
	SUnital	0.687	0.558	0.398	0.455	0.515
	Unital	0.767	0.518	0.508	0.316	0.406
	General	0.779	0.524	0.552	0.384	0.350
$N = 50$	Dephasing	0.127	0.802	0.827	0.911	0.916
	Pauli	0.786	0.322	0.588	0.511	0.638
	SUnital	0.785	0.578	0.326	0.454	0.642
	Unital	0.876	0.561	0.503	0.276	0.488
	General	0.892	0.610	0.658	0.448	0.311
$N = 100$	Dephasing	0.042	0.844	0.911	0.964	0.978
	Pauli	0.826	0.228	0.623	0.672	0.780
	SUnital	0.874	0.613	0.249	0.508	0.793
	Unital	0.925	0.658	0.573	0.260	0.596
	General	0.948	0.715	0.731	0.576	0.266
$N = 1000$	Dephasing	0.003	0.966	0.998	1	1
	Pauli	0.946	0.028	0.930	0.981	0.999
	SUnital	0.992	0.878	0.088	0.868	0.983
	Unital	1	0.974	0.821	0.088	0.913
	General	1	0.995	0.983	0.868	0.077
$N = 10000$	Dephasing	0.001	0.996	1	1	1
	Pauli	0.992	0.006	0.996	1	1
	SUnital	1	0.992	0.011	0.989	1
	Unital	1	0.998	0.930	0.073	0.981
	General	1	1	0.991	0.907	0.092
$N = 100000$	Dephasing	0	0.999	1	1	1
	Pauli	0.999	0.001	1	1	1
	SUnital	1	0.996	0.006	0.998	1
	Unital	1	1	0.932	0.076	0.992
	General	1	1	0.993	0.920	0.086

any added constraints. These features make it possible to use the parametrization in a HMC algorithm, for producing high-quality—in terms of low correlations—samples of CPTP channels from a user-specified distribution. We demonstrated the usefulness of our parametrization in sampling applications taken from quantum process tomography. The method applies to general quantum channel sampling problems.

While our parametrization serves the purpose, it is, of course, just one of the many parametrizations that could be used in a HMC algorithm for sampling from the quantum channel space. For example, it is conceivable that a useful parametrization of a channel can be given in terms of the marginals and the copula of the respective Choi state [31]. This is unexplored territory.

A useful extension of this work will be to discover also an exact parametrization for the case of CPTP and unital

channels. As discussed above, this additional requirement of unitality presents difficulties that can be easily overcome only in the qubit situation. The parametrization for the space of CPTP, unital channels beyond the qubit case, remains an open problem. Note that such a parametrization will give also a possibly useful description of the space of all bipartite mixed quantum states with completely mixed states on both the single-party states; our current parametrization gives the larger space of states where only one of the two single-party states is completely mixed.

ACKNOWLEDGMENTS

This work is supported in part by the Ministry of Education, Singapore (through Grant No. MOE2016-T2-1-130). H.K.N. is also supported by Yale-NUS College (through a startup grant). The Centre for Quantum Technologies is a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore.

APPENDIX A: HAMILTONIAN MONTE CARLO (HMC)

HMC makes use of pseudo-Hamiltonian dynamics in a mock phase space. The parameters of interest are identified as the position variables θ and fictitious momentum variables ϑ are introduced. The Hamiltonian is defined as

$$H(\theta, \vartheta) = \frac{1}{2} \sum_j \vartheta_j^2 - \ln w(\theta), \quad (\text{A1})$$

where $w(\theta)$ is the target distribution. Any reasonable target distribution is permitted and, therefore, one can sample in accordance with any $w(\theta)$.

The HMC algorithm generates a set of sample points which follows the target distribution $w(\theta)$. The HMC algorithm is stated as follows [2]:

HMC algorithm

- (1) Set $j = 1$ and choose an arbitrary starting point $\theta^{(1)}$.
- (2) Generate $\vartheta^{(j)}$ from a multivariate Gaussian distribution with mean zero and unit variance.
- (3) Solve the Hamiltonian equations of motion

$$\frac{d}{dt}\theta_i = \frac{\partial}{\partial \vartheta_i} H, \quad \frac{d}{dt}\vartheta_i = -\frac{\partial}{\partial \theta_i} H \quad (\text{A2})$$

with the initial conditions $(\theta, \vartheta)|_{t=0} = (\theta^{(j)}, \vartheta^{(j)})$ to obtain $(\theta^*, \vartheta^*) = (\theta, -\vartheta)|_{t=T}$.

