

## Efficient unitarity randomized benchmarking of few-qubit Clifford gates

Dirkse, Bas; Helsen, Jonas; Wehner, Stephanie

**DOI**

[10.1103/PhysRevA.99.012315](https://doi.org/10.1103/PhysRevA.99.012315)

**Publication date**

2019

**Document Version**

Final published version

**Published in**

Physical Review A

**Citation (APA)**

Dirkse, B., Helsen, J., & Wehner, S. (2019). Efficient unitarity randomized benchmarking of few-qubit Clifford gates. *Physical Review A*, 99(1), Article 012315. <https://doi.org/10.1103/PhysRevA.99.012315>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

**Efficient unitarity randomized benchmarking of few-qubit Clifford gates**Bas Dirkse,<sup>1,2</sup> Jonas Helsen,<sup>1</sup> and Stephanie Wehner<sup>1</sup><sup>1</sup>*QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*<sup>2</sup>*QuSoft, CWI, and University of Amsterdam, Science Park 123 1098 XG Amsterdam, The Netherlands*

(Received 3 August 2018; published 10 January 2019)

Unitarity randomized benchmarking (URB) is an experimental procedure for estimating the coherence of implemented quantum gates independently of state preparation and measurement errors. These estimates of the coherence are measured by the unitarity. A central problem in this experiment is relating the number of data points to rigorous confidence intervals. In this work we provide a bound on the required number of data points for Clifford URB as a function of confidence and experimental parameters. This bound has favorable scaling in the regime of near-unitary noise and is asymptotically independent of the length of the gate sequences used. We also show that, in contrast to standard randomized benchmarking, a nontrivial number of data points is always required to overcome the randomness introduced by state preparation and measurement errors even in the limit of perfect gates. Our bound is sufficiently sharp to benchmark small-dimensional systems in realistic parameter regimes using a modest number of data points. For example, we show that the unitarity of single-qubit Clifford gates can be rigorously estimated using few hundred data points under the assumption of gate-independent noise. This is a reduction of orders of magnitude compared to previously known bounds.

DOI: [10.1103/PhysRevA.99.012315](https://doi.org/10.1103/PhysRevA.99.012315)**I. INTRODUCTION**

In order to further advance the efforts in building large-scale quantum computers, it is essential to characterize the errors of elementary quantum gates in practical implementations. Randomized benchmarking (RB) [1–4] has in the past years become the standard for assessing the quality of quantum gates [2,5–10]. This is because RB has a simple and efficiently scalable implementation that characterizes gates errors independently of any state preparation and measurement (SPAM) errors. Since the introduction of randomized benchmarking, several variants have been developed [11–15]. One of these variants is unitarity randomized benchmarking (URB) [12,16].

This paper is concerned with the URB protocol proposed in [12]. It provides a method to characterize the coherence of errors in implemented quantum gates that is robust against SPAM errors. This characterization of coherence is quantified by the unitarity, a quantity that is independent of the average gate fidelity measured by standard RB. Being able to estimate the unitarity experimentally provides an extra source of information when optimizing experimental implementations of quantum gates [16]. In particular, the unitarity can help to discriminate whether the dominant error process is coherent (i.e., overrotation or calibration errors) or incoherent (i.e., depolarizing or dephasing noise). This information is useful since these two different types of noise are generally reduced in different ways [16,17]. Additionally, knowing the unitarity of a gate or gate set can be used to get sharper bounds on the credible interval of an interleaved randomized benchmarking experiment [18] and also get improved bounds on the diamond norm error [19–21], which is the relevant metric in the setting of fault-tolerant quantum computing.

The URB protocol is similar to the standard RB protocol and they share many characteristics, like SPAM independent

estimation of its figure of merit. It aims only to provide a partial characterization of the gate set (by estimating the unitarity), instead of characterizing the noise completely, which is what, e.g., channel or gate set tomography aim to do. Since full tomography with rigorous confidence intervals is very resource-intensive [22], in situations where partial noise characterization suffices, more lightweight solutions like RB and URB may be the choice of preference.

In RB-type protocols, the noise-characterizing figure of merit is obtained from the exponential decay rate of the average survival probability with the length of the sequence of gates. For fixed sequence length, the average survival probability is estimated by averaging over a number of randomly sampled gate sequences. An important problem for RB-type procedures is then determining a number of random gate sequences that is practical yet yields a confident estimate of the figure of merit. This problem was realized in the first concrete proposal of RB [4]. Subsequent work focused on resolving this problem in two different, complementary ways. First, statistical tools were applied to allow for confident estimation of the RB decay rate with fewer random gate sequences [23–25]. Second, the underlying distribution from which the RB protocol samples data was analyzed. In particular a sharp bound on the variance of this distribution was derived, which also allows for more resource-efficient estimation of the RB decay rate from measurement data [26,27]. However, no such analysis exists for the related URB protocol.

Here we analyze the statistics of unitarity randomized benchmarking. The aim of this work is to contribute a solution to the following central question: How many random sequences of gates are required in the URB protocol to get a confident estimate of the unitarity from the obtained measurement data? We proceed along the lines of Refs. [26,27] by providing a sharp bound on the variance of the underlying distribution from which the URB protocol samples.

This additional knowledge of the URB sampling distribution allows for more resource-efficient estimation of the unitarity from experimental data. Concretely we demonstrate how our variance bound can be used to bound the required number of random sequences as a function of desired confidence parameters.

In this work, we derive a bound on the variance of the distribution induced by the random sampling of gate sequences in a modified version of the Clifford URB protocol. This modification is based on the adapted RB protocol of Ref. [27]. It requires no experimental overhead while leading to a sharper variance bound (and hence fewer required gate sequences) as well as a simpler fit model for extracting the unitarity. In addition, our statistical analysis reveals the optimal input state and output measurement for minimizing the variance and maximizing the signal strength. We then apply this variance bound using standard concentration inequalities to relate the number of random sequences to desired confidence intervals. Our result is sufficiently sharp to perform the modified URB protocol on few-qubit systems with a modest number of sequences in realistic parameter regimes. It is an improvement of several orders of magnitude in the number of sequences required for fixed confidence, compared to a concentration inequality that does not use the variance (as was first done for RB in Ref. [4]). We show that the variance, and thus number of required gate sequences, scales favorably in the regime of large unitarity, which is the relevant regime for high quality gates. We also show that, in contrast to standard RB [27], a nontrivial number of sequences is always required to overcome the randomness introduced by state preparation and measurement errors even in the limit of perfect gates.

This paper is organized as follows. In the remainder of this section we review the concept of unitarity and the URB protocol to estimate the unitarity of a gate set. We introduce a modification of the protocol based on Ref. [27] for the purpose of improved statistics. Furthermore we explicitly distinguish the two different implementations of the URB protocol and emphasize their benefits and drawbacks. In Sec. II we present our main result [Eqs. (18) and (19)] and illustrate how to apply it using a simulated example. In Sec. III we examine the behavior of our bound in various parameter regimes and discuss the different features of our bound. A brief overview of the proof techniques used to derive our main result is presented in Sec. IV. All technical details of the proof have been delegated to the appendices. In Sec. V we summarize the main conclusions of our work and provide suggestions for future research.

### A. Unitarity

Let us begin with defining the figure of merit that URB estimates. For a quantum channel  $\mathcal{E}$  (here a quantum channel will refer to a completely positive and trace-preserving (CPTP) superoperator), the unitarity is defined as [12]

$$u(\mathcal{E}) = \frac{d}{d-1} \int d\psi \text{Tr} \left\{ \left[ \mathcal{E} \left( |\psi\rangle\langle\psi| - \frac{I}{d} \right) \right]^2 \right\}, \quad (1)$$

where the integration is with respect to the uniform Haar measure on the state space  $\mathcal{H}$ . The prefactor is chosen such that  $0 \leq u \leq 1$ . An equivalent definition of the unitarity can

be given as [12, Proposition 1]

$$u(\mathcal{E}) = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathbf{P}^*} \text{Tr}[\tau \mathcal{E}(\sigma)]^2, \quad (2)$$

where the summation is over the set of all nonidentity, normalized Pauli matrices  $\mathbf{P}^*$ . The normalization is with respect to the Hilbert-Schmidt norm  $\|\sigma\|_2 = \sqrt{\text{Tr}[\sigma^\dagger \sigma]}$ . This alternative definition of the unitarity is often more pleasant to work with. In Example 1 the unitarity of a depolarizing channel is calculated.

The unitarity has some properties that one would intuitively expect a good measure of the coherence of gates to have [12, Proposition 7]. First,  $u = 1$  if and only if  $\mathcal{E}$  is a unitary quantum channel. Second, the unitarity is invariant under unitary transformation. That is, if  $\mathcal{U}, \mathcal{V}$  are unitary quantum channels, then  $u(\mathcal{E}) = u(\mathcal{U}\mathcal{E}\mathcal{V})$ . The unitarity is independent of but related to the average gate fidelity. In fact, the unitarity provides an upper bound on the average gate fidelity [12, Proposition 8],

$$\left( \frac{dF_{\text{avg}} - 1}{d - 1} \right)^2 \leq u. \quad (3)$$

Here  $F_{\text{avg}}$  is the average gate fidelity between the implemented gate and the ideal target gate. This relation expresses the fact that a perfect gate ( $F_{\text{avg}} = 1$ ) must be unitary ( $u = 1$ ). However, the converse does not hold. Indeed, a unitary gate ( $u = 1$ ) can have arbitrary average gate fidelity by considering purely unitary noise (i.e., overrotation). The inequality Eq. (3) is tight, since it holds with equality for a depolarizing channel.

*Example 1.* Let  $\mathcal{E}$  be a depolarizing quantum channel with depolarizing parameter  $p$ :

$$\mathcal{E} : A \mapsto pA + \frac{1-p}{d} \text{Tr}[A]I.$$

Then the unitarity  $u$  of  $\mathcal{E}$  is computed using Eq. (2) as

$$u = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathbf{P}^*} \text{Tr}[p\tau^\dagger p\sigma]^2 = p^2,$$

since  $\text{Tr}[\tau^\dagger \sigma] = \delta_{\sigma, \tau}$ . Note that  $F_{\text{avg}}(\mathcal{E}) = p + \frac{1-p}{d}$ , so that the inequality Eq. (3) is saturated by the depolarizing channel. ■

### B. The URB protocol

This section gives an overview of the URB protocol of Ref. [12] and gives a small modification based on Ref. [27]. The protocol is described for any gate set  $\mathbf{G}$  that is a unitary 2-design [28]. Note that even though the protocol works for all these gate sets, our result of the confidence analysis is only applicable to the Clifford group. In Algorithm 1 we present an outline of the URB protocol, where we distinguish two different implementations (discussed later in this section).

The URB protocol works similar to the standard RB protocol. First one draws a uniformly distributed random sequence of gates (with length  $m$ ) from the gate set  $\mathbf{G}$ . Denote such a sequence

$$\mathcal{G}_j = \mathcal{G}_{j_m} \cdots \mathcal{G}_{j_2} \mathcal{G}_{j_1}, \quad (4)$$

- Fix a gate set  $\mathbf{G}$ , choose a set of sequence lengths  $\mathbb{M}$  to use and determine the number of random sequences  $N_m$  per sequence length  $m \in \mathbb{M}$ .
- 1: **Procedure** URBG,  $\mathbb{M}$ ,  $\{N_m\}$
  - 2: **for all** sequence lengths  $m \in \mathbb{M}$
  - 3:     **repeat**  $N_m$  times
  - 4:         Sample  $m$  random gates  $\mathcal{G}_{j_1}, \dots, \mathcal{G}_{j_m}$  independently and uniformly at random from  $\mathbf{G}$ ;
  - 5:         Compose the sequence  $\mathcal{G}_j = \mathcal{G}_{j_m} \cdots \mathcal{G}_{j_2} \mathcal{G}_{j_1}$ ;
  - 6:         **if** Two-copy implementation
  - 7:             Prepare states  $\rho \approx \frac{I+S}{d(d+1)}$  and  $\hat{\rho} \approx \frac{I-S}{d(d-1)}$ , apply  $\mathcal{G}_j^{\otimes 2}$  to each state and measure  $E \approx S$  a large number of times (where  $S$  denotes the Swap gate);
  - 8:             From this data, estimate the average sequence purity as
$$q_j^{(2)} = (\text{Tr}[E\mathcal{G}_j^{\otimes 2}(\rho)] - \text{Tr}[E\mathcal{G}_j^{\otimes 2}(\hat{\rho})]) = \text{Tr}[E\mathcal{G}_j^{\otimes 2}(\bar{\rho})];$$
  - 9:         **if** Single-copy implementation
  - 10:             **for all** nonidentity Pauli's  $P, Q \neq I$
  - 11:             Prepare states  $\rho_{\mathcal{H}}^{(P)} \approx \frac{I+P}{d}$  and  $\hat{\rho}_{\mathcal{H}}^{(P)} \approx \frac{I-P}{d}$ , apply  $\mathcal{G}_j$  to each state and measure  $E_{\mathcal{H}}^{(Q)} \approx Q$  a large number of times;
  - 12:             From this data, estimate the average sequence purity as
$$q_j^{(1)} = \frac{1}{d^2 - 1} \sum_{P, Q \neq I} \{\text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\rho_{\mathcal{H}}^{(P)})] - \text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\hat{\rho}_{\mathcal{H}}^{(P)})]\}^2;$$
  - 13:         Compute the empirical average over the sampled sequences  $\bar{q}_m = \frac{1}{N_m} \sum_j q_j$ ;
  - 14:     Fit  $\bar{q}_m = Bu^{m-1}$ , where  $B$  is a constant absorbing SPAM errors and  $u$  is the unitarity of the noise map.
- ALG. 1. Outline of the modified unitarity randomized benchmarking protocol.

where each  $j_s$  denotes the randomly drawn gate from  $\mathbf{G}$  at position  $s$ . The subscript  $\mathbf{j}$  denotes the multi-index  $(j_1, j_2, \dots, j_m)$  and therefore indexes the entire sequence. Such a randomly sampled sequence  $\mathcal{G}_j$  is then applied to a state  $\rho$ , after which a two-outcome measurement is performed (in this work the operator  $E$  denotes the Hermitian observable associated with a two-outcome measurement  $\{M, I - M\}$  with outcomes  $\pm 1$ ). However, there are two differences here with respect to the RB protocol. First, there is no global inverse applied at the end of each sequence and second, the expectation value of the measurement outcome is squared. So the URB random variable of interest then becomes  $q_j = \text{Tr}[E\mathcal{G}_j(\rho)]^2$ . Throughout this work, we shall call the URB random variable  $q_j$  the sequence purity (in standard RB, the random variable of interest is typically referred to as the survival probability). The rest of the procedure is then similar: estimate the mean of the sequence purity  $q_j$  using  $N$  random sequences of fixed length, repeat for various sequence lengths and fit to the model

$$\mathbb{E}[q_j] = Bu^{m-1} + A \quad (5)$$

to obtain the unitarity.

Here we analyze a slightly modified version of the protocol of Ref. [12], based on ideas of Refs. [2,24,27]. Every sequence of randomly sampled gates  $\mathcal{G}_j$  is applied to two different input states  $\rho$  and  $\hat{\rho}$ , and half of the difference of their expectation values is taken before squaring. By linearity of quantum mechanics, this is equivalent to performing URB with the traceless input operator

$$\bar{\rho} := \frac{1}{2}(\rho - \hat{\rho}). \quad (6)$$

The factor  $\frac{1}{2}$  is strictly not necessary but is added for better statistical comparison. The key idea behind this is that one effectively works with a traceless input operator  $\bar{\rho}$ . There are two main benefits of this modification. First, it improves the fitting procedure, because the modified fit model for the mean of the sequence purity becomes [see Eq. (53) in Sec. IV B]

$$\mathbb{E}[q_j] = Bu^{m-1}, \quad (7)$$

where the constant  $B$  only depends on the input operator  $\bar{\rho}$  and the measurement observable  $E$ . This is a linear fitting problem in  $u$  by taking the logarithm and can therefore be performed more easily. Second, this modification narrows the distribution of the sequence purity  $q_j$ , improving the confidence in our point estimate  $\bar{q}_m = \frac{1}{N} \sum q_j$  of the exact  $\mathbb{E}[q_j]$ . In the next section we discuss the implementation of the protocol in more detail and emphasize that there are two possible methods to estimate  $q_j$ .

### 1. The two different implementations

In this section we discuss two different possible implementations of the URB protocol (as briefly discussed in Ref. [12]), which are illustrated in Fig. 1. The choice of implementation depends on whether the experimenter has access to two identical copies of the system or not. The implementations differ in the way the sequence purity  $q_j$  is computed and what the ideal input operator  $\bar{\rho}$  and measurement  $E$  are. By ideal operators, we mean the operators that maximize the signal strength [the proportionality factor  $B$  in the fit model Eq. (7)] from which the unitarity is estimated. We will then show that the two implementations are closely related.

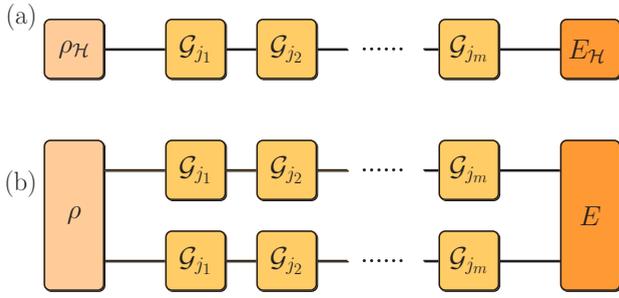


FIG. 1. Schematic difference between the single-copy implementation (a) and the two-copy implementation (b) of the unitarity randomized benchmarking protocol. Each line represents a system on the base Hilbert space  $\mathcal{H}$ . In the single-copy implementation, the expected value of the measurement  $\text{Tr}[E_{\mathcal{H}}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}})]$  needs to be squared to obtain  $q_j$ , whereas in the two-copy implementation  $q_j = \text{Tr}[E\mathcal{G}_j^{\otimes 2}(\bar{\rho})]$  yields the direct outcome.

Let us start by discussing the two-copy implementation [Fig. 1(b)]. As the name suggests, this requires two copies of the system  $\mathcal{H}$  under investigation. The use of two copies follows from the mathematical equivalence:

$$q_j = \text{Tr}[E\mathcal{G}_j(\bar{\rho})]^2 = \text{Tr}[E^{\otimes 2}\mathcal{G}_j^{\otimes 2}(\bar{\rho}^{\otimes 2})]. \quad (8)$$

If the experimenter has access to two identical copies of the system  $\mathcal{H}$ , the input and measurement operator can be entangled across the two copies of the system. The sequence  $\mathcal{G}_j$  is then applied to each half of the system  $\mathcal{H} \otimes \mathcal{H}$ . This yields the sequence purity of the two-copy implementation as

$$q_j^{(2)} = \text{Tr}[E\mathcal{G}_j^{\otimes 2}(\bar{\rho})], \quad (9)$$

where  $\bar{\rho}, E \in \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  are now operators on the two copies of the system. Since  $E$  is a two-valued measurement with outcomes  $(\pm 1)$  and  $\bar{\rho}$  is half the difference between two physical states, it is not hard to show that the sequence purity lies in the interval  $q_j^{(2)} \in [-1, 1]$ . In Sec. II C we show that this interval can be narrowed under mild assumptions. In the two-copy implementation it is implicitly assumed that the experimenter can operate identically on each subsystem without any cross-talk between the two subsystems. Moreover, the experimenter should be able to prepare and measure over the two copies of the system. Experimentally the input and measurement operators  $\bar{\rho}, E \in \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  should be as close to the ideal operators as possible. The ideal operators are given by (see Appendix B 2 for more details and proof)

$$\rho_{\text{id}} = \frac{I + S}{d(d+1)}, \quad \hat{\rho}_{\text{id}} = \frac{I - S}{d(d-1)}, \quad E_{\text{id}} = S, \quad (10)$$

where  $I$  is the identity and  $S$  is the Swap operator on  $\mathcal{H} \otimes \mathcal{H}$ , and  $d$  is the dimension of  $\mathcal{H}$ . The state  $\rho_{\text{id}}$  ( $\hat{\rho}_{\text{id}}$ ) is the maximally mixed state on the symmetric (anti-symmetric) subspace of  $\mathcal{H} \otimes \mathcal{H}$ . Note that the maximally mixed state on a subspace can be prepared by uniformly sampling pure states from an orthonormal basis of this subspace. The operator  $E_{\text{id}}$  is the Hermitian observable associated with a two-valued measurement that discriminates between symmetric (outcome 1) and anti-symmetric states (outcome  $-1$ ).

In the single-copy implementation, the experimenter must obtain an estimate of the sequence purity  $q_j$  using only a single

copy of the system  $\mathcal{H}$ . From Eq. (8), it can be seen that  $q_j = \text{Tr}[E_{\mathcal{H}}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}})]^2$  is the sequence purity given the operators  $\bar{\rho}_{\mathcal{H}}, E_{\mathcal{H}} \in \mathcal{L}(\mathcal{H})$ . Here the subscript  $\mathcal{H}$  is to emphasize that the operators are on a single copy of  $\mathcal{H}$ . Throughout this paper we will just write  $\bar{\rho}$  and  $E$  for operators on  $\mathcal{H} \otimes \mathcal{H}$  and indicate operators on a single copy explicitly by adding a subscript  $\mathcal{H}$ . There are two disadvantages in defining the single-copy sequence purity using one pair of input and measurement operators  $\bar{\rho}_{\mathcal{H}}, E_{\mathcal{H}} \in \mathcal{L}(\mathcal{H})$ . First, the proportionality factor  $B$  in Eq. (7) is upper bounded by  $\frac{1}{d^2-1}$ , where  $d$  is the dimension of  $\mathcal{H}$  [12]. This means that the signal strength decreases exponentially with the system size. Second, the variance of the sequence purity is large. This leads to large uncertainty in the estimated average sequence purity  $\bar{q}_m$ . These disadvantages can be resolved by using multiple different pairs of input and measurement operators [12]. The ideal set of operators is chosen in such a way that summing the expectation values squared for each pair of operators leads to effectively simulating the ideal operators of Eq. (10). Let us make this more precise. Define the single-copy sequence purity as

$$q_j^{(1)} = \frac{1}{d^2-1} \sum_{P, Q \neq I} \text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})]^2, \quad (11)$$

where the sum is over all nonidentity multiqubit Pauli operators  $P, Q$ . Each  $\bar{\rho}_{\mathcal{H}}^{(P)}$  and  $E_{\mathcal{H}}^{(Q)}$  are different input and measurement operator settings indexed by the nonidentity Pauli operators  $P$  and  $Q$ , respectively. For each pair  $P, Q$ , the expectation value  $\text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})]$  is to be estimated experimentally. This expectation can be shown to lie in the interval  $[-1, 1]$  by definition of  $E$  and  $\bar{\rho}$ , so that the expectation value squared lies in the unit interval. Therefore the single-copy sequence purity can in principle lie anywhere in the interval  $q_j^{(1)} \in [0, d^2-1]$ , since each summand lies in the unit interval and the summation runs over  $(d^2-1)^2$  terms. However in Sec. II C we show that this interval can be narrowed significantly under mild assumptions. Since the sum runs twice over all nonidentity Pauli operators, estimating the sequence purity  $q_j^{(1)}$  requires  $(d^2-1)^2$  different settings. This is a number that grows exponentially in the number of qubits comprising the system. We also emphasize that simply squaring and summing up estimates of  $\text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})]$  to obtain an estimate of  $q_j^{(1)}$  yields a positively biased estimator for  $q_j^{(1)}$ . This may lead to overestimating the unitarity. See Sec. IV A 2 for more details on how to correctly estimate  $q_j^{(1)}$ . The states  $\rho_{\mathcal{H}}^{(P)}, \hat{\rho}_{\mathcal{H}}^{(P)}$  and measurement  $E_{\mathcal{H}}^{(Q)}$  should be implemented as closely as possible to the ideal operators

$$\rho_{\mathcal{H},\text{id}}^{(P)} = \frac{I + P}{d}, \quad \hat{\rho}_{\mathcal{H},\text{id}}^{(P)} = \frac{I - P}{d}, \quad E_{\mathcal{H},\text{id}}^{(Q)} = Q. \quad (12)$$

The ideal state  $\rho_{\mathcal{H},\text{id}}^{(P)}$  ( $\hat{\rho}_{\mathcal{H},\text{id}}^{(P)}$ ) is the maximally mixed state on the positive (negative) eigenspace of the Pauli operator  $P$ , and the measurement  $E_{\mathcal{H},\text{id}}^{(Q)}$  is the two-valued measurement that discriminates between the positive (outcome 1) and negative (outcome  $-1$ ) eigenspace of the Pauli operator  $Q$ .

Next we show that the single copy can be interpreted as a special case of the two-copy implementation [this is not surprising in view of Eq. (8)]. To do so, we show that in the single-copy implementation, one effectively works with

two-copy operators of the form

$$\begin{aligned}\bar{\rho}_{\text{eff}} &= \frac{d}{d^2-1} \sum_{P \neq I} \bar{\rho}_{\mathcal{H}}^{(P)} \otimes \bar{\rho}_{\mathcal{H}}^{(P)}, \\ \bar{E}_{\text{eff}} &= \frac{1}{d} \sum_{Q \neq I} \bar{E}_{\mathcal{H}}^{(Q)} \otimes \bar{E}_{\mathcal{H}}^{(Q)}.\end{aligned}\quad (13)$$

Here  $\bar{E}$  ( $\bar{E}_{\mathcal{H}}^{(Q)}$ ) is the traceless part of the observable  $E$  ( $E_{\mathcal{H}}^{(Q)}$ ), defined as

$$\bar{E} := E - \text{Tr}[E] \frac{I}{d}, \quad \bar{E}_{\mathcal{H}} := E_{\mathcal{H}} - \text{Tr}[E_{\mathcal{H}}] \frac{I_{\mathcal{H}}}{d}.\quad (14)$$

The key point is that replacing the observable  $E$  with  $\bar{E}$  makes no difference, since  $\text{Tr}[E \mathcal{G}_j^{\otimes 2}(\bar{\rho})] = \text{Tr}[\bar{E} \mathcal{G}_j^{\otimes 2}(\bar{\rho})]$ . This follows directly from Eq. (14), since  $\text{Tr}[I \mathcal{G}_j^{\otimes 2}(\bar{\rho})] = 0$  by the tracelessness of  $\bar{\rho}$  and the trace-preserving property of  $\mathcal{G}_j^{\otimes 2}$ . Analogously, in the single-copy implementation, the traceless measurement  $\bar{E}_{\mathcal{H}}^{(Q)}$  can be used instead of the observable  $E_{\mathcal{H}}^{(Q)}$ . Throughout the paper, a bar over the measurement operator will mean the traceless component as defined by Eq. (14).

