Practical and reliable error bars for quantum process tomography

Le Phuc Thinh,[1,2] Philippe Faist,[3] Jonas Helsen,[1] David Elkouss,[1] and Stephanie Wehner[1]

[1] *QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*
[2] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore*
[3] *Institute for Quantum Information and Matter, California Institute of Technology, Pasadena CA, 91125, U.S.A.*

(Dated: August 2, 2018)

Current techniques in quantum process tomography typically return a single point estimate of an unknown process based on a finite albeit large amount of measurement data. Due to statistical fluctuations, however, other processes close to the point estimate can also produce the observed data with near certainty. Unless appropriate error bars can be constructed, the point estimate does not carry any sound operational interpretation. Here, we provide a solution to this problem by constructing a confidence region estimator for quantum processes. Our method enables reliable estimation of essentially any figure-of-merit for quantum processes on few qubits, including the diamond distance to a specific noise model, the entanglement fidelity, and the worst-case entanglement fidelity, by identifying error regions which contain the true state with high probability. We also provide a software package—QPtomographer—implementing our estimator for the diamond norm and the worst-case entanglement fidelity. We illustrate its usage and performance with several simulated examples. Our tools can be used to reliably certify the performance of e.g. error correction codes, implementations of unitary gates or more generally any noise process affecting a quantum system.

Usage: Secondary publications and information retrieval purposes.

PACS numbers: May be entered using the `\pacs{#1}` command.

Structure: You may use the `description` environment to structure your abstract; use the optional argument of the `\item` command to give the category of each item.

I. INTRODUCTION

Quantum technologies are improving at an ever faster pace, not only by a concentrated academic effort but increasingly via collaborations with industry. Quantum technologies require very precise manipulation and control of quantum systems, fueling the development of theoretical tools for precise calibration and characterization of quantum devices [1]. Notably, quantum state tomography and quantum process tomography (also known as quantum process tomography) can infer the quantum state or the quantum process that describes a quantum device, providing a natural "quantum debugger" [2].

Quantum state tomography aims to reconstruct the unknown state of a system with reference to a set of calibrated measurement apparatuses. Many reconstruction techniques—formally known as *estimators*—and their statistical properties have been developed and understood. These estimators can be roughly categorized into two groups based on the information they return about the unknown state. Point estimators take tomographic data from experiments and return a *single* quantum state, i.e. a density matrix, that best approximates the true unknown underlying physical state. Examples in this category are linear inversion and maximum likelihood estimators [3–5]. By contrast, region estimators return a *set* of quantum states in order to account for the uncertainty associated with the reconstruction. For state tomography many region estimators have been constructed, for instance, confidence regions [6–8] and Bayesian regions [9, 10]. Good region estimators have the advantage of providing robust statements associated with any chosen failure probability, that is, one can control the level of confidence with which the statement is made. Moreover, unlike point estimators, region estimators have sound operational interpretation under the influence of statistical fluctuations from finite data. Consequently, region estimators are suitable for the certification of quantum hardware for practical applications.

Many tools for quantum process tomography are adapted from quantum state tomography, for instance via the Choi-Jamiołkowski state-process correspondence [11]. Beyond traditional process tomography [12], there are also more advanced tools such as randomized benchmarking [13–16], gate-set tomography [17] and compressed sensing [18], that display certain advantages, such as a reduced number of required measurements. In the case of region estimators, some subtleties prevent a straightforward application of the corresponding tools for quantum states to quantum processes. Indeed, the set of quantum process is in one-to-one correspondence with only a subset of all bipartite states, namely those whose reduced state on one system is maximally mixed; this constraint has to be incorporated explicitly in the region estimator. In this paper, we enrich the statistical toolbox for quantum process tomography by providing a confidence region estimator for quantum process inspired by the state tomography method of Christandl and Renner [6].

Often in certifying specific applications, we are not interested in the full knowledge of the quantum process; a

property of the unknown channel suffices. For example, in quantum key distribution we are often interested in how close the final state output by the protocol is to the ideal key-state; this is captured for instance by the fidelity or the trace distance of the real state to the ideal state [19]. Likewise, in quantum computing a relevant figure-of-merit that enables fault-tolerant computation is the error threshold captured by the diamond distance or the worst-case entanglement fidelity of the real implemented gate relative to the ideal gate [20]. Note, though, that even a single figure of merit may serve as a full characterization of a process: A bound on the diamond distance or the entanglement fidelity to a given fixed channel confines the true channel to a small region in channel space. For these reasons, and because this significantly simplifies our analysis, we focus on estimators for quantum processes that report confidence intervals for a given figure of merit.

Summary of main results:

Our main contribution is three-fold:

(i) A confidence region estimator for channel tomography through the use of the Christandl-Renner-Faist estimator for states and the Choi-Jamiolkowski isomorphism between quantum states and quantum processs. We call this *the biparite-state sampling method*.

(ii) A new confidence region estimator to *directly* (without first tomographing the Choi state associated with the channel) estimate quantum processs and its proof of correctness. We call this *the channel space sampling method*.

(iii) A software package called *QPtomographer* [21] accompanying our theoretical results for analysing experimental data. Our software returns *quantum error bars* which captures all the information about the unknown channel derivable from the tomographic data and enables the user to construct confidence regions for any confidence level of interest.

By comparing the differences of the two estimators, we obtain a better understanding about the relationship between probability measures on state space and channel space which may be of independent interest. Because the estimators return a confidence region, they will work without any assumption on the prior distribution of the unknown process.

To illustrate how to use our result, we consider the scenario of *certifying a quantum memory* (an example of quantum property testing [22]). This corresponds to certifying that a quantum device (approximately) implements the identity channel. We consider three possible figures-of-merit: the diamond distance to the identity channel, the entanglement fidelity and the worst-case entanglement fidelity [23, 24]. Our method yields a reliable estimation of these figures-of-merit.

The paper is organized as follows. We first demonstrate in section II how one can use our method to obtain reliable information in a tomography experiment. The correctness

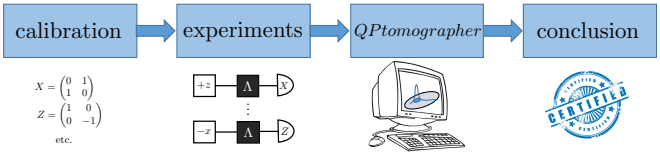

FIG. 1. The workflow of rigorous process tomography. Our data analysis *QPtomographer* supports both prepare-and-measure and ancilla-assisted experimental schemes. The conclusion is guaranteed without any prior information on the unknown quantum process.

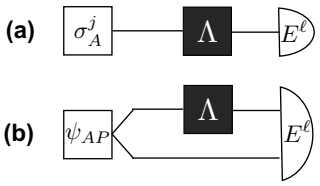

FIG. 2. Illustrations of (a) prepare-and-measure and (b) ancilla-assisted tomographic schemes for an unknown channel $\Lambda_{A\to B}$. In prepare-and-measure, one can only prepare input state σ_A^j which is fed into an unknown channel whose output state is measured by some POVM with elements E_B^ℓ. In ancilla-assisted, one can prepared entangled input state with some reference system P, and measure jointly the output using some POVM with elements E_{BP}^ℓ.

of our tools is justified in section III where we present the main results. Then we study the behavior of our numerical implementations in section IV before concluding our paper with future directions (section V). We leave the formal statements and detailed derivations of our results to the Appendices.

II. SETUP AND WORKFLOW

In this section, we detail the main workflow associated with the tomographic tools we have developed in our paper via a concrete example.

Suppose an experimental team has developed a working quantum memory (single qubit) and would like to certify its performance for usage within a quantum communication protocol such as entanglement distillation. In this context, one way of measuring the performance is the diamond norm distance to the identity process. The workflow for this example is illustrated in Fig. 1. We remark that there are other quantities of interest which do not assume an i.i.d structure, such as for example estimating the capacity as in [25].

The quantum memory's performance can be determined as follows. We assume that we have access to a given number of uses of the quantum memory. The number of uses can be chosen freely, noting that it affects the final error bars.

Moreover, in order to find out what the unknown process was, we need additional access to state preparation

and measurement devices which are information-complete (at least in the physical degrees of freedom where the unknown process acts). In this example, the set of state preparations are the Pauli eigenstates $|\pm x\rangle, |\pm y\rangle, |\pm z\rangle$, while the set of measurement devices are Pauli X, Y, Z measurements. We assume that each use of the quantum process are independent, and that the same unknown quantum process is applied for each run of the experiment, yielding statistics which are independent and identically distributed (i.i.d.). While here we consider a *prepare-and-measure* scenario as depicted in Fig. 2**(a)**, it is also possible to consider an *ancilla-assisted* scheme (Fig. 2**(b)**).

The first step (see Fig. 1) involves calibrating the state preparation and measurement devices to have $|\pm x\rangle, |\pm y\rangle, |\pm z\rangle$ state preparations and X, Y, Z measurements. After this calibration procedure has succeeded, one performs a chosen number $n = 45000$ of individual experiments. Each experiment consists of the following steps

- Prepare an input state by executing one of the devices $|\pm x\rangle, |\pm y\rangle, |\pm z\rangle$ (perhaps at random).
- Apply the (unknown) quantum memory to the said input state.
- Measure the output state using one of the possible X, Y, Z measurement devices (perhaps at random).
- Record the outcome of this experiment in a dataset E.

We remark that the preparation and measurement should yield sufficient data in the sense that all combination of input states and measurements should be chosen (perhaps at random).

Such a dataset E can then be analyzed by our software *QPtomographer*. One provides to our software the information about the measurement settings and the observed dataset. Then, using a Metropolis-Hastings sampling method, the software determines a specific type of distribution of the figure-of-merit Fig. 3 along with corresponding *quantum error bars* (v_0, Δ, γ). The value v_0 is the location of the maximum in Fig. 3, while Δ and γ measure the spread of the error. In our example, the analysis based on the input data set E with $n = 45000$ measurement records returned the quantum error bars

$$(v_0 = 0.058, \Delta = 0.006, \gamma = 0.00019),$$

which determine the parameters of an appropriate fit function (red curve of Fig. 3). The quantum error bars contain all the information about the error analysis. Namely, they (i) form a concise description of the error, (ii) provide an intuitive idea of the magnitude of the error, and (iii) can easily determine confidence regions for the quantum state or quantum process [7]. In this sense, quantum error bars are perfectly analogous to classical error bars: The latter are indeed a concise, intuitive description of the error from which one easily determines rigorous confidence intervals. For this reason it is a natural object to report at the end of a process tomography procedure.

If one wishes to actually derive rigorous confidence regions for the diamond norm distance, one may proceed as follows. First, one fixes a confidence level, say $\alpha = 99\%$, which sets the corresponding error parameter as $\epsilon = 1 - \alpha = 10^{-2}$. By Theorem 2, for $n = 45000$ (size of our dataset E) and $d_A = d_B = 2$, we need to find a region of diamond norm distance values with weight at least

$$1 - \frac{\epsilon}{2}\binom{2n + d_A^2 d_B^2 - 1}{d_A^2 d_B^2 - 1}^{-2} \geq 1 - 10^{-151},$$

With reference to Fig. 3, this means we need to find the x-position such that the area under the curve exceeds $1 - 10^{-151}$. A numerical integration leads to a region at least as large as $[0, 0.24]$. Together with the enlargement by

$$\delta = \sqrt{\frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln\binom{2n + d_A^2 d_B^2 - 1}{d_A^2 d_B^2 - 1}\right)} = 0.1$$

(to exclude nearby channels which could result in the same observed dataset with high probability) the final confidence region is $[0, 0.34]$. This means we have certified that the diamond norm distance of the unknown quantum memory to an ideal quantum memory is at most 0.34 with 99% confidence. In general, increasing the number n of measurement data points will shrink this confidence interval (due to the exponential decays in the diamond distance density, see also Appendix E).

We emphasize that the unnaturally large size of the regions is due in large part to a technical difficulty in the proofs of our bounds that is dealt with by employing tools that are known not to be tight in this context. For this reason, the quantum error bars are more informative than the actual final confidence regions.

This concludes the general workflow associated with our tomographic tools. The next section explain at a high level how our software transform tomographic data into confidence regions.

III. MAIN RESULTS

Our software package *QPtomographer* is built on top of two rigorously proven theoretical constructions. These are confidence region estimators based on the bipartite-state sampling method or the channel-space sampling method. The bipartite-state method works in the ancilla-assisted scheme, while the channel-space method works in both ancilla-assisted and prepare-and-measure schemes. This section gives a high level overview of the constructions together with the main ideas behind the proof of correctness, and leave the details to Appendix B and Appendix C, respectively. We begin with a brief motivation for confidence region estimators.

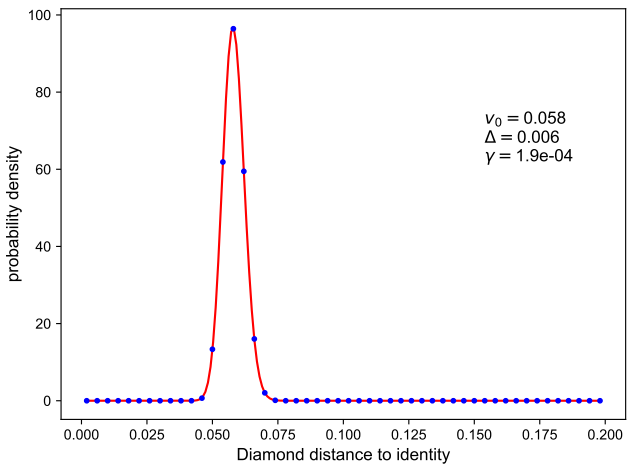

FIG. 3. Typical output from *QPtomographer*. The (blue) dots form the estimated distribution of values of the diamond norm distance to the identity channel as determined by the Metropolis-Hastings random walk. These are well-fitted to the (red) curve, which is compactly described by the triple of numbers (v_0, Δ, γ) which we called *quantum error bars*. Here v_0 is the position of the peak, Δ is half width at relative height $1/e$, and γ is a measure of skewness. These data encode information about the performance of the quantum memory, and enable us to construct confidence intervals certifying its quality.

A. Confidence region estimators

In the limit of infinite data (i.e. the number of records in dataset E is infinity), it is possible to exactly compute the probabilities of each measurement outcome from E and reconstruct the unknown channel by linear inversion on these observed probabilities [2]. However, in the practical scenario of finite data (i.e. dataset E contains n records) statistical fluctuations will imply the failure of all point estimation methods such as linear inversion or maximum likelihood estimation. This is due to the fact that channels close to the point estimate can produce the same dataset with high probability.

In order to make statistically rigorous and operationally sound statements on the unknown channel in this regime, we turn to region estimators, which are generalisations of the process of constructing error bars. We will look at a type of region estimators known as *confidence* region estimators. These are maps from data E to subsets $S_E \subseteq \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ of the set of quantum processes with the property that for all $\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\Pr_E[\Lambda \in S_E] \geq \alpha, \tag{1}$$

where α is a prefixed confidence level and the probability is evaluated over the random data E according to the distribution $\Pr[E|\Lambda]$. It is important to note that confidence is a property of the entire estimator (the procedure $E \mapsto S_E$) and not of any particular subset S_E produced by the estimator.

The operational meaning of confidence region estimators can be understood as follows. Suppose the black box implementing the unknown channel Λ_{true} is in fact prepared by a referee, who knows exactly which channel the black box applies. We proceed with a sequence of state preparations, applications of the channel and measurements of the output states to obtain a dataset E. Then we apply the estimator on E to get S_E. Repeating this procedure a large number of times, say $N = 10^5$, if r denotes different repetitions then we obtain different datasets $E(r = 1), ..., E(r = 10^5)$ with corresponding conclusions that the true channel Λ_{true} should be in the region $S_{E(r=1)}, ..., S_{E(r=10^5)}$. Now since the referee knows exactly the unknown channel, the referee can evaluate the *proportion of correct conclusions*

$$\frac{\left|\{r : r = 1, ..., N \text{ and } \Lambda \in S_{E(r)} \text{ is true } \}\right|}{N}.$$

If the estimator used is a confidence region estimator with confidence level $\alpha = 0.99$, then in the limit of $N \to \infty$ this proportion is at least 0.99. This is the meaning of confidence: the correct conclusion is guaranteed for a large number of uses of the estimator, regardless of the unknown channel. Note that for a specific use of the estimator which returns S_E, we *cannot* draw the conclusion that $\Lambda \in S_E$.

An alternative justification of confidence regions comes from a Bayesian point of view: Bayesian tomography uses outcomes of measurements to update a prior distribution about the quantum state to a posterior distribution. While this posterior clearly depends on the prior, it is known that when enough data is collected, the posterior distribution is no longer sensitive to the exact prior which was originally used (as long as the original prior has full support). Now consider a high-weight region of a posterior distribution, which is also known as a *credible region*. We may ask to what extent this region remains a credible region if we change the underlying prior. It turns out that for a large enough number of measurements, we may find regions which are credible regions for *any* prior, except for some exceptionally unlikely measurement datasets [6]. Such regions are precisely confidence regions.

B. Our confidence region estimators

Our method of constructing region estimators uses the information about the underlying unknown channel via the *likelihood function* defined generically for an observed dataset E as

$$\mathcal{L}(\Lambda|E) = \Pr(E|\Lambda), \tag{2}$$

where the probability of the dataset E under the assumption that the unknown channel is Λ is given by Born's rule. The specific form of the likelihood function depends on the scenarios and assumptions we postulate, c.f. Appendices B,C. The likelihood function can be seen as

giving a ranking about which channel best produce the observed dataset. We now present our methods of process tomography.

Bipartite-state sampling method: the main idea behind this method is that a quantum process is in correspondence with bipartite Choi states via the Choi-Jamiolkowski isomorphism. Hence, we can construct confidence regions for quantum states using the method of Christandl-Renner, and then perform an additional classical post-processing step to recover a confidence region for quantum processes.