- (4) Calculate the acceptance ratio

$$a = \min \left\{ 1, e^{H(\theta^{(j)}, \vartheta^{(j)}) - H(\theta^*, \vartheta^*)} \right\}. \quad (\text{A3})$$

- (5) Draw a random number b uniformly from $[0, 1]$. If $b < a$, set $\theta^{(j+1)} = \theta^*$; otherwise, set $\theta^{(j+1)} = \theta^{(j)}$.

- (6) Set $j = j + 1$. If j equals the desired number of samples, escape the loop; otherwise, return to step 2.

Note that the ϑ distribution in step 2 is proportional to the kinetic-energy factor in $e^{-H(\theta, \vartheta)}$; the HMC algorithm enforces a θ distribution proportional to the potential-energy factor in

$e^{-H(\theta, \vartheta)}$, which is $e^{\ln w(\theta)} = w(\theta)$, the target distribution. If the differential equations in (A2) can be solved exactly, then the acceptance ratio $a = 1$. In practice, the differential equations must be discretized. This is done by the leapfrog method. Due to the discretization error, the acceptance ratio will not be 1 generally. The leapfrog method should be implemented such that the acceptance ratio is around the optimal value of 65% [18].

APPENDIX B: PARAMETRIZING QUTRIT CHANNELS

Here, we report an explicit application of the parameterization of Sec. III, for the case of qutrit channels. We start with the permutation matrix P that reshuffles φ_i 's into ψ_i 's, with the identically zero entries located below the generically nonzero ones. A P that can accomplish this is one such that

$$P = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 25 \\ 26 \\ 27 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \\ 11 \\ 19 \\ 20 \\ 21 \\ 2 \\ 3 \\ 4 \\ 12 \\ 13 \\ 14 \\ 22 \\ 23 \\ 24 \\ 5 \\ 6 \\ 7 \\ 15 \\ 16 \\ 17 \\ 25 \\ 26 \\ 27 \\ 8 \\ 9 \\ 18 \end{pmatrix}. \quad (\text{B1})$$

After the permutation, we have

$$\psi_1 = \begin{pmatrix} \psi_{1,1} \\ \vdots \\ \psi_{1,6} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \psi_2 = \begin{pmatrix} \psi_{2,1} \\ \vdots \\ \psi_{2,15} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \psi_3 = \begin{pmatrix} \psi_{3,1} \\ \vdots \\ \psi_{3,24} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (\text{B2})$$

To parametrize the ψ_i 's such that they are orthonormal, we first parametrize ψ_3 , of unit length,

$$\psi_3 = \begin{pmatrix} e^{i\phi_1} \sin \theta_1 \sin \theta_2 \dots \sin \theta_{22} \sin \theta_{23} \\ e^{i\phi_2} \cos \theta_1 \sin \theta_2 \dots \sin \theta_{22} \sin \theta_{23} \\ \vdots \\ e^{i\phi_{23}} \cos \theta_{22} \sin \theta_{23} \\ e^{i\phi_{24}} \cos \theta_{23} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (\text{B3})$$

Recalling that φ_{33} [see Eq. (12)] is a d^2 -entry real column, and with the P given above, $\psi_{3,4}$, $\psi_{3,5}$, $\psi_{3,6}$, $\psi_{3,13}$, $\psi_{3,14}$, $\psi_{3,15}$, $\psi_{3,22}$, $\psi_{3,23}$, $\psi_{3,24}$ are real. Thus, ϕ_4 , ϕ_5 , ϕ_6 , ϕ_{13} , ϕ_{14} , ϕ_{15} , ϕ_{22} , ϕ_{23} , ϕ_{24} are set to zero. Then, we define $\{v_n\}_{n=1}^{14}$ which lie in the orthogonal subspace of ψ_3 as follows:

$$v_1 = \begin{pmatrix} e^{i\phi_1} \cos \theta_1 \\ -e^{i\phi_2} \sin \theta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} e^{i\phi_1} \sin \theta_1 \cos \theta_2 \\ e^{i\phi_2} \cos \theta_1 \cos \theta_2 \\ -e^{i\phi_3} \sin \theta_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \quad (\text{B4})$$

$$v_{14} = \begin{pmatrix} e^{i\phi_1} \sin \theta_1 \sin \theta_2 \dots \sin \theta_{13} \cos \theta_{14} \\ e^{i\phi_2} \cos \theta_1 \sin \theta_2 \dots \sin \theta_{13} \cos \theta_{14} \\ \vdots \\ e^{i\phi_{14}} \cos \theta_{13} \cos \theta_{14} \\ -e^{i\phi_{15}} \sin \theta_{14} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (\text{B4})$$