The key idea of Eq. (13) is that  $\bar{\rho}_{\text{eff}}$  and  $\bar{E}_{\text{eff}}$  are constructed such that computing  $q_j^{(1)}$  with Eq. (11) is mathematically equivalent to computing  $q_j^{(2)}$  with Eq. (9) using the effective operators Eq. (13),

$$\begin{aligned}q_j^{(1)} &= \frac{1}{d^2-1} \sum_{P, Q \neq I} \text{Tr}[\bar{E}_{\mathcal{H}}^{(Q)} \mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})]^2 \\ &= \text{Tr}[\bar{E}_{\text{eff}} \mathcal{G}_j^{\otimes 2}(\bar{\rho}_{\text{eff}})] = q_j^{(2)}.\end{aligned}\quad (15)$$

In particular the ideal effective operators  $\bar{\rho}_{\text{eff}, \text{id}}$  and  $\bar{E}_{\text{eff}, \text{id}}$  [defined by Eq. (13) for the ideal single-copy operators Eq. (12)] are equal to the ideal two-copy operators Eq. (10),

$$\bar{\rho}_{\text{eff}, \text{id}} = \bar{\rho}_{\text{id}} \quad \text{and} \quad \bar{E}_{\text{eff}, \text{id}} = \bar{E}_{\text{id}}.\quad (16)$$

This follows from the fact that [12]

$$S = \frac{1}{d} \sum_P P \otimes P.\quad (17)$$

Note that the sum is here over all Pauli matrices including the identity. As a result of this, the rest of the paper will exclusively deal with the two-copy operators  $\bar{\rho}$ ,  $E \in \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$ . The results can be interpreted for the single-copy protocol by considering the effective operators Eq. (13).

The two-copy implementation of the protocol as previously discussed, can only be implemented if the experimenter has access to two different, but identical copies of the system under examination. These two systems must be simultaneously accessible for entangled state preparation and measurements, but the unitary control on each subsystem needs to be fully disjoint (i.e., without crosstalk) and identical (meaning noise must be identical on each subsystem). These assumptions are hard if not impossible to fulfill in any experimental system. We emphasize, however, that the two-copy implementation is introduced as a mathematical tool for the analysis of the URB protocol and its equivalence to the more realistic single-copy protocol was shown.

This concludes our review of the URB protocol, including the proposed modification of traceless input operators

and emphasizing the two different implementations (which we have named the single- and two-copy implementation, respectively). Next, we will present our main result. We will show how a concentration inequality can be used to relate the required resources (the number of sequences  $N$ ) to parameters that quantify the confidence in the estimate of the average sequence purity  $\bar{q}_m$ . To do so, we will present a sharp bound  $\sigma^2$  on the variance of the sequence purity  $\mathbb{V}[q_j^{(K)}]$  and present a bound  $L$  on the length of the interval in which the sequence purity  $q_j^{(K)}$  lies. These bounds are independent of  $K$  (the choice between single or two-copy implementation). Therefore, if no implementation-specific details are discussed, the sequence purity is just denoted  $q_j$ .

## II. SUMMARY OF RESULTS

In this section the main contribution of the paper is summarized. The main result is a sharp bound on the number of sequences  $N$  required to obtain the average sequence purity  $\bar{q}_m$  given fixed sequence length  $m$  with a certain *a priori* determined confidence. In Sec. II A we review a result from statistics to quantify the relation between the number of sequences  $N$  and the confidence. This relation requires some knowledge on the distribution of the sequence purity  $q_j$ . A bound on the variance and a bound on the interval length of the sequence purity are needed. In Sec. II B we present a bound on the variance of the URB sequence purity  $q_j$  for benchmarking the Clifford gate set. This is the main contribution of this work. In Sec. II C we present a bound on the length of the interval in which  $q_j$  must lie. Finally in Sec. II D we give some examples on how to use our results.

### A. Relation between the confidence parameters and the number of sequences

Using concentration inequalities from statistics, the confidence in the estimate  $\bar{q}_m$  can be expressed as the probability that it deviates at most  $\epsilon$  from the exact mean  $\mathbb{E}[q_j]$ . If this probability  $\mathbb{P}[|\bar{q}_m - \mathbb{E}[q_j]| \geq \epsilon] \leq \delta$  is to be bounded by  $\delta$ , then the number of required data points  $N$  is related to the confidence parameters  $\epsilon, \delta$  by [29]

$$2 \left[ \left( \frac{L}{L - \epsilon} \right)^{\frac{L^2 - \epsilon L}{\sigma^2 + L^2}} \left( \frac{\sigma^2}{\sigma^2 + \epsilon L} \right)^{\frac{\sigma^2 + \epsilon L}{\sigma^2 + L^2}} \right]^N \leq \delta.\quad (18)$$

In this expression  $\sigma^2$  is a bound on the variance  $\mathbb{V}[q_j]$  and  $L$  is a bound on the length of the interval in which  $q_j$  lies. Given  $\sigma^2$  and  $L$ , there are two ways to apply this inequality. It can either be solved (numerically) for  $\epsilon$ , given fixed  $N$  and  $\delta$ , or it can be solved for  $N$  given  $\epsilon, \delta$ . In any case, it provides a direct relation between the number of required sequences  $N$  and the confidence parameters  $\epsilon, \delta$ , given  $L$  and  $\sigma^2$ . So in order to apply Eq. (18), the bounds  $L$  and  $\sigma^2$  are needed.

In the next section we will present a sharp bound  $\sigma^2$  on the variance of the sequence purity  $\mathbb{V}[q_j]$ . This bound is the key ingredient in using Eq. (18) and it is the main contribution of this paper.

### B. Bound on the variance of the sequence purity

In this section we present a bound  $\sigma^2$  on the variance of the sequence purity  $\mathbb{V}[q_{\mathbf{j}}]$  that is valid under the following assumptions:

(1) The gate set under investigation is the  $d$ -dimensional Clifford group, denoted  $\mathbf{C}(d)$ . Here  $d = 2^q$  for a  $q$ -qubit system. This assumption is necessary for deriving a variance bound. Even though the expected value  $\mathbb{E}[q_{\mathbf{j}}]$  of the URB sequence purity is independent of the chosen gate set (as long as it is a unitary 2-design), the variance is not. The Clifford group was chosen as the default gate set.

(2) Gate errors are independent of the gate. This is known as the gate-independent error model. In this model, the implemented noisy gate is  $\tilde{\mathcal{G}} = \mathcal{G}\Lambda$ , where  $\mathcal{G} \in \mathbf{C}(d)$  is the ideal Clifford gate and  $\Lambda$  is an arbitrary quantum channel describing the noise. Crucially,  $\Lambda$  does not depend on the specific gate  $\mathcal{G} \in \mathbf{C}(d)$ . This assumption is necessary for deriving the fit model for URB [12]. Consequently our variance bound also employs this assumption. The URB protocol has not been analyzed in a gate dependent noise setting.

(3) The noise map  $\Lambda$  is assumed to be unital if  $q \geq 2$  (or equivalently if  $d \geq 4$ ). A quantum channel  $\Lambda$  is unital if the maximally mixed state is a fixed point of the map,  $\Lambda(I) = I$ . If the system under investigation  $\mathcal{H}$  is a single-qubit system ( $d = 2$ ), then this assumption is not necessary. Our result thus holds for any single-qubit quantum channel  $\Lambda$ . This assumption enters in our derivation of the variance bound. It is not a fundamental assumption but rather a condition under which we were able to derive a useful, sharp bound.

At this point, we emphasize that  $\mathbb{V}[q_{\mathbf{j}}]$  is the between-sequence variance, i.e., the variance of  $q_{\mathbf{j}}$  due to the randomly sampled sequence indexed by  $\mathbf{j}$ . In particular this means that given a sequence  $\mathbf{j}$ , we assume that  $q_{\mathbf{j}}$  can be determined with arbitrary precision. In reality  $q_{\mathbf{j}}$  can only be estimated due to the probabilistic nature of quantum mechanics by taking many single-shot measurements of the same sequence  $\mathbf{j}$ . In Sec. IV A we relax this assumption by splitting the total variance into the sum of the between-sequence variance (the variance due to randomly sampled  $\mathbf{j}$ ) and the within-sequence variance (the variance due to uncertainty in  $q_{\mathbf{j}}$  for fixed  $\mathbf{j}$ ).

Under the assumptions stated above, the following bound on the variance  $\mathbb{V}[q_{\mathbf{j}}]$  is derived (see Theorem 1 in Appendix B):

$$\begin{aligned} \mathbb{V}[q_{\mathbf{j}}^{(K)}] &\leq \sigma^2 \\ &= \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 [c_1(d) + c_2(d) \|\bar{E}_{\text{err}}\|_{\infty}^2 \\ &\quad + c_3(d) \|\bar{\rho}_{\text{err}}\|_1^2] + \|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_{\infty}^2, \end{aligned} \quad (19)$$

which is independent of the used implementation (single- or two-copy, corresponding to  $K = 1, 2$ ). Here  $u$  is the unitarity of  $\Lambda$ ,  $m$  is the sequence length,  $\|\bar{E}_{\text{err}}\|_{\infty}^2$ ,  $\|\bar{\rho}_{\text{err}}\|_1^2$  are quantities depending on the quality of state preparation and measurement and  $c_i$  are constants that solely depend on the dimension  $d$ . The values of  $c_i$  for small  $d$  are tabulated in Table I. For precise definitions of these quantities, see Theorem 1 in Appendix B. The error operators have the following

TABLE I. Evaluation of the constants  $c_i(d)$  for various small-dimensional systems. The last row indicates the asymptotic behavior.

$d$	$c_1(d)$	$c_2(d)$	$c_3(d)$
2	$\frac{11}{12}$	$\frac{13}{9}$	$\frac{5}{2}$
4	$\frac{179}{60}$	54.675	48.053
8	1.6322	81.445	119.31
16	1.1443	110.64	296.88
32	1.0354	173.80	891.69
$\rightarrow \infty$	$O(1)$	$O(d)$	$O(d^2)$

definitions:

$$\begin{aligned} \bar{\rho}_{\text{err}} &= \bar{\rho} - \frac{\text{Tr}[\bar{\rho}_{\text{id}}\bar{\rho}]}{\|\bar{\rho}_{\text{id}}\|_2^2} \bar{\rho}_{\text{id}} = \bar{\rho} - (d^2 - 1) \text{Tr}[\bar{\rho}_{\text{id}}\bar{\rho}] \bar{\rho}_{\text{id}}, \\ \bar{E}_{\text{err}} &= \bar{E} - \frac{\text{Tr}[\bar{E}_{\text{id}}\bar{E}]}{\|\bar{E}_{\text{id}}\|_2^2} \bar{E}_{\text{id}} = \bar{E} - \frac{\text{Tr}[\bar{E}_{\text{id}}\bar{E}]}{d^2 - 1} \bar{E}_{\text{id}}, \end{aligned} \quad (20)$$

where the ideal operators  $\bar{\rho}_{\text{id}}$ ,  $\bar{E}_{\text{id}}$  are defined in Eq. (10), and a bar over the measurement operator indicates its traceless component  $\bar{E} = E - \frac{\text{Tr}[E]}{d^2} I$  [as defined in Eq. (14)]. Recall that  $\bar{\rho}$  was defined as the difference between two states Eq. (6). The error operators are defined in such a way that they are orthogonal to the ideal operators with respect to the Hilbert-Schmidt inner product:

$$\text{Tr}[\bar{\rho}_{\text{err}}\bar{\rho}_{\text{id}}] = \text{Tr}[\bar{E}_{\text{err}}\bar{E}_{\text{id}}] = 0. \quad (21)$$

The norms on the error operators are the trace norm and operator norm, respectively, defined for all  $A \in \mathbf{L}(\mathcal{H} \otimes \mathcal{H})$  as

$$\begin{aligned} \|A\|_1 &= \text{Tr}[\sqrt{A^\dagger A}] = \sum_i s_i(A), \\ \|A\|_{\infty} &= \sup_{0 \neq x \in \mathcal{H}^{\otimes 2}} \frac{\|Ax\|_2}{\|x\|_2} = \max_i \{s_i(A)\}, \end{aligned} \quad (22)$$

with  $s_i(A)$  the  $i$ th singular value of  $A$  and  $\|x\|_2$  the euclidean norm on  $\mathcal{H}^{\otimes 2}$ . Note that in the single-copy case the quantities  $\|\bar{\rho}_{\text{err}}\|_1^2$ ,  $\|\bar{E}_{\text{err}}\|_{\infty}^2$  as defined in Eq. (20) are to be estimated using  $\bar{\rho}_{\text{eff}}$  and  $\bar{E}_{\text{eff}}$  as defined in Eq. (13).

The variance bound of Eq. (19) has some appealing qualitative features. The first feature is that the first term is proportional to  $(1 - u)^2$ . This means that the first term goes to zero quadratically as the unitarity  $u$  of the error map  $\Lambda$  approaches 1. The fact that the second term is constant with respect to both  $u$  and  $m$  is unavoidable, as will be discussed in Sec. III B. The second appealing feature is the fact that the bound is asymptotically independent of the sequence length  $m$ . Thus the variance bound is useful in any regime of  $m$ . In Sec. III the dependence of the variance bound and the resulting number of sequences on various parameters is discussed in greater detail.

In the next section we present a bound  $L$  in the length of the interval in which the sequence purity  $q_{\mathbf{j}}$  lies. This is the final ingredient needed in order to apply Eq. (18).

### C. Bound on the interval of the sequence purity

In this section we present the improved bound  $L$  on the length of the interval in which the sequence purity  $q_{\mathbf{j}}^{(K)}$  lies. Even though the actual interval depends on  $K$ , the length

of these intervals is the same. Thus the bound  $L$  on the interval length of the sequence purity is independent of the implementation indexed by  $K$ . The improved bound is derived under the mild assumption that the experimental control is sufficiently good such that  $\text{Tr}[\bar{\rho}_{\text{id}}\bar{\rho}] \geq 0$  and  $\text{Tr}[\bar{E}_{\text{id}}\bar{E}] \geq 0$  (analogous assumption holds for the single-copy input and measurement operators). These conditions are satisfied only if the conditions

$$\text{Tr}[\rho_{\text{id}}\rho] \geq \text{Tr}[\hat{\rho}_{\text{id}}\rho], \quad \text{Tr}[\hat{\rho}_{\text{id}}\hat{\rho}] \geq \text{Tr}[\rho_{\text{id}}\hat{\rho}], \quad (23)$$

$$\text{Tr}[E\bar{\rho}_{\text{id}}] \geq 0 \quad (24)$$

are satisfied. Eq. (23) can be interpreted as requiring that the implemented states  $\rho$ ,  $\hat{\rho}$  have more overlap with their corresponding ideal state than with the noncorresponding ideal states. Eq. (24) is equivalent to  $\text{Tr}[\bar{E}\bar{E}_{\text{id}}] \geq 0$  since  $\bar{E}_{\text{id}} = (d^2 - 1)\bar{\rho}_{\text{id}}$  and  $\text{Tr}[\bar{\rho}_{\text{id}}\bar{E}] = \text{Tr}[E\bar{\rho}_{\text{id}}]$ . Eq. (24) has the interpretation that the measurement  $\{M, I - M\}$  associated with the observable  $E = 2M - I$  assigns the correct outcome (+1 for  $\rho_{\text{id}}$  and -1 for  $\hat{\rho}_{\text{id}}$ ) with at least probability  $\frac{1}{2}$ , or alternatively, that the measurement can correctly discriminate the maximally mixed state on the symmetric subspace ( $\rho_{\text{id}}$ ) from the maximally mixed state on the anti-symmetric subspace ( $\hat{\rho}_{\text{id}}$ ). These are very reasonable assumptions for any practical quantum information device.

In Lemma 12 of Appendix B 2 we show that under the stated assumption, the sequence purity lies in the interval

$$q_j^{(1)} \in [0, 1 + \|\bar{\rho}_{\text{err}}\|_1 + \|\bar{E}_{\text{err}}\|_\infty + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_\infty], \quad (25)$$

$$q_j^{(2)} \in [-\|\bar{\rho}_{\text{err}}\|_1 - \|\bar{E}_{\text{err}}\|_\infty - \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_\infty, 1]. \quad (26)$$

Therefore it follows that

$$L = 1 + \|\bar{\rho}_{\text{err}}\|_1 + \|\bar{E}_{\text{err}}\|_\infty + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_\infty \quad (27)$$

for both implementations. The idea of the proof of Lemma 12 is to decompose the input and measurement operators  $\bar{\rho}$  and  $\bar{E}$  into their ideal and error components according to Eq. (20). This gives rise to four terms. The ideal term  $\text{Tr}[E_{\text{id}}\mathcal{G}_j^{\otimes 2}(\bar{\rho}_{\text{id}})]$  can be bounded in the interval  $[0, 1]$ . The other terms are then bounded in magnitude using Hölder's inequality, which contributes the last three terms in Eq. (27).

#### D. Examples

Perhaps the best way to gain insight in the use of Eq. (18), Eq. (19), and Eq. (27) is by example. In Example 2 we calculate the required number of sequences for a fixed choice of all relevant parameters. In Example 3 we simulate a URB experiment using fixed number of sequences and compute the confidence interval around each estimate  $\bar{q}_m$ . We compare the results of these examples with a previously known bound (first used in Ref. [4]). This bound does not use the variance, but just uses the boundedness of the sequence purity  $q_j$ . It claims that  $\mathbb{P}[|\bar{q}_m - \mathbb{E}[q_j]| \geq \epsilon] \leq \delta$ , whenever [29]

$$2e^{-2N\frac{\epsilon^2}{L^2}} \leq \delta. \quad (28)$$

The number of sequences  $N$  is merely a function of the confidence parameters  $\epsilon$ ,  $\delta$  and the interval length  $L$ . In particular it does not depend on the variance of  $q_j$ .

*Example 2.* Suppose that a URB experiment is performed on the single-qubit Clifford group ( $d = 2$ ). The choice of implementation (single-copy or two-copy) is irrelevant for this example since both the variance bound Eq. (19) and the interval length bound Eq. (27) are independent of the choice of implementation. The only difference in practice is how to estimate the SPAM parameters  $\|\bar{\rho}_{\text{err}}\|_1^2$ ,  $\|E_{\text{err}}\|_\infty^2$ . Furthermore suppose that an *a priori* estimate of the unitarity is  $u = 0.98$  and an estimate for the SPAM parameters is  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$ . Then, after choosing appropriate sequence lengths to use in the experiment, an upper bound on the variance as a function of the sequence length can be computed using Eq. (19). The interval length can be bounded using Eq. (27). Using  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$ , this yields  $L = 1.02 + 0.2\sqrt{2} \approx 1.303$ . Finally, choosing an interval  $\epsilon$  and confidence  $\delta$ , Eq. (18) gives the required number of sequences  $N$  (at fixed length  $m$ ). Concretely, setting  $\epsilon = 0.02$ ,  $\delta = 0.01$  and all other parameters as discussed, the number of sequences required for sequences of length  $m = 10$ , is  $N = 242$ . For sequence length  $m = 30$ , the required number is  $N = 366$ , whereas  $m = 100$  requires  $N = 452$ . The long sequence length limit (when  $u^{2(m-1)} \ll 1$ ), yields  $N = 457$ .

Let us compare these numbers with the previously known bound Eq. (28) that does not use the variance of  $q_j$ . Given our choices of  $\epsilon = 0.02$ ,  $\delta = 0.01$ , and  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$  [from which  $L = 1.02 + 0.2\sqrt{2} \approx 1.303$  is computed using Eq. (27)], the bound Eq. (28) yields  $N = 11\,242$  required sequences. We emphasize that this number is independent of  $u$  or  $m$ . In this scenario, our bound gives approximately two orders of magnitude improvement.

*Example 3.* In Fig. 2 we compare the 99% confidence intervals  $\epsilon$  (for fixed  $N = 250$  and  $\delta = 0.01$ ) around the empirical average sequence purity  $\bar{q}_m$  calculated with and without our variance bound at several different sequence lengths. The empirical average sequence purity  $\bar{q}_m$  data is based on a simulated single-qubit Clifford URB experiment. The length of the confidence interval  $\epsilon$  without variance (larger blue bars) is computed from Eq. (28). Then the choice of  $N = 250$  and  $\delta = 0.01$  yields  $\epsilon = 0.134$ . On the other hand, the length of the confidence interval  $\epsilon$  with variance (smaller red bars in the plot) is computed from Eq. (18) by solving the equation for  $\epsilon$ , using our sharp variance bound Eq. (19). In the evaluation of Eq. (19), the *a priori* estimates  $u = 0.98$  and  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$  were used. Then Eq. (27) yields  $L = 1.02 + 0.2\sqrt{2}$ . Using our sharp variance bound, the values of the confidence interval vary between  $\epsilon = 0.019$  (for  $m = 8$ ) and  $\epsilon = 0.029$  (for  $m = 174$ ). This is approximately an order of magnitude larger than the confidence interval without variance  $\epsilon = 0.134$ .

In this simulated experiment the Clifford gates are implemented with a fixed error channel  $\Lambda$  that is generated by taking a convex combination of the identity channel (with high weight) and a random CPTP map (sampled using QETLAB [30]). Similarly, the noisy input states and measurement operator are simulated by taking a convex combination of the ideal operators and randomly generated operators (generated

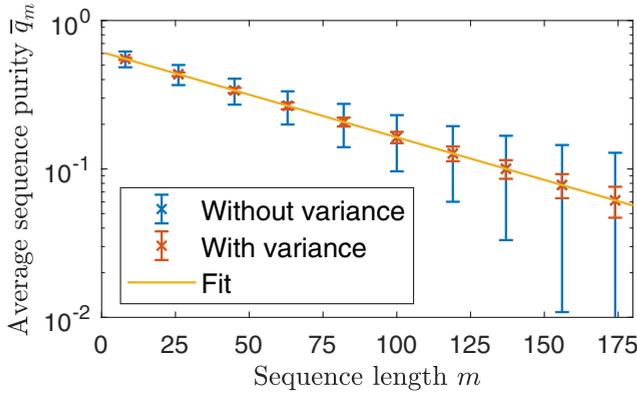


FIG. 2. Comparison of the 99% confidence intervals around the average sequence purity  $\bar{q}_m$  calculated with and without our variance bound at several different sequence lengths. The plot is based on a simulated URB experiment of the single-qubit Clifford group with  $N = 250$  samples per sequence length  $m$ . The empirical average sequence purity  $\bar{q}_m$  (marked with a cross) is plotted versus the sequence length  $m$  on a semilogarithmic scale. The larger (blue) bars indicate the 99% confidence interval without variance Eq. (28) and the smaller (red) bars indicate the 99% confidence interval of Eq. (18) based on our sharp variance bound Eq. (19). Here we used *a priori* estimates of the unitarity and SPAM parameters of  $u = 0.98$  and  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0.02$ , respectively. Then Eq. (27) yields  $L = 1.02 + 0.2\sqrt{2}$ . For completeness, a least-squares fit according to the model  $\bar{q}_m = Bu^{m-1}$  [see Eq. (7)] is shown in the yellow solid line. This yields  $u \approx 0.987$ .

using QETLAB). For this particular realization of an error map  $\Lambda$ , the data points seem to be even more accurate than our confidence interval might suggest based on their proximity to the fit. This is due to the fact that this particular error channel is well-behaved. We emphasize that our bound is valid for any unital or single-qubit error map. In particular this means that our bound is valid for the worst case realizations of  $\Lambda$ . It is unclear what error map  $\Lambda$  maximizes the variance of the sequence purity.

We emphasize that the point of this simulated example is not to prescribe a direct method for extracting the confidence in the unitarity, as this generally depends on the fitting model and the way the uncertainty in the average sequence purity are propagated into the uncertainty of the unitarity. Moreover, more advanced statistical tools may be used to extract the unitarity from the obtained (in this case simulated) data, like e.g., Refs. [23,25]. The goal of this example is to illustrate the significant gain in confidence of the average sequence purity when the simple concentration inequalities of Hoeffding are applied [29]. The point is that the additional knowledge of a variance bound on the underlying distribution of the sequence purity  $q_j$  can be used by statistical tools to extract the unitarity with improved confidence.

In the next section we explore the behavior of our bound in various parameter regimes.

### III. DISCUSSION

This section is devoted to discussing the variance bound and the interval length of the sequence purity in more detail.

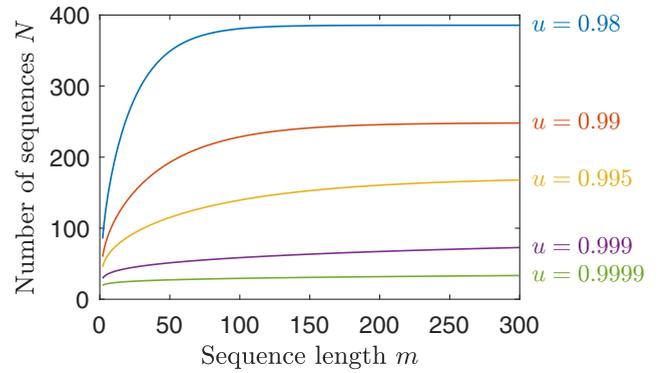


FIG. 3. Number of sequences  $N$  versus the sequence length  $m$  for various values of the unitarity  $u$  when benchmarking the single-qubit Clifford group ( $d = 2$ ). Confidence parameters are  $\epsilon = 0.02$  and  $\delta = 0.01$ . The SPAM parameters are  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0$ . By Eq. (27) then  $L = 1$  is used. The number of sequences is asymptotically independent of the sequence length. This is consistent with our variance bound Eq. (19).

In particular we discuss the variance bound in several different parameter regimes in more detail and aim to provide a better understanding of the parameters that ultimately determine the statistical confidence of the measurements. In Sec. III A we discuss the dependence of the variance bound Eq. (19) on the unitarity  $u$  and the sequence length  $m$ . In Sec. III B we discuss the dependence on the SPAM parameters  $\|\bar{\rho}_{\text{err}}\|_1^2$  and  $\|\bar{E}_\infty\|_1^2$ . Here we also show by example that the variance of the sequence purity does not go to zero in the presence of SPAM errors. In Sec. III C the dependence of the variance bound on the system size is discussed.

#### A. Dependence on unitarity and sequence length

First, we discuss the dependence of the number of required sequences  $N$  on the sequence length  $m$ . In Fig. 3 this dependence is plotted for various values of  $u$  in the absence of SPAM errors (i.e.,  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|E_{\text{err}}\|_\infty^2 = 0$ ). The confidence parameters were fixed at  $\delta = 0.01$  and  $\epsilon = 0.02$ . It can be seen from the figure that  $N$  approaches a constant as  $m$  increases. This is consistent with our variance bound Eq. (19), where the factor depending on  $m$  is

$$\frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2. \quad (29)$$

This approaches a constant in the limit of large sequence lengths. This limit is approximately achieved when  $u^{2(m-1)} \ll 1$ . The exact limit is given by

$$\lim_{m \rightarrow \infty} \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 = \frac{1 - u}{1 + u}. \quad (30)$$

In the presence of SPAM errors, the asymptotic constant is larger than in its absence, but the behavior is similar. Since the variance approaches a constant, so does the required number of sequences for fixed values of the confidence parameters. From here on out, the “large sequence limit” means the regime of  $m$  where  $u^{2(m-1)} \ll 1$  so that the variance bound (and thus the number of sequences) is approximately independent of  $m$ .