Let us now first assume the use of an ancilla-assisted tomographic scheme Fig. 2**(b)**, which loosely corresponds to physically performing the Choi-Jamiolkowski isomorphism in the laboratory. This means having access to a full rank bipartite entangled state $|\psi_{AP}\rangle$ as input to the channel, and performing tomography on the output state $\rho_{BP} := \Lambda_{A\to B}(\psi_{AP})$ which is the unknown Choi state associated with the unknown channel.

Treating ρ_{BP} as the unknown state in a state tomography problem, we now apply the Christandl-Renner method of constructing confidence regions from tomographic data. Recall that the Christandl-Renner confidence region is constructed from the measure

$$\mathrm{d}\mu_E(\sigma_{AB}) := c_E^{-1}\mathrm{tr}(\sigma_{AB}^{\otimes n}E)\mathrm{d}\sigma_{AB} \tag{3}$$

where $c_E = \int \mathrm{tr}(\sigma_{AB}^{\otimes n}E)\mathrm{d}\sigma_{AB}$ is the normalizing constant, and $\mathrm{d}\sigma_{AB}$ is the uniform distribution on bipartite density matrices (obtained by tracing out a Haar random pure state on a larger space). Note that $\mathrm{tr}(\sigma_{AB}^{\otimes n}E)$ is the likelihood function for the outcome E given the state σ_{AB} in this scenario. Confidence regions for the unknown ρ_{AB} can be constructed from $\mathrm{d}\mu_E(\sigma_{AB})$ as the following proposition asserts.

Theorem 1 (Christandl & Renner [6], informal)**.** *Let n be the number of systems measured by a POVM during tomography and $1-\epsilon$ be the desired confidence level. Let $S_{\mu_E} \subseteq \mathrm{D}(\mathcal{H}_{AB})$ be any set of bipartite states with high weight under the probability measure $\mathrm{d}\mu_E(\sigma_{AB})$. Then the enlargement in purified distance $S^{\delta}_{\mu_E}$ where*

$$\delta = \sqrt{\frac{2}{n}\left(\ln\frac{2}{\epsilon} + 2\ln s_{2n,d_{AB}^2}\right)} \tag{4}$$

with $s_{n,d} := \binom{n+d-1}{d-1}$ is a confidence region of confidence level $1-\epsilon$.

Intuitively, we can think of the enlargement as a way to exclude nearby states/channels (relative to a proposed region of states/channels) that can give rise to the same observed dataset E with nonzero probability.

The confidence region $S^{\delta}_{\mu_E}$ contains bipartite quantum states which are not Choi states. This is due to the fact that the method of Christandl and Renner does not *a priori* allow the Choi state constraint $\mathrm{tr}_B(\sigma_{AB}) = \mathbb{1}_A/d_A$. Hence, we have to invent an additional post-processing step to map $S^{\delta}_{\mu_E}$ to a region consisting of exclusively Choi states. By the Choi-Jamiolkowski isomorphism we then have a confidence region for the unknown quantum process. The detailed explanation is left to Appendix B.

Channel-space sampling method: this method is a new construction of confidence region that directly returns channel-space confidence regions. Compared to the bipartite-state method, the channel-space method works in both the prepare-and-measure and ancilla-assisted tomographic schemes and takes into account the *a priori* knowledge that we are estimating a quantum process. This leads to computational efficiency relative to the bipartite-state method because the additional post-processing step of the bipartite-state method is not required here.

The estimator is constructed from the probability measure on the set of quantum processs $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\mathrm{d}\nu_E(\Lambda) := c_E'^{-1}\mathcal{L}(\Lambda|E)\mathrm{d}\nu(\Lambda) \tag{5}$$

where $\mathcal{L}(\Lambda|E)$ is the likelihood for the event E given a channel Λ, $c'_E = \int \mathcal{L}(\Lambda|E)\mathrm{d}\nu(\Lambda)$ serves as a normalizing constant and $\mathrm{d}\nu(\Lambda)$ is the Haar-induced measure on $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$. The likelihood function is adapted depending on prepare-and-measure or ancilla-assisted tomographic scheme and is defined as the probability of obtaining the dataset E given a channel Λ. Informally, this measure captures the information of the unknown channel as revealed by the observed dataset E in an unbiased manner (that is without using any prior knowledge on the unknown).

Given this measure, we obtain

Theorem 2 (informal)**.** *Let n be the number of channel uses during tomography and $1-\epsilon$ be the desired confidence level. Let $R_{\nu_E} \subseteq \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ be a set of channels with high weight under the probability measure $\mathrm{d}\nu_E(\Lambda)$. Then the enlargement in purified distance (for quantum processs, induced from states) $R^{\delta}_{\nu_E}$ where*

$$\delta = \sqrt{\frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d_{AB}^2}\right)}. \tag{6}$$

with $s_{n,d} := \binom{n+d-1}{d-1}$ is a confidence region with confidence level $1-\epsilon$.

Confidence interval for figures-of-merit: in practice, we choose the region in Theorem 2 for any chosen figure-of-merit to be the subset of channels whose figure-of-merit is better than a certain threshold. For the diamond norm distance to the ideal channel, we consider

$$R = \left\{\Lambda : 1/2\|\Lambda - \Lambda^{\mathrm{ideal}}\|_{\diamond} \leq \gamma_E\right\}, \tag{7}$$

and for the worst-case entanglement fidelity we consider

$$R = \left\{\Lambda : F_{\mathrm{worst}}(\Lambda) \geq \gamma_E\right\}. \tag{8}$$

We can work directly with the figure-of-merit by push-forwarding the measure $\mathrm{d}\nu_E(\Lambda)$ to the space of figures-of-merit, which is typically the reals $\mathbb{R}$ or the interval $[0,1]$,

and obtain the histogram $h(v)$ over different values of the figure-of-merit; the enlargement of these regions under the purified distance is translated into a loss in the value of the figures-of-merit: $\gamma_E \to \gamma_E + d_A\delta/2$ for diamond distance and $\gamma_E - d_A\delta$ for worst-case entanglement fidelity. The loss vanishes with increasing number of channel uses (as evident in Equation 4 and Equation 6), which allows reliable estimation of the figure-of-merit.

C. Numerical implementations

The previous section outlined the theoretical results underpinning our software package. We observe a reduction from the problem of constructing confidence regions to a problem of approximating the measures $\mathrm{d}\nu_E(\Lambda)$ or $\mathrm{d}\mu_E(\sigma_{AB})$. Solving this latter problem is the objective of the numerical implementations.

Computing $\mathrm{d}\nu_E(\Lambda)$ and $\mathrm{d}\mu_E(\sigma_{AB})$: in order to approximate a probability measure, we will take the Monte-Carlo approach of producing its samples, i.e. producing a histogram approximating a measure. More samples lead to better approximation but require more computational resources. Sampling according to $\mathrm{d}\mu_E(\sigma_{AB})$ (i.e. the biparite-state method) has been implemented in [7], and sampling according to $\mathrm{d}\nu_E(\Lambda)$ (i.e. the channel space method) can be obtained by similar methods. More precisely, $\mathrm{d}\nu_E(\Lambda)$ can be approximated by Metropolis-Hastings sampling [26] on channel space, which reduces to the ability of sample a "uniformly random quantum process" according to $\mathrm{d}\nu(\Lambda)$. To do this, it suffices to sample a unitary operator at random according to the Haar measure, by Stinespring dilation (see Appendix A 2). Crucially, because we use the Metropolis-Hasting algorithm, it is not necessary to calculate the normalizing constants c_E and c'_E which are difficult to obtain in practice. The parameters required to run the Metropolis-Hastings algorithm are the initial starting point and a jump distribution (a distribution from which we know how to produce samples). For the jump distribution, we have implemented two versions which we call e^{iH} and elementary rotation.

The Metropolis-Hastings algorithm starts with an initial point U_0 in the sample space, which we take to be the identity unitary operator, and conducts a random walk around this space. For each iteration, starting from current location U the jump distribution produces a candidate U' (depending on the current location) for a sample—a unitary matrix—which could potentially comes from $\mathrm{d}\nu_E(\Lambda)$. This candidate is accepted to be a sample of $\mathrm{d}\nu_E(\Lambda)$ with acceptance probability a, and upon acceptance the current location is updated to this point. The acceptance probability is defined to be the likelihood ratio (i.e. probability ratio) of U' to produce the observed dataset E with respect to the the current location U. This can be computed as the state preparations and measurements are known from calibration, and the dataset E is given from the experiment. The sequence of points $\{U_i\}$ visited in this fashion, albeit correlated, are asymptotically distributed according to $\mathrm{d}\nu_E(\Lambda)$ [26].

Extracting information for a given figure-of-merit: in terms of a given figure-of-merit f, the distribution $\mathrm{d}\nu(\Lambda)$ can be represented as a density function $h(v)$ for any possible value v of the figure-of-merit associated with the unknown channel Λ. For all practical purposes, our goal is to obtain a compact description of this density. Clearly, this function is well approximated by the sequence of values $\{f(U_i)\}$ derived from the output of the Metropolis-Hastings algorithm by simply evaluating the figure-of-merit at each point U_i. We organize $\{f(U_i)\}$ into bins of some size to produce a histogram approximating $h(v)$. This histogram is further subjected to a statistical fit analysis to obtain quantum error bars (v_0, Δ, γ), which contain enough information to reconstruct a good approximation of $h(v)$.

We consider two fit models in this paper. The fit model given in Ref. [7]

$$\ln \mu^{\mathrm{fit},\#1}(v) = -a_2 v^2 - a_1 v + m \ln v + c \tag{9}$$

does not have great agreement in our numerical examples (section IV) to the histogram bins. This leads us to develop an empirical model

$$\ln \mu^{\mathrm{fit},\#2}(v) = -a_2 v^2 - a_1 v + m (\ln v)^p + c \ , \tag{10}$$

which fits better to our examples (section IV). In any case, it is important to note that the functions $\mu(v)$ and $h(v)$ both decay exponentially fast (for the same reasons as in Ref. [7]). Hence, when trying to find high-weight regions it is not crucial to know the shape of the function exactly; rather, any imprecision on the shape of the function incurring an error on the estimated weight of a region, can be compensated by only a small increase in the region size (a property of the exponential function). Hence, whenever unspecified, we report quantum error bars as given using fit model #1 and as presented in Ref. [7], keeping in mind that in a paranoid setting one would have to adjust the confidence regions for the corresponding error. In summary, the reported quantum error bars are computed from the fit parameters of the fit model #1 as:

$$v_0 = \frac{1}{4a_2}\left[-a_1 + \sqrt{a_1^2 + 8a_2 m}\right] \ ; \tag{11a}$$

$$\Delta = \left(a_2 + \frac{m}{2v_0^2}\right)^{-1/2} \ ; \tag{11b}$$

$$\gamma = m\,\frac{\Delta^4}{6v_0^3} \ . \tag{11c}$$

See Appendix D and section IV for more details.

D. Relation between our two sampling methods

There is a connection between our two estimators, which we explain in detail in Appendix F. The essential difference

between the bipartite sampling method and the channel-space method can be traced back to how one uses the prior information about the input state. In the former, nothing is assumed about the exact input state other than what can be inferred directly from the measurement data (of course, still under the physical assumption of a pure entangled input); in the latter, the exact input state is assumed with certainty, and is used in the construction of the estimator (as manifestly visible in the likelihood function).

IV. APPLICATION: EXAMPLES

A. One-qubit example

We now illustrate in more details the use of our software package *QPtomographer* by continuing the quantum memory example. The generic procedure is described in Algorithm 2 and Algorithm 3.

Algorithm 1 quantum process Tomography

1: Perform data collection via Algorithm 2 or Algorithm 3
2: Generate random samples (channel space or bipartite)
3: Compute histogram of figure-of-merit
4: Fit analysis of histogram
5: **return** quantum error bars

Algorithm 2 Ancilla-Assisted (see Fig. 2**(b)**)

1: **input** a pure entangled state and a collection of measurements
2: **for** $i = 1$ to n **do**
3: Choose a measurement from the set
4: Apply the channel to the entangled input state
5: Measure the output state with the chosen measurement
6: Record the observed outcome
7: **end for**
8: **return** dataset E storing the measurement and outcomes for each repetition

Algorithm 3 Prepare-and-Measure (see Fig. 2**(a)**)

1: **input** a set of states and a collection of measurements
2: **for** $i = 1$ to n **do**
3: Choose an input state and a measurement from the set
4: Apply the channel to this input state
5: Measure the output state with the chosen measurement
6: Record the input choice and the observed outcome
7: **end for**
8: **return** dataset E storing input state, output measurement and outcomes for each repetition

The output of our classical data analysis is called "quantum error bars" which contain all the information about the figure-of-merit that can be obtained from the tomographic dataset. From here, it is easy to construct confidence regions for any specified confidence level.

Step 1. Data collection:

Consider the scenario of testing the performance of a quantum memory $\Lambda_{A\to B}$. The ideal channel we wish to implement is the identity channel $\mathcal{I}$. Suppose that the real channel implemented in the experiment the depolarizing channel

$$\Lambda_{A\to B}(\rho) = p\,\rho + (1-p)\,d_B^{-1}\mathbb{1}_B\ , \tag{12}$$

acting on one qubit ($d_A = d_B = 2$), with the parameter $p = 0.9$. In other words, the experiment is slightly off from the ideal implementation by some white noise.

Furthermore, we consider the ancilla-assisted scheme, and assume that the input to the channel is half of a pure entangled state $|\psi\rangle_{AP} = (\sigma_A^{1/2}\otimes\mathbb{1})\,d_A^{1/2}|\hat{\Phi}\rangle_{AP}$, where we choose

$$\sigma_A = \begin{pmatrix} 0.6 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}\ , \tag{13}$$

which mimics an input state which deviates slightly from the maximally mixed state. Note that the entangled input state has full Schmidt rank.

Since we do not have an actual experiment, we have to simulate Pauli measurements on the joint state ρ_{BP} after application of the channel $\Lambda_{A\to B}$, with 2 possible outcomes for each of the 3 measurement settings. For each measurement setting, 500 measurement outcomes were simulated. These constitutes the information contained in the (simulated) observed dataset E with $n = 45000$.

We now subject this dataset to an analysis which we aim to measure three figures-of-merit corresponding to our unknown channel: the diamond distance to the identity channel, the average entanglement fidelity and the worst case entanglement fidelity. Refer to the Appendix A for the precise definitions.

Step 2 and 3. Random sampling and histogram:

We use the methods developed in section III B to estimate the three figures-of-merit. The calculation of all three functions was done in C++ using the SCS toolbox [27, 28]. A simple Python interface was used to control the execution of the program. All numerics were run on a 2016 Macbook Pro with 4 physical/8 virtual cores using our code provided at [21].

First, we demonstrate the *bipartite-state method* described in Appendix B. This consists in running the random walk as implemented in Ref. [7], using directly the function (B14) as figure-of-merit. The random walk was used to sample a total of 32768 data points, using a binning analysis as described in Ref. [29], with a step size of ~ 0.001, a sweep size of ~ 1000 and using 2048 thermalization sweeps. Again, two choices of the jump distribution give similar results.

Second, we run the *channel-space method* of analysis as presented in Appendix C. The random walk is run on the space of all quantum processs, as described in

Appendix D, until 32768 data points have been collected. Two ways of performing the random walk (e^{iH} versus elementary rotation) yield similar results, with elementary rotation finishing faster than e^{iH}. Samples from the random walk allow to construct a numerical estimate of a specific distribution of the figure-of-merit, which contains all the necessary information in order to construct confidence regions.

The results are shown in Fig. 4 as the histogram (dot) points with legend label "his.". The histogram points correspond to the numerical estimation of $h(v)$ given by (C42) and $\mu(v)$ given by (B15).

Step 4: Fit analysis of histograms:

In each of these methods, the data—the points underlying the histograms—is fit to two different models as discussed. If good fit is achieved, we can take these models as a description of the histogram points, and therefore also a description of the functions $h(v)$ and $\mu(v)$.

In our example, we discovered that the fit model #1 as described in Equation 9 does not have great agreement with the underlying histogram bins, as underscored by goodness-of-fit values (reduced χ^2) of the order of ~ 25. This is because our (diamond distance, worst-case entanglement fidelity) figure-of-merit does not satisfy the requirements of the "heuristic derivation" in Ref. [7], and it is thus no surprise that the fit model does not align perfectly well with the data. Using the empirical model #2 yields much better agreement (solid curves in Fig. 4), with goodness-of-fit values (reduced χ^2) of ~ 2. Nevertheless, we reported quantum error bars using fit model #1.

Step 5. Quantum error bars and confidence regions:

The quantum error bars (v_0, Δ, γ) are a simple translation from the parameters of the fit model #1. The steps towards a confidence region for diamond norm has been illustrated in section II. In theory we have the guarantee that collecting a larger dataset will yield smaller regions converging to the true value. Unfortunately, the confidence interval for diamond norm distance returned by our method is unreasonably large for the current example: for 99% confidence level we are able to bound the diamond norm by 0.34 as compared to the true value of 0.05. We believe that this is due to operator inequality involved in bounding the failure probability (Proposition 1). Further research is needed to provide better construction of confidence regions (i.e. more efficient in terms of the number of data samples n).

B. Two-qubits example

Now we consider a two-qubit example to illustrate the practicality of our method in this situation. This example also shows that the channel-space and the biparite-state sampling methods do not in general produce the same histogram.