To make ψ_1 and ψ_2 orthogonal to ψ_3 , we set them to be in the span of $\{v_n\}$:

$$\begin{pmatrix} | & | \\ \psi_1 & \psi_2 \\ | & | \end{pmatrix} \equiv V \tilde{\Psi} = \begin{pmatrix} | & | & \dots & | \\ v_1 & v_2 & & v_{14} \\ | & | & & | \end{pmatrix} \begin{pmatrix} | & | \\ \tilde{\psi}_1 & \tilde{\psi}_2 \\ | & | \end{pmatrix}. \quad (\text{B5})$$

The orthonormality of ψ_1 and ψ_2 is equivalent to the orthonormality of $\tilde{\psi}_1$ and $\tilde{\psi}_2$. We simply need to repeat the previous procedure. We parametrize $\tilde{\psi}_2$ to be of unit length

$$\tilde{\psi}_2 = \begin{pmatrix} e^{i\tilde{\phi}_1} \sin \tilde{\theta}_1 \sin \tilde{\theta}_2 \dots \sin \tilde{\theta}_{12} \sin \tilde{\theta}_{13} \\ e^{i\tilde{\phi}_2} \cos \tilde{\theta}_1 \sin \tilde{\theta}_2 \dots \sin \tilde{\theta}_{12} \sin \tilde{\theta}_{13} \\ \vdots \\ e^{i\tilde{\phi}_{13}} \cos \tilde{\theta}_{12} \sin \tilde{\theta}_{13} \\ e^{i\tilde{\phi}_{14}} \cos \tilde{\theta}_{13} \end{pmatrix}. \quad (\text{B6})$$

Next, we define $\{u_n\}_{n=1}^4$, each orthogonal to ψ_2 :

$$u_1 = \begin{pmatrix} e^{i\tilde{\phi}_1} \cos \tilde{\theta}_1 \\ -e^{i\tilde{\phi}_2} \sin \tilde{\theta}_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} e^{i\tilde{\phi}_1} \sin \tilde{\theta}_1 \cos \tilde{\theta}_2 \\ e^{i\tilde{\phi}_2} \cos \tilde{\theta}_1 \cos \tilde{\theta}_2 \\ -e^{i\tilde{\phi}_3} \sin \tilde{\theta}_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \quad (B7)$$

$$u_4 = \begin{pmatrix} e^{i\tilde{\phi}_1} \sin \tilde{\theta}_1 \sin \tilde{\theta}_2 \sin \tilde{\theta}_3 \cos \tilde{\theta}_4 \\ e^{i\tilde{\phi}_2} \cos \tilde{\theta}_1 \sin \tilde{\theta}_2 \sin \tilde{\theta}_3 \cos \tilde{\theta}_4 \\ e^{i\tilde{\phi}_4} \cos \tilde{\theta}_2 \sin \tilde{\theta}_3 \cos \tilde{\theta}_4 \\ e^{i\tilde{\phi}_4} \cos \tilde{\theta}_3 \cos \tilde{\theta}_4 \\ -e^{i\tilde{\phi}_5} \sin \tilde{\theta}_4 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (B7)$$

Finally, to have $\tilde{\psi}_1$ normalized and orthogonal to $\tilde{\psi}_2$, we set

$$\tilde{\psi}_1 \equiv U \bar{\psi}_1 = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_4 \\ | & | & & | \end{pmatrix} \bar{\psi}_1, \quad (B8)$$

where

$$\bar{\psi}_1 = \begin{pmatrix} e^{i\tilde{\phi}_1} \sin \tilde{\theta}_1 \sin \tilde{\theta}_2 \sin \tilde{\theta}_3 \\ e^{i\tilde{\phi}_2} \cos \tilde{\theta}_1 \sin \tilde{\theta}_2 \sin \tilde{\theta}_3 \\ e^{i\tilde{\phi}_3} \cos \tilde{\theta}_2 \sin \tilde{\theta}_3 \\ e^{i\tilde{\phi}_4} \cos \tilde{\theta}_3 \end{pmatrix}. \quad (B9)$$

We check that we have the right number of parameters. The parameters used above are $\theta_1, \dots, \theta_{23}, \phi_1, \dots, \phi_{24}$ (nine of these are set identically to zero), $\tilde{\theta}_1, \dots, \tilde{\theta}_{13}, \tilde{\phi}_1, \dots, \tilde{\phi}_{14}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3$, and $\tilde{\phi}_1, \dots, \tilde{\phi}_4$, giving a total of $72 = 3^2(3^2 - 1)$ parameters, as needed for specifying qutrit channels.