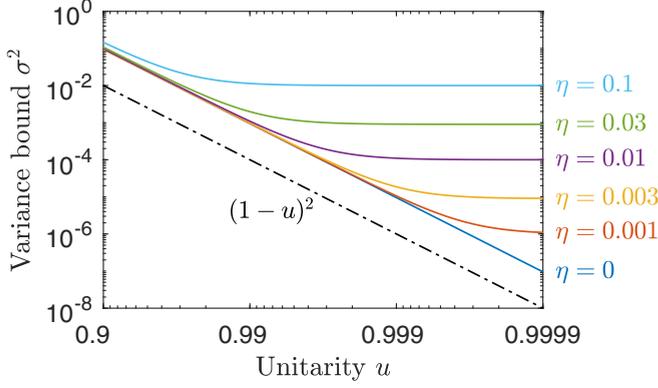


FIG. 4. Semilogarithmic plot of the variance bound  $\sigma^2$  as a function of the unitarity  $u$  for various magnitudes of SPAM errors in the large sequence limit for single-qubit Clifford URB ( $d = 2$ ). The black dash-dotted line is a reference line plotting  $\sigma^2 = (1 - u)^2$ . The differently colored solid lines indicate the various magnitudes of SPAM errors, where  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|\bar{E}_{\text{err}}\|_\infty^2 = \eta$ . There are two regimes. For small SPAM errors and small  $u$ , the variance scales as  $(1 - u)^2$ , whereas for nonzero SPAM errors and large  $u$ , the variance approaches a constant.

Second we discuss the dependence of the variance bound on the unitarity  $u$ . In Fig. 4 the variance bound  $\sigma^2$  is plotted as a function of the unitarity  $u$  for various values of SPAM errors in the long sequence length limit. This figure shows two regimes. In the regime of low unitarity and small SPAM error, the variance is proportional to  $(1 - u)^2$ . This is consistent with Eq. (19), where the variance is dominated by the first term in this regime. However, for nonzero SPAM error and large unitarity, this behavior transitions into a constant variance. In this regime, the variance is dominated by the second, constant term (independent of  $u$ ) in Eq. (19).

The number of required sequences  $N$  shows qualitatively similar behavior, but there are differences. This is due to the fact that  $N$  is a nonlinear function of  $\sigma^2$ . In the regime of constant variance, the number of sequences is also constant. In the regime where the variance bound is proportional to  $(1 - u)^2$ , the number of sequences also decreases as  $N$  increases, but the rate depends also on the choice of  $\epsilon$ .

### B. Dependence on SPAM parameters

In Fig. 5 we show a color plot of the number of sequences  $N$  as a function of the SPAM parameters  $\|\bar{\rho}_{\text{err}}\|_1^2$  and  $\|\bar{E}_{\text{err}}\|_\infty^2$  for fixed unitarity  $u$  in the limit of large sequences. The plot illustrates the qualitative dependence of  $N$  on the magnitude of these SPAM parameters. There are two ways that the SPAM parameters contribute to the number of required sequences  $N$ . First, the variance bound  $\sigma^2$  depends on the SPAM parameters  $\|\bar{\rho}_{\text{err}}\|_1^2$  and  $\|\bar{E}_{\text{err}}\|_\infty^2$  [see Eq. (19)]. Second, the interval length bound  $L$  depends on the square root of these parameters,  $\|\bar{\rho}_{\text{err}}\|_1$  and  $\|\bar{E}_{\text{err}}\|_\infty$  [see Eq. (27)]. Both these bounds increase as the SPAM parameters increase. From the concentration inequality Eq. (18), it follows that the required number of sequences  $N$  for fixed confidence parameters grows with increasing variance and interval length. Both these effects have qualitatively similar behavior. This translate into

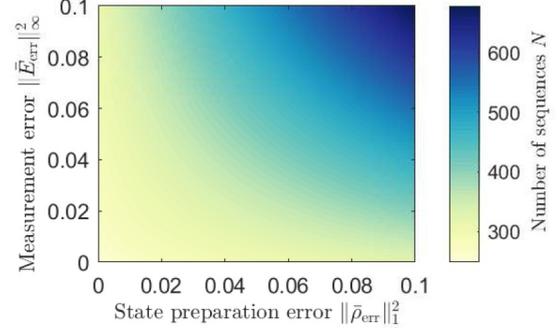


FIG. 5. Color plot of the number of sequences  $N$  as a function of the SPAM parameters  $\|\bar{\rho}_{\text{err}}\|_1^2$  and  $\|\bar{E}_{\text{err}}\|_\infty^2$  in the large sequence length limit for single-qubit Clifford URB ( $d = 2$ ). The parameters  $u = 0.99$  and  $\epsilon = 0.02$ ,  $\delta = 0.01$  were used. This plot illustrates the sensitivity of our result to SPAM errors. In particular, the number of sequences increases most significantly when both state preparation and measurement errors are large.

the illustrated dependence of the number of sequences  $N$  on the SPAM parameters in Fig. 5. In particular, the number of sequences most strongly depends on the product between the two, showing a larger required number in the area where the product  $\|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_\infty^2$  is largest.

The variance bound of Eq. (19) has a constant term  $\|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_\infty^2$ , independent of the unitarity  $u$  and sequence length  $m$ . In particular this means that the variance bound is nonzero in the presence of SPAM error for all sequence lengths  $m$  even in the limit of ideal gates  $\Lambda \rightarrow \mathcal{I}$ . This behavior is also seen in Fig. 4. We argue that this is fundamental to the URB protocol, by showing that the actual variance of the sequence purity  $\mathbb{V}[q_j]$  also has this behavior even when ideal gates are considered. This is done in Example 4. In this example we construct noisy operators  $\bar{\rho}$  and  $\bar{E}$  such that the average sequence purity  $q_j$  is not constant over all possible ideal gate sequences  $\mathcal{G}_j$  (i.e., sequences with  $\Lambda = \mathcal{I}$ ). Thus there exists an error channel (namely  $\Lambda = \mathcal{I}$ ) and noisy operators (namely those constructed in Example 4) such that the variance, and thus the required number of sequences, is nonzero. This behavior is in contrast with standard RB, where all RB gate sequences compose to the identity when  $\Lambda = \mathcal{I}$  (in the RB protocol, a global inverse gate is applied after each sequence). Therefore in standard RB, the survival probability does not depend on the sequence in the absence of gate errors and hence the variance is zero.

*Example 4.* Consider a URB experiment where the gate set under investigation is the single-qubit Clifford group  $\mathbf{C}(2)$ . Suppose that the gates are implemented perfectly, i.e.,  $\Lambda = \mathcal{I}$ . Furthermore assume that the state and measurement operators are given by

$$\rho, \hat{\rho} = \frac{I \otimes I \pm X \otimes X}{4} \quad \text{and} \quad E = X \otimes X, \quad (31)$$

where  $I$  is the identity and  $X$  is the Pauli- $X$  matrix on the single-qubit Hilbert space  $\mathcal{H} \simeq \mathbb{C}^2$ . Since  $\Lambda = \mathcal{I}$ , the sequence  $\mathcal{G}_j$  of  $m$  independently and uniformly distributed Clifford gates reduces to a single Clifford gate  $\mathcal{G}_i$  uniformly drawn from  $\mathbf{C}(2)$ . The group  $\mathbf{C}(2)$  has 24 elements, eight of which map  $X \mapsto \pm X$ . Whether such a map sends  $X$  to  $+X$

or  $-X$  is irrelevant, since if  $\mathcal{G}$  maps  $X \mapsto \pm X$  then  $\mathcal{G}^{\otimes 2}$  maps  $X^{\otimes 2} \mapsto X^{\otimes 2}$  in either case. The other 16 Clifford gates send  $X \mapsto \pm Y$  or  $X \mapsto \pm Z$ , where again the sign is irrelevant. Thus, given that  $\bar{\rho} = \frac{X \otimes X}{4}$ , a fraction  $\frac{8}{24}$  of all sequences  $\mathcal{G}_j$  will satisfy  $\mathcal{G}_j^{\otimes 2}(\bar{\rho}) = \frac{X \otimes X}{4}$  while the others will send  $\bar{\rho}$  either to  $\frac{Y \otimes Y}{4}$  or  $\frac{Z \otimes Z}{4}$ . Since  $\text{Tr}[E(\frac{X \otimes X}{4})] = 1$  and  $\text{Tr}[E(\frac{Y \otimes Y}{4})] = \text{Tr}[E(\frac{Z \otimes Z}{4})] = 0$ , the following probability distribution on  $q_j^{(2)}$  is obtained:

$$\mathbb{P}[q_j^{(2)} = 1] = \frac{1}{3} \quad \text{and} \quad \mathbb{P}[q_j^{(2)} = 0] = \frac{2}{3}. \quad (32)$$

Clearly then  $\mathbb{E}[q_j^{(2)}] = \frac{1}{3}$  and  $\mathbb{V}[q_j^{(2)}] = \frac{2}{9} > 0$ . This example shows that the variance  $\mathbb{V}[q_j]$  of the sequence purity can not go to zero as the unitarity  $u \rightarrow 1$ . ■

Given noisy implementations  $\bar{\rho}$  and  $E$  in the two-copy implementation, the SPAM parameters  $\|\rho_{\text{err}}\|_1^2$  and  $\|\bar{E}_{\text{err}}\|_{\infty}^2$  defined in Eq. (20) can in principle be estimated by relating them to the ideal states and measurements of Eq. (10). In practice, this requires (partial) knowledge of the noisy operators  $\bar{\rho}$  and  $E$ . If a full (tomographic) description of  $\rho$ ,  $\hat{\rho}$ ,  $E$  is available, then  $\|\bar{\rho}_{\text{err}}\|_1^2$  and  $\|\bar{E}_{\text{err}}\|_{\infty}^2$  can be calculated from the definition Eq. (20). However, if only partial knowledge is available (e.g., a lower bound on state preparation fidelity), then the SPAM quantities need to be bounded. For example  $\|\bar{\rho}_{\text{err}}\|_1^2$  can be upper bounded if the fidelity between  $\rho$  ( $\hat{\rho}$ ) and  $\rho_{\text{id}}$  ( $\hat{\rho}_{\text{id}}$ ) is known, by application of the Fuchs–Van de Graaff inequality [31]. In the single-copy implementation, slightly more work is needed. The SPAM parameters are then defined with respect to  $\bar{\rho}_{\text{eff}}$  and  $\bar{E}_{\text{eff}}$  Eq. (13). However, only (partial) knowledge of the physical operators  $\rho_{\mathcal{H}}$  and  $E_{\mathcal{H}}$  are available. Noise on these physical operators needs to be translated to noise on the effective operators  $\bar{\rho}_{\text{eff}}$  and  $\bar{E}_{\text{eff}}$ .

### C. Dimension-dependent constants

In this section, the dependence of the variance bound Eq. (19) and consequently the number of sequences on the system size is examined. An undesirable feature of the variance bound is the asymptotic growth of the constants  $c_2(d)$  and  $c_3(d)$  with the dimension  $d = 2^q$  of the  $q$ -qubit system. This means that for large systems, the bound becomes loose and ultimately vacuous. This is illustrated in Fig. 6, where the number of sequences  $N$  is plotted as a function of the system size  $q$  on a semilogarithmic scale (for fixed unitarity  $u$  and large sequence length  $m$ ). The number of sequences is plotted in the absence of SPAM error, with state preparation or measurement error only and with both errors simultaneously. This is done to distinguish the different contributions of the constants  $c_1$ ,  $c_2$ , and  $c_3$  in Eq. (19). In the absence of SPAM error, only  $c_1$  is relevant. This constant takes its maximum at  $q = 2$  and asymptotically goes to 1. However with measurement error, the number of sequences needed grows exponentially with the system size. With state preparation error, this expectational growth is even faster. This is consistent with the asymptotic limits of the constants  $c_2 = O(d)$  and  $c_3 = O(d^2)$ , since  $d = 2^q$ . In particular, this figure shows that our variance bound is prohibitively loose for  $q \geq 6$  (assuming  $u = 0.99$  and large  $m$ ), since the first order bound Eq. (28) yields a smaller number of sequences  $N$  as indicated by the black dash-dotted line in the figure.

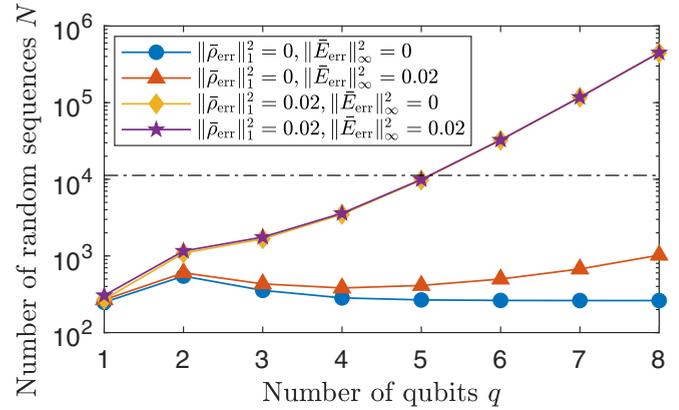


FIG. 6. Number of sequences  $N$  as a function of the number of qubits  $q$  comprising the system for different values of the SPAM parameters. A fixed unitarity  $u = 0.99$  and the large sequence length limit are used. The interval bound  $L$  is computed using Eq. (27) as a function of the SPAM quantities (see legend). The confidence parameters  $\epsilon = 0.02$ ,  $\delta = 0.01$  were used. The dashed line indicates the first-order bound Eq. (28) corresponding to  $\|\bar{\rho}_{\text{err}}\|_1^2 = \|\bar{E}_{\text{err}}\|_{\infty}^2 = 0.02$ . For the given confidence and SPAM parameters, our bound gives an improvement of the required number of sequences up to five-qubit systems.

We believe that the unbounded growth of our variance bound with the system size is an artifact of the proof rather than a fundamental property. The sequence purity  $q_j$  is a bounded, discrete random variable, where the bound  $L$  does not depend on the dimension  $d$ . Therefore the exact variance  $\mathbb{V}[q_j]$  can not asymptotically grow with the system dimension  $d$ . The bound of Eq. (19) is, however, sharp enough for practical use in few-qubit systems.

## IV. METHODS

This section gives an high-level overview of the methods used for deriving our main result Eq. (18) and Eq. (19). In Sec. IV A we focus on the statistical aspect of our result related to Eq. (18). We also relate the between-sequence variance  $\mathbb{V}[q_j]$  (the quantity which we bounded in this work) to the within-sequence variance that arises due to the fact that  $q_j$  can be estimated only by collecting a finite sample of single-shot measurements for a given sequence. In Sec. IV B we discuss the derivation of the fit model (as derived in Ref. [12]) and derive an expression for the variance  $\mathbb{V}[q_j]$ . In Sec. IV C we give an outline of the proof of our variance bound Eq. (19).

### A. Estimation theory

Ultimately, the URB protocol leads to the complex statistical estimation problem of determining  $u$  and the confidence thereof, given a large set of realizations of the sequence purity  $q_j$  (for multiple sequence lengths  $m$ ). There are several ways one can go about this problem (see e.g., Ref. [25] for a Bayesian inference approach). In this paper we take a frequentist approach and determine a confidence interval for the point estimates  $\bar{q}_m$  of  $\mathbb{E}[q_j]$ . These confidence intervals (for different values of  $m$ ) can then be taken into account when fitting the point estimates  $\bar{q}_m = Bu^{m-1}$  to the fit model. The

main contribution of this work is improving the confidence interval of  $\bar{q}_m$  by bounding the variance of the sequence purity  $q_j$ . This variance bound provides strictly more information on the distribution of  $q_j$  than what was known before [12] and could therefore also be of value when using other estimation techniques to extract the unitarity  $u$  from the set of measurement outcomes.

The intuitive idea is that estimating the mean of a bounded distribution of random variables requires fewer samples when the distribution is narrowly peaked around the mean. Since the variance is a measure of the spread of the distribution, it is intuitive that having knowledge of the variance improves the confidence in the estimate of the mean. This idea is made precise in statistics by concentration inequalities. Here we use a concentration inequality due to Hoeffding [29]. Given a collection of  $N$  independent and identically distributed (i.i.d.) random variables  $X_i$ , sampled from a distribution on a length  $L$  interval with mean  $\mu$  and variance  $\sigma^2$ , the following statement holds for all  $0 \leq \epsilon \leq L$ :

$$\mathbb{P}[|\bar{X} - \mu| \geq \epsilon] \leq 2 \left[ \left( \frac{L}{L - \epsilon} \right)^{\frac{L^2 - \epsilon L}{\sigma^2 + L^2}} \left( \frac{\sigma^2}{\sigma^2 + \epsilon L} \right)^{\frac{\sigma^2 + \epsilon L}{\sigma^2 + L^2}} \right]^N, \quad (33)$$

where  $\bar{X} = \frac{1}{N} \sum_i X_i$  is the empirical mean. This is essentially Eq. (18) using the fact that  $q_j$  are i.i.d. random variables. The point is that if one wishes to bound this probability by  $\delta$ , then upper bounding the right-hand side by  $\delta$  gives a means to relate  $N$ ,  $\delta$ , and  $\epsilon$ . Instead of the exact (unknown) variance of the distribution of  $q_j$ , an upper bound is used.

The fact that our variance bound Eq. (19) depends on the unitarity  $u$ , the quantity that one ultimately attempts to estimate, may seem strange and circular. But this is actually a feature of statistics, which is more apparent in the Bayesian view. One may have an *a priori* distribution of the unitarity  $u$  of the gate set and given some experimental data (the complete URB data set) one can construct a more concentrated *a posteriori* distribution on the unitarity. In the frequentist view, an *a priori* lower bound to the unitarity can be known with very high confidence. Then performing URB will improve the estimate of the unitarity and increase the confidence in this estimate. In principle this procedure can be done by doing several successive URB experiments, further increasing the confidence in the outcome. Note that a first lower bound can always be obtained from the average gate fidelity [by application of Eq. (3)], which is estimated using standard RB.

Finally there is one subtlety that deserves some attention. The protocol requires the experimenter to measure  $\text{Tr}[E\mathcal{G}_j^{\otimes 2}(\rho)]$ , but actually this is an expectation value of the measurement operator  $E$  (a Hermitian observable) given the state  $\mathcal{G}_j^{\otimes 2}(\rho)$ . This expectation value must be learned from multiple single-shot measurements of preparing the state, apply gates, and measure. The outcome is inherently probabilistic (with a Bernoulli distribution) by the laws of quantum mechanics and either a click or no click is observed with the probability given by Born's rule. To estimate the expectation value  $\text{Tr}[E\mathcal{G}_j^{\otimes 2}(\rho)]$ , a large number of single-shot measurements must be taken and the proportion of clicks is an estimate  $\text{Tr}[E\mathcal{G}_j^{\otimes 2}(\rho)]$ . In reality then, there is also

some uncertainty in each data point  $q_j$ , which propagates into increased uncertainty in the average  $\bar{q}_m$ . So far we have assumed the uncertainty in  $\bar{q}_m$  is dominated by the uncertainty due to the randomly sampled sequences and not due to the uncertainty in determining each sequence purity  $q_j$ . This assumption is motivated by experiments in which it is hard to store many sequences, but easy to repeat single-shot measurements of the same sequence. In these experiments it is then easy to do enough single-shot measurements of each  $q_j$ , such that the uncertainty in  $\bar{q}_m$  is dominated by the uncertainty due to the randomly sampled sequences. This assumption is not fundamental, however, but is related to classical hardware control of the experimenter. In the next section we will discuss the validity of this assumption, estimate the number of required single-shot measurements and show how this assumption can be dropped if one wishes to explicitly take into account finite sampling uncertainty.

### 1. Finite sampling statistics

In the previous section it was discussed that the quantity  $q_j$  is actually not directly accessible, but must be estimated by performing a large number of single-shot measurements. Born's rule states that given a (two-valued) POVM measurement  $\{M, I - M\}$  and a state  $\rho$ , the probability of getting outcome 1 (associated with  $M$ ) is given by  $\text{Tr}[M\rho]$  and outcome 0 (associated with  $I - M$ ) is  $1 - \text{Tr}[M\rho]$ . This can be used to construct a probability distribution for a single-shot measurement of  $q_j^{(K)}$ , given a fixed sequence indexed by  $\mathbf{j}$ . The distribution is determined by the definition of  $q_j^{(K)}$  and depends on the choice of implementation. Recall that  $q_j$  is calculated using the difference of two states  $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$ .

Let us denote  $\bar{q}_j$  an unbiased estimator for the exact  $q_j$  given a fixed sequence indexed by  $\mathbf{j}$ . Then there is uncertainty in  $\bar{q}_j$  due to the uniformly distributed random sequences  $\mathbf{j}$  and due to the fact that  $\bar{q}_j$  is itself a random variable for fixed  $\mathbf{j}$  (since it is an estimator for the exact  $q_j$ ). The contribution of each source of uncertainty can be quantified by the law of total variance [32], which states that

$$\mathbb{V}[\bar{q}_j] = \mathbb{E}[\mathbb{V}[\bar{q}_j|\mathbf{j}]] + \mathbb{V}[\mathbb{E}[\bar{q}_j|\mathbf{j}]] = \mathbb{E}[\mathbb{V}[\bar{q}_j|\mathbf{j}]] + \mathbb{V}[q_j]. \quad (34)$$

Here the quantity  $\mathbb{V}[\bar{q}_j|\mathbf{j}]$  is referred to as the within-sequence variance (for the given sequence  $\mathbf{j}$ ). It is the variance of the sequence purity  $\bar{q}_j$  given fixed  $\mathbf{j}$  solely due to the finite sampling statistics. The quantity  $\mathbb{V}[q_j]$  is the between-sequence variance of  $q_j$  and is solely due to the fact that the sequences  $\mathbf{j}$  are sampled from a uniform distribution. This equation expresses that the total variance is the sum of the expected within-sequence variance (expected over the uniformly distributed random sequences) and the between-sequence variance. The quantity  $\mathbb{V}[q_j]$  was bounded in this work Eq. (19).

To examine the term  $\mathbb{E}[\mathbb{V}[\bar{q}_j|\mathbf{j}]]$  in Eq. (34), an expression or bound on the within-sequence variance  $\mathbb{V}[\bar{q}_j|\mathbf{j}]$  as a function of the number of single-shot repetitions is required. We will show how this is done for the two-copy implementation, leaving the more cumbersome (but in principle not more difficult) single-copy implementation as an open problem. Define the single-shot random variable by  $x_r$ , where the subscript  $r$  indexes the different single-shot realizations (for

$r = 1, \dots, R$  for some large  $R$ ), by the following distribution:

$$\mathbb{P}[x_r = y | \mathbf{j}] = \begin{cases} a(1-b), & \text{if } y = 1, \\ ab + (1-a)(1-b), & \text{if } y = 0, \\ (1-a)b, & \text{if } y = -1. \end{cases} \quad (35)$$

Here  $a = \text{Tr}[M\mathcal{G}_j^{\otimes 2}(\rho)]$ ,  $b = \text{Tr}[M\mathcal{G}_j^{\otimes 2}(\hat{\rho})]$  and  $M = \frac{1}{2}(I + E)$  is the POVM element associated with the two-valued measurement  $E$ . The outcome  $x_r = 1$  is interpreted as measuring a click only for  $\rho$ , outcome  $x_r = 0$  corresponds to a click for both or neither states and outcome  $x_r = -1$  is associated with a click only for  $\hat{\rho}$ . This is indeed the single-shot outcome measurement outcome of a  $q_j^{(2)}$  measurement, since

$$q_j^{(2)} = \mathbb{E}[x_r | \mathbf{j}] = a - b = \text{Tr}[E\mathcal{G}_j^{\otimes 2}(\bar{\rho})]. \quad (36)$$

The natural unbiased estimator of  $q_j^{(2)}$  is then given by

$$\bar{q}_j^{(2)} = \frac{1}{R} \sum_{r=1}^R x_r. \quad (37)$$

The within-sequence variance  $\mathbb{V}[\bar{q}_j^{(2)} | \mathbf{j}]$  is related to the variance of  $x_r$  [which can be computed given the probability distribution Eq. (35)] using the fact that  $x_r$  are i.i.d. and mutually uncorrelated random variables:

$$\mathbb{V}[\bar{q}_j^{(2)} | \mathbf{j}] = \mathbb{V}\left[\frac{1}{R} \sum_{r=1}^R x_r | \mathbf{j}\right] = \frac{1}{R^2} \sum_{r=1}^R \mathbb{V}[x_r | \mathbf{j}] = \frac{1}{R} \mathbb{V}[x_r | \mathbf{j}].$$

This follows the definition of the variance and linearity of the expected value. The variance of  $x_r$  [computed from the distribution Eq. (35)] is then

$$\mathbb{V}[x_r | \mathbf{j}] = [a(1-a) + b(1-b)] \leq \frac{1}{2}, \quad (38)$$

where the upper bound is trivially obtained by maximizing over  $0 \leq a, b \leq 1$ . The within-sequence variance thus satisfies

$$\mathbb{V}[\bar{q}_j^{(2)} | \mathbf{j}] = \frac{1}{R} [a(1-a) + b(1-b)] \leq \frac{1}{2R}. \quad (39)$$

Hence for the two-copy implementation, the total variance is bounded by

$$\mathbb{V}[\bar{q}_j^{(2)}] \leq \sigma^2 + \frac{1}{2R}, \quad (40)$$

where  $R$  is the number of single-shot measurements taken per sequence and  $\sigma^2$  is the variance bound of Eq. (19).

It may seem that the modification of the protocol to use the difference of two states  $\bar{\rho}$  means that twice as many single-shot measurements must be taken. This is, however, not the case [27]. To see this, let  $\mathbb{V}_\rho$  be the variance associated with a single measurement setting on the state  $\rho$ . Then for the difference of two states, the variance associated with that measurement satisfies

$$\mathbb{V}_{\bar{\rho}} = \mathbb{V}_{\frac{1}{2}(\rho - \hat{\rho})} \leq \frac{1}{4}(\mathbb{V}_\rho + \mathbb{V}_{\hat{\rho}}) \leq \frac{1}{2} \max(\mathbb{V}_\rho, \mathbb{V}_{\hat{\rho}}). \quad (41)$$

So to the contrary, fewer sequences are required to get an accurate estimate of  $\text{Tr}[E\mathcal{G}_j(\bar{\rho})]$  than of  $\text{Tr}[E\mathcal{G}_j(\rho)]$ . This can explicitly be seen in the two-copy implementation, where the within-sequence variance  $\mathbb{V}_{\bar{\rho}}[\bar{q}_j | \mathbf{j}]$  was computed in Eq. (39). However, if only a single state  $\rho$  were used, then  $\mathbb{P}[x_r = 1] =$

$a$  and  $\mathbb{P}[x_r = -1] = 1 - a$ . Therefore the variance  $\mathbb{V}_\rho[\bar{q}_j | \mathbf{j}] = \frac{1}{R} \mathbb{V}_\rho[x_r | \mathbf{j}] = \frac{4a(1-a)}{R} \leq \frac{1}{R}$ , which is indeed a factor 2 larger than in Eq. (39).