Suppose that the real channel implemented in the experiment the two-qubits depolarizing channel

$$\Lambda_{A\to B}(\rho) = p\,\rho + (1-p)\,d_B^{-1}\mathbb{1}_B\ , \tag{14}$$

with $d_A = d_B = 4$ and we are interested in the diamond distance to the identity channel. Assuming access to state preparation that produces $|\psi\rangle_{AP} = (\sigma_A^{1/2}\otimes\mathbb{1})\,d_A^{1/2}|\hat{\Phi}\rangle_{AP}$ with

$$\sigma_A = \begin{pmatrix} 0.35 & 0 & 0.04 & 0.1i \\ 0 & 0.15 & 0.05 & 0 \\ 0.04 & 0.05 & 0.32 & 0 \\ -0.1i & 0 & 0 & 0.18 \end{pmatrix}\ , \tag{15}$$

and $3^4 = 81$ Pauli measurement settings each having $2^2 = 4$ outcomes. We perform similar analyses on a simulated dataset of size $n = 40500$ which we generated using the state preparations and measurements described above. The result is presented in Fig. 5.

The channel-space sampling method's $h(v)$ is peaked at lower values of the figure-of-merit, as can be seen in Fig. 5. We observe that, in this case, the knowledge of the input state significantly shifts the corresponding histogram distribution towards lower values of the figure-of-merit, allowing to construct smaller confidence regions. Based on several examples studied, this is not always the case; with less noise (smaller p), for instance, the curve for $\mu(v)$ and the curve for $h(v)$ get closer to each other.

On a technical level, we show that the Hilbert-Schmidt measure over the bipartite states factorizes as a measure over states on the input system and the relevant measure over all channels (Appendix F). Hence, a large uncertainty over the input state may enlarge the resulting region as opposed to considering a region only on the channel space for a fixed known input state. However, it is not impossible that under some lucky circumstances a finite distribution width on the input state helps add more weight to regions of a higher figure of merit, effectively shrinking the region. Indeed, it could happen that the input state assumed in the channel-space method is far from the optimal state for distinguishing the channels in terms of the diamond norm; in such a case a prior which is more "smeared out" over different input states might result in smaller quantum error bars for the diamond norm. We believe that this is why neither method performs globally better than the other. See Appendix F for further details on the relationship between the two methods.

V. CONCLUSIONS

One might think that carrying over the notion of quantum error bars in quantum state tomography to quantum process tomography is as straightforward as converting quantum states to channels via the Choi-Jamiołkowski isomorphism. However, our study reveals a more complicated structure. We find that different analysis methods are suited to different experimental process tomography

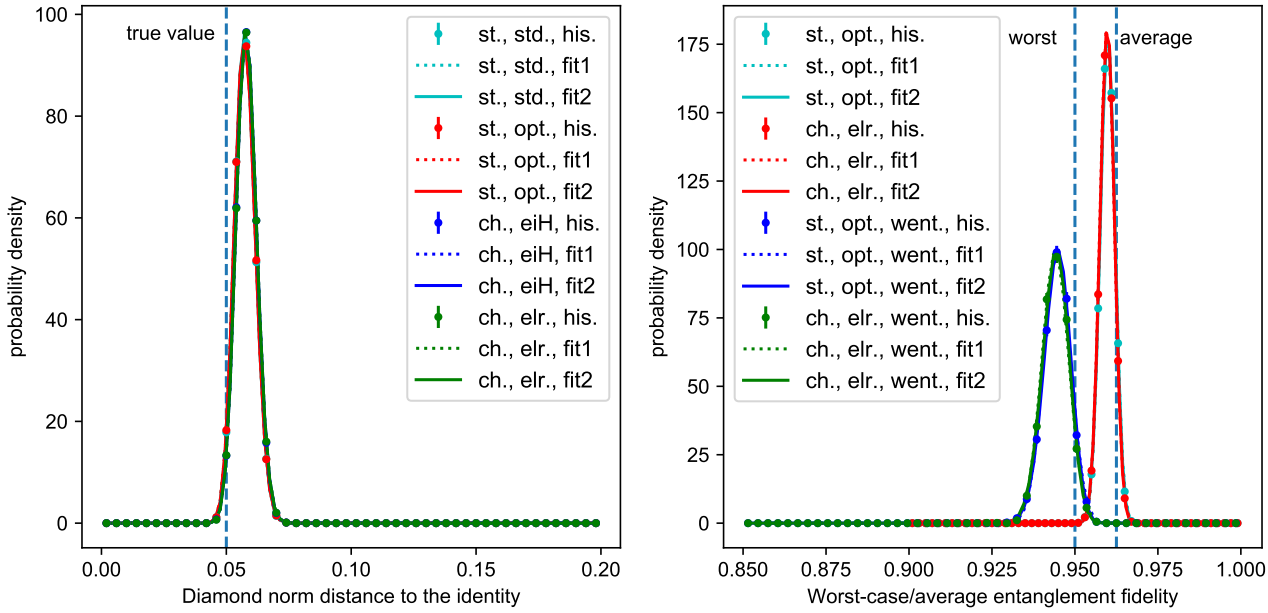

FIG. 4. Distribution of the figures-of-merit for the single-qubit process example (left: diamond norm distance, right: entanglement fidelities) relevant to construct confidence regions. Vertical dashed lines are true figure-of-merit values of the unknown channel. The biparite-state sampling method (legend st., shorthand for *state*) consists in estimating the diamond norm while ignoring the information about the exact input state to the channel. In contrast, the channel-space method (legends ch., shorthand for *channel*) uses the information about the input state to obtain better bounds on the figures-of-merit. Within each method, we also plot the results obtain from different jump distributions (legends eiH, elr. for channel-space, and std., opt. for bipartite-state) used in the Metropolis-Hastings random walk. The dotted curves are the fits of the raw histogram bins (legend his.) according to our fit model #1, with corresponding *quantum error bars* (v_0, Δ, γ) [7], while the solid curves are fits using our improved, empirical fit model #2. These plots should be understood as tools to construct confidence regions, i.e., given a threshold on the x-axis, one may easily calculate from these curves the confidence with which one may ascertain the true figure-of-merit (see main text).

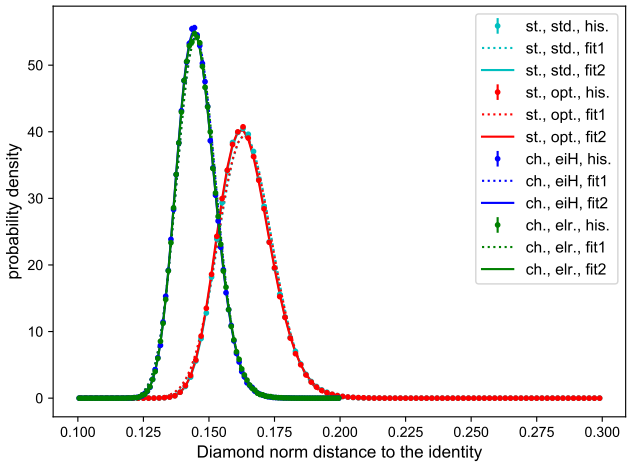

FIG. 5. Distribution of the diamond norm distance for the two-qubit process example. The difference between the two methods for the entanglement fidelity is not a contradiction, rather, in this situation the channel space method gives better tomographic results compared to the biparite-state method. The reason is due to the additional use of prior information about the input state in the channel space method.

setups. In the experimentally more realistic prepare-and-measure scheme, a judicious use of the prior knowledge about the input state to the process allows us in typical situations to obtain tighter quantum error bars for the process. These results are obtained by developing a new method, along with corresponding proofs, which are specific to process tomography. On the other hand, in the case of the ancilla-assisted scheme, we can directly apply the methods developed for quantum state tomography, harnessing them to directly yield reliable statements about the quantum process itself, while ignoring any information the measurements provide about the input state used to probe the process.

We hence provide a fully-fledged and practical toolbox named *QPtomographer*, with solid theoretical foundations, for quantum process tomography of arbitrary quantum processes, using any experimental quantum process tomography setup, and given measurement outcomes from any measurement settings. Our software package facilitates the numerical analysis in practice by automating the implementation of the Metropolis-Hastings random walk, as well as the calculation of the diamond norm, by simple high-level Python function calls, while transparently delegating the computation-intensive routines to heavily

optimized C++ code which makes use of modern programming techniques including template metaprogramming and exploiting hardware SIMD instructions.

On the spectrum of characterization tools for quantum devices, our method can be seen as lying on the opposite end of randomized benchmarking [13–16]. While slightly more involved, our technique can be applied to any choice of state preparations and measurements, and can be applied to any individual process. By determining the diamond norm or the worst-case entanglement fidelity to any given ideal process, we provide individual full characterization of the processes implemented by individual gates. More generally our methods allow the reliable estimation of any specific property of the quantum process.

We note that our method is currently limited to processes acting on few qubits, as our confidence region produces unreasonably large regions, and the algorithm stores dense representations of the quantum process. However, we expect that our methods will be used to certify individual components of complex setups, for instance, individual 2-qubit gates. Because we estimate robust, composable figures-of-merit such as the worst-case entanglement fidelity or the diamond norm, the composition of individually certified components is still certified to function accurately.

Finally, we may ask whether the channel method is always superior to the bipartite sampling method. As noted above, the additional prior knowledge about the input state which the channel-space method enjoys in contrast to the bipartite sampling method is not sufficient to guarantee this. We leave a more precise understanding of the relation between our two methods open for future study.

ACKNOWLEDGMENTS

TLP, JH, DE and SW are supported by an ERC Starting Grant (SW), an NWO VIDI Grant (SW), and an NWO Zwaartekracht grant (QSC). PhF acknowledges support from the Swiss National Science Foundation through the Early PostDoc.Mobility Fellowship No. P2EZP2_165239 hosted by the Institute for Quantum Information and Matter (IQIM) at Caltech, as well as from the National Science Foundation.

Appendix A: Notations & preliminaries

We begin by setting up some notations and recalling standard definitions. For more information on states and processes see [30–32].

1. Quantum processes and figures-of-merit

Let $\mathcal{H}_A$ be the Hilbert space of dimension d_A associated with the quantum system denoted A. By $\mathrm{D}(\mathcal{H}_A)$ we mean the subset of $\mathrm{End}(\mathcal{H}_A)$—the set of linear transformations on $\mathcal{H}_A$—consisting of density matrices $\rho_A \geq 0$ (positive semidefinite) with $\mathrm{tr}(\rho_A) = 1$. Composite systems are described by tensor product constructions, for instance $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ is the Hilbert space of composite system AB.

Quantum measurements on quantum system are the positive operator valued measures or POVMs on $\mathcal{H}$. For finite number of outcomes, a POVM is a set of positive operators—the effects—that sum to the identity operator on $\mathcal{H}$. We will overload the notation E to mean an outcome label, and also the effect E (i.e. an operator/matrix) in the POVM. This is equivalent to the usual "observables" formulation of measurement, i.e. a hermitian operator. For example, a Z measurement/observable has two outcomes $E = +1$ and $E = -1$ with associated effects $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, respectively.

A quantum process $\Lambda_{A\to B}$ mapping a quantum system A to a quantum system B is a completely positive trace-preserving linear map from $\mathrm{End}\left(\mathcal{H}_A\right)$ to $\mathrm{End}\left(\mathcal{H}_B\right)$. In general we will denote quantum processs by capital greek letters. We will often drop the subscripts when the quantum systems are clear from the context.

The set of all possible quantum processs is denoted $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$, and it is in one-one correspondence with the set of bipartite Choi states $\mathscr{C}(\mathcal{H}_{AB})$ via the Choi-Jamiolkowski isomorphism

$$\begin{aligned} J : \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) &\to \mathrm{End}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right) \\ \Lambda_{A\to B} &\mapsto (I_A \otimes \Lambda_{\bar{A}\to B})(|\hat{\Phi}\rangle\langle\hat{\Phi}|_{A\bar{A}}) \end{aligned} \tag{A1}$$

where $|\hat{\Phi}\rangle := \frac{1}{\sqrt{d_A}} \sum_k |k\rangle_A |k\rangle_{\bar{A}}$ is the maximally entangled state on $\mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ and I_A is the identity channel acting on the system A. Explicitly, the set of Choi matrices is defined as the image of $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ under the Choi-Jamiolkowski isomorphism and has the following compact description

$$\mathscr{C}(\mathcal{H}_{AB}) = \{\rho \in \mathrm{D}(\mathcal{H}_{AB}) : \mathrm{tr}_B(\rho_{AB}) = \mathbb{1}_A/d_A\} \ . \tag{A2}$$

Throughout the appendix, we will use the *convention* that Λ_{AB} is the Choi state associated with the channel $\Lambda_{A\to B}$.

The action of the channel can be recovered from its Choi state by the inverse of Choi-Jamiolkowski isomorphism

$$\Lambda(\rho) = d_A \mathrm{tr}_A(\Lambda_{AB} \cdot \rho_A^{\intercal} \otimes \mathbb{1}_B)\,, \tag{A3}$$

where $\intercal$ is the transpose with respect to the basis of $\mathcal{H}_A$ defining the maximally entangled state.

Recall that the fidelity between two states σ, σ' is defined as

$$F(\sigma, \sigma') := \left\|\sqrt{\sigma}\sqrt{\sigma'}\right\|_1 = \mathrm{tr}\sqrt{\sqrt{\sigma}\sigma'\sqrt{\sigma}}\,, \tag{A4}$$

and the purified distance between quantum states is defined as $P(\sigma, \sigma') := \sqrt{1 - F(\sigma, \sigma')^2}$. Then the purified distance between channels is defined as

$$P(\Psi_{A\to B}, \Psi'_{A\to B}) := P(\Psi_{AB}, \Psi'_{AB})\,. \tag{A5}$$

a. *Diamond distance*

We first introduce the familiar diamond distance. The diamond distance from the real or actual implementation $\Lambda_{A\to B}$ to the ideal or target implementation Λ_{ideal} is denoted as

$$f_\diamond(\Lambda_{A\to B}) = \frac{1}{2}\|\Lambda_{A\to B} - \Lambda^{\mathrm{ideal}}_{A\to B}\|_\diamond. \tag{A6}$$

This function $f_\diamond : \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) \to [0, 1]$ (or equivalently from $\mathscr{C}(\mathcal{H}_{AB})$) can be cast as a semidefinite program [33]

Primal problem

$$\begin{aligned} \text{maximize:} \quad & \left\langle d_A\Lambda_{AB} - d_A\Lambda^{\mathrm{ideal}}_{AB}, W\right\rangle \\ \text{subject to:} \quad & W \le \mathbb{1}_{\mathcal{B}} \otimes \rho, \\ & W \ge 0, \\ & \rho \in \mathrm{D}(\mathcal{X}). \end{aligned}$$

Dual problem

$$\begin{aligned} \text{minimize:} \quad & \|\mathrm{tr}_B(Z)\|_\infty \\ \text{subject to:} \quad & Z \ge d_A\Lambda_{AB} - d_A\Lambda^{\mathrm{ideal}}_{AB}, \\ & Z \ge 0. \end{aligned}$$

b. *Entanglement fidelity*

The entanglement fidelity is another measure of how close a given channel is to the identity channel. More specifically, it measures how well a channel preserves the maximally entangled state.

The *entanglement fidelity* of a channel $\Lambda_{A\to B}$ with $B \simeq A$ is defined as

$$F_e(\Lambda) = F^2(\Lambda_{\bar{A}\to B}(\hat{\Phi}_{A\bar{A}}), \hat{\Phi}_{AB})\ , \tag{A7}$$

recalling that $|\hat{\Phi}\rangle_{A\bar{A}}$ is the normalized maximally entangled state between the systems A and $\bar{A}$.

Because $\Lambda_{AB} = \Lambda_{\bar{A}\to B}(\hat{\Phi}_{A\bar{A}})$ is the normalized Choi state corresponding to the channel $\Lambda_{A\to B}$, the entanglement fidelity of the channel $\Lambda_{A\to B}$ is in fact exactly the fidelity of the corresponding normalized Choi state to the maximally entangled state:

$$F_e(\Lambda) = F^2(\Lambda_{AB}, \hat{\Phi}_{AB})\ . \tag{A8}$$

c. Worst-case entanglement fidelity

The worst-case entanglement fidelity is a better measure of the reliability of the channel to simulate the identity channel, if we have to worry about any possible input state being fed into the channel. In effect, the worst-case entanglement fidelity measures how well the channel preserves any given state on a system and any purification. It is defined as

$$F_{\text{worst}}(\Lambda_{A\to B}) = \inf_{\sigma_{A\bar{A}}} F^2(\Lambda_{\bar{A}\to B}(\sigma_{A\bar{A}}), \sigma_{AB}) \ , \tag{A9}$$

where the optimization ranges over all bipartite quantum states σ_{AB} defined over the input $\bar{A}$ and a reference system $A \simeq \bar{A}$. The optimization variable, which appears in both slots of the fidelity F, may be restricted to pure states without loss of generality.