APPENDIX C: ITERATIVE ALGORITHM FOR ESTIMATING THE MARGINAL LIKELIHOOD

To estimate the marginal likelihood reliably, we follow the procedure in Ref. [25]. For the following discussion, we assume

$$0 \leq f(p) \leq 1 \quad (C1)$$

for the sake of simplicity. First, we note that the integrands in (41) are ill suited for MC integration due to the presence of the Dirac delta factors. We consider the antiderivatives

$$P_{r,0}(F) = \int (dp) w_r(p) \eta(F - f(p)) \quad (C2)$$

and

$$P_{r,D}(F) = \frac{1}{L(D)} \int (dp) w_r(p) \eta(F - f(p)) L(D|p). \quad (C3)$$

With a sample of $w_r(p)$ and $\frac{w_r(p)L(D|p)}{L(D)}$, we can evaluate the antiderivatives for various values of F and fit them with several-parameters functions. From the fitted functions, we can then calculate the derivatives

$$W_{r,0}(F) = \frac{\partial}{\partial F} P_{r,0}(F) = \int (dp) w_r(p) \delta(F - f(p)) \quad (C4)$$

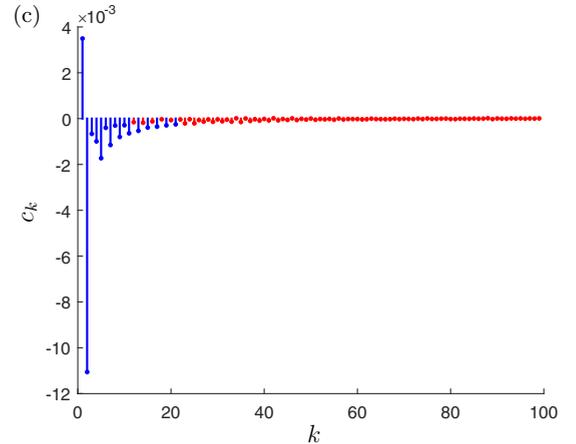
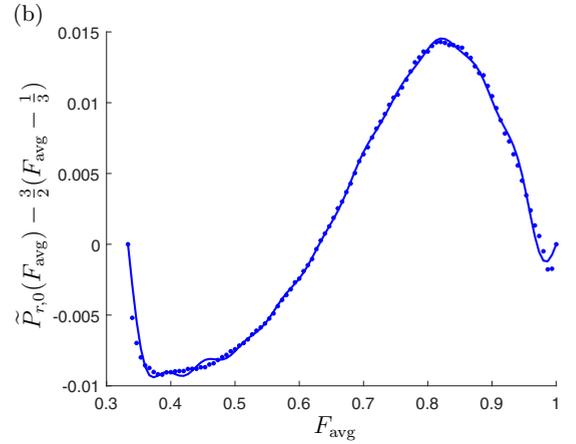
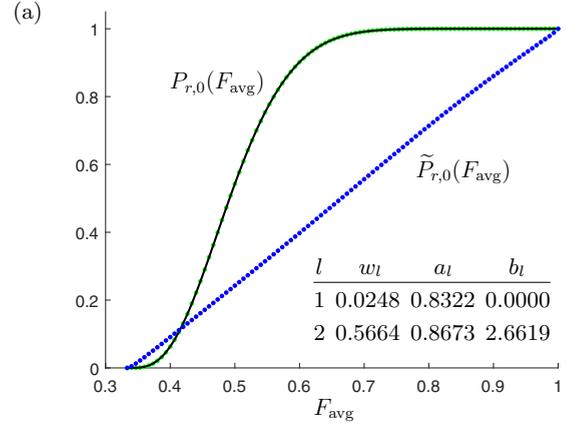


FIG. 6. Average gate fidelity. (a) The green dots depict the MC values of $P_{r,0}(F_{\text{avg}})$, and the black curve is fitted to them; the inset table reports the fitting parameters. The values of $\tilde{P}_{r,0}(F_{\text{avg}})$ are traced out by the blue dots. (b) The blue dots show the MC values of $\tilde{P}_{r,0}(F_{\text{avg}})$ after subtracting the straight line $\frac{3}{2}(F_{\text{avg}} - \frac{1}{3})$. The blue curve, a truncated Fourier series, is fitted to the dots. (c) Fourier amplitudes for $\tilde{P}_{r,0}(F_{\text{avg}})$. The high-frequency noise is removed from the fit in (b) by discarding the red amplitudes whose magnitude is less than 2% of that of largest amplitude.

and

$$\begin{aligned} W_{r,D}(F) &= \frac{\partial}{\partial F} P_{r,D}(F) \\ &= \frac{1}{L(D)} \int (dp) w_r(p) \delta(F - f(p)) L(D|p) \end{aligned} \quad (C5)$$