## 2. The unbiased estimator of the sequence purity in the single-copy implementation

In the single-copy implementation care must be taken in defining an appropriate estimator of  $q_j^{(1)}$ . Analogously to the above, one can define a random variable  $x_r^{PQ}$  associated with a single-shot measurement of  $\text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})]$  for a fixed sequence indexed by  $\mathbf{j}$ , depending on the Pauli's  $P$  and  $Q$ . Then

$$\mathbb{E}[x_r^{PQ} | \mathbf{j}] = \text{Tr}[E_{\mathcal{H}}^{(Q)}\mathcal{G}_j(\bar{\rho}_{\mathcal{H}}^{(P)})], \quad (42)$$

so that

$$q_j^{(1)} = \frac{1}{d^2 - 1} \sum_{P, Q \neq I} \mathbb{E}[x_r^{PQ} | \mathbf{j}]^2. \quad (43)$$

If we denote  $\bar{x}_{PQ} = \frac{1}{R} \sum_{r=1}^R x_r^{PQ}$ , then one could try to estimate  $q_j^{(1)}$  by  $\bar{q}_j^{(1)} = \frac{1}{d^2 - 1} \sum_{P, Q \neq I} \bar{x}_{PQ}^2$ . This estimate is biased, however, and overestimates the actual value of  $q_j^{(1)}$ , since

$$\begin{aligned} \mathbb{E}[\bar{x}_{PQ}^2 | \mathbf{j}] &= \mathbb{E}[\bar{x}_{PQ} | \mathbf{j}]^2 + \mathbb{V}[\bar{x}_{PQ} | \mathbf{j}] \\ &= \mathbb{E}[\bar{x}_{PQ} | \mathbf{j}]^2 + \frac{1}{R} \mathbb{V}[x_r^{PQ} | \mathbf{j}]. \end{aligned} \quad (44)$$

To remedy this, one can make use of the unbiased estimator

$$\bar{q}_j^{(1)} = \frac{1}{d^2 - 1} \sum_{P, Q \neq I} \bar{x}_{PQ}^2 - \frac{1}{R} s_{PQ}^2, \quad (45)$$

where

$$s_{PQ}^2 = \frac{1}{R - 1} \sum_{r=1}^R (x_r^{PQ} - \bar{x}_{PQ})^2 \quad (46)$$

is the unbiased estimate of  $\mathbb{V}[x_r^{PQ} | \mathbf{j}]$ . It is important to take this into consideration when performing a Clifford URB experiment using the single-copy implementation, since overestimating  $q_j^{(1)}$  can lead to an overestimate of the unitarity obtained from the experiment.

## B. Fit model and variance expression

In this section we first briefly review the derivation of the fit model of URB (as derived in Ref. [12]), slightly adapted with our modification of a traceless input operator  $\bar{\rho}$ . Then we derive an expression for the variance of the sequence purity. We do so using slightly different notation, picking an orthonormal basis for the space of linear operators  $\mathcal{L}(\mathcal{H})$  (in particular we use the normalized Pauli operators). We can then vectorize any operator with respect to that basis, which we will denote with a bracket-like notation  $\rho \rightarrow |\rho\rangle\rangle$  and  $E \rightarrow \langle\langle E |$ . Quantum channels can then be viewed as matrices on these vectors, i.e.,  $\mathcal{E}(\rho) \rightarrow |\mathcal{E}(\rho)\rangle\rangle = \mathcal{E}|\rho\rangle\rangle$ , where we use boldface notation for the matrix representation of a quantum channel. The Hilbert-Schmidt inner product  $\text{Tr}[E^\dagger \rho]$ , carries over as the vector inner product with respect to any basis, so that  $\text{Tr}[E^\dagger \rho] = \langle\langle E | \rho \rangle\rangle$ . Finally composition of channels

$\mathcal{E}_1\mathcal{E}_2 \rightarrow \mathbf{E}_1\mathbf{E}_2$  carries over as matrix multiplication. This notation is known as the natural representation, Liouville representation, or Pauli transfer matrix representation [26,33]. See Appendix A 1 b for more details.

Using this notation, the expected value of the sequence purity  $\mathbb{E}[q_j]$  can be written as

$$\mathbb{E}[q_j] = \frac{1}{|\mathbf{C}(d)|^m} \sum_j \langle\langle E | \mathcal{G}_j^{\otimes 2} | \bar{\rho} \rangle\rangle = \langle\langle E | (\mathcal{G}_{\text{avg}}^{(2)} \Lambda^{\otimes 2})^m | \bar{\rho} \rangle\rangle, \quad (47)$$

where

$$\mathcal{G}_{\text{avg}}^{(n)} = \frac{1}{|\mathbf{C}(d)|} \sum_{\mathcal{G} \in \mathbf{C}(d)} \mathcal{G}^{\otimes n}. \quad (48)$$

The key idea behind deriving the fitting model is that  $\mathcal{G}_{\text{avg}}^{(2)}$  is the orthogonal projection onto the vector space  $W = \text{Span}\{I, S\} \subset \mathbf{L}(\mathcal{H} \otimes \mathcal{H})$ . This is a result from representation theory of finite groups; see Lemma 2 in Appendix A 2 for details. It is for this reason that the ideal state and measurement operators of Eq. (10) are elements of the subspace  $W$ . The operators  $I$  and  $S$  do not form an orthogonal basis for  $W$ , but the following orthonormal basis can be constructed:

$$B_1 = \frac{I}{d} = \sigma_0 \otimes \sigma_0, \quad (49)$$

$$B_2 = \frac{S - \frac{I}{d}}{\sqrt{d^2 - 1}} = \frac{1}{\sqrt{d^2 - 1}} \sum_{\sigma \in \mathbf{P}^*} \sigma \otimes \sigma, \quad (50)$$

where  $\sigma_0$  is the Hilbert-Schmidt normalized identity on  $\mathcal{H}$  and  $\sigma \in \mathbf{P}^*$  are the  $d^2 - 1$  traceless normalized Pauli operators on  $\mathcal{H}$ . Since  $\mathcal{G}_{\text{avg}}^{(2)}$  is an orthogonal projection, it follows that  $(\mathcal{G}_{\text{avg}}^{(2)})^2 = \mathcal{G}_{\text{avg}}^{(2)}$ . Therefore we can rewrite

$$\mathbb{E}[q_j] = \langle\langle E | \mathcal{M}^{m-1} \Lambda^{\otimes 2} | \bar{\rho} \rangle\rangle, \quad (51)$$

where  $\mathcal{M} = \mathcal{G}_{\text{avg}}^{(2)} \Lambda^{\otimes 2} \mathcal{G}_{\text{avg}}^{(2)}$ . It can be shown that  $\mathcal{M}$  (which has only support on  $W$ ) has the following matrix entries [12]:

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ \frac{\|\alpha(\Lambda)\|^2}{\sqrt{d^2-1}} & u(\Lambda) \end{bmatrix}, \quad (52)$$

in the basis  $\{B_1, B_2\}$ , with  $\alpha$  the nonunitarity vector of  $\Lambda$  [see Eq. (A11) in Appendix A 1 b for details]. In particular this means that  $u(\Lambda) = \langle\langle B_2 | \Lambda^{\otimes 2} | B_2 \rangle\rangle$ , which might not be too surprising in view of Eq. (2). Since the input state  $\bar{\rho}$  is traceless and quantum channels are trace preserving, Eq. (51) is evaluated as

$$\mathbb{E}[q_j] = \langle\langle E | B_2 \rangle\rangle \langle\langle B_2 | \bar{\rho} \rangle\rangle u^{m-1} = B u^{m-1}, \quad (53)$$

where the final channel  $\Lambda^{\otimes 2}$  has been absorbed into the state as state preparation error. The robustness to state preparation and measurement errors stems from the fact that every component of  $\bar{\rho}$  and  $E$  outside the subspace  $W$  is projected out by the procedure.

In very similar fashion the variance, defined as  $\mathbb{V}[q_j] = \mathbb{E}[q_j^2] - \mathbb{E}[q_j]^2$ , can be computed. Using  $\text{Tr}[A]^2 = \text{Tr}[A^{\otimes 2}]$ , the mixed-product property of the tensor product [i.e.,  $(A \otimes$

$B)(C \otimes D) = (AC) \otimes (BD)$ ] and linearity, we write

$$\begin{aligned} \mathbb{E}[q_j^2] &= \frac{1}{|\mathbf{C}(d)|^m} \sum_j \langle\langle E^{\otimes 2} | \mathcal{G}_j^{\otimes 4} | \bar{\rho}^{\otimes 2} \rangle\rangle \\ &= \langle\langle E^{\otimes 2} | (\mathcal{G}_{\text{avg}}^{(4)} \Lambda^{\otimes 4})^m | \bar{\rho}^{\otimes 2} \rangle\rangle \\ &= \langle\langle E^{\otimes 2} | \mathcal{N}^{m-1} \Lambda^{\otimes 4} | \bar{\rho}^{\otimes 2} \rangle\rangle, \end{aligned} \quad (54)$$

where  $\mathcal{N} = \mathcal{G}_{\text{avg}}^{(4)} \Lambda^{\otimes 4} \mathcal{G}_{\text{avg}}^{(4)}$ , using the fact that  $\mathcal{G}_{\text{avg}}^{(4)}$  is also an orthogonal projection (Lemma 2 of Appendix A 2), and

$$\mathbb{E}[q_j]^2 = \langle\langle E^{\otimes 2} | (\mathcal{M}^{\otimes 2})^{m-1} \Lambda^{\otimes 4} | \bar{\rho}^{\otimes 2} \rangle\rangle. \quad (55)$$

Putting it together yields the following expression for the variance:

$$\mathbb{V}[q_j] = \langle\langle E^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}^{\otimes 2} \rangle\rangle, \quad (56)$$

where the final channel  $\Lambda^{\otimes 4}$  has again been absorbed into the state as state preparation error. One of the key ingredients of understanding this expression is finding the subspace of  $\mathbf{L}(\mathcal{H}^{\otimes 4})$  onto which  $\mathcal{G}_{\text{avg}}^{(4)}$  projects. The next section elaborates on this idea.

### C. Sketch of proof on variance bound

In this section we discuss and sketch the main ideas for the proof of our variance bound Eq. (19). A complete proof is given in Appendix B, Theorem 1. We actually prove a slightly stronger statement:

$$\begin{aligned} \mathbb{V}[q_j] &\leq \|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_\infty^2 + \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 \\ &\quad \times [\alpha^2 \beta^2 c_1(d) + \alpha^2 c_2(d) \|\bar{E}_{\text{err}}\|_\infty^2 + \beta^2 c_3(d) \|\bar{\rho}_{\text{err}}\|_1^2], \end{aligned} \quad (57)$$

where

$$\alpha = \frac{\text{Tr}[\bar{\rho}_{\text{id}} \bar{\rho}]}{\|\bar{\rho}_{\text{id}}\|_2^2} = (d^2 - 1) \text{Tr}[\bar{\rho}_{\text{id}} \bar{\rho}], \quad (58)$$

$$\beta = \frac{\text{Tr}[\bar{E}_{\text{id}} \bar{E}]}{\|\bar{E}_{\text{id}}\|_2^2} = \frac{\text{Tr}[\bar{E}_{\text{id}} \bar{E}]}{d^2 - 1}. \quad (59)$$

These quantities arise in the decomposition of the operators  $\bar{\rho}$ ,  $\bar{E}$  into an ideal and error parts as

$$\bar{\rho} = \alpha \bar{\rho}_{\text{id}} + \bar{\rho}_{\text{err}} \quad \text{and} \quad \bar{E} = \beta \bar{E}_{\text{id}} + \bar{E}_{\text{err}}. \quad (60)$$

It can be shown that  $-1 \leq \alpha, \beta \leq 1$  (see Appendix B, Lemma 11), so that Eq. (57) indeed implies Eq. (19). The quantities  $\alpha, \beta$  are generally unknown to the experimenter and therefore easily eliminated from the variance bound. Finally we remark that the bound on the interval length  $L$  [given in Eq. (27)] can also be slightly improved if additional information on  $\alpha$  or  $\beta$  is known. See Appendix B, Lemma 12 for a precise statement.

Our analysis departs from the expression of the variance Eq. (56). First, let us note that fully characterizing the operator  $\mathcal{N}$  seems infeasible. This was possible for the operator  $\mathcal{M}$ , since it only has support on the two-dimensional subspace  $W$ . The dimension of the support of  $\mathcal{N}$  (the dimension of the

space onto which  $\mathcal{G}_{\text{avg}}^{(4)}$  projects) is given, however, by [34–36]

$$|\text{Rge}(\mathcal{N})| = \begin{cases} 15 & \text{if } d = 2; \\ 29 & \text{if } d = 4; \\ 30 & \text{otherwise.} \end{cases} \quad (61)$$

Therefore calculating the  $|\text{Rge}(\mathcal{N})|^2$  matrix entries of  $\mathcal{N}$  seems infeasible. A different approach is thus needed. We use a telescoping series expansion (see Lemma 4 in Appendix A 3 for a proof)

$$\mathcal{N}^m - (\mathcal{M}^{\otimes 2})^m = \sum_{s=1}^m \mathcal{N}^{m-s} [\mathcal{N} - \mathcal{M}^{\otimes 2}] (\mathcal{M}^{\otimes 2})^{s-1} \quad (62)$$

in Eq. (56). The main idea of this is to study the middle operator  $\mathcal{N} - \mathcal{M}^{\otimes 2}$  carefully and sharply bound the relevant matrix entries of this operator. The action of  $(\mathcal{M}^{\otimes 2})^{s-1}$  is well understood because the full 2-by-2 matrix description of  $\mathcal{M}$  is known [given in Eq. (52)]. Finally the action of the remaining higher powers  $\mathcal{N}^{m-s-1}$  are bounded more trivially, since less information is computed about  $\mathcal{N}$ . Let us make these ideas more precise now.

In the previous it was discussed that the operator  $\mathcal{M}$  only has support on the subspace  $W = \text{Span}\{I, S\} = \text{Span}\{B_1, B_2\}$ . Therefore the analysis of the variance expression is quite different for the components of  $\bar{\rho}$  and  $E$  on the subspace  $W$  and its orthogonal complement. In fact, this lead to the decomposition of the operators  $\bar{\rho}$ ,  $\bar{E}$  into an ideal and error parts as

$$\bar{\rho} = \alpha \bar{\rho}_{\text{id}} + \bar{\rho}_{\text{err}} \quad \text{and} \quad \bar{E} = \beta \bar{E}_{\text{id}} + \bar{E}_{\text{err}}, \quad (63)$$

where the bar over  $E$  indicates its traceless component. In fact, the identity component of  $E$  does not contribute at all to  $q_j$  (and therefore to its mean and variance), because the input operator is traceless and all applied maps  $\mathcal{G}_j$  are trace preserving. So the traceless ideal components are in the traceless subspace of  $W$  (spanned by  $B_2$ ) and the error components are in the orthogonal complement  $W^\perp$ . In principle, plugging the above expansion into Eq. (56) yields 16 different terms after distributing the tensor powers in  $\bar{\rho}$  and  $E$  over the sum. However, 12 factors containing mixed tensor products of ideal and error components (e.g.,  $\bar{E}_{\text{id}} \otimes \bar{E}_{\text{err}}$ ) vanish. This is due to the structure of the space onto which  $\mathcal{G}_{\text{avg}}^{(4)}$  projects (see Appendix B 1 for more details). Thus we expand Eq. (56) as

$$\mathbb{V}[q_j] = \alpha^2 \beta^2 \langle \langle \bar{E}_{\text{id}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{id}}^{\otimes 2} \rangle \rangle \quad (64)$$

$$+ \alpha^2 \langle \langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{id}}^{\otimes 2} \rangle \rangle \quad (65)$$

$$+ \beta^2 \langle \langle \bar{E}_{\text{id}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle \rangle \quad (66)$$

$$+ \langle \langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle \rangle. \quad (67)$$

Each of these terms is bounded separately. Here we will demonstrate the ideas of our proof using the term of Eq. (65). The two terms Eqs. (64) and (66) are similar (only a few technical details are different; see Theorem 1 in Appendix B 2 for precise treatment of all terms). Using the telescoping series

Eq. (62) term Eq. (65) can be written as

$$\begin{aligned} (65) &= \alpha^2 \sum_{s=1}^{m-1} \langle \langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] (\mathcal{M}^{\otimes 2})^{s-1} | \bar{\rho}_{\text{id}}^{\otimes 2} \rangle \rangle \\ &= \alpha^2 \sum_{s=1}^{m-1} u^{2(s-1)} \langle \langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] | \bar{\rho}_{\text{id}}^{\otimes 2} \rangle \rangle, \end{aligned} \quad (68)$$

where the second line follows from the fact that  $\mathcal{M}|B_2\rangle = u|B_2\rangle$  and  $\bar{\rho}_{\text{id}} = \frac{1}{\sqrt{d^2-1}} B_2$  [see Eq. (B23) in Appendix B]. The next step is analyzing

$$\mathcal{N} - \mathcal{M}^{\otimes 2} \left| \frac{1}{d^2-1} B_2^{\otimes 2} \right\rangle = \frac{1}{d^2-1} \sum_i a_i |A_i\rangle, \quad (69)$$

where  $a_i = \langle \langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2^{\otimes 2} \rangle \rangle$  and  $|A_i\rangle$  is a basis for the space  $\text{Rge}(\mathcal{G}_{\text{avg}}^{(4)})$  on which  $\mathcal{N}$  has support. To find the basis  $|A_i\rangle$  explicitly, the following ideas from representation theory are used (see Appendix B 1 for details).

The map  $\mathcal{G} \mapsto \mathcal{G}^{\otimes n}$  is a group representation of the Clifford group  $\mathbf{C}(d)$  for any  $n$ . A fundamental result in group representation theory [37] (Lemma 2 in Appendix A 2) is that  $\mathcal{G}_{\text{avg}}^{(n)}$  is the orthogonal projection onto the trivial subspace of the representation  $\mathcal{G} \mapsto \mathcal{G}^{\otimes n}$ . For  $n = 2$ , the trivial subspace was found to be the space  $W$  [12], giving rise to the fit model of Eq. (53). The task at hand here is to find the trivial subspace for  $n = 4$ . To do so, the following is used. If  $(V, R)$  is an irreducible, real representation of a group  $\mathbf{C}(d)$ , then [37]

$$\left( \text{Span} \left\{ \sum_{v \in V} v \otimes v \right\}, R \otimes R \right) \quad (70)$$

is the only trivial representation of  $V \otimes V$  of the group  $\mathbf{C}(d)$  (see Lemma 3 in Appendix A 2). This allows us to calculate all trivial subrepresentations of  $\mathcal{G} \mapsto \mathcal{G}^{\otimes 4}$ , using a complete description of the irreducible representations of  $\mathcal{G} \mapsto \mathcal{G}^{\otimes 2}$ . These were found in Refs. [34,36]. Therefore Eq. (70) provides a method to compute the  $|A_i\rangle$  using the explicit description of the irreducible spaces of  $\mathcal{G} \mapsto \mathcal{G}^{\otimes 2}$  found in Ref. [36].

Hence, the following expansion is obtained for Eq. (65), using the expansion Eq. (69):

$$(65) = \frac{\alpha^2}{d^2-1} \sum_{s=1}^{m-1} u^{2(s-1)} \sum_i a_i \langle \langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle \rangle, \quad (71)$$

where  $a_i = \langle \langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2^{\otimes 2} \rangle \rangle$  are the coefficients of the expansion. The factor  $\frac{1}{d^2-1}$  is later absorbed into the constant  $c_2(d)$  in the final result. Up until this point, equality still holds. Now we are finally in a position to start bounding the term Eq. (65). To do so, we upper bound each  $a_i$ . These bounds involve constants depending on the dimension  $d$  [which are all absorbed into  $c_2(d)$ ] and are proportional to  $(1-u)^2$ . Finally the inner product containing  $\mathcal{N}^{m-s-1}$  is upper bounded by a constant depending on the dimension and proportional to  $\|\bar{E}_{\text{err}}\|_\infty^2$  (and in particular independent of  $m$  or  $s$ ). This then

gives a total bound on the term Eq. (65),

$$(65) \leq \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 \alpha^2 c_2(d) \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad (72)$$

where we used the geometric series

$$\sum_{s=1}^{m-1} u^{2(s-1)} = \frac{1 - u^{2(m-1)}}{1 - u^2}. \quad (73)$$

The terms Eq. (64) and Eq. (66) can be bounded by repeating all these steps, using a different telescoping series expansion where the factors  $(\mathcal{M}^{\otimes 2})^{s-1}$  and  $\mathcal{N}^{m-s}$  are interchanged in Eq. (62). The analysis is then performed by simplifying the inner product from left to right. This involves a few technicalities, but no new ideas. In the end, only the bound on the final inner product with  $\mathcal{N}^{m-s-1}$  and the proportionality constants  $c_1(d)$ ,  $c_3(d)$  differ, as can be seen from the result Eq. (57). Finally for the final term Eq. (67), there is not much more to do than

$$(67) = \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \leq \|\bar{E}_{\text{err}}\|_{\infty}^2 \|\bar{\rho}_{\text{err}}\|_1^2, \quad (74)$$

using Hölder's inequality and the fact that  $\mathcal{N}$  is contractive in the induced trace norm [38], i.e.,  $\|\mathcal{N}\|_{1 \rightarrow 1} \leq 1$  (see Proposition 19 in Appendix C).

## V. CONCLUSION AND FUTURE WORK

In this work we have shown a significant reduction in the required number of random sequences for unitarity randomized benchmarking (URB) than previously could be justified. This reduction is achieved by analyzing the statistics of the protocol. In particular, we have provided a bound on the variance of the sequence purity. Application of a concentration inequality yields the reduction in number of sequences, provided that the variance bound is sharp enough. We have shown that in realistic parameter regimes, the required number of sequences is in the order of hundreds, when benchmarking few-qubit Clifford gates. This brings benchmarking the unitarity of few-qubit Clifford gates into the realm of experimental feasibility.

The main ingredient of this result was a sharp bound on the variance of the sequence purity. The analysis was done for a slightly modified version of the protocol. This modification leads to better guarantees on the confidence and additionally yields a linear fitting problem. Our variance bound has the attractive property that it scales quadratically in  $1 - u$ , where  $u$  is the unitarity, up to constant contribution due to state preparation and measurement (SPAM) errors. This implies that fewer sequences are required to estimate highly coherent gates. We show that the constant contribution due to SPAM errors is a fundamental property of URB (and therefore not an artifact of our bound). Furthermore our bound is asymptotically independent of the sequence length and is therefore applicable in both short and long sequence lengths. Finally our bound grows exponentially in the number of qubits comprising the system. We argue that this is an artifact of the bound, which could be improved upon. As a result, our bound becomes vacuous for large systems. However, we have shown that our bound is sharp enough to benchmark few-qubit systems (say, up to five qubits).

During the analysis of the URB protocol, we have emphasized two different implementation techniques. We have explicitly shown their optimal state preparation and measurement settings for practical implementation. We highlighted the benefits and drawbacks of each implementation and showed the statistical difference between the two.

### A. Future work

There are a few caveats in the analysis of this work, which arise from the assumptions under which the bound holds. Each of these assumptions as summarized in Sec. II is an open avenue for future research. First and foremost, the assumption of the gate independent error model is rather strong and never completely satisfied in practical implementations of gates. The analysis of the URB protocol so far has been restricted to the gate-independent noise model [12]. There are three somewhat independent open problems with the URB protocol when one wants to generalize the model to (Markovian) gate-dependent errors. First, the behavior of the protocol must be studied. This means that the validity and deviation of the fit model must be studied under this more general noise model. Second, the statistics of the protocol can be studied in the gate-dependent error model. This aims to provide an answer to the question how many resources are required to extract the unitarity from measurement data in this more general noise model, provided that a generalized fit model is found. Finally one can attempt to relate the URB decay rate(s) in the gate-dependent setting to physically relevant quantities (like the unitarity) of the gates comprising the gate set. All three of these problems relating to gate-dependent errors are tough problems and many research focused on answering analogous questions for standard RB. For standard RB, progress has been made in terms of understanding the fit model and relating the decay rate to a physically interpretable infidelity in the gate-dependent error model [39–41]. However, statistical analyses of standard RB only apply to the gate-independent error model [24,26,27]. We suspect that some of the progress made in analyzing gate-dependent RB can be modified and applied to URB, but we have left this for future work.

A second interesting avenue is exploring how unitarity randomized benchmarking behaves when the assumption of unitary 2-design is relaxed [42]. This would give rise to a protocol that can benchmark the unitarity of different gate sets that do not form a 2-design. Interesting examples are the Dihedral group [43,44], subgroups of monomial unitary matrices [45] and subgroups of the Clifford group [46,47], where progress have been made for standard RB. Note that the first two of these gate sets are particularly interesting since they contain the  $T$ -gate. A general framework for standard RB given an arbitrary gate set is provided in [48]. An interesting open question is whether these techniques can be applied to URB.

Finally it is interesting if the current limitations of our bound can be improved upon. In particular an open question is how to improve this bound to be asymptotically independent of the dimension, a caveat that currently renders our bound impractical for large system ( $q \gg 5$ ). Similarly we wonder if our bound can be generalized to general multiqubit noise models that need not be unital. These lines of future work could improve the applicability of our bound.

## ACKNOWLEDGMENTS

The authors would like to thank Michael Walter for inspiring discussions on the topic. B.D., J.H., and S.W. are funded by NWA, a NWO VIDI grant, an ERC Starting Grant QINTERNET, and NWO Zwaartekracht QSC.

## APPENDIX A: PRELIMINARIES

The appendices are devoted to proving the upper bound Eq. (19) [actually we prove Eq. (57), which implies Eq. (19)] on the variance of the sequence purity for Clifford Unitarity Randomized Benchmarking. To do so, this appendix first provides an overview of the preliminaries and sets the formal notation used in the rest of the appendices. The material covered in this appendix is not a new result. In Appendix B then the variance bound of Eq. (19) is proven. It also contains the proof of the interval of the sequence purity Eq. (27). Finally, all technical lemmas used in the proof of the variance bound are collected in Appendix C. The material in Appendices B and C is the main result of this work.

## 1. Notation and definitions

In this subsection we summarize all notation used in the paper and the appendices. Suppose our principle system under investigation is a  $q$ -qubit system. Its state space is then represented by a  $d$ -dimensional Hilbert space  $\mathcal{H}$ , where  $d = 2^q$ . Typically  $\mathcal{H}$  is identified with  $\mathbb{C}^d$ . General vector spaces are typically denoted  $V$ . The dimension of a vector space is denoted  $|V| = \dim(V)$ . Hence  $d = 2^q = |\mathcal{H}|$ . The set of linear operators between two vector spaces  $V_1, V_2$  is denoted  $\mathbf{L}(V_1, V_2)$  [some references write  $\text{Hom}(V_1, V_2)$ ]. We write  $\mathbf{L}(V)$  as shorthand for  $\mathbf{L}(V, V)$  [in the literature also written as  $\text{End}(V)$ ]. It is convenient to think of  $\mathbf{L}(\mathcal{H})$  as a Hilbert space in itself, equipped with the Hilbert-Schmidt inner product. This inner product is defined as  $\langle A, B \rangle_{\text{HS}} = \text{Tr}[A^\dagger B]$  for any  $A, B \in \mathbf{L}(\mathcal{H})$ . It induces the Hilbert-Schmidt norm  $\|A\|_2 = \sqrt{\langle A, A \rangle_{\text{HS}}}$ . This is in fact a special case of the more general Schatten  $p$ -norms (for  $1 \leq p \leq \infty$ ), which are defined as

$$\|A\|_p^p = \text{Tr}[(A^\dagger A)^{\frac{p}{2}}] = \|s(A)\|_p^p = \sum_i s_i(A)^p. \quad (\text{A1})$$

Here  $s(A)$  denotes the vector of singular values  $s_i(A)$  of  $A$ . The Hilbert-Schmidt norm corresponds to  $p = 2$ . Other important special cases are the trace norm ( $p = 1$ ) and the operator norm to ( $p = \infty$ ).