Now we show that the worst-case entanglement fidelity can be computed by evaluating a simple semidefinite program. That a semidefinite program formulation of the worst-case entanglement fidelity can be used in the context of quantum error correction to find suitable recovery procedures for fixed input were put forth in refs. [34, 35]. We build upon those constructions to optimize over the input state, while in our case the problem is simplified as there is no recovery operation. Using our notation, we write

$$F_{\text{worst}}(\Lambda_{A\to B}) = \inf_{|\,\phi\rangle_{A\bar{A}}} F^2(\Lambda_{\bar{A}\to B}(\phi_{A\bar{A}}), \phi_{AB}) = \inf_{T_A\colon \operatorname{tr}(TT^\dagger)=1} \operatorname{tr}(\Lambda_{\bar{A}\to B}(T_A \tilde{\Phi}_{A\bar{A}} T_A^\dagger)\, T_A\, \tilde{\Phi}_{AB} T_A^\dagger) \ , \tag{A10}$$

where we have defined the non-normalized maximally entangled state $|\,\tilde{\Phi}\rangle_{AB} = d_A^{1/2}\,|\,\hat{\Phi}\rangle_{AB}$. The last equality comes from the fact that any bipartite pure state $|\,\phi\rangle_{A\bar{A}}$ can be parametrized by a complex matrix T_A satisfying $\operatorname{tr}(T_A T_A^\dagger) = 1$ via $|\,\phi\rangle_{A\bar{A}} = T_A\,|\,\tilde{\Phi}\rangle_{A\bar{A}}$, with moreover $\operatorname{tr}_{\bar{A}}(\phi_{A\bar{A}}) = T_A T_A^\dagger$ (indeed, choose T_A with matrix elements $\langle i|T|j\rangle_A = \langle i, j\,|\,\phi\rangle$). Then, with $T'_A := T_A^\dagger$, and noting that all $T'_A T'^\dagger_A$ with $\operatorname{tr}(T'_A T'^\dagger_A) = 1$ can be written as a density matrix $\rho_A = T'_A T'^\dagger_A$, we have

$$\begin{aligned}
(\text{A10}) &= \inf_{T'_A\colon \operatorname{tr}(T'^\dagger T')=1} \operatorname{tr}(T'_A\, T'^\dagger_A\, \Lambda_{\bar{A}\to B}(\tilde{\Phi}_{A\bar{A}})\, T'_A\, T'^\dagger_A\, \tilde{\Phi}_{AB}) \\
&= \inf_{\rho_A \geqslant 0\colon \operatorname{tr}(\rho_A)=1} \langle \tilde{\Phi}\,|_{AB}\, \rho_A\, \Lambda_{\bar{A}\to B}(\tilde{\Phi}_{A\bar{A}})\, \rho_A\,|\,\tilde{\Phi}\rangle_{AB} \ .
\end{aligned} \tag{A11}$$

This is a minimization over a positive semidefinite quadratic form in $\rho_A\,|\,\tilde{\Phi}\rangle_{AB}$, so it is (quite surprisingly) a convex optimization in terms of ρ_A. We know that positive semidefinite quadratic optimizations may be written as semidefinite programs. Indeed, for any positive semidefinite matrix $Q = MM^\dagger$, we have that $\langle\psi|Q|\psi\rangle \leqslant \mu$ if and only if $\left[\begin{smallmatrix} \mathbb{1} & M^\dagger|\,\psi\rangle \\ \langle\psi\,|\,M & \mu \end{smallmatrix}\right] \geqslant 0$. So, finally, we may write the worst-case entanglement fidelity as a semidefinite program in terms of the real variable μ and the positive semidefinite variable $\rho_A \geqslant 0$:

$$\begin{aligned}
F_{\text{worst}}(\Lambda_{A\to B}) \quad = \quad & \text{minimize:} && \mu && , \\
& \text{subject to:} && \operatorname{tr}(\rho_A) = 1 && \\
& && \left[\begin{array}{c|c} \mathbb{1} & M_{AB}^\dagger \rho_A\,|\,\tilde{\Phi}\rangle_{AB} \\ \hline \langle\tilde{\Phi}\,|_{AB}\, \rho_A M_{AB} & \mu \end{array}\right] \geqslant 0 &&
\end{aligned} \tag{A12}$$

where M_{AB} is a factorization of the nonnormalized Choi matrix of the process, satisfying

$$M_{AB}\, M_{AB}^\dagger = d_A \Lambda_{AB} = \Lambda_{\bar{A}\to B}(\tilde{\Phi}_{A\bar{A}}) \ . \tag{A13}$$

The factorization can be obtained using a Cholseky or LDLT factorization, for instance; or more generally by computing any matrix square root. The unitary freedom of the matrix square root decomposition (i.e., the freedom of redefining $M \to MU$) is irrelevant here.

2. Haar induced measures

Later, we will base our confidence region estimators on the following two "uniform" measures. They are both measures induced by the unique Haar measure on the unitary group $\mathbb{U}(\mathcal{H})$ acting on some Hilbert space.

The first measure is defined on the set of mixed quantum states [36]. Since any density matrix has a (nonunique) purification, the space $\mathrm{D}(\mathcal{H}_{AB})$ admits a purification space $\mathrm{Pure}(\mathcal{H}_{ABA'B'})$ whose elements are rank one density operators on $\mathcal{H}_{ABA'B'}$ with $A'B'$ being an isomorphic copy of AB. The Haar measure $dU_{ABA'B'}$ then induces a measure on $\mathrm{Pure}(\mathcal{H}_{ABA'B'})$ via the relation $|\psi\rangle\langle\psi| = U|\psi_0\rangle\langle\psi_0|U^\dagger$ for an arbitrary pure state $|\psi_0\rangle$, which induces a measure $\mathrm{d}\sigma_{AB}$ on $\mathrm{D}(\mathcal{H}_{AB})$ by partial tracing.

The second measure is defined on the set of quantum processs, or equivalently on the set of bipartite Choi states. Let

$$\mathscr{PC} = \left\{|\Psi\rangle \in \mathcal{H}_{ABA'B'} : \mathrm{tr}_{BA'B'}(|\Psi\rangle\langle\Psi|) = d_A^{-1}\mathbb{1}_A\right\} \tag{A14}$$

be the set of purifications of arbitrary Choi states. Without loss of generality, let us define a fixed reference pure state in $\mathscr{PC}$

$$|\Psi_0\rangle := \frac{1}{d_A}\sum_{i=1}^{d_A}|i\rangle_A|v_i\rangle_{BA'B'} \tag{A15}$$

with $\{|v_i\rangle_{BA'B'}\}$ some fixed orthonormal set of vectors. Then for all $|\Psi\rangle \in \mathscr{PC}$, there exists a unitary $U_{BA'B'}$ such that $|\Psi\rangle = \mathbb{1}_A \otimes U_{BA'B'}|\Psi_0\rangle$. This relation transfer the unique Haar measure $\mathrm{d}U_{BA'B'}$ on the unitary group $\mathbb{U}(\mathcal{H}_{BA'B'})$ to a measure on $\mathscr{PC}$ which we will denote as $\mathrm{d}\nu(|\Psi\rangle)$. Again, by partial tracing the system $A'B'$, this measure induces the measure $\mathrm{d}\nu(\Psi_{AB})$ on Choi states $\mathscr{C}(\mathcal{H}_{AB})$ (also denoted as $\mathrm{d}\nu(\Lambda_{AB})$ by changing the dummy variable). Finally, taking the inverse of the Choi-Jamiolkowski isomorphism gives the induced measure $\mathrm{d}\nu(\Psi_{A\to B})$ (or in a different notation $\mathrm{d}\nu(\Lambda_{A\to B})$) on channel space $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ which is the starting point of the channel space sampling method.

The relation between these measures will be discussed in Appendix F when we compare the two region estimators.

3. The i.i.d. hypothesis

In this paper, *we work under the assumption of i.i.d. (independent and identically distributed) channels.* This means any time we use the experimental device, it is assumed that one and the same transformation $\Lambda_{A\to B}$ has been applied. Experimentally, this assumption is well justified if the same experimental conditions can be reproduced because the abstract channel is a function of the working parameters of the physical device. The i.i.d. hypothesis also gives a clear operational meaning to the question: to which object does the tomographic statement apply? It is one and the same $\Lambda_{A\to B}$ which does not vary from past to future uses.

Even though we work under the i.i.d. assumption, we note that this can be weakened to permutation invariant through the use of the quantum de Finetti theorem for channels [37].

Before proceeding further, we give a clarifying remark about our notation. We usually consider n uses of a channel Λ. Under the i.i.d. assumption we can describe this situation by tensor product construction giving a composite channel $\Lambda^{\otimes n}$ acting on the composite Hilbert space $\mathcal{H}_A^{\otimes n}$ and transform the system to $\mathcal{H}_B^{\otimes n}$. As usual, by measurement on $\mathcal{H}_B^{\otimes n}$ and by knowing the input state on $\mathcal{H}_A^{\otimes n}$ we can perform tomography of the unknown channel. Our convention has been to denote a measurement on $\mathcal{H}_B^{\otimes n}$ by a POVM $\{E\}$ with E standing for both the labels of the various outcomes and the actual operators/matrices. This captures both i.i.d. measurements and entangled measurements in the following sense. Suppose $n = 2$ and we perform X and Z on each subsystem. This can be equivalently described by two POVMs $\{|+x\rangle\langle+x|, |-x\rangle\langle-x|\}$ and $\{|+z\rangle\langle+z|, |-z\rangle\langle-z|\}$, and then by tensor product construction combined into a single POVM on the composite Hilbert space. However, this is not the only measurement that one can do: one can perform the Bell measurement projecting into the four maximally entangled states. Our description and notation is flexible for arbitrary measurement one can perform.

Appendix B: The bipartite-state sampling method

This method requires experimentalists to work in the ancilla-assisted scheme (see Fig. 2**(b)**): we select a full Schmidt rank entangled state ψ_{AP}, a collection of bipartite measurements $E^{(\ell)}$ with corresponding effects $E_k^{(\ell)}$, and assume the experiment can implement the channel $\Lambda \otimes \mathcal{I}$, where $\mathcal{I}$ is the identity map. Again, the collection of measurement should be informationally complete if one wishes to infer full information about the channel. We assume knowledge of the state preparations and measurements in the form of matrices in the computational basis. This means the pure entangled state has the form

$$|\psi\rangle_{AP} = \sum_i s_i|i\rangle_A|i\rangle_P = \sqrt{d_A}\psi_A^{1/2}|\hat{\Phi}\rangle_{AP} = \sqrt{d_A}\psi_P^{1/2}|\hat{\Phi}\rangle_{AP}, \tag{B1}$$

where ψ_A, ψ_P are the respective reduced states on A and P of $|\psi_{AP}\rangle\langle\psi_{AP}|$ and $|\hat{\Phi}\rangle_{AP}$ the maximally entangled state on $\mathcal{H}_{AP}$. Note that not all pure state on AP has this form, but we assume it without loss of generality by redefining $|\hat{\Phi}\rangle_{AP}$ if necessary.

The tomography procedure proceeds according to Algorithm 2. In each round, we prepare $|\psi\rangle_{AP}$ and we apply the unknown channel $\Lambda_{A\to B}\otimes\mathcal{I}_{P\to P}$. We then perform a measurement on the bipartite output system BP using a setting of our choice, yielding an outcome POVM effect $E_k^{(\ell)}$. The dataset stores all the outcomes of different rounds.

In other words, the ancilla-assisted scheme actually realizes the (theoretical) Choi-Jamiołkowski isomorphism in the laboratory under the assumption of the input state and the channel.

The likelihood function for this scheme is given by

$$\mathcal{L}_{\mathrm{AA}}(\Lambda|E) = \prod_{k,\ell}\left[\mathrm{tr}(\Lambda_{A\to B}(\psi_{AP})\,E_k^{(\ell)})\right]^{n_{k,\ell}}\,, \tag{B2}$$

where $n_{k,\ell}$ is the number of times the POVM effect $E_k^{(\ell)}$ appears in the dataset E. Since

$$\Lambda_{A\to B}(\psi_{AP}) = d_A\psi_P^{1/2}\,\Lambda_{BP}\,\psi_P^{1/2} \tag{B3}$$

where Λ_{BP} is the corresponding Choi state, we have

$$\mathcal{L}_{\mathrm{AA}}(\Lambda|E) = \prod_{k,\ell}\left[d_A\mathrm{tr}(\Lambda_{BP}\,\psi_P^{1/2}\,E_k^{(\ell)}\,\psi_P^{1/2})\right]^{n_{k,\ell}} = d_A^n\mathrm{tr}\left(\Lambda_{BP}^{\otimes n}\bigotimes_{k,\ell}\psi_P^{1/2}\,E_k^{(\ell)}\,\psi_P^{1/2}\right)\,, \tag{B4}$$

where $\bigotimes_{k,\ell}$ ranges over the observed dataset E

Since quantum processs correspond to bipartite quantum states via the Choi-Jamiolkowski isomophism, we can generalize quantum state tomography methods to quantum processs. Here, we *directly* apply the existing procedure of Faist and Renner [7] designed for quantum states to infer information about quantum processs. The main result in this section is Theorem 3. We first recall the procedure of constructing confidence region estimators for quantum states, phrased in terms of bipartite states in anticipation with the connection to quantum processs.

1. Christandl-Renner confidence regions

Given access to n copies of an unknown state ρ_{AB}, we can perform a (joint or collective) POVM measurement on $\rho_{AB}^{\otimes n}$ and upon receiving the dataset E, the Christandl-Renner procedure outputs a distribution

$$\mathrm{d}\mu_E(\sigma_{AB}) := c_E^{-1}\mathrm{tr}(\sigma_{AB}^{\otimes n}E)\mathrm{d}\sigma_{AB} \tag{B5}$$

where $c_E = \int \mathrm{tr}(\sigma_{AB}^{\otimes n}E)\mathrm{d}\sigma_{AB}$ and $\mathrm{d}\sigma_{AB}$ is the uniform distribution on bipartite density matrices. Confidence regions for the unknown ρ_{AB} can be constructed from $\mathrm{d}\mu_E(\sigma_{AB})$ as the following proposition asserts.

Theorem (1 of main text)**.** *Let n be the number of systems measured by a POVM during tomography and $1-\epsilon$ be the desired confidence level. For each effect E in the POVM, let $S_{\mu_E}\subseteq \mathrm{D}(\mathcal{H}_{AB})$ be a set of states such that*

$$\int_{S_{\mu_E}}\mathrm{d}\mu_E(\sigma_{AB}) \geq 1 - \frac{\epsilon}{2}s_{2n,d_{AB}^2}^{-1}\,, \tag{B6}$$

where $s_{n,d} = \binom{n+d-1}{d-1} \leq (n+1)^{d-1}$ and let $S_{\mu_E}^{\delta}$ be the enlargement of S_{μ_E} defined as

$$S_{\mu_E}^{\delta} := \{\sigma_{AB} : \exists\sigma' \in S_{\mu_E}\, with\, P(\sigma,\sigma') \leq \delta\}\,. \tag{B7}$$

Then the mapping $E\mapsto S_{\mu_E}^{\delta}$ is a confidence region estimator for the unknown ρ_{AB} with confidence level $1-\epsilon$ if

$$\delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 2\ln s_{2n,d_{AB}^2}\right)\,. \tag{B8}$$

In other words, for any $\rho_{AB}\in\mathrm{D}(\mathcal{H}_{AB})$,

$$\Pr_E[\rho_{AB}\in S_{\mu_E}^{\delta}] \geq 1-\epsilon\,, \tag{B9}$$

where the probability is taken over the random dataset E with distribution $\mathrm{tr}(\rho_{AB}^{\otimes n}E)$.

2. Mapping channel tomography to biparite-state tomography

Consider the ancilla-assisted scheme. In order to learn what the channel $\Lambda_{A\to B}$ is, we may carry out the experiment as described in Algorithm 2, and use the outcome measurements to perform full tomography on the output state $\rho_{BP} := \Lambda_{A\to B}(\psi_{AP})$. We may then ask, what does this tell us about the unknown channel $\Lambda_{A\to B}$?

Observe that if we knew the output state ρ_{BP} *exactly* (limit of infinite data) and assume the input state ψ_{AP} has full rank, then we could read out the true channel: its Choi state is simply given as $\Lambda_{BP} = d_A^{-1}\rho_P^{-1/2}\,\rho_{BP}\,\rho_P^{-1/2}$. Indeed, we have

$$\rho_{BP} := \Lambda_{A\to B}(\psi_{AP}) = d_A\psi_P^{1/2}\,\Lambda_{A\to B}(\hat{\Phi}_{AP})\,\psi_P^{1/2} = d_A\rho_P^{1/2}\,\Lambda_{BP}\,\rho_P^{1/2} \tag{B10}$$

since under the assumption that P has undergone identity transformation it follows $\psi_P = \rho_P$. Note that this is the same trick used in Appendix A 1 c to derive the semidefinite program.

Thanks to this observation, we may use the quantum state tomography method of Ref. [7] to construct confidence regions on the space of quantum processs $\Lambda_{A\to B}$, as well as on a figure-of-merit such as the diamond norm to an ideal channel.

To do so, we ignore the knowledge of the exact input state ψ_A, but we assume the global state $|\psi_{AP}\rangle$ has full Schmidt rank (i.e. forgetting the Schmidt coefficients). Upon observing the dataset E, the classical data processing returns a bipartite state region $S^\delta_{\mu_E}$, which contains information about the pair $(\Lambda_{A\to B}, \psi_A)$ [38]. The interpretation of $S^\delta_{\mu_E}$ is given by Theorem 1, and together with the observation above (see Eq. (B10)) we have

$$\Pr_E\left[d_A\rho_P^{1/2}\Lambda_{BP}\rho_P^{1/2} \in S^\delta_{\mu_E}\right] \geq 1-\epsilon\,, \tag{B11}$$

where the probability is taken over all possible dataset E with distribution $\mathrm{tr}((\rho_P^{1/2}\Lambda_{BP}\rho_P^{1/2})^{\otimes n}E)$. To recover information about the channel Λ, for each $\rho_{BP} \in S^\delta_{\mu_E}$ we apply the (completely positive) transformation T defined as

$$\begin{aligned} T:\ \ &\mathrm{End}(\mathcal{H}_{AB}) \to \mathrm{End}(\mathcal{H}_{AB}) \\ &\rho_{BP} \mapsto d_A^{-1}\rho_P^{-1/2}\rho_{BP}\rho_P^{-1/2}. \end{aligned} \tag{B12}$$

Observe that T maps any ρ_{BP} with full rank marginal ρ_P to a Choi state. Also, the set $S^\delta_{\mu_E}$ only contains ρ_{BP} with full rank marginal ρ_P because we only sample according to the uniform measure $\mathrm{d}\sigma_{AB}$ (i.e. the set of rank-deficient ρ_{BP} has measure zero). This means the image of $S^\delta_{\mu_E}$ under T will be a set of Choi matrices which can be interpreted via Choi-Jamiolkowski as a region of quantum processs (completely positive and trace-preserving maps). We conclude

$$\Pr_E\left[\Lambda_{BP} \in T(S^\delta_{\mu_E})\right] \geq 1-\epsilon\,, \tag{B13}$$

which implies $T(S^\delta_{\mu_E})$ are confidence regions for quantum processs.