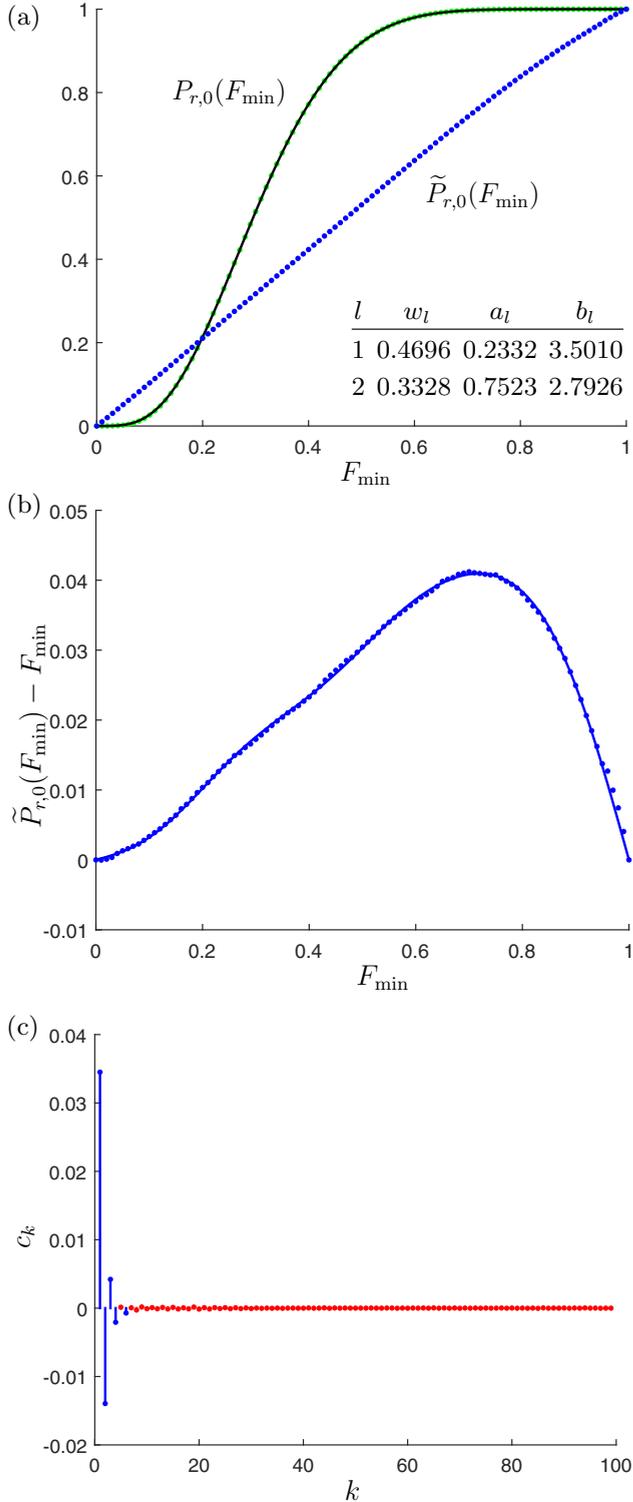


FIG. 7. Minimum gate fidelity. (a) The green dots depict the MC values of $P_{r,0}(F_{\min})$, and the black curve is fitted to them; the inset table reports the fitting parameters. The values of $\tilde{P}_{r,0}(F_{\min})$ are traced out by the blue dots. (b) The blue dots show the MC values of $\tilde{P}_{r,0}(F_{\min})$ after subtracting the straight line F_{\min} . The blue curve, a truncated Fourier series, is fitted to the dots. (c) Fourier amplitudes for $\tilde{P}_{r,0}(F_{\min})$. The high-frequency noise is removed from the fit in (b) by discarding the red amplitudes whose magnitude is less than 2% of that of largest amplitude.

and obtain the marginal likelihood by

$$L(D|F) = \frac{W_{r,D}(F)}{W_{r,0}(F)}. \quad (\text{C6})$$

A problem arises when $P_{r,0}(F)$ is very close to a constant over some range of values of F . The common situation is that $P_{r,0}(F)$ is very close to zero for a range of values near $F = 0$ and very close to one for a range of values near $F = 1$. MC integration is not precise enough to distinguish $P_{r,0}(F) \gtrsim 0$ from $P_{r,0}(F) = 0$ and $P_{r,0}(F) \lesssim 1$ from $P_{r,0}(F) = 1$. As a result, the estimated value of $W_{r,0}(F)$ will be equal to zero over those range of values. We cannot get a reliable estimation of $L(D|F)$ in this situation since $W_{r,0}(F)$ is the denominator in Eq. (C6). To overcome this problem, we note that we can do the replacement

$$w_r(p) \rightarrow w_r(p)g(f(p)) \quad (\text{C7})$$

with an arbitrary function $g(F) > 0$ without changing the value of $L(D|F)$.