The normalized Pauli matrices form an orthonormal basis of  $\mathbf{L}(\mathcal{H})$  with respect to the Hilbert-Schmidt inner product. The set of normalized Pauli's is denoted

$$\mathbf{P} := \left\{ \frac{P}{\sqrt{d}} \mid P \in \{I, X, Y, Z\}^{\otimes q} \right\}, \quad (\text{A2})$$

where  $I, X, Y, Z$  denote the usual (unnormalized) Pauli matrices. The set of traceless Pauli matrices is denoted  $\mathbf{P}^* = \mathbf{P} \setminus \{\sigma_0\}$ , where  $\sigma_0 := \frac{1}{\sqrt{d}} I^{\otimes q}$  is the normalized identity. Elements of  $\mathbf{P}$  are denoted by the Greek symbols  $\sigma, \tau$ . For two normalized Pauli matrices  $\sigma, \tau \in \mathbf{P}$ , we define the normalized matrix product  $\sigma \cdot \tau := \frac{1}{\sqrt{d}} \sigma \tau$ . This ensures that  $\|\sigma \cdot \tau\|_2 = 1$  so that  $\sigma \cdot \tau \in \pm \mathbf{P}$ . The tensor product between two Pauli matrices can then be conveniently omitted, so that  $\sigma \tau := \sigma \otimes \tau$ . This

is used for brevity when writing many tensor products of normalized Pauli matrices. From here on out, we will omit the tensor product. Finally for every normalized Pauli  $\tau \in \mathbf{P}$ , we define  $C_\tau$  as the set of all elements of  $\mathbf{P}^*$  that commute with  $\tau$ , except for  $\tau$  itself [36]:

$$C_\tau := \{\sigma \in \mathbf{P}^* : \sigma \cdot \tau = \tau \cdot \sigma\}. \quad (\text{A3})$$

In Ref. [36] it is shown that  $|C_\tau| = \frac{d^2-4}{2}$ .

The Clifford group, denoted  $\mathbf{C}(d)$ , has a natural action by conjugation on the set of Pauli matrices  $\mathbf{P}$ . Informally speaking, the Clifford group sends Pauli matrices to Pauli matrices under conjugation. More formally speaking, the Clifford group is the normalizer of the Pauli group (the group generated by  $\mathbf{P}$ ) in the unitary group, up to global phase:

$$\mathbf{C}(d) := \{U \in \mathbf{U}(d) : U\sigma U^\dagger \in \pm \mathbf{P}, \forall \sigma \in \mathbf{P}\} / \mathbf{U}(1). \quad (\text{A4})$$

An alternative description of the Clifford group is given in terms of its generators. The group is generated as

$$\mathbf{C}(d) = \langle \{H_i, S_i, \text{CNOT}_{ij} \mid i, j = 1, \dots, q, i \neq j\} \rangle / \mathbf{U}(1), \quad (\text{A5})$$

where  $H_i$  is the Hadamard gate and  $S_i$  is the  $\frac{\pi}{4}$ -phase gate on qubit  $i$ , and  $\text{CNOT}_{ij}$  is the CNOT gate on qubits  $i, j$ . For a more detailed introduction into the Pauli and Clifford group, see [49] and references therein. The size of the Clifford group is [50]

$$|\mathbf{C}(d)| = \prod_{j=1}^q 2(4^j - 1)4^j = 2^{O(q^2)}. \quad (\text{A6})$$

## a. States, measurements, and quantum channels

In quantum mechanics, quantum states are described by density operators. A density operator  $\rho \in \mathbf{L}(\mathcal{H})$  satisfies two properties. It is positive semidefinite (denoted  $\rho \geq 0$ ) and has  $\text{Tr}[\rho] = 1$ . POVM elements  $M \in \mathbf{L}(\mathcal{H})$  are positive semidefinite operators with all eigenvalues smaller than one. This means that  $I - M$  is also positive semidefinite and a POVM therefore satisfies  $0 \leq M \leq I$ . A general POVM measurement is described by a collection of POVM elements  $\{M_1, \dots, M_n\}$  that satisfy  $\sum_{i=1}^n M_i = I$ . Denote the measurement outcome associated with  $M_i$  as  $m_i$ . Then given a state  $\rho$ , the probability to observe outcome  $m_i$  is  $\text{Tr}[M_i \rho]$ . The Hermitian observable  $E \in \mathbf{L}(\mathcal{H})$  associated with this measurement is then  $E = \sum_{i=1}^n m_i M_i$ . Therefore the expectation value of the measurement, given the state  $\rho$ , is  $\langle E \rangle_\rho = \text{Tr}[E \rho]$ . In this work, we will only consider two-valued measurements, with associated outcomes  $\pm 1$ . Such a measurement is thus described by the POVM measurement  $M, I - M$  and the corresponding observable is  $E = M - (I - M) = 2M - I$ .

Operations on quantum states that transform one state into the other are described by quantum channels. In general, transformations of linear operators  $A \in \mathbf{L}(\mathcal{H})$  are described by a linear operator  $\mathcal{E} : \mathbf{L}(\mathcal{H}) \rightarrow \mathbf{L}(\mathcal{H})$ . These linear operators are sometimes called superoperators, to distinguish them from linear operators  $A \in \mathbf{L}(\mathcal{H})$ . A quantum channel is a superoperator  $\mathcal{E} : \mathbf{L}(\mathcal{H}) \rightarrow \mathbf{L}(\mathcal{H})$  that is

(1) completely positive (CP), i.e.,  $(\mathcal{E} \otimes \mathcal{I})(A) \geq 0$  for all  $0 \leq A \in \mathbf{L}(\mathcal{H}^{\otimes 2})$ , where  $\mathcal{I}$  is the identity channel and

(2) trace preserving (TP), i.e.,  $\text{Tr}[\mathcal{E}(A)] = \text{Tr}[A]$  for all  $A \in \mathcal{L}(\mathcal{H})$ .

Intuitively, this means that density operators are mapped to density operators. Thus quantum channels (CPTP superoperators) are indeed the operators that map quantum states to quantum states. Here generic quantum channels are denoted  $\mathcal{E}$  or  $\Lambda$ . A quantum channel is said to be unitary (denoted  $\mathcal{G}$ ) if  $\mathcal{G}(A) = GAG^\dagger$  for some unitary  $G \in \mathcal{L}(\mathcal{H})$  and for all  $A \in \mathcal{L}(\mathcal{H})$ . So unitary quantum channels (also called unitaries or gates) are denoted with a calligraphic  $\mathcal{G}$  and their counterparts in  $\mathcal{L}(\mathcal{H})$  are denoted  $G$ . Unital maps are superoperators  $\mathcal{E}$  that satisfy  $\mathcal{E}(I) = I$ . Note that all unitaries are unital, but the converse is not true [consider the completely depolarizing channel  $\mathcal{E}(A) = \frac{\text{Tr}[A]}{d}I$ ]. The space of superoperators is typically equipped with the induced Schatten norms, defined as

$$\|\mathcal{E}\|_{p \rightarrow q} = \sup_{A \in \mathcal{L}(\mathcal{H})} \{\|\mathcal{E}(A)\|_q : \|A\|_p = 1\}. \quad (\text{A7})$$

Important special cases are  $p = q = 1$ , which yields the induced trace norm and  $p = q = 2$  which results in the operator norm ( $\|\mathcal{E}\|_\infty = \|\mathcal{E}\|_{2 \rightarrow 2}$ ). For more details on states, measurements and quantum channels, the reader is referred to textbooks like Refs. [33,51]. In the next section, we will discuss the Liouville representation of states, measurements, and quantum channels.

### b. Liouville representation

Here we expand on the definition of the Liouville representation (also known as the natural or affine representation or the Pauli transfer matrix) [26,33] introduced in the main text. This representation exploits the fact that the Pauli matrices form an orthogonal basis for the set of linear operators with respect to the Hilbert-Schmidt inner product. We can then think of linear operators  $A \in \mathcal{L}(\mathbb{C}^d)$  as column vectors or row vectors with entries determined by the inner product with respect to a Pauli basis operator. Formally, we introduce a linear map  $|\cdot\rangle\rangle : \mathcal{L}(\mathbb{C}^d) \rightarrow \mathbb{C}^{d^2}$  defined by  $|\sigma_i\rangle\rangle = e_i$ , where  $\sigma_i$  is the  $i$ th normalized Pauli matrix in  $\mathbf{P}$  and  $e_i$  is the  $i$ th canonical basis vector of  $\mathbb{C}^{d^2}$ . The map is then extended to  $\mathcal{L}(\mathcal{H})$  by linearity, so that

$$|A\rangle\rangle = \sum_{\sigma_i \in \mathbf{P}} \langle \sigma_i, A \rangle_{\text{HS}} |\sigma_i\rangle\rangle. \quad (\text{A8})$$

The adjoint is then defined via  $\langle\langle A| = |A\rangle\rangle^\dagger$ . As a result, the inner product carries over as

$$\langle\langle A|B\rangle\rangle = \langle A, B \rangle_{\text{HS}} = \text{Tr}[A^\dagger B], \quad \forall A, B \in \mathcal{L}(\mathbb{C}^d). \quad (\text{A9})$$

Quantum channels  $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \rightarrow \mathcal{L}(\mathbb{C}^d)$  can then be viewed as matrices acting on the vectors  $|A\rangle\rangle$ . This matrix, called the Liouville matrix, is a map  $\mathcal{E} : \mathbb{C}^{d^2} \rightarrow \mathbb{C}^{d^2}$  defined by  $\mathcal{E}_{ij} = \langle\langle \sigma_i | \mathcal{E}(\sigma_j) \rangle\rangle$  (with  $\sigma_i, \sigma_j \in \mathbf{P}$ ). The Liouville matrix  $\mathcal{E}$  corresponding to the quantum channel  $\mathcal{E}$  is denoted in bold font to distinguish the two. The Liouville matrix representation of quantum channels naturally respects the vectorization  $|\cdot\rangle\rangle$ , the product (channel composition is identified with matrix multiplication), the adjoint and the tensor product. That is, for superoperators  $\mathcal{E}_1, \mathcal{E}_2 : \mathcal{L}(\mathbb{C}^d) \rightarrow \mathcal{L}(\mathbb{C}^d)$  and linear operators

$A, B, Q \in \mathcal{L}(\mathbb{C}^d)$ , the following relations hold:

$$\begin{aligned} |\mathcal{E}_2 \mathcal{E}_1(A)\rangle\rangle &= \mathcal{E}_2 |\mathcal{E}_1(A)\rangle\rangle = \mathcal{E}_2 \mathcal{E}_1 |A\rangle\rangle, \\ |\mathcal{E}_2 \otimes \mathcal{E}_1(A \otimes B)\rangle\rangle &= \mathcal{E}_2 \otimes \mathcal{E}_1 |A \otimes B\rangle\rangle = \mathcal{E}_2 \otimes \mathcal{E}_1 |A\rangle\rangle |B\rangle\rangle \\ &= \mathcal{E}_2 |A\rangle\rangle \otimes \mathcal{E}_1 |B\rangle\rangle, \\ |\mathcal{E}_1^\dagger(A)\rangle\rangle &= \mathcal{E}_1^\dagger |A\rangle\rangle, \\ \text{Tr}[Q^\dagger \mathcal{E}_1(A)] &= \langle\langle Q | \mathcal{E}_1(A) \rangle\rangle = \langle\langle Q | \mathcal{E}_1 |A\rangle\rangle. \end{aligned} \quad (\text{A10})$$

Note that with slight Dirac-notation-like ambiguity, the (not necessarily Hermitian operator)  $\mathcal{E}_1$  is always applied to the ket  $|A\rangle\rangle$  and not to the bra  $\langle\langle Q |$  in the last line. A quantum channel has a special block form of its Liouville matrix by imposing the trace-preserving property. If the first basis element of  $\mathbf{P}$  is  $\sigma_0 = \frac{1}{\sqrt{d}}$ , a quantum channel can be written as

$$\mathcal{E} = \begin{bmatrix} 1 & 0 \\ \alpha(\mathcal{E}) & \mathcal{E}_u \end{bmatrix}, \quad (\text{A11})$$

where  $\alpha(\mathcal{E})$  is the nonunitarity vector (of length  $d^2 - 1$ ) and  $\mathcal{E}_u$  is the unital block (of size  $d^2 - 1$  by  $d^2 - 1$ ) of  $\mathcal{E}$ . The trace-preserving property implies that no traceless Pauli matrix in  $\mathbf{P}^*$  can be mapped to  $\sigma_0$ , since  $\langle\langle \sigma_0 | \mathcal{E}(\tau) \rangle\rangle = \frac{\text{Tr}[\mathcal{E}(\tau)]}{\sqrt{d}} = 0$  for all  $\tau \in \mathbf{P}$ . Similarly  $\langle\langle \sigma_0 | \mathcal{E}(\sigma_0) \rangle\rangle = \frac{\text{Tr}[\mathcal{E}(\sigma_0)]}{\sqrt{d}} = 1$ . This justifies the first row of Eq. (A11). In terms of this decomposition, the definition of the unitarity Eq. (2) can be rewritten as

$$\begin{aligned} u(\mathcal{E}) &= \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathbf{P}^*} \langle\langle \tau | \mathcal{E} | \sigma \rangle\rangle^2 = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathbf{P}^*} \langle\langle \sigma | \mathcal{E}_u^\dagger \mathcal{E}_u | \sigma \rangle\rangle \\ &= \frac{1}{d^2 - 1} \sum_{\sigma \in \mathbf{P}^*} \langle\langle \sigma | \mathcal{E}_u \mathcal{E}_u^\dagger | \sigma \rangle\rangle = \frac{\text{Tr}[\mathcal{E}_u^\dagger \mathcal{E}_u]}{d^2 - 1} = \frac{\text{Tr}[\mathcal{E}_u \mathcal{E}_u^\dagger]}{d^2 - 1}, \end{aligned} \quad (\text{A12})$$

where  $\mathcal{E}_u$  is slight abuse of notation for  $1 \oplus \mathcal{E}_u$ .

## 2. Representation theory

Here we give a brief overview of the required representation theory of finite groups. This section will briefly provide some definitions and the results used in this work. For more details the reader can refer to textbooks like Refs. [37,52]. Let  $\mathbf{G}$  denote a finite group,  $V$  some finite-dimensional complex vector space. Let  $\mathbf{GL}(V)$  denote the general linear group on  $V$  (i.e., the set of invertible linear operators on  $V$ ). Then a representation  $(V, R)$  is a map  $R : \mathbf{G} \rightarrow \mathbf{GL}(V)$  that satisfies  $R(g)R(h) = R(gh)$  for all  $g, h \in \mathbf{G}$ . If  $V$  is equipped with an inner product (making it a Hilbert space) and  $R(g)$  is unitary for all  $g \in \mathbf{G}$ , then  $(V, R)$  is called a unitary representation of  $\mathbf{G}$ . If  $R$  is an injective map, then the representation is faithful. If the map  $R$  is clear from the context, the representation is just referred to as  $V$ .

A subspace  $W \subseteq V$  is called a subrepresentation of  $V$  if  $R(g)W \subseteq W$  for all  $g \in \mathbf{G}$ . If  $W = 0$  and  $W = V$  are the only subrepresentations of  $V$ , then  $V$  is an irreducible representation (often called irrep). Consider two representations  $V_1, V_2$  of  $\mathbf{G}$ . Then a mapping  $\varphi : V_1 \rightarrow V_2$  is called an intertwining operator if  $\varphi R_1(g) = R_2(g)\varphi$ . Intuitively, an intertwining operator preserves the structure of a representation. The representations  $V_1$  and  $V_2$  are called equivalent (denoted

$V_1 \cong V_2$ ) if there exists an intertwining operator  $\varphi$  that is an isomorphism between the vector spaces. A fundamental result in representation theory of finite groups is that a representation  $(V, R)$  can always be written as the direct sum of irreps.

*Lemma 1 (Maschke's Theorem [37]).* Let  $(V, R)$  be a finite-dimensional, nonzero representation of a finite group  $\mathbf{G}$ . Then  $(V, R)$  decomposes uniquely (up to isomorphisms and ordering) as

$$\begin{aligned} V &= \bigoplus_{i=1}^k (\mathbb{C}^{n_i} \otimes V_i) = \bigoplus_{i=1}^k V_i^{\oplus n_i} \quad \text{and} \\ R &= \bigoplus_{i=1}^k (I_{n_i} \otimes R_i) = \bigoplus_{i=1}^k R_i^{\oplus n_i}, \end{aligned} \quad (\text{A13})$$

where the set  $\{(V_i, R_i) : i = 1, \dots, k\}$  contains mutually inequivalent, nonzero, irreducible representations occurring with multiplicity  $n_i$  in the decomposition of  $(V, R)$  and  $I_{n_i}$  is the identity on a  $n_i$ -dimensional vector space.

As an example consider the Clifford group  $\mathbf{G} = \mathbf{C}(d) \subset \mathbf{L}(\mathcal{H})$ . Then the map  $R_1 : G \mapsto \mathcal{G}$  that associates the quantum channel  $\mathcal{G}$  with the abstract group element  $G \in \mathbf{C}(d)$  is a representation of  $\mathbf{C}(d)$  on the space  $V_1 = \mathbf{L}(\mathcal{H})$ . In fact  $G$  is itself a representation (the defining representation) on  $\mathcal{H}$ . The Liouville representation is also a representation on the space  $V_2 = \mathbb{C}^{d^2}$  via the map  $R_2 : G \mapsto \mathcal{G}$ . The Liouville representation  $(V_2, R_2)$  and the quantum channel representation  $(V_1, R_1)$  are equivalent representations of the Clifford group  $\mathbf{C}(d)$ . The intertwining operator that establishes this equivalence is given by  $\varphi = |\cdot\rangle\rangle : V_1 \rightarrow V_2$  (defined in Appendix A 1 b), the map that sends a linear operator  $A \in V_1$  to the corresponding Liouville vector  $|A\rangle\rangle \in V_2$ . The intertwining property  $R_2\varphi = \varphi R_1$  is then explicitly expressed as  $\mathcal{G}|A\rangle\rangle = |\mathcal{G}(A)\rangle\rangle$  for all  $A \in V_1$  and  $G \in \mathbf{C}(d)$ .

A crucial ingredient to the URB protocol is constructing the projector onto the trivial subrepresentations of a representation  $(V, R)$ . This is achieved in the following result.

*Lemma 2 (Projection onto trivial subrepresentations [37]).* Let  $(V, R)$  be any representation of a group  $\mathbf{G}$  and let  $V^{\mathbf{G}} := \{v \in V : R(g)v = v, \forall g \in \mathbf{G}\}$  denote the subspace on which  $\mathbf{G}$  acts trivially. Define the map  $\phi : V \rightarrow V$  by

$$\phi = \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} R(g). \quad (\text{A14})$$

Then  $\phi$  is an intertwining operator and moreover  $\phi$  is the orthogonal projection onto  $V^{\mathbf{G}}$ .

The next lemma is crucial for the variance analysis, as it provides a method to identify the subspace of trivial representations  $(V \otimes V^*)^{\mathbf{G}}$ , given a decomposition of  $V$  into irreps.

*Lemma 3.* Let  $(V, R_V)$  and  $(W, R_W)$  be unitary, irreducible finite-dimensional representations of a finite-dimensional group  $\mathbf{G}$  and let  $\{v_i\}, \{w_i\}$  be an orthonor-

mal basis for  $V, W$ , respectively. If  $V \cong W$  are equivalent representations (and the basis vectors are labeled such that the intertwining map  $\varphi$  between  $V$  and  $W$  maps  $v_i \mapsto w_i$ ), then the  $(V \otimes W^*, R_{V \otimes W^*})$  has one and only one trivial subrepresentation

$$(V \otimes W^*)^{\mathbf{G}} = \text{Span} \left\{ \sum_i v_i \otimes w_i^\dagger \right\}. \quad (\text{A15})$$

If  $V$  and  $W$  are not equivalent, then

$$(V \otimes W^*)^{\mathbf{G}} = \emptyset. \quad (\text{A16})$$

*Proof.* The proof makes use of the canonical isomorphism  $\alpha : V \otimes W^* \rightarrow \mathbf{L}(W, V)$  defined by  $v \otimes w^\dagger \mapsto vw^\dagger$  (extended by linearity), where  $V^*$  is the dual space of  $V$  (carrying the dual representation) and  $vw^\dagger$  acts on  $x \in W$  by  $vw^\dagger x := v \langle w, x \rangle$  (with  $\langle \cdot, \cdot \rangle$  the inner product on  $W$ ). Now  $\alpha$  is an intertwining operator [37]. Therefore it follows that

$$\alpha((V \otimes W^*)^{\mathbf{G}}) = (\mathbf{L}(W, V))^{\mathbf{G}}, \quad (\text{A17})$$

since  $\alpha$  preserve the structure of the representation. The subspace  $(\mathbf{L}(W, V))^{\mathbf{G}}$  of trivial subrepresentations of  $\mathbf{L}(W, V)$  is precisely the space of intertwining operators between the representations  $W$  and  $V$  [37]. Thus a trivial representation of  $V \otimes W^*$  corresponds to an intertwining operator from  $W$  to  $V$ . Schur's Lemma states that [37]

$$|(\mathbf{L}(W, V))^{\mathbf{G}}| = \begin{cases} 1 & \text{if } V \cong W \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A18})$$

So if  $V$  and  $W$  are inequivalent  $\alpha((V \otimes W^*)^{\mathbf{G}}) = (\mathbf{L}(W, V))^{\mathbf{G}} = \emptyset$ . And if  $V \cong W$ , let  $\phi \in (\mathbf{L}(W, V))^{\mathbf{G}}$  be the intertwining isomorphism with  $\|\phi\|_\infty = 1$ . Then letting  $v_i = \phi(w_i)$ , we can write  $\phi = \sum_i v_i w_i^\dagger$ , so that

$$\alpha((V \otimes W^*)^{\mathbf{G}}) = \text{Span} \left\{ \sum_i v_i w_i^\dagger \right\}, \quad (\text{A19})$$

which yields the result after applying  $\alpha^{-1}$ .  $\blacksquare$

*Corollary 1.* If moreover the representation  $V = W$  is real and thus orthogonal, then (using  $V^* \cong V$ ) it follows that

$$(V \otimes V)^{\mathbf{G}} = \text{Span} \left\{ \sum_i v_i \otimes v_i \right\}. \quad (\text{A20})$$

*Corollary 2.* Let  $V$  be a finite-dimensional vector space carrying a group representation. By Lemma 1 there exists a decomposition  $V = \bigoplus_{i=1}^k V_i^{\oplus n_i}$  into mutually inequivalent irreducible representations. Denote  $V_{i_s}$  the  $s$ th copy of the space  $V_i$  ( $s = 1, \dots, n_i$ ) and denote  $\{v_j^{(i_s)} : j = 1, \dots, |V_i|\}$  an orthonormal basis of  $V_{i_s}$  that respect the isomorphisms between equivalent spaces (meaning that  $v_j^{(i_s)} \mapsto v_j^{(i_{s'})}$  under the intertwining isomorphism between  $V_{i_s}$  and  $V_{i_{s'}}$ ). Then the trivial subrepresentations of  $V \otimes V$  are given by

$$(V \otimes V)^{\mathbf{G}} = \text{Span} \left\{ \sum_{j=1}^{|V_i|} v_j^{(i_s)} \otimes v_j^{(i_{s'})} \mid \forall s, s' = 1, \dots, n_i, \forall i = 1, \dots, k \right\}. \quad (\text{A21})$$

*Proof.* Let us start by writing

$$V \otimes V = \bigoplus_{i,i'=1}^k \bigoplus_s^{n_i} \bigoplus_{s'}^{n_{i'}} (V_{i_s} \otimes V_{i'_s}). \quad (\text{A22})$$

Each trivial subrepresentation is found by application of Lemma 3 to each term in this decomposition. This makes use of the fact that  $V_{i_s} \cong V_{i'_s}$  are equivalent if and only if  $i' = i$  by virtue of the decomposition. ■

In Appendix B this machinery is used to find the trivial subrepresentations of the Liouville tensor-4 representation of the Clifford group  $\mathbf{C}(d)$ . But first a section is given with some preliminary technical lemmas from literature that are required in the proof of our variance bound.

### 3. Technical lemmas from literature

In this section we review a few lemmas from literature that are required for our variance bound. Some lemmas are stated without proof and the reader is then referred to the reference for a proof. The first lemma is a telescoping series for expanding the variance expression. It is applied to quantum channels, but here presented in more general form.

*Lemma 4 (Telescoping Series [27]).* Let  $A$  be an associative algebra with unit. Then for  $a, b \in A$  and  $m \in \mathbb{N}_+$ ,

$$a^m - b^m = \sum_{s=1}^m a^{m-s} (a-b) b^{s-1} = \sum_{s=1}^m b^{m-s} (a-b) a^{s-1}. \quad (\text{A23})$$

*Proof.* By direct computation, it follows that

$$\begin{aligned} \sum_{s=1}^m a^{m-s} (a-b) b^{s-1} &= \sum_{s=1}^m a^{m-s+1} b^{s-1} - a^{m-s} b^s \\ &= \sum_{s=0}^{m-1} a^{m-s} b^s - \sum_{s=1}^m a^{m-s} b^s = a^m b^0 - a^0 b^m = a^m - b^m \end{aligned}$$

and

$$\begin{aligned} \sum_{s=1}^m b^{s-1} (a-b) a^{m-s} &= \sum_{s=1}^m b^{s-1} a^{m-s+1} - b^s a^{m-1} \\ &= \sum_{s=0}^{m-1} b^s a^{m-s} - \sum_{s=1}^m b^s a^{m-1} = b^0 a^m - a^0 b^m = a^m - b^m. \end{aligned}$$

Note that the set of quantum channels form an associative algebra with unit, so that this lemma indeed applies to quantum channels.

Next we present a lemma that bounds the induced Schatten  $p \rightarrow p$  norm of a quantum channel.

*Lemma 5 (Pérez-García. et al. [38]).* Let  $\mathcal{E}$  be a CPTP quantum channel on a  $d$ -dimensional Hilbert space  $\mathcal{H}$ , with  $d = 2^q$  for a  $q$ -qubit system. Then for all  $p \in [1, \infty]$ ,

$$\|\mathcal{E}\|_{p \rightarrow p} = \max_{A \in \mathcal{L}(\mathcal{H})} \{\|\mathcal{E}(A)\|_p : \|A\|_p = 1\} \leq d^{1-\frac{1}{p}} \quad (\text{A24})$$

and

$$\|\mathcal{E}\|_{p \rightarrow p}^H := \max_{\substack{0 \neq A \in \mathcal{L}(\mathcal{H}) \\ \text{Tr}[A]=0, A=A^\dagger}} \left\{ \frac{\|\mathcal{E}(A)\|_p}{\|A\|_p} \right\} \leq \left( \frac{d}{2} \right)^{1-\frac{1}{p}}. \quad (\text{A25})$$

If in addition  $\mathcal{E}$  is unital ( $\mathcal{E}(I) = I$ ), then  $\|\mathcal{E}\|_{p \rightarrow p} \leq 1$  for all  $p \in [1, \infty]$ .

The following three lemmas are used to bound the quantities  $a_i$  Eq. (69). First, we state a technical lemma used in Ref. [27], which can be restated as

*Lemma 6 (Helsen. et al. [27]).* Let  $\mathcal{E}$  be a CPTP map on a  $d$ -dimensional Hilbert space. Then

$$0 \leq \frac{1}{d^2-1} \sum_{\sigma \in \mathcal{P}^*} \langle \langle \sigma | \mathcal{E} | \sigma \rangle \rangle^2 - f^2 \leq \frac{d^2-2}{d^2} (1-f)^2, \quad (\text{A26})$$

where

$$f = \frac{1}{d^2-1} \sum_{\sigma \in \mathcal{P}^*} \langle \langle \sigma | \mathcal{E} | \sigma \rangle \rangle \quad (\text{A27})$$

is the randomized benchmarking decay parameter of  $\mathcal{E}$ .