3. Regions for figures-of-merit

The confidence region on channel space constructed in the last section contains full information on the unknown channel. But if one is only interested in a property of the channel, for instance how close is it to an ideal process, then obtaining confidence region for a given figure-of-merit suffices. We now present how one can do this using pushforward of measures.

Given a figure-of-merit for quantum processs f_{channel} (defined on channel space), we associate a function f defined on the set of bipartite states as

$$f(\rho_{BP}) := f_{\mathrm{channel}}(J^{-1}(d_A^{-1}\rho_P^{-1/2}\,\rho_{BP}\,\rho_P^{-1/2})), \tag{B14}$$

which is just f_{channel} acting on the channel $J^{-1}(d_A^{-1}\rho_P^{-1/2}\,\rho_{BP}\,\rho_P^{-1/2})$ obtained from ρ_{BP} via the mapping T. This allows us to directly use the tools of Ref. [7] to obtain confidence intervals for the figure-of-merit f which will yield the same result as f_{channel}. Explicitly, for any $v \in \mathbb{R}$

$$\mu(v) = \int \mathrm{d}\mu_E(\sigma_{AB})\,\delta(f(\sigma_{AB}) - v) \tag{B15}$$

is the probability density of the pushforward of $\mathrm{d}\mu_E(\sigma_{AB})$ along f. This density provides confidence region for a figure-of-merit as certified by the following proposition.

Theorem 3. *Let μ_E be given as in* (B5)*, and let $\mu(v)$ be defined as in* (B15)*. Then for any threshold value $v_{\text{thres}} > 0$, the region*

$$R^{v_{\text{thres}},\delta} = \{\rho_{AB} : f(\rho_{AB}) \leqslant v_{\text{thres}} + O(\delta)\} \tag{B16}$$

of states representing channels at least $v_{\text{thres}} + O(\delta)$-close to the reference channel, is a confidence region of confidence level $1-\epsilon$ where

$$\epsilon = \text{poly}(n) \left[1 - \int_0^{v_{\text{thres}}} \mu(v)\, \mathrm{d}v\right] . \tag{B17}$$

In summary, for ancilla-assisted tomography scheme, determining the histogram $\mu(v)$ in (B15) gives us all the necessary information to construct confidence regions of any confidence level in terms of the figure-of-merit $f_{\text{channel}}(\Lambda_{A\to B})$.

Diamond distance to ideal and worst-case entanglement fidelity:
The methodology outlined in the previous paragraphs can be specialized to the diamond distance to an ideal reference channel $\Lambda^{\text{ideal}}_{B\to P}$. Here we take

$$f_{\text{channel}}(\Lambda_{B\to P}) = f_\diamond(\Lambda_{B\to P}) = \frac{1}{2}\|\Lambda_{B\to P} - \Lambda^{\text{ideal}}_{B\to P}\|_\diamond \tag{B18}$$

to be the desired figure-of-merit on channel space. This induces a figure-of-merit in the space of bipartite quantum states

$$f(\rho_{BP}) = \frac{1}{2}\max\left\{\left\langle \rho_P^{-1/2}\rho_{BP}\rho_P^{-1/2} - d_A\Lambda^{\text{ideal}}_{BP}, W\right\rangle : W \leq \mathbb{1}_{\mathcal{B}}\otimes\bar{\rho}, W \geq 0, \bar{\rho}\in \mathrm{D}(\mathcal{X})\right\} . \tag{B19}$$

One is left to perform a numerical computation of $\mu(v)$ for the above function f, as explained in details in Ref. [7].

Appendix C: The channel-space sampling method

This method applies to either the ancilla-assisted scheme explained in the previous Appendix, or the prepare-and-measure scheme where no entanglement is required. In the prepare-measure scheme (see Fig. 2**(a)**), we select a collection of input states σ^j, and select a collection of measurements $E^{(\ell)} = \{E_k^{(\ell)}\}$. This set of state preparation and measurement (SPAM) should be informationally complete if one wish to fully reconstruct the unknown channel. The SPAM is represented as certain set of matrices in the computational basis $\{|i\rangle : i = 0, ..., d_A - 1\}$.

The data collection procedure goes as follows: in each round, we choose an input state σ^j, we choose a measurement ℓ on output, we send σ^j through the channel, and record the measurement outcome k on the output. The dataset E consists of all pairs $(\sigma^j, E_k^{(\ell)})$ chosen and observed for each round.

Typically one can choose the states j in order, i.e., first perform measurements on σ^1, then on σ^2, etc. The choice of the output measurement setting ℓ is allowed to depend on j. Since we are under i.i.d. channel assumption, at each round it is the same unknown channel Λ which is applied, and that previous outcomes have no influence on new rounds.

The likelihood function for a dataset E in this scenario is defined using the matrix representations of the SPAM according to Born's rule

$$\mathcal{L}_{\text{PM}}(\Lambda|E) = \prod_{j,k,\ell}\left[\text{tr}(\Lambda(\sigma^j)\, E_k^{(\ell)})\right]^{n_{j,k,\ell}} , \tag{C1}$$

where $n_{j,k,\ell}$ is the number of times the given pair $(\sigma^j, E_k^{(\ell)})$ appears in the dataset E. Using (A3), we rewrite the likelihood function as

$$\mathcal{L}_{\text{PM}}(\Lambda|E) = \prod_{j,k,\ell}\left[d_A\text{tr}(\Lambda_{AB}\ (\sigma_A^j)^{\intercal}\otimes E_k^{(\ell)})\right]^{n_{j,k,\ell}} = d_A^n\text{tr}\left(\Lambda_{AB}^{\otimes n}\bigotimes_{j,k,\ell}(\sigma_A^j)^{\intercal}\otimes E_k^{(\ell)}\right), \tag{C2}$$

where $\bigotimes_{j,k,\ell}$ ranges over the observed dataset E.

The method in the previous section maps a channel tomography problem into a (constrained) biparite-state tomography problem. One may ask if this is the only solution. In this section, we provide an alternative construction natively on the channel space. This has consequence on the numerical implementation: we no longer need to samples from biparite-state space. Instead, we can directly sample "random channels" which leads to improved numerical efficiency. The main results in this section are Theorems 2 and 4.

1. Regions on channel space

Inspired by the Christandl-Renner construction [6], we define the following confidence region estimator for quantum processs. Our confidence region is constructed from the probability measure on the space of quantum processs $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\mathrm{d}\nu_E(\Lambda) := c_E'^{-1}\mathcal{L}(\Lambda|E)\mathrm{d}\nu(\Lambda) \tag{C3}$$

where $\mathcal{L}(\Lambda|E)$ is either prepare-measure or ancilla assisted likelihood and $c'_E = \int \mathcal{L}(\Lambda|E)\mathrm{d}\nu(\Lambda)$ serves as a normalizing constant and $\mathrm{d}\nu(\Lambda)$ is the induced measure on $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ defined in Appendix A 2.

The main result in this Section is

Theorem (2 of maintext)**.** *Let n be the number of channel uses during tomography and $1-\epsilon$ be the desired confidence level. For each dataset E, let $R_{\nu_E} \subseteq \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ be a set of channels such that*

$$\int_{R_{\nu_E}} \mathrm{d}\nu_E(\Lambda) \geq 1 - \frac{\epsilon}{2} s^{-2}_{2n,d^2_{AB}} \,, \tag{C4}$$

where $s_{n,d} = \binom{n+d-1}{d-1} \leq (n+1)^{d-1}$ and let $R^\delta_{\nu_E}$ be the enlargement

$$R^\delta_{\nu_E} := \{\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) : \exists \Lambda' \in R_{\nu_E} \text{ with } P(\Lambda,\Lambda') \leq \delta\}\,. \tag{C5}$$

Then the mapping $E \mapsto R^\delta_{\nu_E}$ is a confidence region estimator for the unknown $\Lambda_{A\to B}$ with confidence level $1-\epsilon$ if

$$\delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d^2_{AB}}\right). \tag{C6}$$

In other words, for all channel $\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\Pr_E[\Lambda_{A\to B} \in R^\delta_{\nu_E}] \geq 1-\epsilon\,, \tag{C7}$$

where the probability is over the random dataset E with distribution $\Pr(E|\Lambda) = \mathcal{L}(\Lambda|E)$.

Before starting the proof, we will need the following results.

Proposition 1. *For any channel $\Lambda_{A\to B}$, if $|\Lambda\rangle \in \mathcal{H}_{ABA'B'}$ is a purification of its Choi state then*

$$|\Lambda\rangle\langle\Lambda|^{\otimes n} \leq s^2_{n,d^2_{AB}} \int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n} = s^2_{n,d^2_{AB}} \int \mathrm{d}U\, U^{\otimes n}_{BA'B'}|\Psi_0\rangle\langle\Psi_0|^{\otimes n} U^{\dagger\otimes n}_{BA'B'} \tag{C8}$$

where $s_{n,d} := \binom{n+d-1}{d-1}$.

Proof. The main idea of this proof is to discretize the Haar integral using Caratheodory's theorem, and dominate the left hand side by a trivial operator inequality.

By definition, the operator

$$\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n} \tag{C9}$$

lies in the convex hull of the set $\{|\Psi\rangle\langle\Psi|^{\otimes n} : |\Psi\rangle \in \mathscr{P}\mathscr{C}\}$, whose linear span (in the ambient space $\mathrm{End}(\mathcal{H}^{\otimes n}_{ABA'B'})$) has dimension D. By Caratheodory's theorem, there exists a convex combination $(q_i, |\Psi_i\rangle\langle\Psi_i|^{\otimes n})$ with size at most $D+1$ such that

$$\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n} = \sum_{i=1}^{D+1} q_i |\Psi_i\rangle\langle\Psi_i|^{\otimes n}\,. \tag{C10}$$

Among the probability weights q_i there exists a largest element denoted $q_{\max}$ and its associated purified Choi state $|\Psi_{\max}\rangle\langle\Psi_{\max}|$, from which we split off this term in the finite sum as

$$\sum_{i=1}^{D+1} q_i|\Psi_i\rangle\langle\Psi_i|^{\otimes n} = q_{\max}|\Psi_{\max}\rangle\langle\Psi_{\max}|^{\otimes n} + \sum_{i\neq\max} q_i|\Psi_i\rangle\langle\Psi_i|^{\otimes n}\,. \tag{C11}$$

By left-invariance of the measure $\mathrm{d}\nu(|\Psi\rangle)$ and the (unitary) structure of the set $\mathscr{P}\mathscr{C}$, we can without loss of generality assume that $\Psi_{\max} = \Lambda$. More precisely, let $W_{BA'B'}$ be a unitary transformation bringing $|\Psi_{\max}\rangle$ to $|\Lambda\rangle$, we have (leaving the system label $BA'B'$ implicit)

$$W^{\otimes n}\left(\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n}\right)W^{\dagger\otimes n} = q_{\max}W^{\otimes n}|\Psi_{\max}\rangle\langle\Psi_{\max}|^{\otimes n}W^{\dagger\otimes n} + \sum_{i\neq\max} q_i W^{\otimes n}|\Psi_i\rangle\langle\Psi_i|^{\otimes n}W^{\dagger\otimes n}. \quad \text{(C12)}$$

Using linearity of integration and translational invariance of the integrating measure, this equation simplifies to

$$\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n} = q_{\max}|\Lambda\rangle\langle\Lambda|^{\otimes n} + \sum_{i\neq\max} q_i|\Psi_i'\rangle\langle\Psi_i'|^{\otimes n}, \quad \text{(C13)}$$

where $|\Psi_i'\rangle$ is some other vector in $\mathscr{P}\mathscr{C}$.

Now since all operators in the convex combination are positive-semidefinite, we obtain

$$\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n} \geq q_{\max}|\Lambda\rangle\langle\Lambda|^{\otimes n}. \quad \text{(C14)}$$

By the property of the maximum weight $q_{\max}$, namely $q_{\max} \geq 1/(D+1)$, we get

$$|\Lambda\rangle\langle\Lambda|^{\otimes n} \leq (D+1)\int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\Psi\rangle)\,|\Psi\rangle\langle\Psi|^{\otimes n}. \quad \text{(C15)}$$

Finally, $\mathrm{span}\{|\Psi\rangle\langle\Psi|^{\otimes n} : |\Psi\rangle \in \mathscr{P}\mathscr{C}\} \subseteq \mathrm{span}\{|\Psi\rangle\langle\Psi|^{\otimes n} : |\Psi\rangle \in \mathcal{H}_{ABA'B'}\}$ and the latter is identified as a subspace of $\mathrm{End}(\mathrm{Sym}^n(\mathcal{H}_{ABA'B'}))$, the operator space on the symmetric subspace of $\mathcal{H}_{ABA'B'}^{\otimes n}$. Together with the constraint that trace is 1, we have $D \leq s_{n,d_{ABA'B'}}^2 - 1$ where the dimension of the symmetric subspace is $s_{n,d} := \binom{n+d-1}{d-1}$. This completes the proof of the operator inequality. □

Proof of Theorem 2. Our proof technique follows closely that of [6], with the main technical difficulty being incorporating the *a priori* constraint $\mathrm{tr}_B(\Lambda_{AB}) = \mathbb{1}_A/d_A$. This allows the reduction of numerical sampling from biparite-state space to channel space.

For any region estimator, our construction $E \mapsto R_{\nu_E}^\delta$ in particular, the failure probability of the reconstruction typically depends on the underlying unknown channel

$$P_{\mathrm{fail}}(\Lambda_{A\to B}) = \Pr_E[\Lambda_{A\to B} \notin R_{\nu_E}^\delta] := \sum_E \Pr(E|\Lambda)\chi(\Lambda_{A\to B}; \overline{R_{\nu_E}^\delta}), \quad \text{(C16)}$$

where $\Pr(E|\Lambda)$ is the probability of obtaining dataset E, and $\chi(\Lambda_{A\to B}; \overline{R_{\nu_E}^\delta})$ is the indicator function of the set $\overline{R_{\nu_E}^\delta} := \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) \setminus R_{\nu_E}^\delta$ (i.e. the complement set). Recall that

$$\Pr(E|\Lambda) = \begin{cases} d_A^n \mathrm{tr}\left(\Lambda_{AB}^{\otimes n} \bigotimes_{j,k,\ell}(\sigma_A^j)^\intercal \otimes E_k^{(\ell)}\right) & \text{in prepare-and-measure scheme} \\ d_A^n \mathrm{tr}\left(\Lambda_{BP}^{\otimes n} \bigotimes_{k,\ell} \psi_P^{1/2} E_k^{(\ell)} \psi_P^{1/2}\right) & \text{in ancilla-assisted scheme} \end{cases} \quad \text{(C17)}$$

Our goal will be bounding this failure probability independently of $\Lambda_{A\to B}$ by using the operator inequality we have just developed.

Before starting the actual calculations, observe that $\Pr(E|\Lambda)$ for both schemes are functions of the type $\mathrm{tr}(\Lambda^{\otimes n}\otimes\cdots)$ where $\otimes\cdots$ is the operator constructed from the observed dataset E from information about the state preparation and measurement schemes. In the following, we do not utilise the exact form of $\otimes\cdots$ for each schemes and thus the calculation works for both schemes. We choose to put $\otimes\cdots$ as the operator corresponding to the prepare-and-measure scheme for concreteness.

Via the Choi-Jamiolkowski isomorphism, the failure probability reads

$$P_{\mathrm{fail}}(\Lambda_{AB}) = \Pr_E[\Lambda_{AB} \notin R_{\nu_E}^\delta] := \sum_E d_A^n \mathrm{tr}[\Lambda_{AB}^{\otimes n}\,\rho_{A^n}^\intercal \otimes E_{B^n}]\chi(\Lambda_{AB}; \overline{R_{\nu_E}^\delta}), \quad \text{(C18)}$$

where we have abused the notation $R_{\nu_E}^\delta$ to mean both the set in channel space $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$ and in Choi state space $\mathscr{C}(\mathcal{H}_{AB})$. This can be rewritten in terms of an arbitrary purification of the Choi state Λ_{AB}

$$P_{\mathrm{fail}}(\Lambda_{AB}) = \sum_E d_A^n \mathrm{tr}[|\Lambda\rangle\langle\Lambda|_{ABA'B'}^{\otimes n}\,\rho_{A^n}^\intercal \otimes E_{B^n}]\chi(|\Lambda\rangle_{ABA'B'}; \overline{Q_{\nu_E}^\delta}), \quad \text{(C19)}$$

where $\overline{Q^\delta_{\nu_E}} := \mathrm{tr}^{-1}_{A'B'}(\overline{R^\delta_{\nu_E}})$ contains all the purifications of matrices in $\overline{R^\delta_{\nu_E}}$. In the following, we will bound (C19) independent of $|\Lambda\rangle \in \mathscr{PC}$.

We first analyze the indicator function of the set $\overline{Q^\delta_{\nu_E}}$, which is by definition

$$\chi(|\Lambda\rangle_{ABA'B'}; \overline{Q^\delta_{\nu_E}}) = \begin{cases} 1 & \text{if } |\Lambda\rangle_{ABA'B'} \in \overline{Q^\delta_{\nu_E}} \\ 0 & \text{otherwise}\,. \end{cases} \tag{C20}$$

Without the knowledge of $|\Lambda\rangle_{ABA'B'}$, the condition $|\Lambda\rangle_{ABA'B'} \in \overline{Q^\delta_{\nu_E}}$ can only be *physically checked* by a measurement POVM with effects T and $\mathbb{1} - T$ acting on the quantum state $|\Lambda\rangle_{ABA'B'}$. Upon the observation of the effect T, we decide that $|\Lambda\rangle_{ABA'B'} \in \overline{Q^\delta_{\nu_E}}$ and similarly for $\mathbb{1} - T$. In other words, we are approximating $\chi(|\Lambda\rangle_{ABA'B'}; \overline{Q^\delta_{\nu_E}})$ by a quantum measurement. Here we construct such an approximation using Holevo's covariant measurement [39].