The procedure for obtaining a reliable estimation of $L(D|F)$ is as follows:

(1) Sample according to $w_r(p)$. Use this sample to calculate $P_{r,0}(F)$. Fit a several-parameters function to $P_{r,0}(F)$ and obtain $W_{r,0}(F)$ by differentiating the fitted function.

(2) Sample according to $\tilde{w}_r(p) = \frac{w_r(p)}{W_{r,0}(f(p))}$. Use this sample to calculate

$$\tilde{P}_{r,0}(F) = \int (dp) \tilde{w}_r(p) \eta(F - f(p)). \quad (\text{C8})$$

Fit a several-parameters function to $\tilde{P}_{r,0}(F)$ and obtain $\tilde{W}_{r,0}(F)$ by differentiating the fitted function.

(3) Sample according to $\frac{\tilde{w}_r(p)L(D|p)}{L(D)} = \frac{w_r(p)L(D|p)}{W_{r,0}(f(p))L(D)}$. Use this sample to calculate

$$\tilde{P}_{r,D}(F) = \int (dp) \tilde{w}_r(p) \eta(F - f(p)) L(D|p). \quad (\text{C9})$$

Fit a several-parameters function to $\tilde{P}_{r,D}(F)$ and obtain $\tilde{W}_{r,D}(F)$ by differentiating the fitted function.

(4) Obtain the marginal likelihood from

$$L(D|F) = \frac{\tilde{W}_{r,D}(F)}{\tilde{W}_{r,0}(F)}. \quad (\text{C10})$$

The reason that we can have a reliable estimation of $L(D|F)$ using $\tilde{W}_{r,0}(F)$ obtained in step 2 is as follows. Suppose the exact value of $W_{r,0}(F)$ is known, $\tilde{W}_{r,0}(F)$ will be equal to 1 and $\tilde{P}_{r,0}(F)$ will be equal to F . If the exact values of $W_{r,0}(F)$ are not known, but we have a good approximation for $W_{r,0}(F)$ from step 1 and use it for the calculation of $\tilde{P}_{r,0}(F)$ in step 2, the $\tilde{P}_{r,0}(F)$ that we obtain will still be quite close to F and $\tilde{W}_{r,0}(F)$ will be nonzero for all range of F values.

In step 1, $P_{r,0}(F)$ can be fitted with a linear combination of regularized incomplete beta functions

$$I_{a,b}(x) = \frac{\int_0^x t^{a-1}(1-t)^{b-1} dt}{\int_0^1 t^{a-1}(1-t)^{b-1} dt}, \quad (\text{C11})$$

that is,

$$P_{r,0}(F) = w_1 I_{a_{\min}, b_1}(F) + w_2 I_{a_1, b_{\min}}(F) + w_3 I_{a_2, b_2}(F) + \dots + \left(1 - \sum_{i=1}^{N-1} w_i\right) I_{a_{N-1}, b_{N-1}}(F), \quad (\text{C12})$$

with the fitting parameters a_1, \dots, a_{N-1} , b_1, \dots, b_{N-1} , w_1, \dots, w_{N-1} . a_{\min} and b_{\min} are fixed by the power laws satisfied by $P_{r,0}(F)$ near $F = 0$ and 1:

$$P_{r,0}(F) \propto F^{a_{\min}} \quad \text{for } F \gtrsim 0 \quad (\text{C13})$$

and

$$1 - P_{r,0}(F) \propto (1 - F)^{b_{\min}} \quad \text{for } F \lesssim 1. \quad (\text{C14})$$

In step 2, a truncated Fourier series of the form

$$\tilde{P}_{r,0}(F) \simeq F + c_1 \sin(\pi F) + c_2 \sin(2\pi F) + c_3 \sin(3\pi F) + \dots \quad (\text{C15})$$

is usually a good fitting function. In step 3, $\tilde{P}_{r,D}(F)$ can be fitted with a smoothing spline.