Here this lemma is applied to channels of the form

$$\mathcal{E}_1 = \begin{bmatrix} 1 & 0 \\ 0 & \Lambda_u \Lambda_u^\dagger \end{bmatrix} \quad \text{and} \quad \mathcal{E}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \Lambda_u^\dagger \Lambda_u \end{bmatrix}, \quad (\text{A28})$$

where  $\Lambda_u$  is the unital block of the error map  $\Lambda$  under investigation, since then  $f(\mathcal{E}_1) = f(\mathcal{E}_2) = u(\Lambda)$ . It is not clear that these superoperators are even a quantum channel (i.e., that they are CPTP). Therefore the following lemma provides a necessary condition on  $\Lambda$  for which Eq. (A28) are CPTP maps.

*Lemma 7.* Let  $\Lambda$  be a CPTP quantum channel on a  $d$ -dimensional Hilbert space. Then the channels  $\mathcal{E}_1, \mathcal{E}_2$  defined in Eq. (A28) are CPTP if either  $d = 2$  or if  $\Lambda$  is unital (or both). Moreover  $\|\mathcal{E}_1\|_{2 \rightarrow 2}, \|\mathcal{E}_2\|_{2 \rightarrow 2} \leq 1$ .

*Proof.* If  $d = 2$  (that is, if  $\Lambda$  is a single-qubit channel), then the unital part of  $\Lambda$ , defined as

$$\hat{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & \Lambda_u \end{bmatrix}, \quad (\text{A29})$$

is CPTP [53, Theorem IV.1]. For the general  $d$ -dimensional case, it is assumed that  $\Lambda$  is unital, so that  $\Lambda = \hat{\Lambda}$ . So in either case,  $\hat{\Lambda}$  is CPTP and unital. It can be shown that the adjoint of a CPTP and unital map is also CPTP and unital [33, Proposition 2.18 and Theorem 2.26], i.e.,  $\hat{\Lambda}^\dagger$  is CPTP and unital. Therefore  $\mathcal{E}_1 = \hat{\Lambda} \hat{\Lambda}^\dagger$  and  $\mathcal{E}_2 = \hat{\Lambda}^\dagger \hat{\Lambda}$  are also CPTP and unital. Lemma 5 then ensures that  $\|\mathcal{E}_1\|_{2 \rightarrow 2} \leq 1$  and  $\|\mathcal{E}_2\|_{2 \rightarrow 2} \leq 1$ . ■

Third is a lemma from matrix analysis. It is a characterization of positive semidefinite matrices in terms of its principal minors. This lemma was used on  $\mathbf{I} - \hat{\Lambda} \hat{\Lambda}^\dagger$  to bound its off-diagonal terms.

*Lemma 8 (Sylvester's criterion).* Let  $A \in \mathcal{L}(\mathbb{C}^{d^2})$  be a Hermitian matrix. Then  $A$  is positive semidefinite if and only if all of its principal minors are nonnegative.

*Proof.* See, e.g., Ref. [54, Corollary 7.1.5 and Theorem 7.2.5]. ■

Next we present two results, also from matrix analysis, that are used several times to bound inner products. The first is a trace inequality and the second is Hölder's inequality.

*Lemma 9.* Let  $A, B \in \mathcal{L}(\mathcal{H})$  be two linear operators on a  $d$ -dimensional Hilbert space  $\mathcal{H}$ . Denote their singular values as  $s_i(A), s_i(B)$  respectively with  $i = 1, \dots, d$ , both in decreasing order. Finally let  $s(A)$  and  $s(B)$  denote vectors with entries  $s_i(A)$  and  $s_i(B)$ . Then

(1) (Von Neumann's trace inequality)  $\text{Re}(\text{Tr}[AB]) \leq \sum_{i=1}^d s_i(A)s_i(B)$  and

(2) (Hölder's inequality)  $\sum_{i=1}^d |s_i(A)s_i(B)| \leq \|s(A)\|_p \|s(B)\|_q = \|A\|_p \|B\|_q$ , for any pair  $p, q \in [1, \infty]$  such that  $p^{-1} + q^{-1} = 1$ .

Since singular values are nonnegative, combining the statements yields  $\text{Re}(\text{Tr}[AB]) \leq \|A\|_p \|B\|_q$  for any pair  $p, q \in [1, \infty]$  such that  $p^{-1} + q^{-1} = 1$ .

*Proof.* Statement 1 is proven for example in [54, Theorem 8.7.6] and statement 2 is proven in Ref. [55, Theorem 31.3].

*Corollary 3.* If  $A, B \in \mathcal{L}(\mathcal{H})$  are Hermitian, then  $\text{Tr}[AB]^* = \text{Tr}[(AB)^\dagger] = \text{Tr}[B^\dagger A^\dagger] = \text{Tr}[BA] = \text{Tr}[AB]$ , so that  $\text{Tr}[AB]$  is real. Therefore  $\text{Tr}[AB] \leq \|A\|_p \|B\|_q$  for any  $p, q \in [1, \infty]$  satisfying  $p^{-1} + q^{-1} = 1$ .

Finally some of our bounds use the fact that the mean of squares is larger than the square of the mean. We show this well-known fact below.

*Lemma 10 (Mean of squares is larger than square of mean).* Let  $\{x_i\} \subset \mathbb{R}$  be a collection of  $N$  real numbers. Then

$$\left( \frac{1}{N} \sum_{i=1}^N x_i \right)^2 \leq \frac{1}{N} \sum_{i=1}^N x_i^2. \quad (\text{A30})$$

*Proof.* By direct computation, it follows that

$$\begin{aligned} & \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right) - \left( \frac{1}{N} \sum_{i=1}^N x_i \right)^2 \\ &= \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right) - 2 \left( \frac{1}{N} \sum_{i=1}^N x_i \right) \left( \frac{1}{N} \sum_{k=1}^N x_k \right) + \left( \frac{1}{N} \sum_{k=1}^N x_k \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[ x_i^2 - 2x_i \left( \frac{1}{N} \sum_{k=1}^N x_k \right) + \left( \frac{1}{N} \sum_{k=1}^N x_k \right)^2 \right] \\ &= \frac{1}{N} \sum_{i=1}^N \left[ x_i - \left( \frac{1}{N} \sum_{k=1}^N x_k \right) \right]^2 \geq 0, \end{aligned} \quad (\text{A31})$$

since it is the sum of real numbers squared, proving the result. ■

## APPENDIX B: VARIANCE BOUND AND INTERVAL LENGTH BOUND

This section is devoted to rigorously proving the variance bound Eq. (19). Along the way we also prove the interval length bound Eq. (27). The key ingredient of the variance bound proof is finding the trivial subrepresentations of the

Liouville tensor-4 representation of the Clifford group  $\mathbf{C}(d)$ . This is done in the first subsection. Then the variance bound Eq. (19) is proven. The technical lemmas used in this proof are collected in Appendix C.

### 1. Trivial subrepresentations of the tensor-4 Liouville representation of the Clifford group

This section is concerned with presenting the trivial subrepresentations of the representation  $G \mapsto \mathcal{G}^{\otimes 4}$  of the Clifford group  $\mathbf{C}(d)$ . This representation is equivalent to  $G \mapsto \mathcal{G}^{\otimes 4}$  by the intertwining isomorphism  $|\cdot\rangle\rangle$ . Therefore both are considered the same and with slight abuse of notation we refer to them both as the same representation, which we will call the tensor-4 Liouville representation.

The key idea is to apply Lemma 3 and its corollaries to find the trivial subrepresentations of the tensor-4 representation  $G \mapsto \mathcal{G}^{\otimes 4}$ . This requires a full description of the Liouville tensor-2 representation  $G \mapsto \mathcal{G}^{\otimes 2}$  in terms of its irreducible components. This was studied in [35,36]. Let us denote  $V = \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  as the space that carries the tensor-2 representation. The present problem is therefore to find the trivial subrepresentations of  $V \otimes V$ , given a decomposition of  $V$  into irreducible representations. In an earlier result [34] the multiplicity of the trivial representation in  $V \otimes V$  was calculated. They found that

$$|(V \otimes V)^{\mathbf{C}(d)}| = \begin{cases} 15 & \text{if } d = 2; \\ 29 & \text{if } d = 4; \\ 30 & \text{otherwise,} \end{cases} \quad (\text{B1})$$

which is a justification of Eq. (61). First we will discuss the decomposition of  $V$  into irreducible representations [36], and next we will apply Lemma 3 to find  $(V \otimes V)^{\mathbf{C}(d)}$  explicitly.

The full decomposition of the Liouville tensor-2 representation  $(V, R)$  given by  $R : \mathbf{C}(d) \rightarrow \text{GL}(V) : G \mapsto \mathcal{G}^{\otimes 2}$  is studied in [36]. We will review the result of this work here, following their notation. A summary of the relevant subspaces is given in Table II. First, the representation  $V$  is decomposed in the following subrepresentations, defined by

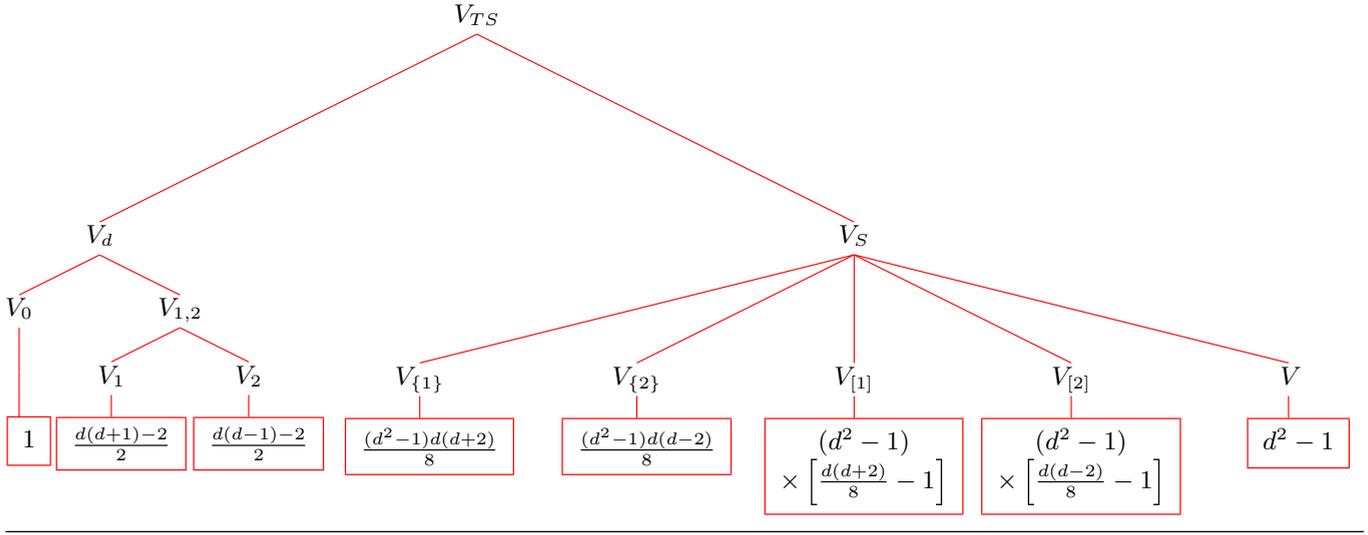
$$\begin{aligned} V_S &:= \text{Span} \left\{ \frac{\sigma\tau + \tau\sigma}{\sqrt{2}} : \sigma, \tau \in \mathbf{P}^*, \sigma \neq \tau \right\}, \\ V_A &:= \text{Span} \left\{ \frac{\sigma\tau - \tau\sigma}{\sqrt{2}} : \sigma, \tau \in \mathbf{P}^*, \sigma \neq \tau \right\}, \\ V_d &:= \text{Span} \{ \sigma\sigma : \sigma \in \mathbf{P}^* \}, \\ V_{r,l} &:= \text{Span} \{ \sigma_0\sigma, \sigma\sigma_0 : \sigma \in \mathbf{P}^* \}, \\ V_{\text{id}} &:= \text{Span} \{ B_1 = \sigma_0\sigma_0 \}. \end{aligned} \quad (\text{B2})$$

Recall that the tensor symbol is omitted for brevity (so  $\sigma\tau$  means  $\sigma \otimes \tau$  here). Each of these spaces carries a subrepresentation and furthermore  $V = V_{\text{id}} \oplus V_{r,l} \oplus V_d \oplus V_S \oplus V_A$ . Finally let us define the traceless, symmetric subspace as

$$V_{TS} := V_S \oplus V_d. \quad (\text{B3})$$

Since the ideal input and measurement operators for the URB protocol  $\bar{\rho}_{\text{id}}, \bar{E}_{\text{id}}$  [as defined in Eq. (10); see also Eq. (B23)] are elements of  $V_{TS}$  and since  $\Lambda^{\otimes 2}(V_{TS}) \subseteq V_{TS}$  by the trace-preserving property of  $\Lambda$  and the symmetry with respect to

TABLE II. Hierarchy of subspaces contained within the traceless, symmetric subspace  $V_{TS}$ , carrying the relevant subrepresentation of the Liouville tensor-4 representation  $G \mapsto \mathcal{G}^{\otimes 4}$ . Every child node is a subspace (that also carries a subrepresentation) of its parent node and all child nodes direct sum to their parent. Leaf nodes represent the final irreducible subspaces and their dimension are shown in the box below each leaf node. Definitions of the composite spaces are given in the main text ((B2), (B3), and (B4); for the definitions of the irreducible spaces, see [36]). For  $d = 2$  or  $d = 4$  there are certain subspaces with  $|V_i| \leq 0$ . This means that such a subspace is empty and therefore not present in the decomposition. Summing the dimensions of the child nodes together, yields the following sizes for the decomposable spaces:  $|V_{1,2}| = d^2 - 2$ ,  $|V_d| = d^2 - 1$ ,  $|V_S| = \frac{1}{2}(d^2 - 1)(d^2 - 2)$  and  $|V_{TS}| = \frac{1}{2}d^2(d^2 - 1)$ .



swapping the two copies of  $\mathcal{H}$ , the only relevant subspace of  $V$  is  $V_{TS}$ . Therefore we continue our analysis of  $V_{TS}$ .

The space  $V_d$  can be broken up into the two subrepresentations:

$$V_0 := \text{Span} \left\{ B_2 = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \sigma \sigma \right\} \quad \text{and} \quad V_{1,2} := V_d \setminus V_0. \tag{B4}$$

In the single-qubit case ( $q = 1$ ), the spaces  $V_S$  and  $V_{1,2}$  are irreducible, therefore fully characterizing  $V_{TS} = V_0 \oplus V_{1,2} \oplus V_S$ . However, if  $q \geq 2$  the space  $V_{1,2}$  breaks into two irreps, indexed by the index set  $\mathcal{Z}_{1,2}$ . For  $q = 2$ ,  $V_S$  breaks into four irreps, while for  $q \geq 3$  it breaks into five irreps, which will be indexed by  $\mathcal{Z}_S$ . So the space  $V_{TS}$  breaks up into the following number of irreps:

$$|\mathcal{Z}_{TS}| = \begin{cases} 3 & \text{if } q = 1; \\ 7 & \text{if } q = 2; \\ 8 & \text{if } q \geq 3, \end{cases} \tag{B5}$$

where  $\mathcal{Z}_{TS} := \mathcal{Z}_d \cup \mathcal{Z}_S = \{0\} \cup \mathcal{Z}_{1,2} \cup \mathcal{Z}_S$ . A summary of all the subspaces of  $V_{TS}$  that carry subrepresentations is given in Table II, together with the dimensions of the spaces. In Ref. [36] it is shown that all irreducible representations contained in  $V_{TS} = V_d \oplus V_S$  indexed by  $\mathcal{Z}_{TS}$  are mutually inequivalent. Therefore it follows from Lemma 3 that there are precisely  $|\mathcal{Z}_{TS}|$  trivial subrepresentations contained in  $V_{TS} \otimes V_{TS}$ . The lemma also provides an explicit method of finding them, given a basis for  $V_i$  from Ref. [36].

Let  $\mathcal{B}_i$  denote an orthonormal basis for  $V_i$ , for  $i \in \mathcal{Z}_{TS}$ . Then since all irreps indexed by  $\mathcal{Z}_{TS}$  are mutually inequivalent, Lemma 3 gives an explicit way to compute the trivial

subreps of  $(V_{TS} \otimes V_{TS})$  as

$$A_i = \frac{1}{\sqrt{|V_i|}} \sum_{v_i \in \mathcal{B}_i} v_i v_i, \quad \forall i \in \mathcal{Z}_{TS}, \tag{B6}$$

where the normalization constant is to normalize  $A_i$  with respect to the Hilbert-Schmidt norm  $\|A_i\|_2 = 1$ . In the multi-qubit case where  $V_{1,2}$  and  $V_S$  are not irrep, it is still useful to define

$$A_j = \frac{1}{\sqrt{|V_j|}} \sum_{i \in \mathcal{Z}_j} \sqrt{|V_i|} A_i, \quad j \in \{S; d; 1, 2\}. \tag{B7}$$

In fact, this allows us to explicitly find  $A_{1,2}$  from  $A_d$  and  $A_0$ . Using the basis for  $V_0$ ,  $V_d$  and  $V_S$  [in Eqs. (B2) and (B4)], we therefore explicitly find

$$A_0 := B_2 B_2 = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathcal{P}^*} \sigma \sigma \tau \tau, \tag{B8}$$

$$A_{1,2} := \frac{1}{\sqrt{d^2 - 2}} \left( \sum_{\sigma \in \mathcal{P}^*} \sigma^{\otimes 4} - A_0 \right), \tag{B9}$$

$$A_S := \sqrt{\frac{1}{2(d^2 - 1)(d^2 - 2)}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \sigma \tau \sigma \tau + \sigma \tau \tau \sigma. \tag{B10}$$

No explicit expression is needed for any  $i \in \mathcal{Z}_S$  or  $i \in \mathcal{Z}_{1,2}$  if  $V_S$  and  $V_{1,2}$  are reducible (which happens in the multi-qubit case), because bounds are defined in terms of  $A_S$  and  $A_{1,2}$ . The only exception to this is  $i = [\text{adj}] \in \mathcal{Z}_S$ . The space

$V_{[\text{adj}]} \subset V_S$ , which carries an irrep, is defined by [36]

$$V_{[\text{adj}]} = \text{Span} \left\{ v_{\tau}^{[\text{adj}]} = \frac{1}{2\sqrt{|C_{\tau}|}} \times \sum_{\sigma \in C_{\tau}} \sigma(\sigma \cdot \tau) + (\sigma \cdot \tau)\sigma \mid \tau \in \mathbf{P}^* \right\}, \quad (\text{B11})$$

where  $\cdot$  indicates the normalized matrix product and where  $C_{\tau}$  is the set of all elements of  $\mathbf{P}^*$  that commute with  $\tau$  as defined in Eq. (A3). The corresponding trivial subrepresentation, as computed using Eq. (B6), is

$$A_{[\text{adj}]} = \frac{1}{2(d^2 - 4)\sqrt{d^2 - 1}} \times \sum_{\tau \in \mathbf{P}^*} \left[ \sum_{\sigma \in C_{\tau}} (\sigma \cdot \tau)\sigma + \sigma(\sigma \cdot \tau) \right]^{\otimes 2}. \quad (\text{B12})$$

In the next section, we use the trivial subrepresentations of the Liouville tensor-4 representation to prove our variance bound.

## 2. Statement and proof of the variance bound and interval length bound

In this section we will state and prove our main theorem on the variance bound and prove the interval in which the average sequence purity is found. We also show the optimality of the ideal input and measurement operators. First, we will recapture some of the most important definitions and results discussed in the main text. The point of departure is the expression for the variance of Eq. (56)

$$\mathbb{V}[q_{\mathbf{j}}] = \langle\langle \bar{E}^{\otimes 2} | \mathcal{M}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}^{\otimes 2} \rangle\rangle, \quad (\text{B13})$$

where the operators are defined as

$$\mathcal{M} := \mathcal{G}_{\text{avg}}^{(2)} \Lambda^{\otimes 2} \mathcal{G}_{\text{avg}}^{(2)}, \quad \mathcal{N} := \mathcal{G}_{\text{avg}}^{(4)} \Lambda^{\otimes 4} \mathcal{G}_{\text{avg}}^{(4)},$$

$$\mathcal{G}_{\text{avg}}^{(n)} := \frac{1}{|C(d)|} \sum_{\mathcal{G} \in C(d)} \mathcal{G}^{\otimes n}. \quad (\text{B14})$$

Here  $q_{\mathbf{j}}$  is the sequence purity due to the sequence  $\mathbf{j}$ . As discussed in the main text,  $\mathcal{M}$  only has support on the space  $W = \text{Span}\{B_1, B_2\} \subset \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  [12], where

$$B_1 = \frac{I}{d} = \sigma_0 \sigma_0, \quad (\text{B15})$$

$$B_2 = \frac{S - B_1}{\sqrt{d^2 - 1}} = \frac{1}{\sqrt{d^2 - 1}} \sum_{\sigma \in \mathbf{P}^*} \sigma \sigma. \quad (\text{B16})$$

In particular the matrix elements of  $\mathcal{M}$  with respect to this basis [see also Eq. (52)] as

$$\mathcal{M} = \begin{bmatrix} 1 & 0 \\ \frac{\|\alpha(\Lambda)\|^2}{\sqrt{d^2 - 1}} & u(\Lambda) \end{bmatrix}. \quad (\text{B17})$$

From this it follows that [see also Eq. (A12)]

$$\mathcal{M}|B_2\rangle = u|B_2\rangle, \quad (\text{B18})$$

which implies that  $\langle\langle B_2 | \mathcal{M} | B_2 \rangle\rangle = \langle\langle B_2 | \Lambda^{\otimes 2} | B_2 \rangle\rangle = u$ , since  $\mathcal{G}_{\text{avg}}^{(2)}|B_2\rangle = |B_2\rangle$  and  $B_2$  is normalized. This is used in the analysis of Eq. (B13).

In Eq. (B13) the measurement  $E$  is replaced with its traceless counterpart  $\bar{E}$ , which is defined as

$$\bar{E} := E - \frac{\text{Tr}[E]}{d^2} I = E - \langle\langle B_1 | E \rangle\rangle B_1. \quad (\text{B19})$$

Since  $\bar{\rho}$  is traceless by construction and  $\mathcal{G}_{\mathbf{j}}$  is trace-preserving, it follows that  $q_{\mathbf{j}} = \langle\langle E | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | \bar{\rho} \rangle\rangle = \langle\langle \bar{E} | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | \bar{\rho} \rangle\rangle$ . This justifies the replacement of  $E$  by  $\bar{E}$  in all expectation value and variance expressions. In our analysis it is advantageous to think of  $\bar{E}$  instead of  $E$ , since then  $\bar{E}_{\text{id}}, \bar{\rho}_{\text{id}} \propto B_2$ . The ideal state and measurement operators were defined in Eq. (10). For completeness, they are

$$E_{\text{id}} = S = B_1 + \sqrt{d^2 - 1} B_2, \quad (\text{B20})$$

$$\rho_{\text{id}} = \frac{I + S}{d(d+1)} = \frac{1}{d} B_1 + \frac{\sqrt{d^2 - 1}}{d(d+1)} B_2, \quad (\text{B21})$$

$$\hat{\rho}_{\text{id}} = \frac{I - S}{d(d-1)} = \frac{1}{d} B_1 - \frac{\sqrt{d^2 - 1}}{d(d-1)} B_2, \quad (\text{B22})$$

from which it follows that

$$\bar{E}_{\text{id}} = \sqrt{d^2 - 1} B_2 \quad \text{and} \quad \bar{\rho}_{\text{id}} = \frac{\rho_{\text{id}} - \hat{\rho}_{\text{id}}}{2} = \frac{1}{\sqrt{d^2 - 1}} B_2. \quad (\text{B23})$$

The implemented operators  $\bar{\rho}$  and  $E$  can then be decomposed into an ideal part and an error part as

$$\alpha := \frac{\langle\langle \bar{\rho}_{\text{id}} | \bar{\rho} \rangle\rangle}{\langle\langle \bar{\rho}_{\text{id}} | \bar{\rho}_{\text{id}} \rangle\rangle} = (d^2 - 1) \langle\langle \bar{\rho}_{\text{id}} | \bar{\rho} \rangle\rangle, \quad \bar{\rho}_{\text{err}} := \bar{\rho} - \alpha \bar{\rho}_{\text{id}}, \quad (\text{B24})$$

$$\beta := \frac{\langle\langle \bar{E}_{\text{id}} | \bar{E} \rangle\rangle}{\langle\langle \bar{E}_{\text{id}} | \bar{E}_{\text{id}} \rangle\rangle} = \frac{1}{d^2 - 1} \langle\langle \bar{E}_{\text{id}} | \bar{E} \rangle\rangle, \quad \bar{E}_{\text{err}} := \bar{E} - \beta \bar{E}_{\text{id}}. \quad (\text{B25})$$

This decomposition is chosen such that  $\text{Tr}[\bar{\rho}_{\text{id}} \bar{\rho}_{\text{err}}] = \text{Tr}[\bar{E}_{\text{id}} \bar{E}_{\text{err}}] = 0$ . It can be shown that the ideal operators  $\bar{\rho}_{\text{id}}, \bar{E}_{\text{id}}$  are in fact ideal, in the sense that they maximize the prefactor  $B$  in the fit model  $\mathbb{E}[q_{\mathbf{j}}] = B u^{m-1}$  (and also minimize the variance as we will see). The prefactor  $B$  is given by [see Eq. (53)]

$$B = \langle\langle E | \mathcal{G}_{\text{avg}}^{(2)} | \bar{\rho} \rangle\rangle = \langle\langle \bar{E} | B_2 \rangle\rangle \langle\langle B_2 | \bar{\rho} \rangle\rangle = \alpha \beta. \quad (\text{B26})$$

The ideal operators  $\bar{\rho}_{\text{id}}, \bar{E}_{\text{id}}$  will yield  $B = 1$ . The following lemma shows that this is in fact optimal.

*Lemma 11 (Optimality of ideal operators).* The prefactor  $B$  in the fit model for URB as given in Eq. (B26) satisfies  $|B| \leq 1$  for all input and measurement operators  $\bar{\rho}, E$ .