Let k be the number of copies of $|\Lambda\rangle \in \mathcal{H}_{ABA'B'}$ used in the approximation, i.e. we are given $|\Lambda\rangle\langle\Lambda|^{\otimes k}$. If we ignore the fact that $|\Lambda\rangle \in \mathscr{PC}$, we can use the Holevo's continuous POVM $\{s_{k,d^2_{AB}} |\phi\rangle\langle\phi|^{\otimes k} \mathrm{d}\phi\}$ to distinguish $|\Lambda\rangle \in \mathcal{H}_{ABA'B'}$ among the set of pure states. Here, $\mathrm{d}\phi$ is the uniform spherical measure on the set of pure states of $\mathcal{H}_{ABA'B'}$ and $s_{k,d^2_{AB}}$ is the dimension of the symmetric subspace of $\mathcal{H}^{\otimes k}_{ABA'B'}$. Coarse graining this measurement, we can distinguish $|\Lambda\rangle_{ABA'B'} \in \overline{Q^\delta_{\nu_E}}$ versus $|\Lambda\rangle_{ABA'B'} \in Q^\delta_{\nu_E}$ by the following POVM with two effects (analogous to Ref. [6])

$$T_{Q^{\delta/2}_{\nu_E}} := s_{k,d^2_{AB}} \int_{\overline{Q^{\delta/2}_{\nu_E}}} |\phi\rangle\langle\phi|^{\otimes k} \mathrm{d}\phi\,, \text{ and } \mathbb{1} - T_{Q^{\delta/2}_{\nu_E}}\,. \tag{C21}$$

We now check that this POVM indeed approximates $\chi(|\Lambda\rangle; \overline{Q^\delta_{\nu_E}})$. For all $|\Lambda\rangle \in \overline{Q^\delta_{\nu_E}}$, using the definition of $\chi(|\Lambda\rangle; \overline{Q^\delta_{\nu_E}})$

$$\chi(|\Lambda\rangle; \overline{Q^\delta_{\nu_E}}) - \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) = 1 - s_{k,d^2_{AB}} \int_{\overline{Q^{\delta/2}_{\nu_E}}} \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} |\phi\rangle\langle\phi|^{\otimes k}) \mathrm{d}\phi\,. \tag{C22}$$

Since $|\Lambda\rangle\langle\Lambda|^{\otimes k}$ is supported on the symmetric subspace, we reinterpret the constant 1 above as

$$1 = \mathrm{tr}\left(|\Lambda\rangle\langle\Lambda|^{\otimes k} s_{k,d^2_{AB}} \int |\phi\rangle\langle\phi|^{\otimes k} \mathrm{d}\phi \right)\,, \tag{C23}$$

which implies for all $|\Lambda\rangle \in \overline{Q^\delta_{\nu_E}}$

$$\chi(|\Lambda\rangle; \overline{Q^\delta_{\nu_E}}) - \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) = s_{k,d^2_{AB}} \left(\int \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} |\phi\rangle\langle\phi|^{\otimes k}) \mathrm{d}\phi - \int_{\overline{Q^{\delta/2}_{\nu_E}}} \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} |\phi\rangle\langle\phi|^{\otimes k}) \mathrm{d}\phi \right) \tag{C24}$$

$$= s_{k,d^2_{AB}} \int_{Q^{\delta/2}_{\nu_E}} \mathrm{tr}(|\Lambda\rangle\langle\Lambda|\,|\phi\rangle\langle\phi|)^k \mathrm{d}\phi \tag{C25}$$

$$\leq s_{k,d^2_{AB}} \max_{|\phi\rangle \in Q^{\delta/2}_{\nu_E}} F(\Lambda_{AB}, \mathrm{tr}_{A'B'} |\phi\rangle\langle\phi|)^k\,. \tag{C26}$$

By the definition of the sets

$$R^{\delta/2}_{\nu_E} := \{\Psi \in \mathscr{C}(\mathcal{H}_{AB}) : \exists \Psi' \in R_{\nu_E} \text{ with } P(\Psi, \Psi') \leq \delta/2\}\,, \tag{C27}$$

and

$$\overline{R^\delta_{\nu_E}} := \mathscr{C}(\mathcal{H}_{AB}) \setminus \{\Psi \in \mathscr{C}(\mathcal{H}_{AB}) : \exists \Psi' \in R_{\nu_E} \text{ with } P(\Psi, \Psi') \leq \delta\}\,, \tag{C28}$$

we have for $\Lambda_{AB} \in \overline{R^\delta_{\nu_E}}$ and $\phi_{AB} := \mathrm{tr}_{A'B'} |\phi\rangle\langle\phi| \in R^{\delta/2}_{\nu_E}$

$$F(\Lambda_{AB}, \phi_{AB}) = \sqrt{1 - P(\Lambda_{AB}, \phi_{AB})^2} \leq \sqrt{1 - (\delta/2)^2} \leq e^{-\delta^2/2}\,, \tag{C29}$$

using the reverse triangle inequality for purified distance. In summary, we obtain the approximation

$$\chi(|\Lambda\rangle; \overline{Q^\delta_{\nu_E}}) - \mathrm{tr}(|\Lambda\rangle\langle\Lambda|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) \leq \epsilon_1 := s_{k,d^2_{AB}} e^{-k\delta^2/2}\,. \tag{C30}$$

Now we can start bounding the failure probability. Inserting (C30) into (C19), we have an intermediate bound

$$P_{\text{fail}}(\Lambda_{AB}) \leq \epsilon_1 + \sum_E d_A^n \text{tr}[|\,\Lambda\rangle\langle\Lambda\,|^{\otimes n}_{ABA'B'}\, \rho^{\intercal}_{A^n} \otimes E_{B^n}] \text{tr}[|\,\Lambda\rangle\langle\Lambda\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}] \tag{C31}$$

$$= \epsilon_1 + \sum_E d_A^n \text{tr}[|\,\Lambda\rangle\langle\Lambda\,|^{\otimes (n+k)}_{ABA'B'}\, \rho^{\intercal}_{A^n} \otimes E_{B^n} \otimes T_{Q^{\delta/2}_{\nu_E}}]\,. \tag{C32}$$

Using the operator inequality in the Proposition 1, namely

$$|\,\Lambda\rangle\langle\Lambda\,|^{\otimes (n+k)}_{ABA'B'} \leq s^2_{n+k,d^2_{AB}} \int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\,\Psi\rangle)\, |\,\Psi\rangle\langle\Psi\,|^{\otimes (n+k)}\,, \tag{C33}$$

we can bound the right hand side *independent of the unknown* Λ_{AB} as follows

$$P_{\text{fail}}(\Lambda_{AB}) \leq \epsilon_1 + s^2_{n+k,d^2_{AB}} \sum_E \int_{\mathscr{P}\mathscr{C}} \mathrm{d}\nu(|\,\Psi\rangle) d_A^n \text{tr}[|\,\Psi\rangle\langle\Psi\,|^{\otimes n} \rho^{\intercal}_{A^n} \otimes E_{B^n}] \text{tr}[|\,\Psi\rangle\langle\Psi\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}] \tag{C34}$$

$$= \epsilon_1 + s^2_{n+k,d^2_{AB}} \sum_E c'_E \int \mathrm{d}\nu_E(\Psi) \text{tr}[|\,\Psi\rangle\langle\Psi\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}]\,, \tag{C35}$$

where the last equality follows from the definition of the *a posteriori* measure $\mathrm{d}\nu_E(\Psi)$. For each measurement outcome E, the integral can split into two parts based on the set R_{ν_E} from which the kernels are uniformly bounded as follows:

$$\int_{R_{\nu_E}} \mathrm{d}\nu_E(\Psi) \text{tr}(|\,\Psi\rangle\langle\Psi\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) \leq s_{k,d^2_{AB}} (1-(\delta/2)^2)^{k/2} \leq s_{k,d^2_{AB}} e^{-k\delta^2/2}\,, \tag{C36}$$

using the definition of $T_{Q^{\delta/2}_{\nu_E}}$ and the fidelity bound $F(\Psi_{AB} \in R_{\nu_E}, \phi_{AB} \in \overline{R^{\delta/2}_{\nu_E}}) \leq \sqrt{1-(\delta/2)^2}$, and

$$\int_{\overline{R_{\nu_E}}} \mathrm{d}\nu_E(\Psi) \text{tr}(|\,\Psi\rangle\langle\Psi\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) \leq \int_{\overline{R_{\nu_E}}} \mathrm{d}\nu_E(\Psi)\,, \tag{C37}$$

since $\text{tr}(|\,\Psi\rangle\langle\Psi\,|^{\otimes k} T_{Q^{\delta/2}_{\nu_E}}) \leq 1$. Choose $k = n$, the fact that $\sum_E c'_E \leq 1$, and combine all the inequalities together we have

$$P_{\text{fail}}(\Lambda_{AB}) \leq \epsilon_1 + s^2_{n+k,d^2_{AB}} \epsilon_1 + s^2_{n+k,d^2_{AB}} \sum_E c'_E \int_{\overline{R_{\nu_E}}} \mathrm{d}\nu_E(\Psi) \tag{C38}$$

$$= s_{n,d^2_{AB}} e^{-n\delta^2/2} + s^2_{2n,d^2_{AB}} s_{n,d^2_{AB}} e^{-n\delta^2/2} + s^2_{2n,d^2_{AB}} \sum_E c'_E \int_{\overline{R_{\nu_E}}} \mathrm{d}\nu_E(\Psi) \tag{C39}$$

$$\leq s^3_{2n,d^2_{AB}} e^{-n\delta^2/2} + s^2_{2n,d^2_{AB}} \sum_E c'_E \int_{\overline{R_{\nu_E}}} \mathrm{d}\nu_E(\Psi)\,. \tag{C40}$$

If we choose R_{ν_E} and δ such that

$$\int_{R_{\nu_E}} \mathrm{d}\nu_E(\Psi) \geq 1 - \frac{\epsilon}{2} s^{-2}_{2n,d^2_{AB}} \;\text{ and }\; \delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d^2_{AB}}\right) \tag{C41}$$

then $P_{\text{fail}}(\Lambda_{AB}) \leq \epsilon/2 + \epsilon/2 = \epsilon$ as desired. The proof of the Proposition is complete. □

2. Regions for figures-of-merit

The construction of confidence region on channel-space can be pushed-forward to obtain confidence regions for any figure-of-merit of channels we are interested in. The idea is exactly the same as reference [7] and we include it here for completeness. Let $f_{\text{channel}} : \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) \to \mathbb{R}$ be an arbitrary figure-of-merit of channels. The measure $\mathrm{d}\nu_E(\Lambda)$ can be pushed-forward by f_{channel} to a measure on $\mathbb{R}$, which can then be represented as a density function $h(v)$ with respect to the Lebesgue measure of $\mathbb{R}$. Concretely, we have

$$h(v) = \int \mathrm{d}\nu_E(\Lambda) \delta(f_{\text{channel}}(\Lambda) - v), \tag{C42}$$

where $\delta(f_{\text{channel}}(\Lambda) - v)$ is the Dirac delta measure on $\mathbb{R}$ at the point mass $v \in \mathbb{R}$. And for some subset of values V, the measure of V is given by

$$\int_{f_{\text{channel}}^{-1}(V)} \mathrm{d}\nu_E(\Lambda) = \int_V h(v)\mathrm{d}v \tag{C43}$$

where $\mathrm{d}v$ is the Lebesgue measure on $\mathbb{R}$. The density $h(v)$ allows us to construct confidence interval for the property we desired.

Proposition 2. *Let f_{channel} be a figure-of-merit and choose a confidence level $1-\epsilon$. For each dataset E, let $V_{\nu_E} \subseteq \mathbb{R}$ be a region of values such that*

$$\int_{V_{\nu_E}} h(v)\mathrm{d}v \geq 1 - \frac{\epsilon}{2} s_{2n,d_{AB}^2}^{-2}\,, \tag{C44}$$

and let $V_{\nu_E}^{\delta}$ be defined as

$$V_{\nu_E}^{\delta} := \{v \in \mathbb{R} : \exists v' \in V_{\nu_E} \text{ with } |v - v'| \leq \omega_{f_{\text{channel}}}(\delta)\}\,, \tag{C45}$$

where $\omega_f(\delta) := \sup_{P(\Lambda,\Lambda')\leq\delta} |f(\Lambda) - f(\Lambda')|$. Then the mapping $E \mapsto V_{\nu_E}^{\delta}$ is a confidence region estimator for the figure-of-merit f_{channel} with confidence level $1-\epsilon$ if

$$\delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d_{AB}^2}\right) \tag{C46}$$

In other words, for all channel $\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\Pr_E[f_{\text{channel}}(\Lambda) \in V_{\nu_E}^{\delta}] \geq 1 - \epsilon. \tag{C47}$$

Proof. It is clear from the fact that as defined, $V_{\nu_E}^{\delta} \supseteq f_{\text{channel}}(f_{\text{channel}}^{-1}(V_{\nu_E})^{\delta})$. □

For each figure-of-merit of interest, we can derive a bound on $\omega_f(\delta)$ by simple inequalities for distance measures. For diamond distance, we have the following result.

Proposition 3. *For each dataset E, let $\gamma_E \in [0,1]$ be such that*

$$\int_0^{\gamma_E} h(v)\mathrm{d}v \geq 1 - \frac{\epsilon}{2} s_{2n,d_{AB}^2}^{-2}\,, \tag{C48}$$

Then the mapping $E \mapsto [0, \gamma_E + d_1\delta/2]$ is a confidence region estimator for the diamond distance to ideal with confidence level $1-\epsilon$ if

$$\delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d_{AB}^2}\right) \tag{C49}$$

In other words, for all channel $\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\Pr_E\left[\frac{1}{2}||\Lambda_{A\to B} - \Lambda_{A\to B}^{\text{ideal}}||_\diamond \leq \gamma_E + d_A\delta/2\right] \geq 1 - \epsilon\,, \tag{C50}$$

where the probability is over the random dataset E with distribution $\Pr(E|\Lambda) = \mathcal{L}(\Lambda|E)$.

Proof. Continuing from the previous Proposition, we set $V_E := [0, \gamma_E]$; it remains for us to obtain a bound on $\omega_{f_\diamond}(\delta)$. Using the reverse triangle inequality and SDP reformulation of diamond norm, we have

$$\begin{aligned}
|f_\diamond(\Lambda) - f_\diamond(\Lambda')| &= \frac{1}{2}\left|||\Lambda - \Lambda^{\text{ideal}}||_\diamond - ||\Lambda' - \Lambda^{\text{ideal}}||_\diamond\right| \\
&\leq \frac{1}{2}||\Lambda_{A\to B} - \Lambda'_{A\to B}||_\diamond \qquad \text{(C51)} \\
&\leq \frac{1}{2}||d_A(\Lambda_{AB} - \Lambda'_{AB})||_1\,, \qquad \text{(C52)}
\end{aligned}$$

where the last inequality utilises the duality between Schatten 1-norm and Schatten ∞-norm to bound the objective function of the diamond norm SDP. Since the purified distance dominates the trace distance, we obtain

$$\frac{1}{2}||\Lambda_{AB} - \Lambda'_{AB}||_1 \leq \frac{1}{2}P(\Lambda_{AB}, \Lambda'_{AB})\,, \tag{C53}$$

which implies $\omega_{f_\diamond}(\delta) \leq d_A\delta/2$. □

For worst-case entanglement fidelity, we have the following result.

Proposition 4. *For each dataset E, let $\gamma_E \in [0,1]$ be such that*

$$\int_0^{\gamma_E} h(v)\mathrm{d}v \geq 1 - \frac{\epsilon}{2} s_{2n,d_{AB}^2}^{-2}\,, \tag{C54}$$

Then the mapping $E \mapsto [0, \gamma_E - d_A\delta]$ is a confidence region estimator for the diamond distance to ideal with confidence level $1-\epsilon$ if

$$\delta^2 = \frac{2}{n}\left(\ln\frac{2}{\epsilon} + 3\ln s_{2n,d_{AB}^2}\right) \tag{C55}$$

In other words, for all channel $\Lambda \in \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$

$$\Pr_E\left[F_{\mathrm{worst}}(\Lambda_{A\to B}) \geq \gamma_E - d_A\delta\right] \geq 1-\epsilon\,, \tag{C56}$$

where the probability is over the random dataset E with distribution $\Pr(E|\Lambda) = \mathcal{L}(\Lambda|E)$.

Proof. We set $V_E := [\gamma_E, 1]$. Let ρ_A be an optimizer of $F_{\mathrm{worst}}(\Lambda')$, since ρ_A will give an upper bound on $F_{\mathrm{worst}}(\Lambda)$ we have with $f = F_{\mathrm{worst}}$

$$|f(\Lambda) - f(\Lambda')| = |F_{\mathrm{worst}}(\Lambda) - F_{\mathrm{worst}}(\Lambda')| \leq |\langle\tilde{\Phi}|\rho_A(d_A\Lambda_{AB})\rho_A|\tilde{\Phi}\rangle - \langle\tilde{\Phi}|\rho_A(d_A\Lambda'_{AB})\rho_A|\tilde{\Phi}\rangle| \tag{C57}$$

$$= d_A\left|\left\langle \rho_A|\tilde{\Phi}\rangle\langle\tilde{\Phi}|\rho_A, \Lambda_{AB} - \Lambda'_{AB}\right\rangle\right| \leq d_A\|\rho_A|\tilde{\Phi}\rangle\langle\tilde{\Phi}|\rho_A\|_\infty\|\Lambda_{AB} - \Lambda'_{AB}\|_1 \leq d_A\delta\,, \tag{C58}$$

using Holder inequality for Schatten norms and $\|\Lambda_{AB} - \Lambda'_{AB}\|_1 \leq P(\Lambda_{AB}, \Lambda'_{AB})$. □

Appendix D: Metropolis-Hastings algorithm in channel space

The previous two sections describe the construction of confidence region estimators for quantum processes, which utilize distributions $\mathrm{d}\mu_E(\sigma)$ and $\mathrm{d}\nu_E(\Lambda)$. We now describe how one can numerically estimate such distributions so that the densities $\mu(v)$ and $h(v)$ can be approximated.