APPENDIX D: INTERMEDIATE RESULTS FOR THE ESTIMATION OF THE MARGINAL LIKELIHOOD

1. Average gate fidelity

The green dots in Fig. 6(a) show the values of $P_{r,0}(F_{\text{avg}})$ obtained by a MC integration with 1 000 000 sample points. The MC integration is not precise enough to distinguish $P_{r,0}(F_{\text{avg}}) \gtrsim 0$ from $P_{r,0}(F_{\text{avg}}) = 0$ near $F_{\text{avg}} = \frac{1}{3}$ and to distinguish $P_{r,0}(F_{\text{avg}}) \lesssim 1$ from $P_{r,0}(F_{\text{avg}}) = 1$ near $F_{\text{avg}} = 1$. Therefore, a reliable approximation for $W_{r,0}(F_{\text{avg}}) = \frac{\partial}{\partial F_{\text{avg}}} P_{r,0}(F_{\text{avg}})$ cannot be obtained. To overcome this problem, we follow the procedure stated in Appendix C. First, we fit the green dots with a three-term fitting function of the form of Eq. (C12) with $F = \frac{3}{2}(F_{\text{avg}} - \frac{1}{3})$, $a_{\min} = 3$, and $b_{\min} = \frac{21}{2}$. The black curve is the fitted curve of $P_{r,0}(F_{\text{avg}})$. The fitting parameters are shown

in the inset table. $\tilde{P}_{r,0}(F_{\text{avg}})$ is obtained from a MC integration with 1 500 000 sample points and shown as the blue dots in Fig. 6(a).

The $\tilde{P}_{r,0}(F_{\text{avg}})$ is quite close to the straight line $\frac{3}{2}(F_{\text{avg}} - \frac{1}{3})$. The $\tilde{P}_{r,0}(F_{\text{avg}})$ after subtracting the straight line $\frac{3}{2}(F_{\text{avg}} - \frac{1}{3})$ is shown as the blue dots in Fig. 6(b). The blue curve shows the fitting curve, a truncated Fourier series whose Fourier amplitudes are reported in Fig. 6(c).

$P_{r,D}(F_{\text{avg}})$ is evaluated by a MC integration with 1 500 000 sample points and it can be fitted with a smoothing spline. The marginal likelihood shown in Fig. 2 is obtained from the ratio of $\tilde{W}_{r,D}(F_{\text{avg}})$ and $\tilde{W}_{r,0}(F_{\text{avg}})$.

2. Worst-case fidelity of a unital qubit channel

The green dots in Fig. 7(a) show the values of $P_{r,0}(F_{\min})$ from a MC integration with 1 000 000 sample points. The MC integration is not precise enough to distinguish $P_{r,0}(F_{\min}) \gtrsim 0$ from $P_{r,0}(F_{\min}) = 0$ near $F_{\min} = 0$ and to distinguish $P_{r,0}(F_{\min}) \lesssim 1$ from $P_{r,0}(F_{\min}) = 1$ near $F_{\min} = 1$. Therefore, a reliable approximation for $W_{r,0}(F_{\min}) = \frac{\partial}{\partial F_{\min}} P_{r,0}(F_{\min})$ cannot be obtained. To overcome this problem, we follow the procedure stated in Appendix C. First, we fit the green dots with a three-term fitting function of the form in Eq. (C12) with $F = F_{\min}$, $a_{\min} = 4$, and $b_{\min} = \frac{15}{2}$. The black curve is fitted to the numerical values for $P_{r,0}(F_{\min})$. The fitting parameters are shown in the inset table. $\tilde{P}_{r,0}(F_{\min})$ is obtained by a MC integration with 1 500 000 sample points and shown as the blue dots in Fig. 7(a).

The values of $\tilde{P}_{r,0}(F_{\min})$ are quite close to the straight line F_{\min} . The corresponding values after subtracting this straight line make up the blue dots in Fig. 7(b). The blue fitting curve is a truncated Fourier series with the Fourier amplitudes of Fig. 7(c).

$\tilde{P}_{r,D}(F_{\min})$ is evaluated by a MC integration with 1 500 000 sample points and it can be fitted with a smoothing spline. The marginal likelihood shown in Fig. 4 is the ratio of $\tilde{W}_{r,D}(F_{\min})$ and $\tilde{W}_{r,0}(F_{\min})$.