*Proof.* Let us write the two-valued measurement  $E$  with outcomes  $\pm 1$  in terms of its POVM elements  $\{M, I - M\}$ , so that  $E = M - (I - M) = 2M - I$ . By definition  $M$  satisfies  $0 \leq M \leq I$ . Since  $\mathcal{G}_{\text{avg}}^{(2)}$  is a CPTP map and  $\rho, \hat{\rho} \geq 0$  are quantum states, it follows that  $\mathcal{G}_{\text{avg}}^{(2)}(\rho), \mathcal{G}_{\text{avg}}^{(2)}(\hat{\rho}) \geq 0$ . Using the fact that  $\text{Tr}[AB] \geq 0$  for all positive semidefinite operators  $A, B \geq 0$ , it follows that

$$0 = \langle\langle 0 | \mathcal{G}_{\text{avg}}^{(2)} | \rho \rangle\rangle \leq \langle\langle M | \mathcal{G}_{\text{avg}}^{(2)} | \rho \rangle\rangle \leq \langle\langle I | \mathcal{G}_{\text{avg}}^{(2)} | \rho \rangle\rangle = 1. \quad (\text{B27})$$

In terms of the measurement  $E$ , this means that  $-1 \leq \langle\langle E | \mathcal{G}_{\text{avg}}^{(2)} | \rho \rangle\rangle \leq 1$ . Analogously, this holds for  $\hat{\rho}$ . Since  $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$  it follows that  $-1 \leq B = \langle\langle E | \mathcal{G}_{\text{avg}}^{(2)} | \bar{\rho} \rangle\rangle \leq 1$ . ■

*Corollary 4.* The quantities  $\alpha, \beta$  as defined in Eq. (B24) and Eq. (B25) satisfy  $-1 \leq \alpha, \beta \leq 1$ .

*Proof.* Lemma 11 and Eq. (B26) show that  $-1 \leq \alpha\beta \leq 1$  for all  $\bar{\rho}, E$ . Note that  $\alpha$  only depends on  $\bar{\rho}$  and  $\beta$  only on  $E$ . Therefore if we fix  $\bar{\rho} = \bar{\rho}_{\text{id}}$  (which implies  $\alpha = 1$ ), then we have  $-1 \leq \beta \leq 1$ . Analogously fixing  $E = E_{\text{id}}$  (which implies  $\beta = 1$ ) yields  $-1 \leq \alpha \leq 1$ . ■

Very similar reasoning also gives the bound on the interval in which the sequence purity  $q_{\mathbf{j}}^{(K)}$  lies [see Eq. (27)]. This bound will be proven in the following lemma.

*Lemma 12 (Bound on interval lengths).* Let  $q_{\mathbf{j}}^{(K)}$  denote the sequence purity of the  $K$ -copy implementation due to the random sequence  $\mathbf{j}$  as defined in Eq. (11) and Eq. (9):

$$q_{\mathbf{j}}^{(1)} = \frac{1}{d^2 - 1} \sum_{P, Q \neq I} \langle \langle E_{\mathcal{H}}^{(Q)} | \mathcal{G}_{\mathbf{j}} | \bar{\rho}_{\mathcal{H}}^{(P)} \rangle \rangle^2 \quad \text{and} \\ q_{\mathbf{j}}^{(2)} = \langle \langle E | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | \bar{\rho} \rangle \rangle. \quad (\text{B28})$$

Assume that  $\alpha, \beta \geq 0$  (equivalent to  $\text{Tr}[\bar{\rho}_{\text{id}} \bar{\rho}] \geq 0$  and  $\text{Tr}[\bar{E}_{\text{id}} \bar{E}] \geq 0$  stated in Sec. II C). Then for all operators  $\bar{\rho}, E$  [which are the effective operators in the single-copy implementation, see Eq. (13)], all CPTP error maps  $\Lambda$  and all sequences of Clifford gates indexed by  $\mathbf{j}$ ,

$$q_{\mathbf{j}}^{(1)} \in [0, \alpha\beta + \beta \|\bar{\rho}_{\text{err}}\|_1 + \alpha \|\bar{E}_{\text{err}}\|_{\infty} + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_{\infty}], \quad (\text{B29})$$

$$q_{\mathbf{j}}^{(2)} \in [-\beta \|\bar{\rho}_{\text{err}}\|_1 - \alpha \|\bar{E}_{\text{err}}\|_{\infty} - \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_{\infty}, 1]. \quad (\text{B30})$$

*Corollary 5.* The interval length for  $q_{\mathbf{j}}^{(1)}$  and  $q_{\mathbf{j}}^{(2)}$  can be bounded independent of  $\alpha, \beta$  by using that  $\alpha, \beta \leq 1$  (Lemma 11) as  $L = 1 + \|\bar{\rho}_{\text{err}}\|_1 + \|\bar{E}_{\text{err}}\|_{\infty} + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_{\infty}$ .

*Proof.* Starting with the two-copy implementation, let us write  $E = M - (I - M) = 2M - I$ , where  $0 \leq M \leq I$  is a POVM element (the measurement  $E$  is described by the POVM set  $\{M, I - M\}$ , assigning outcome 1 to  $M$  and  $-1$  to  $I - M$ ). Then using the fact that  $\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho) \geq 0$  is positive semidefinite, it follows that

$$0 = \text{Tr}[0\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq \text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq \text{Tr}[I\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] = 1,$$

expressing that  $\text{Tr}[M\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)]$  is indeed the probability associated with obtaining outcome  $M$ . Therefore  $-1 \leq \text{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\rho)] \leq 1$ . Exactly the same argument holds for  $\hat{\rho}$ , so that [recall that  $\bar{\rho} = \frac{1}{2}(\rho - \hat{\rho})$ ]

$$-1 \leq q_{\mathbf{j}}^{(2)} = \text{Tr}[E\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho})] \leq 1. \quad (\text{B31})$$

The lower bound can be improved by using the decomposition Eq. (B24) and Eq. (B25) to write  $\bar{\rho} = \alpha\bar{\rho}_{\text{id}} + \bar{\rho}_{\text{err}}$  and  $\bar{E} = \beta\bar{E}_{\text{id}} + \bar{E}_{\text{err}}$ . Then

$$q_{\mathbf{j}}^{(2)} = \alpha\beta \text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \alpha \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] \\ + \beta \text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] + \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]. \quad (\text{B32})$$

The first term satisfies  $\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] \leq 1$  by Eq. (B31) (which holds for all  $E, \bar{\rho}$  so in particular for  $E_{\text{id}}, \bar{\rho}_{\text{id}}$ ). However, we also find that

$$\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] = \text{Tr}[B_2\mathcal{G}_{\mathbf{j}}^{\otimes 2}(B_2)] \\ = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathcal{P}^*} \text{Tr}[\sigma\mathcal{G}_{\mathbf{j}}(\tau)]^2 \geq 0. \quad (\text{B33})$$

The remaining three terms in Eq. (B32) are bounded using Proposition 19, which yields (using  $\alpha, \beta \geq 0$ )

$$\alpha |\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})]| \leq \alpha \|\bar{E}_{\text{err}}\|_{\infty} \|\bar{\rho}_{\text{id}}\| = \alpha \|\bar{E}_{\text{err}}\|_{\infty}, \\ \beta |\text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]| \leq \beta \|\bar{E}_{\text{id}}\|_{\infty} \|\bar{\rho}_{\text{err}}\|_1 = \beta \|\bar{\rho}_{\text{err}}\|_1, \\ |\text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})]| \leq \|\bar{E}_{\text{err}}\|_{\infty} \|\bar{\rho}_{\text{err}}\|_1. \quad (\text{B34})$$

So by combining Eqs. (B32), (B33), and (B34), we find that

$$q_{\mathbf{j}}^{(2)} \geq 0 - \alpha \|\bar{E}_{\text{err}}\|_{\infty} - \beta \|\bar{\rho}_{\text{err}}\|_1 - \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_{\infty}. \quad (\text{B35})$$

The above argument also holds in the single-copy implementation if we let  $E = E_{\text{eff}}$  and  $\bar{\rho} = \bar{\rho}_{\text{eff}}$  as defined in Eq. (13). However, now we use it to upper bound  $q_{\mathbf{j}}^{(1)}$ . It follows that

$$q_{\mathbf{j}}^{(1)} = \alpha\beta \text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] + \alpha \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{id}})] \\ + \beta \text{Tr}[\bar{E}_{\text{id}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] + \text{Tr}[\bar{E}_{\text{err}}\mathcal{G}_{\mathbf{j}}^{\otimes 2}(\bar{\rho}_{\text{err}})] \\ \leq \alpha\beta + \beta \|\bar{\rho}_{\text{err}}\|_1 + \alpha \|\bar{E}_{\text{err}}\|_{\infty} + \|\bar{\rho}_{\text{err}}\|_1 \|\bar{E}_{\text{err}}\|_{\infty}. \quad (\text{B36})$$

The lower bound  $q_{\mathbf{j}}^{(1)} \geq 0$  follows directly from the fact that it is defined as the sum of real numbers squared. ■

So far we have recaptured the essential definitions and notations, shown optimality of the ideal operators and proven a bound in the interval in which the sequence purity  $q_{\mathbf{j}}$  lies. Next we will state our variance bound Eq. (19) and give the complete proof.

*Theorem 1 (Variance bound).* Let  $\mathcal{H}$  be a  $d$ -dimensional Hilbert space, with  $d = 2^q$  for a  $q$ -qubit system. Let  $E \in V = \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  be the Hermitian observable associated with a two-valued measurement with outcomes  $\pm 1$  and  $\rho, \hat{\rho} \in V = \mathcal{L}(\mathcal{H} \otimes \mathcal{H})$  be two quantum states on two copies of the system. Consider the URB experiment (using the states and measurement  $\rho, \hat{\rho}, E$ ) of the Clifford group  $\mathcal{C}(d)$ , assuming that a noisy implementation of  $\mathcal{G} \in \mathcal{C}(d)$  is given by  $\bar{\mathcal{G}} = \mathcal{G}\Lambda$ , where  $\Lambda$  is a CPTP map. In this experiment the sequence purity is  $q_{\mathbf{j}} = \langle \langle \bar{E} | (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho} \rangle \rangle$ , with  $\mathcal{M}$  defined in Eq. (B14).

Under the assumption that  $d = 2$  or  $\Lambda$  is unital (that is,  $\Lambda(I) = I$ ), the following bound on the variance  $\mathbb{V}[q_{\mathbf{j}}]$  holds:

$$\mathbb{V}[q_{\mathbf{j}}] \leq \sigma^2 = \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 [\alpha^2 \beta^2 c_1(d) \\ + \alpha^2 c_2(d) \|\bar{E}_{\text{err}}\|_{\infty}^2 + \beta^2 c_3(d) \|\bar{\rho}_{\text{err}}\|_1^2] \\ + \|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad (\text{B37})$$

where  $u$  is the unitarity of  $\Lambda$ ,  $m$  is the length of the sequence indexed by  $\mathbf{j}$ ,  $c_i(d)$  are functions only of the dimension  $d$  and  $\alpha, \beta, \bar{\rho}_{\text{err}}$ , and  $\bar{E}_{\text{err}}$  are defined in Eqs. (B24) and (B25). Precise definitions of the dimension-dependent functions  $c_i(d)$  will be given in the proof, but closed form expressions are

messy and therefore not written down explicitly. Asymptotically, these functions satisfy

$$c_1(d) = O(1), \quad c_2(d) = O(d), \quad c_3(d) = O(d^2). \quad (\text{B38})$$

*Proof.* We start from the derived expression for the variance Eq. (B13). First, let us decompose the state and measurement operators in ideal and error components as [see Eqs. (B23)–(B25)]

$$\bar{\rho} = \alpha \bar{\rho}_{\text{id}} + \bar{\rho}_{\text{err}} \quad \text{and} \quad \bar{E} = \beta \bar{E}_{\text{id}} + \bar{E}_{\text{err}}. \quad (\text{B39})$$

Define again  $W = \text{Span}\{B_1, B_2\} \subset V$ , with  $B_1, B_2$  defined in Eqs. (B15) and (B16), respectively. Then the ideal components  $\bar{\rho}_{\text{id}}$  and  $\bar{E}_{\text{id}}$  are in  $W$  and the error components  $\bar{\rho}_{\text{err}}$  and  $\bar{E}_{\text{err}}$  are in the orthogonal complement  $W^\perp$ . Plugging this expansion into Eq. (B13) in principle yields 16 terms.

However, the 12 terms with an ideal component tensor error component (e.g.,  $\bar{\rho}_{\text{err}} \otimes \bar{\rho}_{\text{id}}$ ) vanish, because both

$$(\mathcal{G}_{\text{avg}}^{(2)})^{\otimes 2}(W \otimes W^\perp) = (\mathcal{G}_{\text{avg}}^{(2)})^{\otimes 2}(W^\perp \otimes W) = \emptyset, \quad (\text{B40})$$

$$\mathcal{G}_{\text{avg}}^{(4)}(W \otimes W^\perp) = \mathcal{G}_{\text{avg}}^{(4)}(W^\perp \otimes W) = \emptyset. \quad (\text{B41})$$

Eq. (B40) is easy to see because  $\mathcal{G}_{\text{avg}}^{(2)}$  is the orthogonal projection onto  $W$ . Equation Eq. (B41) follows from the fact that  $W$  carries the trivial subrepresentations of the Liouville tensor-2 representation and  $W^\perp$  carries all other necessarily nontrivial subrepresentations. By Lemma 3 the spaces  $W^\perp \otimes W$  and  $W \otimes W^\perp$  (which are representations of the Liouville tensor-4 representation) do not carry trivial subrepresentations. Hence  $\mathcal{G}_{\text{avg}}^{(4)}$ , the projector onto the trivial subrepresentations of the Liouville tensor-4 rep (by Lemma 2), does not project onto any subspace of  $W^\perp \otimes W$  and  $W \otimes W^\perp$ . This justifies the following expression for the variance:

$$\mathbb{V}[q_j] = \alpha^2 \beta^2 \langle\langle B_2^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | B_2^{\otimes 2} \rangle\rangle \quad (\text{B42})$$

$$+ \frac{\alpha^2}{d^2 - 1} \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | B_2^{\otimes 2} \rangle\rangle \quad (\text{B43})$$

$$+ (d^2 - 1) \beta^2 \langle\langle B_2^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \quad (\text{B44})$$

$$+ \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} - (\mathcal{M}^{\otimes 2})^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle, \quad (\text{B45})$$

where the expressions of Eq. (B23) are used for the ideal operators  $\bar{\rho}_{\text{id}}, \bar{E}_{\text{id}}$ . We will analyze each of the four terms separately. The term we start with is Eq. (B43), since this term most clearly conveys the idea of our analysis. Then the terms Eqs. (B42) and (B44) are treated in similar fashion, but with a small additional technicality. Finally the term Eq. (B45) is treated in a totally different fashion.

The analysis of Eq. (B43) starts by using Lemma 4 (telescoping series lemma), so that we can write this term as

$$\begin{aligned} (\text{B43}) &= \frac{\alpha^2}{d^2 - 1} \sum_{s=1}^{m-1} \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] (\mathcal{M}^{\otimes 2})^{s-1} | B_2^{\otimes 2} \rangle\rangle \\ &= \frac{\alpha^2}{d^2 - 1} \sum_{s=1}^{m-1} u^{2(s-1)} \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] | B_2^{\otimes 2} \rangle\rangle. \end{aligned} \quad (\text{B46})$$

In the second line we used that  $\mathcal{M}|B_2\rangle\rangle = u|B_2\rangle\rangle$ . The idea is to expand  $\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2^{\otimes 2}\rangle\rangle$  in the basis  $\{A_i : i \in \mathcal{Z}_{TS}\}$  of the subspace  $V_{TS} \otimes V_{TS} \subset \text{Rge}(\mathcal{G}_{\text{avg}}^{(4)}) \subset V \otimes V$ .  $V_{TS}$  is the trace-preserving, symmetric subspace of  $V = \mathbb{L}(\mathcal{H}^{\otimes 2})$ , as defined in Appendix B 1. The restriction of  $\mathcal{G}_{\text{avg}}^{(4)}$  to  $V_{TS} \otimes V_{TS}$  is justified by the fact that  $\Lambda^{\otimes 2}(B_2) \in V_{TS}$ . Hence we expand

$$\mathcal{N} - \mathcal{M}^{\otimes 2}|B_2^{\otimes 2}\rangle\rangle = \sum_{i \in \mathcal{Z}_{TS}} a_i |A_i\rangle\rangle, \quad \text{where} \quad a_i := \langle\langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2^{\otimes 2} \rangle\rangle. \quad (\text{B47})$$

Therefore Eq. (B43) can be written as

$$(\text{B43}) = \frac{\alpha^2}{d^2 - 1} \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} a_i \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\rangle. \quad (\text{B48})$$

For the terms Eqs. (B42) and (B44), something similar is done. The telescoping series (Lemma 4) is now written in the other way. Therefore we can write Eq. (B42) as

$$(\text{B42}) = \alpha^2 \beta^2 \sum_{s=1}^{m-1} \langle\langle B_2^{\otimes 2} | (\mathcal{M}^{\otimes 2})^{s-1} [\mathcal{N} - \mathcal{M}^{\otimes 2}] \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle \quad (\text{B49})$$

$$= \alpha^2 \beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \langle\langle B_2^{\otimes 2} | [\mathcal{N} - \mathcal{M}^{\otimes 2}] \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle. \quad (\text{B50})$$

The step from Eqs. (B49) to (B50) is not immediately clear, since

$$\langle\langle B_2 B_2 | (\mathcal{M}^{\otimes 2})^{s-1} = x_{11}^{(s)} \langle\langle B_1 B_1 | + x_{12}^{(s)} \langle\langle B_1 B_2 | + x_{21}^{(s)} \langle\langle B_2 B_1 | + u^{2(s-1)} \langle\langle B_2 B_2 |, \quad (\text{B51})$$

for some coefficients  $x_{11}^{(s)}, x_{12}^{(s)}, x_{21}^{(s)} \in \mathbb{R}$ . However we show that Eq. (B50) is justified, since

$$\langle\langle B_k B_l | [\mathcal{N} - \mathcal{M}^{\otimes 2}] \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle = 0, \quad \text{if } k = 1 \text{ or } l = 1. \quad (\text{B52})$$

This follows from the trace-preserving properties of  $\mathcal{N}$ ,  $\mathcal{M}$ , the tracelessness of  $B_2$  and the fact that  $B_1 = \frac{1}{d}$ . In particular,

$$\langle\langle B_k B_l | \mathcal{N}^{m-s} | B_2^{\otimes 2} \rangle\rangle = \frac{1}{|\mathcal{C}(d)|^{m-s}} \sum_{\mathbf{j}} \langle\langle B_k | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle \langle\langle B_l | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle = 0, \quad (\text{B53})$$

$$\langle\langle B_k B_l | \mathcal{M}^{\otimes 2} \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle = \frac{1}{|\mathcal{C}(d)|^{m-s-1}} \sum_{\mathbf{j}} \langle\langle B_k | \mathcal{M} \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle \langle\langle B_l | \mathcal{M} \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle = 0, \quad (\text{B54})$$

if  $l = 1$  or  $k = 1$ , since  $\langle\langle B_1 | \mathcal{M} \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle = 0$  and  $\langle\langle B_1 | \mathcal{G}_{\mathbf{j}}^{\otimes 2} | B_2 \rangle\rangle = 0$ . This justifies Eq. (B50). Next we use a similar expansion

$$\langle\langle B_2^{\otimes 2} | \mathcal{N} - \mathcal{M}^{\otimes 2} = \sum_{i \in \mathcal{Z}_{TS}} b_i \langle\langle A_i |, \quad \text{where } b_i := \langle\langle B_2^{\otimes 2} | \mathcal{N} - \mathcal{M}^{\otimes 2} | A_i \rangle\rangle. \quad (\text{B55})$$

Therefore we arrive at

$$(\text{B42}) = \alpha^2 \beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} b_i \langle\langle A_i | \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle. \quad (\text{B56})$$

Similarly to the analysis Eq. (B42), we can write Eq. (B44) as

$$(\text{B44}) = (d^2 - 1) \beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} b_i \langle\langle A_i | \mathcal{N}^{m-s-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle. \quad (\text{B57})$$

Finally, we slightly rewrite Eq. (B45) by noting that  $(\text{B45}) = \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle$ , because  $\mathcal{M} | \rho_{\text{err}} \rangle\rangle = 0$ . We therefore arrive at the following expression of the variance:

$$\mathbb{V}[q_{\mathbf{j}}] = \alpha^2 \beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} a_i \langle\langle B_2^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\rangle \quad (\text{B58})$$

$$+ \frac{1}{d^2 - 1} \alpha^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} a_i \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\rangle \quad (\text{B59})$$

$$+ (d^2 - 1) \beta^2 \sum_{s=1}^{m-1} u^{2(s-1)} \sum_{i \in \mathcal{Z}_{TS}} b_i \langle\langle A_i | \mathcal{N}^{m-s-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \quad (\text{B60})$$

$$+ \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle. \quad (\text{B61})$$

This expression is still exact, as we have only expanded each term in the equation.

The variance bound is obtained by bounding the remaining inner products and the quantities  $a_i, b_i$  in this expression. This technical task is delegated to Appendix C, with a number of technical propositions that compute bounds on the quantities above. We summarize the results here. The bounds on  $a_i$  and  $b_i$  for  $i \in \{0; [\text{adj}]; S; 1, 2\}$  are obtained under the assumption

that  $d = 2$  or that  $\Lambda$  is unital in Propositions 13–18. In summary

$$0 = a_0 = b_0, \quad (\text{B62})$$

$$0 \leq a_{1,2}, b_{1,2} \leq \frac{\sqrt{d^2 - 2}}{d^2} (1 - u)^2, \quad (\text{B63})$$

$$0 \leq a_S, b_S \leq \sqrt{\frac{d^2 - 2}{d^2 - 1}} \sqrt{2} (1 - u)^2, \quad (\text{B64})$$

$$0 \leq a_{[\text{adj}]}, b_{[\text{adj}]} \leq \sqrt{d^2 - 1} (1 - u)^2. \quad (\text{B65})$$

In the case of  $d \geq 4$ , bounds on  $a_i$  are needed for  $i \in \mathcal{Z}_{1,2} \cup \mathcal{Z}_S \setminus \{[\text{adj}]\}$  in terms of the above bounds on  $a_S$  and  $a_{1,2}$ . To do so, we use Eq. (B7), which states

$$\sqrt{|V_{1,2}|} a_{1,2} = \sum_{i \in \mathcal{Z}_{1,2}} \sqrt{|V_i|} a_i, \quad \sqrt{|V_S|} a_S = \sum_{i \in \mathcal{Z}_S} \sqrt{|V_i|} a_i. \quad (\text{B66})$$

From this it follows that

$$\sqrt{|V_{1,2}|} a_{1,2} = \sum_{i \in \mathcal{Z}_{1,2}} \sqrt{|V_i|} a_i, \quad \sqrt{|V_S|} a_S = \sum_{i \in \mathcal{Z}_S} \sqrt{|V_i|} a_i, \quad (\text{B67})$$

$$\sqrt{|V_{1,2}|} b_{1,2} = \sum_{i \in \mathcal{Z}_{1,2}} \sqrt{|V_i|} b_i, \quad \sqrt{|V_S|} b_S = \sum_{i \in \mathcal{Z}_S} \sqrt{|V_i|} b_i. \quad (\text{B68})$$

Thus, since  $a_i, b_i \geq 0$  by Proposition 13, these equations imply the following bounds:

$$a_i \leq \sqrt{\frac{|V_{1,2}|}{|V_i|}} a_{1,2}, \quad b_i \leq \sqrt{\frac{|V_{1,2}|}{|V_i|}} b_{1,2}, \quad \forall i \in \mathcal{Z}_{1,2} \quad (\text{B69})$$

$$a_i \leq \sqrt{\frac{|V_S|}{|V_i|}} a_S, \quad b_i \leq \sqrt{\frac{|V_S|}{|V_i|}} b_S, \quad \forall i \in \mathcal{Z}_S \setminus \{\{\text{adj}\}\}. \quad (\text{B70})$$

The size of the relevant spaces (as derived in [36]) was summarized in Table II. The inner products in Eqs. (B58)–(B61) are bounded using Propositions 19–21. Proposition 19 is applicable since  $\mathcal{N}^m$  is a CPTP map for any  $m \in \mathbb{N}$ , since CPTP maps are closed under composition. Now  $\mathcal{N}$  is CPTP because  $\mathcal{N}$  is the convex combination of the CPTP sequences  $\mathcal{G}_j$  and a convex combination of CPTP maps is CPTP. The results of Propositions 19, 20, and 21 are summarized as follows:

$$\langle\langle A_i | \mathcal{N}^{m-s-1} | B_2^{\otimes 2} \rangle\rangle \leq \frac{1}{\sqrt{|V_i|}}, \quad (\text{B71})$$

$$\langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_i \rangle\rangle \leq d^2 \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad (\text{B72})$$

$$\langle\langle A_i | \mathcal{N}^{m-s-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \leq \sqrt{\frac{6}{(d-2)(d-1)}} \|\bar{\rho}_{\text{err}}\|_1^2, \quad (\text{B73})$$

$$\langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \leq \|\bar{E}_{\text{err}}\|_{\infty}^2 \|\bar{\rho}_{\text{err}}\|_1^2, \quad (\text{B74})$$

where we have used that  $\|A^{\otimes k}\|_p = \|A\|_p^k$  for any  $k \in \mathbb{N}$  and  $p \in [1, \infty]$ . Equations (B72) and (B73) have single-qubit specific ( $d = 2$ ) improvements (derived in Proposition 20), using the fact that  $V_{1,2}$  and  $V_S$  actually are irreducible subrepresentations. Since we have explicit expressions for  $A_{1,2}$  and  $A_S$  [Eqs. (B9) and (B10), respectively], their norms can be computed directly. Using this gives the improved single-qubit bounds,

$$\langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_S \rangle\rangle \leq \frac{5}{\sqrt{3}} \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad \langle\langle \bar{E}_{\text{err}}^{\otimes 2} | \mathcal{N}^{m-s-1} | A_{1,2} \rangle\rangle \leq 2\sqrt{2} \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad (\text{B75})$$

$$\langle\langle A_S | \mathcal{N}^{m-s-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \leq \frac{1}{\sqrt{3}} \|\bar{\rho}_{\text{err}}\|_1^2, \quad \langle\langle A_{1,2} | \mathcal{N}^{m-s-1} | \bar{\rho}_{\text{err}}^{\otimes 2} \rangle\rangle \leq \frac{\sqrt{2}}{3} \|\bar{\rho}_{\text{err}}\|_1^2. \quad (\text{B76})$$

Plugging all of these bounds into Eqs. (B58)–(B61) and using the geometric series

$$\sum_{s=1}^{m-1} u^{2(s-1)} = \frac{1 - u^{2(m-1)}}{1 - u^2} \quad (\text{B77})$$

will yield the bound Eq. (B37)

$$\mathbb{V}[q_j] \leq \sigma^2 = \frac{1 - u^{2(m-1)}}{1 - u^2} (1 - u)^2 [\alpha^2 \beta^2 c_1(d) + \alpha^2 c_2(d) \|\bar{E}_{\text{err}}\|_{\infty}^2 + \beta^2 c_3(d) \|\bar{\rho}_{\text{err}}\|_1^2] + \|\bar{\rho}_{\text{err}}\|_1^2 \|\bar{E}_{\text{err}}\|_{\infty}^2, \quad (\text{B78})$$

where

$$c_1(d) = \begin{cases} \frac{\sqrt{2}}{4} \frac{1}{\sqrt{2}} + \sqrt{\frac{2}{3}} \sqrt{2} \frac{1}{\sqrt{3}} = \frac{11}{12}, & \text{if } d = 2, \\ \frac{\sqrt{d^2-2}}{d^2} \sum_{i \in \mathcal{Z}_{1,2}} \frac{\sqrt{|V_{1,2}|}}{|V_i|} + \sqrt{2} \sqrt{\frac{d^2-2}{d^2-1}} \sum_{i \in \mathcal{Z}_S \setminus \{\{\text{adj}\}\}} \frac{\sqrt{|V_S|}}{|V_i|} + \frac{\sqrt{d^2-1}}{\sqrt{|V_{\{\text{adj}\}}|}}, & \text{if } d \geq 4, \end{cases}$$

$$c_2(d) = \begin{cases} \frac{1}{3} \left( \frac{\sqrt{2}}{4} 2\sqrt{2} + \sqrt{\frac{2}{3}} \sqrt{2} \frac{5}{\sqrt{3}} \right) = \frac{13}{9}, & \text{if } d = 2, \\ \frac{d^2}{d^2-1} \left( \frac{\sqrt{d^2-2}}{d^2} \sum_{i \in \mathcal{Z}_{1,2}} \frac{\sqrt{|V_{1,2}|}}{\sqrt{|V_i|}} + \sqrt{2} \sqrt{\frac{d^2-2}{d^2-1}} \sum_{i \in \mathcal{Z}_S \setminus \{\{\text{adj}\}\}} \frac{\sqrt{|V_S|}}{\sqrt{|V_i|}} + \sqrt{d^2-1} \right), & \text{if } d \geq 4, \end{cases}$$

$$c_3(d) = \begin{cases} 3 \left( \frac{\sqrt{2}}{4} \frac{\sqrt{2}}{3} + \sqrt{\frac{2}{3}} \sqrt{2} \frac{1}{\sqrt{3}} \right) = \frac{5}{2}, & \text{if } d = 2, \\ (d^2 - 1) \sqrt{\frac{6}{(d-2)(d-1)}} \left( \frac{\sqrt{d^2-2}}{d^2} \sum_{i \in \mathcal{Z}_{1,2}} \frac{\sqrt{|V_{1,2}|}}{\sqrt{|V_i|}} + \sqrt{2} \sqrt{\frac{d^2-2}{d^2-1}} \sum_{i \in \mathcal{Z}_S \setminus \{\{\text{adj}\}\}} \frac{\sqrt{|V_S|}}{\sqrt{|V_i|}} + \sqrt{d^2-1} \right), & \text{if } d \geq 4. \end{cases}$$

The size of the spaces  $V_i$  in these equations are found in Table II. The asymptotic behavior of the dimension-dependent functions  $c_i(d)$  can be found if all relevant dimensions of the spaces are plugged into the above equations.  $\blacksquare$

## APPENDIX C: BOUNDS ON INDIVIDUAL QUANTITIES IN THE PROOF

This section provides the technical lemmas and propositions referred to in the previous section. They are collected here together in an attempt not to clutter the main line of the proof. Most of these technical lemmas put a bound on quantities arising in the proof of Theorem 1.