The distribution $\mathrm{d}\mu_E(\sigma)$ or the density $\mu(v)$ can be estimated by numerically producing a lot of samples. These can be generated by the Metropolis-Hastings random walk in (bipartite) state space, whose details can be found in Ref. [7]. Here we only discuss the Metropolis-Hastings random walk in channel space.

Recall that in the channel space method, we need to be able to compute the density $h(v)$ for the given figure-of-merit f_{channel}. We do this numerically using Metropolis-Hastings algorithm. The output of this algorithm is a histogram of the figure-of-merit which approximates the continuous density.

Let us recall the Metropolis-Hasting algorithm for continuous sample space [26]. Let $p(x)\mathrm{d}x$ be the target distribution from which we want to sample, and $q(x'|x)\mathrm{d}x'$ be a proposal distribution, all displayed with respect to the same base measure $\mathrm{d}x = \mathrm{d}x'$. We assume that the proposal density function is symmetric $q(x'|x) = q(x|x')$. When the process is at point x, the distribution $q(x'|x)\mathrm{d}x'$ proposes a new point x'. If $p(x')/p(x) \geq 1$ then we jump unconditionally to the new point x'; otherwise, $p(x')/p(x) < 1$ and we jump to x' only with probability $p(x')/p(x)$. The points visited in this fashion, for a large number of iterations, are distributed according to the target distribution. Note that the algorithm only requires computing the ratio $p(x')/p(x)$ and thus does not require determining any normalization factor for $p(x)$.

We want to generate samples from the target distribution

$$\mathrm{d}\nu_E(\Lambda) := c_E'^{-1}\mathcal{L}(\Lambda|E)\mathrm{d}\nu(\Lambda) \tag{D1}$$

where $\mathcal{L}(\Lambda|E)$ is the prepare-and-measure or ancilla-assisted likelihood function and $\mathrm{d}\nu(\Lambda)$ is the induced measure on channel space. Recalling the definition of $\mathrm{d}\nu(\Lambda)$, we thus want to sample from

$$\mathrm{d}\nu_E(U_{BA'B'}) = c_E'^{-1}\mathcal{L}(U_{BA'B'}|E)\mathrm{d}U_{BA'B'}\,, \tag{D2}$$

with $\mathrm{d}U_{BA'B'}$ the invariant Haar measure. Concretely, in the prepare-and-measure scheme we take

$$\mathcal{L}_{\mathrm{PM}}(U|E) = d_A^n \mathrm{tr}\left((U|\Psi_0\rangle\langle\Psi_0|U^\dagger)^{\otimes n}\bigotimes_{j,k,\ell}(\sigma_A^j)^\intercal \otimes E_k^{(\ell)}\right) \tag{D3}$$

and in ancilla-assisted scheme we take

$$\mathcal{L}_{\mathrm{AA}}(U|E) = d_A^n \mathrm{tr}\left((U|\,\Psi_0\rangle\langle\Psi_0\,|U^\dagger)^{\otimes n} \bigotimes_{k,\ell} \psi_P^{1/2}\, E_k^{(\ell)}\, \psi_P^{1/2} \right), \tag{D4}$$

where $|\,\Psi_0\rangle$ is the fixed reference state in (A15). This can be done using the Metropolis-Hastings algorithm, by designing a symmetric proposal distribution over the space of all unitaries $U_{BA'B'}$ and setting $q(U'_{BA'B'}|U_{BA'B'}) \propto \mathcal{L}(U'_{BA'B'}|E)$. To ensure $q(U'|U) = q(U|U')$, let $q(W)\mathrm{d}W$ be a distribution on unitaries on $BA'B'$ such that $q(W) = q(W^\dagger)$. For each point U, if we define $U' := WU$, then we have a symmetric proposal distribution $q(U'|U) = q(WU|U) = q(W) = q(W^{-1}) = q(W^{-1}U'|U') = q(U|U')$, namely $q(WU|U)\mathrm{d}W$ where $\mathrm{d}W$ is the Haar measure. It remains to fix a $q(W)\mathrm{d}W$ with $q(W) = q(W^\dagger)$. We have implemented two choices:

- "e^{iH}-type jumps": We pick a random $d_{BA'B'} \times d_{BA'B'}$ matrix N with each entry independent and normally distributed complex numbers with standard deviation given by the step size. We then calculate $H = N + N^\dagger$ and set $W = e^{iH}$, inducing a measure $q(W)\,dW$. Denoting by dN the measure induced on N by this sampling procedure, observe that $dN = d(-N)$ as the normal distribution is symmetric. Furthermore the Haar measure is invariant under the adjoint, $dW = d(W^\dagger)$, since $d(W^\dagger)$ is also unitarily invariant and is thus also the Haar measure. Hence, $q(W)\,dW = dN = d(-N) = q(W^\dagger)\,d(W^\dagger) = q(W^\dagger)\,dW$ as required.

- "elementary rotation jumps": Choose $m \in \{x, y, z\}$ uniformly at random and choose two indices $i < j$ uniformly at random. Choose $\sin(\alpha)$ at random (normally distributed number whose standard deviation is the step size; truncated to $[-1, 1]$). Define the unitary W_1 as the qubit rotation on the subspace spanned by $\{|\,i\rangle, |\,j\rangle\}$ defined by $e^{i\alpha\,(\vec{e}_m\cdot\vec{\sigma})} = \cos(\alpha)\,\mathbb{1} + i\sin(\alpha)\,(\vec{e}_m\cdot\vec{\sigma})$, where $\vec{e}_m$ is the m-th basis vector in 3D and where $\{\sigma_x, \sigma_y, \sigma_z\}$ are the Pauli matrices. We see that $-\alpha\,(\vec{e}_m\cdot\vec{\sigma})$ is sampled with the same probability as $\alpha\,(\vec{e}_m\cdot\vec{\sigma})$ and hence for the same reason as above, $q(W) = q(W^\dagger)$. In order to keep the acceptance ratio at a reasonable rate, we sample $N_{\text{inner-iter}}$ different instances of W_1, and multiply them together to form the sampled W. One should choose $N_{\text{inner-iter}}$ such that it is possible to keep the acceptance rate around 30%.

Appendix E: Convergence in number of samples $N \to \infty$

We now turn to an example where we clearly observe the convergence of the distributions $h(v)$ and $\mu(v)$ around the known true figure-of-merit. Consider a noisy identity process on a qutrit, of the form

$$\Lambda_{A\to B}(\rho) = p\,\rho + (1-p)\,d_B^{-1}\mathbb{1}_B\ , \tag{E1}$$

with $p = 0.96$ and $d_B = 3$. This gives us a diamond norm to the identity process of

$$\frac{1}{2}\|\Lambda_{A\to B} - \mathcal{I}_{A\to B}\|_\diamond = 0.03556\ . \tag{E2}$$

We consider measurements on the input and output systems given by using the Gell-Mann matrices as observables:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}; \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}; \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix};$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}; \qquad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Each single-system measurement setting has three possible outcomes. For each pair of measurement settings (for the input and the output system) we simulate N measurement outcomes. We choose $N = 10^6$ for our reference experiment, yielding a total of $N_{\text{tot}} = 8^2 \times 10^6 = 6.4 \times 10^7$ measurement outcomes. We denote the corresponding frequency vector by $(n^{\text{Ref}}_{j_A j_B, \ell_A \ell_B})$, where j_i labels the measurement setting on system i and ℓ_i labels the corresponding measurement outcome. We group together all the indices into a collective index k, such that n^{Ref}_k denotes the number of times the joint POVM effect $E^{(k)}_{AB}$ was observed.

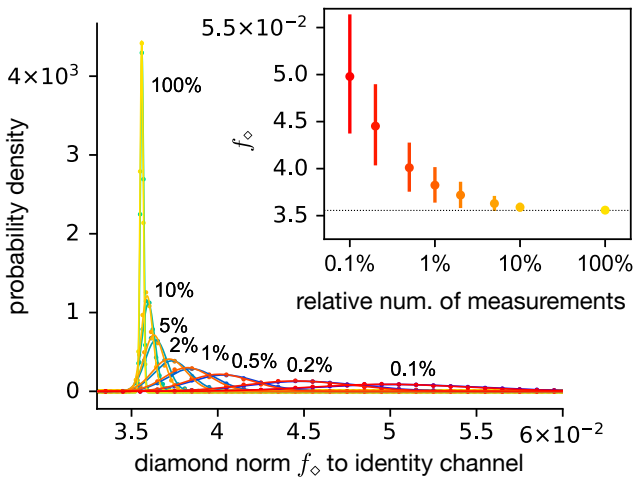

FIG. 6. Convergence of quantum error bars to the true value of figure of merit in the limit of many measurements, for a noisy identity process on a qutrit. Measurements using Gell-Mann matrices as observables on the input and the output systems were simulated with 10^6 outcomes per setting, providing the reference experiment (labeled "100%"). Analyses as in Fig. 4 were then carried out after artificially rescaling the measured frequency counts by various factors (percentage labels), allowing us to compare regimes with different number of measurements while still keeping the estimated expectation values of the measured observables constant to facilitate comparison. As the number of measurements increases, the distribution of $f_\diamond$, the diamond norm to the identity channel, peaks to the known true value of 3.556×10^{-2}. Data points display the numerical histogram (biparite-state sampling method: blue–green; channel-space method: red–yellow) which are fit to our model #1. **Inset:** the quantum error bars (v_0, Δ, γ) obtained from the fit [7] (channel-space method only) are plotted against the number of measurements relative to the reference experiment; markers represent v_0 with an error bar representing $[v_0 - (\Delta - \gamma), v_0 + \Delta + \gamma]$ for each analysis instance. The dotted line indicates the known true value of $f_\diamond$.

The corresponding analysis is depicted in Fig. 6, as the curve labeled "100%". Thanks to the large number of measurements, the distributions $h(v)$ and $\mu(v)$ peak sharply around the true value of $f_\diamond$.

We now ask, how would these distributions look if fewer measurements had been taken? Instead of simulating new outcomes, which would cause the peak to be displaced and would make a comparison more difficult, we artificially rescale the frequency vector n_k^{Ref} by a factor α, i.e., we define $n_k^\alpha = \lfloor \alpha\, n_k^{\mathrm{Ref}} \rfloor$, where by $\lfloor x \rfloor$ we denote the largest integer less than or equal to x. For instance, we may choose $\alpha = 0.01 = 1\%$ to represent an experiment in which only $n \approx \alpha\, N = 10^4$ measurements per setting were sampled, instead of N. While this rescaling of the frequency vector is artificial, the resulting measurement counts are still representative of possible outcomes that one could have sampled if we had simulated directly only $\alpha\, N$ outcomes per setting; crucially, doing so facilitates comparisons between the different settings. The analysis for a selection of values for α (given as percentages) is presented in Fig. 6. The corresponding peaks are indeed seen to converge towards the true value of $f_\diamond$. For each value of α, we calculate the corresponding quantum error bars (v_0, Δ, γ), and plot them against α (Fig. 6, inset). The quantum error bars become a better and tighter description of the true state as the number of measurements increase, as expected.

The quantum error bar Δ is the one which is most akin to a "standard deviation," as in the limit $\gamma \to 0$ the fit model (9) becomes a Gaussian. We may investigate the precise scaling of Δ as a function of the number of measurements by plotting the magnitude of this quantum error bar against the number of measurements in a log-log plot (Fig. 7). We indeed observe a scaling close to $1/\sqrt{n}$, where $n \approx \alpha N$ is the number of measurements, as expected from known results in usual quantum tomography. We expect that by improving the measurement settings, for instance by using adaptive measurements, tighter error bars can be achieved with fewer measurements [40–43].

This depiction allows us again to appreciate the convergence to the true value of $f_\diamond$.

Appendix F: Relations of two methods

In this section, we discuss the theoretical connections between the two methods, specifically the relationship between the densities $\mu(v)$ and $h(v)$. We will use basic notions from measure theory which is available in any standard textbook.

Recall that we use the induced measure $\mathrm{d}\sigma_{AB}$ on density matrices $\mathrm{D}(\mathcal{H}_{AB})$ in the biparite-state sampling and the measure $\mathrm{d}\nu(\Psi)$ on Choi state $\mathscr{C}(\mathcal{H}_{AB})$ in the channel space method. It is helpful for the reader to refresh the definition of these measures in Appendix A 2. The following result connect these probability measures; its proof is delayed till

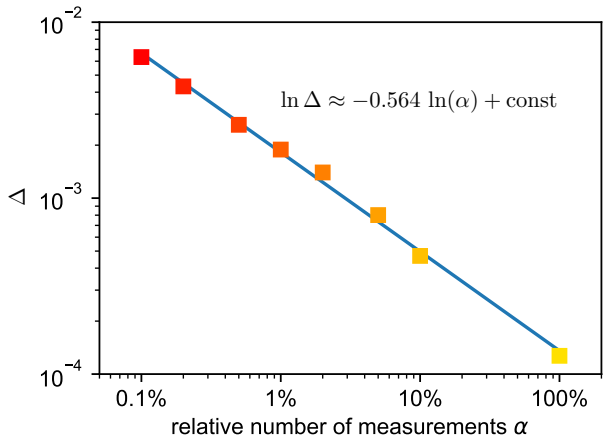

FIG. 7. One of the quantum error bars, Δ, is observed to scale approximately as $1/\sqrt{N}$, where N indicates the number of measurements, as expected in standard (non-adaptive) quantum tomography. The setting is the same as in Fig. 6. By choosing more sophisticated measurement operators, for instance by adapting the measurement settings based on earlier outcomes, the scaling could be improved [40–43].

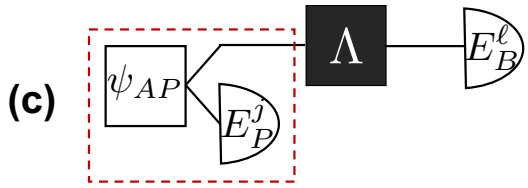

FIG. 8. An intermediate tomographic scheme. Scenario (c) comes from restricting E_k^ℓ acting on BP of Fig. 2**(b)** to be a tensor product measurement. The measurement E_P^j on half of an entangled state in (c) can be seen as a probabilistic state preparation similar to Fig. 2**(a)**.

the end of this Appendix.

Proposition 5. *The measure* $\mathrm{d}\sigma_{AB}$ *factors as* $\mathrm{d}\sigma_A \mathrm{d}\nu(\Lambda_{A\to B})$ *in the sense that for all measurable function* $g(\sigma_{AB})$

$$\int \mathrm{d}\sigma_{AB}\, g(\sigma_{AB}) = \int \mathrm{d}\sigma_A \int \mathrm{d}\nu(\Lambda_{A\to B})\, g(d_A \sigma_A^{1/2} J(\Lambda_{A\to B}) \sigma_A^{1/2}) \tag{F1}$$

$$= \int \mathrm{d}\sigma_A \int \mathrm{d}\nu(\Lambda_{A\to B})\, g(d_A \sigma_P^{1/2} J(\Lambda_{A\to B}) \sigma_P^{1/2})\,, \tag{F2}$$

where $\mathrm{d}\sigma_A$ *is the reduced measure of* $\mathrm{d}\sigma_{AB}$ *via partial tracing and* $\mathrm{d}\nu(\Lambda_{A\to B})$ *is the uniform measure on channel space induced by* $\mathrm{d}U_{BA'B'}$ *and* $\sigma_P = \sigma_A^\intercal$.

We remark that intuitively this result is clear: the probability measure $\mathrm{d}\sigma_{AB}$ can be "conditioned" on different values of $y = \mathrm{tr}_B(\sigma_{AB})$ giving rise to conditional probability measures $\mathrm{d}\nu_y(\sigma_{AB})$ and these are recognised as $\mathrm{d}\nu(\Lambda)$ by unitary invariance. However, the fact that these events which we are conditioning on has measure zero under $\mathrm{d}\sigma_{AB}$ makes the proof more complicated.

Proposition 5 tells us that integrating over all bipartite states according to the measure $\mathrm{d}\sigma_{AB}$ can be done by separately integrating over all possible input states σ_A and over all possible channels $\Lambda_{A\to B}$, by combining them as $\sigma_A^{1/2}\, \Lambda_{AB}\, \sigma_A^{1/2}$ where $\Lambda_{AB} = J(\Lambda_{A\to B})$. Equivalently, this can be done by separately integrating over all possible *transposed* input states σ_A and over all possible channels $\Lambda_{A\to B}$ as in (F2). We can use this intuition to relate the two methods presented above.

In order to connect both quantities, we consider the situation depicted in Fig. 8. Assume that for each repetition $j = 1 \ldots n$ the input ρ_A^j is chosen by a measurement on the pure state $|\psi\rangle_{AP} = \sigma_A^{1/2} |\hat{\Phi}\rangle$ for some given state σ_A, and the outcome POVM effect E_P^j was observed. The measurement on the output state of the channel is chosen from some collection of measurements acting only on system B only. Assuming that the outcome POVM effect E_B^j was observed, the dataset E consists of the pairs (E_P^j, E_B^j) for all n repetitions.

Viewing this scenario as an ancilla-assisted scheme (by moving the measurement on P to the end), we can employ the biparite-state sampling method and calculate $\mu(v)$ by integrating our test function $\delta(f(\rho_{AB}) - v)$ over the *full biparite-state space* according to (B15) and (B5):

$$\mu(v) = c_E^{-1} \int \mathrm{d}\sigma_{AB}\, \mathcal{L}_1(\sigma_{AB}|E)\, \delta(f(\rho_{AB}) - v) \;, \tag{F3}$$

where

$$\mathcal{L}_1(\sigma_{AB}|E) = \mathrm{tr}(\sigma_{AB}^{\otimes n}\, E)\,, \tag{F4}$$

for $E = \otimes_{j=1}^n E_P^j \otimes E_B^j$.