-
- [1] J. Shang, Y.-L. Seah, H. K. Ng, D. J. Nott, and B.-G. Englert, Monte Carlo sampling from the quantum state space. I, *New J. Phys.* **17**, 043017 (2015).
 - [2] Y.-L. Seah, J. Shang, H. K. Ng, D. J. Nott, and B.-G. Englert, Monte Carlo sampling from the quantum state space. II, *New J. Phys.* **17**, 043018 (2015).
 - [3] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Volume of the set of separable states, *Phys. Rev. A* **58**, 883 (1998).
 - [4] K. Życzkowski, Volume of the set of separable states. II, *Phys. Rev. A* **60**, 3496 (1999).
 - [5] K. Życzkowski and H.-J. Sommers, Induced measures in the space of mixed quantum states, *J. Phys. A: Math. Gen.* **34**, 7111 (2001).
 - [6] R. Blume-Kohout, Optimal, reliable estimation of quantum states, *New J. Phys.* **12**, 043034 (2010).
 - [7] F. Huszár and N. M. T. Hounslow, Adaptive Bayesian quantum tomography, *Phys. Rev. A* **85**, 052120 (2012).
 - [8] C. Oh, Y. S. Teo, and H. Jeong, Efficient Bayesian credible-region certification for quantum-state tomography, *Phys. Rev. A* **100**, 012345 (2019).
 - [9] M.-D. Choi, Completely positive linear maps on complex matrices, *Lin. Algebra Appl.* **10**, 285 (1975).
 - [10] A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Rep. Math. Phys.* **3**, 275 (1972).
 - [11] L. P. Thinh, P. Faist, J. Helsen, D. Elkouss, and S. Wehner, Practical and reliable error bars for quantum process tomography, *Phys. Rev. A* **99**, 052311 (2019).
 - [12] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).
 - [13] W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski, Random quantum operations, *Phys. Lett. A* **373**, 320 (2009).
 - [14] W. Bruzda, M. Smaczyński, V. Cappellini, H.-J. Sommers, and K. Życzkowski, Universality of spectra for interacting

- quantum chaotic systems, *Phys. Rev. E* **81**, 066209 (2010).
- [15] R. M. Neal, *Bayesian Learning for Neural Networks*, Lecture Notes in Statistics Vol. 118 (Springer, Heidelberg, 1996).
- [16] A. Hajian, Efficient cosmological parameter estimation with Hamiltonian Monte Carlo technique, *Phys. Rev. D* **75**, 083525 (2007).
- [17] E. K. Porter and J. Carré, A Hamiltonian Monte Carlo method for Bayesian inference of supermassive black hole binaries, *Class. Quantum Grav.* **31**, 145004 (2014).
- [18] R. M. Neal, MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones and X.-L. Meng (Chapman and Hall, Boca Raton, FL, 2011), Chap. 5.
- [19] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, *Phys. Lett. B* **195**, 216 (1987).
- [20] T. Durt, B.-G. Englert, I. Bengtsson, and K. Życzkowski, On mutually unbiased bases, *Int. J. Quantum Phys.* **8**, 535 (2010).
- [21] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek, Maximum-Likelihood Methods in Quantum Mechanics, in *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer, Heidelberg, 2004), Chap. 3.
- [22] An alternative, equally straightforward, parametrization exploits that the two-qubit Choi states for unital channels have real 4×4 matrices in the “magic basis” [30], composed of the singlet state $|s\rangle = 2^{-\frac{1}{2}}(|01\rangle - |10\rangle)$ and the three triplet states $i\sigma_x \otimes \mathbb{1}|s\rangle$, $i\sigma_y \otimes \mathbb{1}|s\rangle$, and $i\sigma_z \otimes \mathbb{1}|s\rangle$.
- [23] J. Shang, H. K. Ng, A. Sehwat, X. Li, and B.-G. Englert, Optimal error regions for quantum state estimation, *New J. Phys.* **15**, 123026 (2013).
- [24] M. Evans, *Measuring Statistical Evidence Using Relative Belief*, Monographs on Statistics and Applied Probability Vol. 144 (CRC Press, Boca Raton, FL, 2015).
- [25] X. Li, J. Shang, H. K. Ng, and B.-G. Englert, Optimal error intervals for properties of the quantum state, *Phys. Rev. A* **94**, 062112 (2016).
- [26] J. Emerson, R. Alicki, and K. Życzkowski, Noise estimation with random unitary operators, *J. Opt. B: Quantum Semiclassical Opt.* **7**, S347 (2005).
- [27] H. Akaike, Information theory and an extension of the maximum likelihood principle, in *Proceedings of the 2nd International Symposium on Information Theory*, edited by B. N. Petrov and F. Cási (Akadémiai Kiadó, Budapest, 1973).
- [28] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* **6**, 461 (1978).
- [29] M. Evans and Y. Guo, Measuring and controlling bias for some Bayesian inferences and the relation to frequentist criteria, [arXiv:1903.01696](https://arxiv.org/abs/1903.01696).
- [30] S. Hill and W. K. Wootters, Entanglement of a Pair of Quantum Bits, *Phys. Rev. Lett.* **78**, 5022 (1997).
- [31] A. Lovas and A. Andai, On the notion of quantum copulas, [arXiv:1902.08460](https://arxiv.org/abs/1902.08460).