On the other hand, we can also view this as a prepare and measure scheme and use the channel space method to compute, the histogram by (C3) and (C42) as an integration over the *space of all channels* only,

$$h(v) = c_E'^{-1} \int \mathrm{d}\nu(\Lambda_{A\to B}) \mathcal{L}_2(\Lambda|E) \delta(f_{\text{channel}}(\Lambda) - v) \;, \tag{F5}$$

where

$$\mathcal{L}_2(\Lambda|E) = \prod_{j=1}^n \mathrm{tr}(\Lambda_{A\to B}(\rho_A^j)\, E_B^j)\,. \tag{F6}$$

We may rewrite each factor term using the Choi-Jamiolkowski state of the channel as

$$\mathrm{tr}(\Lambda_{A\to B}(\rho_A^j)\, E_B^j) = \mathrm{tr}(\sigma_P^{1/2}\, \Lambda_{PB}\, \sigma_P^{1/2}\, (E_B^j \otimes E_P^j)) \tag{F7}$$

(where $\sigma_P = \sigma_A^T$) and thus

$$\mathcal{L}_2(\Lambda|E;\sigma_A) = \mathrm{tr}((\sigma_P^{1/2}\, \Lambda_{PB}\, \sigma_P^{1/2})^{\otimes n}\, E) \;, \tag{F8}$$

now defining the same operator $E = \otimes_{j=1}^n E_P^j \otimes E_B^j$ as before and where σ_A is fixed.

The similarity of (F3) and (F4) with (F5) and (F8) is now more evident. It is worth giving a precise interpretation to both $\mathcal{L}_1(\sigma_{AB}|E)$ and $\mathcal{L}_2(\Lambda|E;\sigma_A)$. The function $\mathcal{L}_1(\sigma_{AB}|E)$ is a probability density on the biparite-state space with respect to $\mathrm{d}\sigma_{AB}$, describing the Bayesian posterior distribution after observing data E for an agent using the uniform prior $\mathrm{d}\sigma_{AB}$ (and thus ignoring any prior information about what the input state actually is). On the other hand, $\mathcal{L}_2(\Lambda|E;\sigma_A)$ is the posterior distribution in the space of all channels, after observing data E for an agent which is using the prior $\mathrm{d}\nu(\Lambda_{A\to B})$. Yet, Prop. 5 tells us that the prior $\mathrm{d}\nu(\Lambda_{A\to B})$ is precisely the same as the prior in the biparite-state space corresponding to knowing with certainty that the input state is exactly σ_A. Indeed, $\mathrm{d}\nu(\Lambda_{A\to B})$ is precisely the measure induced by $d\sigma'_{AB}\delta(\mathrm{tr}_B(\sigma'_{AB}) - \sigma_A)$ on $\Lambda_{A\to B} = J^{-1}(\sigma_A'^{-1/2}\sigma'_{AB}\sigma_A'^{-1/2})$, where $\delta(\mathrm{tr}_B(\sigma'_{AB}) - \sigma_A)$ is a Dirac delta at the point σ_A. That is, with the shorthand $\sigma'_A = \mathrm{tr}_B(\sigma'_{AB})$, we may rewrite (F5) as

$$h(v) = c_E'^{-1} \int d\sigma'_{AB}\delta(\sigma'_A - \sigma_A) \int \mathrm{d}\nu(\Lambda_{A\to B}) \cdot \mathcal{L}_2(\Lambda|E;\sigma'_A)\delta(f_{\text{channel}}(\Lambda_{A\to B}) - v) \;. \tag{F9}$$

Hence, the difference between the bipartite sampling method and the channel-space method, at least in the current scenario, is exactly the prior information about the input state. In the former, nothing is assumed about the input state other than what can be inferred directly from the measurement data; in the latter, the exact input state is assumed with certainty as represented by the first Dirac delta function in (F9).

Finally, we will prove the following result, which is easily seen to imply the Proposition 5.

Proposition 6. *There exists an essentially unique family of probability measures* $\mathrm{d}\nu_y(\sigma_{AB})$ *on* $\mathrm{D}(\mathcal{H}_{AB})$ *indexed by full rank* $y \in \mathrm{D}(\mathcal{H}_A)$ *such that*

$$\int \mathrm{d}\sigma_{AB} g(\sigma_{AB}) = \int \mathrm{d}\sigma_A(y) \int_{\mathrm{tr}_B^{-1}(y)} \mathrm{d}\nu_y(\sigma_{AB}) g(\sigma_{AB})\,, \tag{F10}$$

where $\mathrm{d}\sigma_A(y)$ *is the reduced measure of* $\mathrm{d}\sigma_{AB}$ *and* $\mathrm{tr}_B^{-1}(y)$ *denotes the preimage of* y *under partial tracing* B*. Moreover, each member* $\mathrm{d}\nu_y(\sigma_{AB})$ *of the family is supported on* $\mathrm{tr}_B^{-1}(y)$ *and actually isomorphic to* $\mathrm{d}\nu(\Psi)$ *on* $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$*. These isomorphisms are given by*

$$J_y^{-1} : \mathrm{tr}_B^{-1}(y) \to \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) \tag{F11}$$

$$\sigma_{AB} \mapsto J^{-1}\left(d_A^{-1} y^{-1/2} \sigma_{AB} y^{-1/2}\right) \;. \tag{F12}$$

Proof. Again, it will be convenient to work in the purified picture. By definition, $\mathrm{d}\sigma_{AB}$ originates from the uniform spherical measure $\mathrm{d}|\phi\rangle_{ABA'B'}$ induced by the Haar measure $\mathrm{d}U_{ABA'B'}$ by the relation $|\phi\rangle_{ABA'B'} = U_{ABA'B'}|\Psi_0\rangle$. On the other hand, $\mathrm{d}\nu(\Psi)$ comes from the Haar measure $\mathrm{d}U_{BA'B'}$ via the relation $|\Psi\rangle = U_{BA'B'}|\Psi_0\rangle$.

Consider the partial trace $\mathrm{tr}_{BA'B'} : \mathrm{End}(\mathcal{H}_{ABA'B'}) \to \mathrm{End}(\mathcal{H}_A)$. Two things happen under this mapping.

First, the measure $\mathrm{d}|\phi\rangle_{ABA'B'}$ admits a pushforward along $\mathrm{tr}_{BA'B'}$ denoted as $\mathrm{d}\sigma_A(y)$ living on space $\mathrm{D}(\mathcal{H}_A)$. Note that this measure $\mathrm{d}\sigma_A(y)$ no longer coincides with the Haar induced (or Hilbert-Schmidt induced) uniform measure on $\mathrm{D}(\mathcal{H}_A)$ (since such measure arises uniquely from the Haar measure $\mathrm{d}U_{AA'}$ acting on $\mathcal{H}_{AA'}$).

Second, the space $\mathrm{End}(\mathcal{H}_{ABA'B'})$ is partitioned into fibers $\mathrm{tr}^{-1}_{BA'B'}(y)$ over $y \in \mathrm{End}(\mathcal{H}_A)$. Observe that one of such fibers corresponds to the set of purified Choi states $\mathscr{PC}$: take $y = \mathbb{1}/d_A$. Moreover, if $y \in \mathrm{D}(\mathcal{H}_A)$ is full rank, then the fiber over y is isomorphic to $\mathscr{PC}$. Indeed, the bijection is given by

$$J_y^{-1} : \mathrm{tr}^{-1}_{BA'B'}(y) \to \mathscr{PC} \tag{F13}$$

$$\phi_{ABA'B'} \mapsto d_A^{-1} y^{-1/2} \phi_{ABA'B'} y^{-1/2} \,. \tag{F14}$$

Note that partial tracing out $A'B'$ gives Choi-Jamiolkowski isomorphisms identifying $\mathrm{tr}_B^{-1}(y) \subseteq \mathrm{D}(\mathcal{H}_{AB})$ with the space of all quantum processs:

$$J_y^{-1} : \mathrm{tr}_B^{-1}(y) \to \mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B) \tag{F15}$$

$$\sigma_{AB} \mapsto J^{-1}\left(d_A^{-1} y^{-1/2} \sigma_{AB} y^{-1/2}\right) , \tag{F16}$$

where J^{-1} is the standard Choi-Jamiolkowski isomophism identifying $\mathscr{C}(\mathcal{H}_{AB})$ with $\mathscr{C}(\mathcal{H}_A \to \mathcal{H}_B)$. We stress again that these are isomorphisms only for full rank $y \in \mathrm{D}(\mathcal{H}_A)$.

The probability measure $\mathrm{d}\phi_{ABA'B'}$ then disintegrates [44] into a family of conditional probability measures $\mathrm{d}\nu_y(\phi_{ABA'B'})$ on each fiber (or preimage over y) $\mathrm{tr}^{-1}_{BA'B'}(y)$ such that

$$\int \mathrm{d}\phi_{ABA'B'} g(\phi_{ABA'B'}) = \int \mathrm{d}\sigma_A(y) \int_{\mathrm{tr}^{-1}_{BA'B'}(y)} \mathrm{d}\nu_y(\phi_{ABA'B'}) g(\phi_{ABA'B'}) \tag{F17}$$

for all functions $g(\phi_{ABA'B'})$. Moreover, the family $\{\mathrm{d}\nu_y(\phi_{ABA'B'}) : y \in \mathrm{D}(\mathcal{H}_A)\}$ is $\mathrm{d}\sigma_A(y)$-almost everywhere unique and each member $\mathrm{d}\nu_y(\phi_{ABA'B'})$ is supported on $\mathrm{tr}^{-1}_{BA'B'}(y)$.

Without loss of generality, we only pay attention to full rank $y \in \mathrm{D}(\mathcal{H}_A)$ because the set of rank-deficient density matrices y has measure zero under $\mathrm{d}\phi_{ABA'B'}$. Here, each fiber $\mathrm{tr}^{-1}_{BA'B'}(y)$ has been identified with the space $\mathscr{PC}$. Under this identification, we will show that $\mathrm{d}\nu_y(\phi_{ABA'B'})$ is almost everywhere equivalent to with the uniform measure on channel space $\mathrm{d}\nu(\Psi)$. This follows from unitary invariance of $\mathrm{d}\nu_y(\phi_{ABA'B'})$ and the uniqueness of the Haar measure $\mathrm{d}U_{BA'B'}$. Specifically, since $\mathrm{d}(U^\dagger_{BA'B'} \phi_{ABA'B'} U_{BA'B'}) = \mathrm{d}\phi_{ABA'B'}$ for all $U_{BA'B'}$ we have by change of variables

$$\int \mathrm{d}\phi_{ABA'B'} g(\phi_{ABA'B'}) = \int \mathrm{d}(U^\dagger_{BA'B'} \phi_{ABA'B'} U_{BA'B'}) g(\phi_{ABA'B'}) \tag{F18}$$

$$= \int \mathrm{d}\phi_{ABA'B'} g(U_{BA'B'} \phi_{ABA'B'} U^\dagger_{BA'B'}) \tag{F19}$$

$$= \int \mathrm{d}\sigma_A(y) \int_{\mathrm{tr}^{-1}_{BA'B'}(y)} \mathrm{d}\nu_y(\phi_{ABA'B'}) g(U_{BA'B'} \phi_{ABA'B'} U^\dagger_{BA'B'}) \tag{F20}$$

$$= \int \mathrm{d}\sigma_A(y) \int_{\mathrm{tr}^{-1}_{BA'B'}(y)} \mathrm{d}\nu_y(U^\dagger_{BA'B'} \phi_{ABA'B'} U_{BA'B'}) g(\phi_{ABA'B'}) \tag{F21}$$

$$\tag{F22}$$

where the last equality follows from the fact that the fiber $\mathrm{tr}^{-1}_{BA'B'}(y)$ is invariant under all $U_{BA'B'}$. By uniqueness of the family, we have

$$\mathrm{d}\nu_y(U^\dagger_{BA'B'} \phi_{ABA'B'} U_{BA'B'}) = \mathrm{d}\nu_y(\phi_{ABA'B'}) \text{ for all } U_{BA'B'} \,. \tag{F23}$$

This says that each member $\mathrm{d}\nu_y(\phi_{ABA'B'})$ of the disintegration family is unitary invariant. Due to the uniqueness of the normalized Haar measure we conclude $\mathrm{d}\nu_y(\phi_{ABA'B'}) = \mathrm{d}\nu(\Psi)$. In fact, we obtain correspondences between the objects

$$\mathrm{tr}^{-1}_{BA'B'}(y) \leftrightarrow \mathscr{PC} \tag{F24}$$

$$\mathrm{d}\nu_y(\phi_{ABA'B'}) \leftrightarrow \mathrm{d}\nu(\Psi) \tag{F25}$$

induced by J_y^{-1}.

Taking partial trace of system $A'B'$ yields the statement of the proposition and completes the proof. □

[1] C. W. Helstrom, Journal of Statistical Physics **1**, 231 (1969).
[2] M. Paris and J. Rehacek, *Quantum State Estimation*, 1st ed. (Springer Publishing Company, Incorporated, 2010).
[3] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
[4] T. L. Scholten and R. Blume-Kohout, ArXiv e-prints (2016), arXiv:1609.04385 [quant-ph].
[5] J. Shang, Z. Zhang, and H. K. Ng, Phys. Rev. A **95**, 062336 (2017).
[6] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
[7] P. Faist and R. Renner, Phys. Rev. Lett. **117**, 010404 (2016).
[8] R. Blume-Kohout, ArXiv e-prints (2012), arXiv:1202.5270 [quant-ph].
[9] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, New Journal of Physics **15**, 123026 (2013).
[10] X. Li, J. Shang, H. K. Ng, and B.-G. Englert, Phys. Rev. A **94**, 062112 (2016).
[11] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[12] I. L. Chuang and M. A. Nielsen, Journal of Modern Optics **44**, 2455 (1997).
[13] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Physical Review A **77**, 012307 (2008), arXiv:0707.0963.
[14] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009).
[15] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Phys. Rev. X **4**, 011050 (2014), arXiv:1306.2348.
[16] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, ArXiv e-prints (2017), arXiv:1701.04299 [quant-ph].
[17] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nature Communications **8**, 1 (2017), arXiv:1605.07674.
[18] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, (2017), arXiv:1701.03135.
[19] C. Portmann and R. Renner, ArXiv e-prints (2014), arXiv:1409.3525 [quant-ph].
[20] D. Aharonov and M. Ben-Or, in *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97 (ACM, New York, NY, USA, 1997) pp. 176–188.
[21] P. Faist, L. P. Thinh, J. Helsen, D. Elkouss, and S. Wehner, "The QPTomographer project," `https://github.com/Tomographer/QPtomographer` (2018).
[22] A. Montanaro and R. de Wolf, ArXiv e-prints (2013), arXiv:1310.2035 [quant-ph].
[23] B. Schumacher, Physical Review A **54**, 2614 (1996), arXiv:9604023 [quant-ph].
[24] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Physical Review A **71**, 062310 (2005), arXiv:0408063 [quant-ph].
[25] C. Pfister, M. A. Rol, A. Mantri, M. Tomamichel, and S. Wehner, Nature Communications **9**, 27 (2018).
[26] S. Chib and E. Greenberg, The American Statistician **49**, 327 (1995).
[27] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "SCS: Splitting conic solver, version 1.2.6," `https://github.com/cvxgrp/scs` (2016).
[28] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, Journal of Optimization Theory and Applications **169**, 1042 (2016).
[29] V. Ambegaokar and M. Troyer, American Journal of Physics **78**, 150 (2010), arXiv:0906.0943 [physics.comp-ph].
[30] M. Wolf, Lecture Notes Available at http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf. (2012).
[31] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
[32] J. Watrous, Draft Book Available at https://cs.uwaterloo.ca/ watrous/TQI/ (2016).
[33] J. Watrous, ArXiv e-prints (2009), arXiv:0901.4709 [quant-ph].
[34] N. Yamamoto, S. Hara, and K. Tsumura, Physical Review A **71**, 022322 (2005), arXiv:0606105 [quant-ph].
[35] A. S. Fletcher, P. W. Shor, and M. Z. Win, Physical Review A **75**, 012338 (2007), 0606035 [quant-ph].
[36] K. Zyczkowski and H.-J. Sommers, Journal of Physics A **34**, 7111 (2001).
[37] G. Chiribella, "On quantum estimation, quantum cloning and finite quantum de finetti theorems," in *Theory of Quantum Computation, Communication, and Cryptography: 5th Conference, TQC 2010, Leeds, UK, April 13-15, 2010, Revised Selected Papers*, edited by W. van Dam, V. M. Kendon, and S. Severini (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011) pp. 9–25.
[38] In fact, in addition to reconstructing the channel $\Lambda_{A\to B}$, we may use this procedure to confirm the correct preparation of the input state σ_A.
[39] H. A, *Probabilistic and Statistical Aspects of Quantum Theory*, 1st ed. (North-Holland, Amsterdam, 1982).
[40] T. Sugiyama, P. S. Turner, and M. Murao, Physical Review A **85**, 52107 (2012).
[41] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Physical Review Letters **111**, 183601 (2013).
[42] C. Granade, C. Ferrie, and S. T. Flammia, ArXiv e-prints (2016), arXiv:1605.05039 [quant-ph].
[43] I. A. Pogorelov, G. I. Struchalin, S. S. Straupe, I. V. Radchenko, K. S. Kravtsov, and S. P. Kulik, Phys. Rev. A **95**, 012302 (2017).

[44] P.-A. M. Claude Dellacherie, *Probabilities and Potential*, Vol. 29 (North-Holland Mathematics Studies, 1978).