We start by bounding the  $a_i$ . Only bounds on  $a_0$ ,  $a_S$ ,  $a_{1,2}$ , and  $a_{[\text{adj}]}$  are provided. In the multiqubit case where  $V_S$  and  $V_{1,2}$  are not irreducible representations, the quantities  $a_i$  for  $i \in \mathcal{Z}_S, \mathcal{Z}_d$  are bounded by  $a_S$  and  $a_{1,2}$ . The only exception is  $i = [\text{adj}]$ , for which we provide a separate bound. Let us start with showing that all  $a_i$  and  $b_i$  are nonnegative.

*Proposition 13 (Lower bound on  $a_i$  and  $b_i$ ).* For all CPTP  $\Lambda$  and all  $i \in \mathcal{Z}_{TS}$ , one has

$$a_i = \langle\langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle \geq 0, \quad b_i = \langle\langle B_2 B_2 | \mathcal{N} - \mathcal{M}^{\otimes 2} | A_i \rangle\rangle \geq 0. \quad (\text{C1})$$

*Proof.* If  $i = 0$ , then Proposition 14 will show that  $a_0 = 0$ , which includes this lower bound. For all other  $i \in \mathcal{Z}_{TS} \setminus \{0\}$ , we show that  $\mathcal{M}^{\otimes 2} | A_i \rangle = 0$ . This is because  $\mathcal{M}^{\otimes 2}$  is supported on  $W^{\otimes 2} = \text{Span}\{B_1 B_1, B_1 B_2, B_2 B_1, B_2 B_2\}$ , where  $A_0 = B_2 B_2$ . But  $B_1 B_1, B_1 B_2, B_2 B_1 \in (V_{TS}^{\otimes 2})^\perp$ . Since  $A_i \in V_{TS}^{\otimes 2} \setminus \text{Span}\{A_0\}$  the claim follows. Therefore  $a_i = \langle\langle B_2 B_2 | \mathcal{N} | A_i \rangle\rangle$ . Using the definitions of  $\mathcal{N}$  Eq. (B14) and  $A_i$  Eq. (B6), it follows that

$$a_i = \frac{1}{|V_i| |\mathbf{C}(d)|^2} \sum_{\mathcal{G}, \mathcal{G}' \in \mathbf{C}(d)} \sum_{v_i \in B_i} \langle\langle v_i v_i | \mathcal{G}^{\otimes 4} \Lambda^{\otimes 4} \mathcal{G}'^{\otimes 4} | B_2 B_2 \rangle\rangle = \frac{1}{|V_i| |\mathbf{C}(d)|^2} \sum_{\mathcal{G}, \mathcal{G}' \in \mathbf{C}(d)} \sum_{v_i \in B_i} \langle\langle v_i | \mathcal{G}^{\otimes 2} \Lambda^{\otimes 2} \mathcal{G}'^{\otimes 2} | B_2 \rangle\rangle^2 \geq 0, \quad (\text{C2})$$

which is nonnegative as it is the sum of real numbers squared. Analogously,

$$b_i = \frac{1}{|V_i| |\mathbf{C}(d)|^2} \sum_{\mathcal{G}, \mathcal{G}' \in \mathbf{C}(d)} \sum_{v_i \in B_i} \langle\langle B_2 | \mathcal{G}^{\otimes 2} \Lambda^{\otimes 2} \mathcal{G}'^{\otimes 2} | v_i \rangle\rangle^2 \geq 0. \quad \blacksquare$$

Next we show that  $a_0$  vanishes.

*Proposition 14 (Bound on  $a_0$ ).* Let  $a_0$  be defined by Eq. (B47). Then for all CPTP quantum channels  $\Lambda$ ,  $a_0 = 0$ .

*Proof.* By definition of Eq. (B47) it follows that [using that  $A_0 = B_2 B_2$  by definition of Eq. (B8)]

$$a_0 = \langle\langle A_0 | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle = \langle\langle B_2 B_2 | \Lambda^{\otimes 4} - \Lambda^{\otimes 4} | B_2 B_2 \rangle\rangle = 0, \quad (\text{C3})$$

since  $\mathcal{G}_{\text{avg}}^{(4)} | B_2 B_2 \rangle = (\mathcal{G}_{\text{avg}}^{(2)})^{\otimes 2} | B_2 B_2 \rangle = | B_2 B_2 \rangle$ . \blacksquare

The next proposition gives a bound on  $a_{1,2}$ .

*Proposition 15 (Bound on  $a_{1,2}$ ).* Let  $a_{1,2}$  be defined as in Eq. (B47) and let  $\Lambda$  be a CPTP map. If  $\Lambda$  is a single-qubit channel (i.e., if  $d = 2$ ) or if  $\Lambda$  is unital [i.e.,  $\Lambda(I) = I$ ], then

$$a_{1,2} = \frac{1}{\sqrt{d^2 - 2}} \left( \frac{1}{d^2 - 1} \sum_{\sigma \in \mathbf{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 - u^2 \right) \leq \frac{\sqrt{d^2 - 2}}{d^2} (1 - u)^2. \quad (\text{C4})$$

*Proof.* By the definition Eq. (B47),  $a_{1,2} = \langle\langle A_{1,2} | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle$ , where

$$B_2 B_2 = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathbf{P}^*} \sigma \sigma \tau \tau \quad \text{and} \quad A_{1,2} = \frac{1}{\sqrt{d^2 - 2}} \left( \sum_{\sigma \in \mathbf{P}^*} \sigma^{\otimes 4} - A_0 \right) \quad (\text{C5})$$

were defined in Eqs. (B8) and (B9), respectively. Therefore  $a_{1,2}$  is computed as [recalling that  $A_0 = B_2 B_2$  and using Eq. (B18)]

$$\begin{aligned} a_{1,2} &= \langle\langle A_{1,2} | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle \\ &= \frac{1}{(d^2 - 1) \sqrt{d^2 - 2}} \left( \sum_{\sigma, \hat{\sigma}, \hat{\tau} \in \mathbf{P}^*} \langle\langle \sigma \sigma \sigma \sigma | \Lambda^{\otimes 4} | \hat{\sigma} \hat{\sigma} \hat{\tau} \hat{\tau} \rangle\rangle - \langle\langle B_2 B_2 | \Lambda^{\otimes 4} | B_2 B_2 \rangle\rangle \right) \\ &= \frac{1}{\sqrt{d^2 - 2}} \left( \frac{1}{d^2 - 1} \sum_{\sigma, \hat{\sigma}, \hat{\tau} \in \mathbf{P}^*} \langle\langle \sigma | \Lambda | \hat{\sigma} \rangle\rangle^2 \langle\langle \sigma | \Lambda | \hat{\tau} \rangle\rangle^2 - u^2 \right) \\ &= \frac{1}{\sqrt{d^2 - 2}} \left( \frac{1}{d^2 - 1} \sum_{\sigma \in \mathbf{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 - u^2 \right), \end{aligned} \quad (\text{C6})$$

where in the last step, the following was used:

$$\sum_{\hat{\sigma} \in \mathbf{P}^*} \langle\langle \sigma | \Lambda | \hat{\sigma} \rangle\rangle \langle\langle \tau | \Lambda | \hat{\sigma} \rangle\rangle = \sum_{\hat{\sigma} \in \mathbf{P}^*} \langle\langle \sigma | \Lambda | \hat{\sigma} \rangle\rangle \langle\langle \hat{\sigma} | \Lambda^\dagger | \tau \rangle\rangle = \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle, \quad \forall \sigma, \tau \in \mathbf{P}^*, \quad (\text{C7})$$

abusing notation slightly by writing  $\Lambda_u$  instead of  $1 \oplus \Lambda_u$  and using the fact that  $\sum_{\hat{\sigma} \in \mathcal{P}^*} |\hat{\sigma}\rangle\langle\hat{\sigma}|$  is the projection onto the unital block.

The bound of Eq. (C4) is then shown as follows. The idea is to apply Lemma 6 to the map

$$\mathcal{E} := \begin{bmatrix} 1 & 0 \\ 0 & \Lambda_u \Lambda_u^\dagger \end{bmatrix}, \quad (\text{C8})$$

since this map is constructed such that

$$f(\mathcal{E}) = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathcal{E} | \sigma \rangle\rangle = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle = u(\Lambda) \quad (\text{C9})$$

and

$$\frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 - u(\Lambda)^2 = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathcal{E} | \sigma \rangle\rangle^2 - f(\mathcal{E})^2. \quad (\text{C10})$$

Application of Lemma 6 requires the map  $\mathcal{E}$  to be CPTP. This is guaranteed by Lemma 7, using the assumption that  $\Lambda$  is a single-qubit or unital channel. Therefore Lemma 6 applied to the channel  $\mathcal{E}$  defined above, yields

$$\frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 - u(\Lambda)^2 = \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathcal{E} | \sigma \rangle\rangle^2 - f(\mathcal{E})^2 \leq \frac{d^2 - 2}{d^2} [1 - f(\mathcal{E})]^2 = \frac{d^2 - 2}{d^2} [1 - u(\Lambda)]^2. \quad (\text{C11})$$

Plugging this into Eq. (C6) yields the result.  $\blacksquare$

The next proposition bounds the quantity  $a_S$ .

*Proposition 16 (Bound on  $a_S$ ).* Let  $a_S$  be defined as in Eq. (B47) and let  $\Lambda$  be a CPTP map. If  $\Lambda$  is a single-qubit channel (i.e., if  $d = 2$ ) or if  $\Lambda$  is unital (i.e.,  $\Lambda(I) = I$ ), then

$$a_S = \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2 \leq \sqrt{\frac{d^2 - 2}{d^2 - 1}} \sqrt{2} (1 - u)^2. \quad (\text{C12})$$

*Proof.* First, let us show the evaluation of  $a_S$ . By the definition Eq. (B47),  $a_S = \langle\langle A_S | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle$ , where

$$B_2 B_2 = \frac{1}{d^2 - 1} \sum_{\sigma, \tau \in \mathcal{P}^*} \sigma \sigma \tau \tau \quad \text{and} \quad A_S = \sqrt{\frac{1}{2(d^2 - 1)(d^2 - 2)}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \sigma \tau \sigma \tau + \sigma \tau \tau \sigma \quad (\text{C13})$$

were defined in Eqs. (B8) and (B9) respectively. Therefore  $a_S$  is computed as

$$\begin{aligned} a_S &= \langle\langle A_S | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle \\ &= \frac{1}{\sqrt{2}(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau, \hat{\sigma}, \hat{\tau} \in \mathcal{P}^* \\ \sigma \neq \tau}} \langle\langle \sigma \tau \sigma \tau + \sigma \tau \tau \sigma | \Lambda^{\otimes 4} | \hat{\sigma} \hat{\sigma} \hat{\tau} \hat{\tau} \rangle\rangle \\ &= \frac{1}{\sqrt{2}(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau, \hat{\sigma}, \hat{\tau} \in \mathcal{P}^* \\ \sigma \neq \tau}} 2 \langle\langle \sigma | \Lambda | \hat{\sigma} \rangle\rangle \langle\langle \tau | \Lambda | \hat{\sigma} \rangle\rangle \langle\langle \sigma | \Lambda | \hat{\tau} \rangle\rangle \langle\langle \tau | \Lambda | \hat{\tau} \rangle\rangle \\ &= \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2 \\ &= \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2. \end{aligned} \quad (\text{C14})$$

In the fourth step, the trick of Eq. (C7) was again used. In the final step, it is used that  $\langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2$  is the square of off-diagonal matrix elements of  $\Lambda_u \Lambda_u^\dagger$ , so that  $\langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2 = \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2$ .

The bound is derived as follows. Under the stated assumption that  $\Lambda$  is a single-qubit or unital channel, Lemma 7 guarantees that  $\|\Lambda_u \Lambda_u^\dagger\|_{2 \rightarrow 2} \leq 1$ . Here  $\|\cdot\|_{2 \rightarrow 2}$  is the induced Schatten 2-norm [see Eq. (A7)]. Since  $\langle\langle A | B \rangle\rangle = \text{Tr}[A^\dagger B]$  for any  $A, B \in \mathcal{L}(\mathcal{H})$  [and therefore  $\|A\|_2 = \| |A\rangle\rangle \|_2$  for all  $A \in \mathcal{L}(\mathcal{H})$ ], it follows that  $\|\Lambda_u \Lambda_u^\dagger\|_{2 \rightarrow 2} = \|\Lambda_u \Lambda_u^\dagger\|_{2 \rightarrow 2}$ . But the operator norm

(Schatten  $\infty$ -norm) on matrices is just the induced  $2 \rightarrow 2$  norm, so that it can be concluded that  $\|\Lambda_u \Lambda_u^\dagger\|_\infty = \|\Lambda_u \Lambda_u^\dagger\|_{2 \rightarrow 2} = \|\Lambda_u \Lambda_u^\dagger\|_{2 \rightarrow 2} \leq 1$ . Together with the fact that a matrix of the form  $\Lambda_u \Lambda_u^\dagger$  is itself positive semidefinite, this implies that the matrix  $\mathbf{I} - \Lambda_u \Lambda_u^\dagger \geq 0$  is also positive semidefinite as a matrix (not to be confused with being a positive superoperator). Now the key idea is to bound the off-diagonal elements of the symmetric positive semidefinite matrix  $\mathbf{I} - \Lambda_u \Lambda_u^\dagger$  by the diagonal elements using the Sylvester's Criterion for positive semidefinite matrices (Lemma 8). This criterion states that a Hermitian matrix is positive semidefinite if and only if all of its principal minors are nonnegative. Here we use the only if part, since it has been established that  $\mathbf{I} - \Lambda_u \Lambda_u^\dagger$  is positive semidefinite. In particular we use that the positive semidefiniteness of  $\mathbf{I} - \Lambda_u \Lambda_u^\dagger$  implies that all of its second order minors are nonnegative. This means that

$$\langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \langle\langle \tau | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle - \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2 \geq 0, \quad \forall \sigma, \tau \in \mathcal{P}^*, \sigma \neq \tau. \quad (\text{C15})$$

Plugging this into Eq. (C14) yields

$$\begin{aligned} a_S &\leq \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \sum_{\substack{\sigma, \tau \in \mathcal{P}^* \\ \sigma \neq \tau}} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \langle\langle \tau | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle \\ &= \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \left[ \left( \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \right)^2 - \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 \right]. \end{aligned} \quad (\text{C16})$$

The final step is to use that the mean of squares is larger than the square of the mean (Lemma 10). This means in our setting that

$$\frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 \geq \left( \frac{1}{d^2 - 1} \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \right)^2. \quad (\text{C17})$$

Multiplying by  $-(d^2 - 1)$  and plugging into Eq. (C16) yields the bound

$$a_S \leq \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \left( 1 - \frac{1}{d^2 - 1} \right) \left( \sum_{\sigma \in \mathcal{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \right)^2 \quad (\text{C18})$$

$$= \frac{\sqrt{2}}{(d^2 - 1)^{\frac{3}{2}} (d^2 - 2)^{\frac{1}{2}}} \frac{d^2 - 2}{d^2 - 1} ((d^2 - 1)(1 - u))^2 \quad (\text{C19})$$

$$= \sqrt{\frac{d^2 - 2}{d^2 - 1}} \sqrt{2} (1 - u)^2, \quad (\text{C20})$$

using the definition of  $u$  Eq. (A12) and the fact that  $u(\mathcal{I}) = 1$ . ■

Finally, a bound on  $a_{[\text{adj}]}$  is presented.

*Proposition 17 (Bound on  $a_{[\text{adj}]}$ ).* Let  $a_{[\text{adj}]}$  be defined as in Eq. (B47) and let  $\Lambda$  be a CPTP map. If  $\Lambda$  is a single-qubit channel (i.e., if  $d = 2$ ) or if  $\Lambda$  is unital [i.e.,  $\Lambda(I) = I$ ], then

$$a_{[\text{adj}]} = \frac{2}{(d^2 - 4)(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathcal{C}_\tau} \left( \sum_{\sigma \in \mathcal{C}_\tau} \langle\langle \sigma \cdot \tau | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle \right)^2 \leq \sqrt{d^2 - 1} (1 - u)^2, \quad (\text{C21})$$

where  $\mathcal{C}_\tau$  is the set of all normalized Pauli's that commute with  $\tau$  (except for  $\tau$  and  $\sigma_0$ ), as defined in Eq. (A3).

*Proof.* By the definition Eq. (B47),  $a_{[\text{adj}]} = \langle\langle A_S | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle$ , where

$$B_2 B_2 = \frac{1}{d^2 - 1} \left( \sum_{\hat{\sigma} \in \mathcal{P}^*} \hat{\sigma} \hat{\sigma} \right)^{\otimes 2} \quad \text{and} \quad A_{[\text{adj}]} = \frac{1}{2(d^2 - 4)\sqrt{d^2 - 1}} \sum_{\tau \in \mathcal{P}^*} \left[ \sum_{\sigma \in \mathcal{C}_\tau} (\sigma \cdot \tau) \sigma + \sigma (\sigma \cdot \tau) \right]^{\otimes 2} \quad (\text{C22})$$

were defined in Eqs. (B8) and (B12) respectively. Therefore  $a_{[\text{adj}]}$  is computed as

$$\begin{aligned} a_{[\text{adj}]} &= \frac{1}{2(d^2 - 4)(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathcal{P}^*} \left\langle \left\langle \left[ \sum_{\sigma \in \mathcal{C}_\tau} (\sigma \cdot \tau) \sigma + \sigma (\sigma \cdot \tau) \right]^{\otimes 2} \middle| \Lambda^{\otimes 4} \middle| \left( \sum_{\hat{\sigma} \in \mathcal{P}^*} \hat{\sigma} \hat{\sigma} \right)^{\otimes 2} \right\rangle \right\rangle \\ &= \frac{1}{2(d^2 - 4)(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathcal{P}^*} \left[ \sum_{\sigma \in \mathcal{C}_\tau} \sum_{\hat{\sigma} \in \mathcal{P}^*} \langle\langle (\sigma \cdot \tau) \sigma + \sigma (\sigma \cdot \tau) | \Lambda^{\otimes 2} | \hat{\sigma} \hat{\sigma} \rangle\rangle \right]^2 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \mathbf{P}^*} \left( \sum_{\sigma \in C_\tau} \sum_{\hat{\sigma} \in \mathbf{P}^*} 2 \langle \langle \sigma \cdot \tau | \mathbf{\Lambda} | \hat{\sigma} \rangle \rangle \langle \langle \sigma | \mathbf{\Lambda} | \hat{\sigma} \rangle \rangle \right)^2 \\
 &= \frac{2}{(d^2-4)(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \mathbf{P}^*} \left( \sum_{\sigma \in C_\tau} \langle \langle \sigma \cdot \tau | \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \right)^2, \tag{C23}
 \end{aligned}$$

where in the final line we used again the trick of Eq. (C7). Our bound on this quantity again starts with using the fact that the mean of the squares is larger than the square of the mean (Lemma 10), yielding for all  $\tau \in \mathbf{P}^*$

$$\left( \frac{2}{d^2-4} \sum_{\sigma \in C_\tau} \langle \langle \sigma \cdot \tau | \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \right)^2 \leq \frac{2}{d^2-4} \sum_{\sigma \in C_\tau} \langle \langle \sigma \cdot \tau | \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle^2. \tag{C24}$$

Multiplying with  $\frac{d^2-4}{2}$  and plugging into the above yields

$$a_S \leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \mathbf{P}^*} \sum_{\sigma \in C_\tau} \langle \langle \sigma \cdot \tau | \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle^2. \tag{C25}$$

Now we use the facts that  $\sigma \cdot \tau \neq \sigma$  to write this as

$$a_S \leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \mathbf{P}^*} \sum_{\sigma \in C_\tau} \langle \langle \sigma \cdot \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle^2, \tag{C26}$$

where  $\mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger$  is a positive semidefinite matrix, since  $\|\mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger\|_\infty = \|\mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger\|_{2 \rightarrow 2} \leq 1$  under the stated assumptions on  $\mathbf{\Lambda}$  by Lemma 7 and the fact that a matrix of the form  $\mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger$  is itself positive semidefinite. This allows us again to use Sylvester's criterion (Lemma 8) to bound off-diagonal terms by diagonal terms by using the fact that all minors of degree 2 of  $\mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger$  must be nonnegative:

$$\langle \langle \sigma | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \langle \langle \sigma \cdot \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \cdot \tau \rangle \rangle - \langle \langle \sigma \cdot \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle^2 \geq 0, \quad \forall \tau \in \mathbf{P}^*, \forall \sigma \in C_\tau. \tag{C27}$$

Therefore, we arrive at

$$\begin{aligned}
 a_S &\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau \in \mathbf{P}^*} \sum_{\sigma \in C_\tau} \langle \langle \sigma | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \langle \langle \sigma \cdot \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \cdot \tau \rangle \rangle \\
 &\leq \frac{1}{(d^2-1)^{\frac{3}{2}}} \sum_{\tau, \sigma \in \mathbf{P}^*} \langle \langle \sigma | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \langle \langle \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \tau \rangle \rangle \\
 &= \sqrt{d^2-1} (1-u)^2, \tag{C28}
 \end{aligned}$$

where in the second line the sum over  $\sigma \in C_\tau$  was completed to the sum over  $\sigma \in \mathbf{P}^*$  by adding all the nonnegative terms  $\langle \langle \sigma | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \rangle \rangle \langle \langle \sigma \cdot \tau | \mathbf{I} - \mathbf{\Lambda}_u \mathbf{\Lambda}_u^\dagger | \sigma \cdot \tau \rangle \rangle$  with  $\sigma \in \mathbf{P}^* \setminus C_\tau$  for each  $\tau \in \mathbf{P}^*$ . All these terms are nonnegative because they are the product of diagonal elements of positive-semidefinite matrices, which must be nonnegative. ■

This completes the set of propositions to bound the quantities  $a_i$ . The quantities  $b_i$  are strongly related to the quantities  $a_i$ , and we will show that they satisfy the same upper bounds. More precisely, the next proposition establishes that all bounds on  $a_i$  also hold for  $b_i$ , for  $i \in \{1, 2; S; 0; [\text{adj}]\}$ .

*Proposition 18 (Bounds on  $b_i$ ).* Let  $\mathbf{\Lambda}$  be a CPTP map. Assume that  $d = 2$  or that  $\mathbf{\Lambda}$  is unital. Let  $a_i = \langle \langle A_i | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle \rangle$  and  $b_i = \langle \langle B_2 B_2 | \mathcal{N} - \mathcal{M}^{\otimes 2} | A_i \rangle \rangle$  as above. Then

$$b_0 = a_0 = 0, \tag{C29}$$

$$b_{1,2} = a_{1,2} \leq \frac{\sqrt{d^2-2}}{d^2} (1-u)^2, \tag{C30}$$

$$b_S = a_S \leq \sqrt{\frac{d^2-2}{d^2-1}} \sqrt{2} (1-u)^2, \tag{C31}$$

$$b_{[\text{adj}]} \leq \sqrt{d^2-1} (1-u)^2. \tag{C32}$$

*Proof.* The equality  $b_0 = a_0 = \langle\langle B_2 B_2 | \mathcal{N} - \mathcal{M}^{\otimes 2} | B_2 B_2 \rangle\rangle$  immediately follows from the fact that  $A_0 = B_2 B_2$ . Thus  $b_0 = 0$  by Proposition 14. In general,  $b_i$  can be written as

$$b_i = \langle\langle B_2 B_2 | \mathcal{N} - \mathcal{M}^{\otimes 2} | A_i \rangle\rangle = \langle\langle A_i | \mathcal{N}^\dagger - (\mathcal{M}^{\otimes 2})^\dagger | B_2 B_2 \rangle\rangle. \quad (\text{C33})$$

Now since  $\mathcal{G}_{\text{avg}}^{(n)}$  are orthogonal projections,  $(\mathcal{G}_{\text{avg}}^{(n)})^\dagger = \mathcal{G}_{\text{avg}}^{(n)}$ . Therefore  $\mathcal{N}^\dagger = \mathcal{G}_{\text{avg}}^{(4)}(\Lambda^\dagger)^{\otimes 4} \mathcal{G}_{\text{avg}}^{(4)}$  and  $\mathcal{M}^\dagger = \mathcal{G}_{\text{avg}}^{(2)}(\Lambda^\dagger)^{\otimes 2} \mathcal{G}_{\text{avg}}^{(2)}$ . Thus,  $b_i$  and  $a_i$  are related by  $b_i(\Lambda) = a_i(\Lambda^\dagger)$ . That is,  $b_i$  can be obtained from  $a_i$  by replacing  $\Lambda$  with  $\Lambda^\dagger$  in the exact expressions.

We first show that this implies  $b_{1,2} = a_{1,2}$  and  $b_S = a_S$ . This follows from the two identities [using only the trick of Eq. (C7) over and over again]

$$\sum_{\sigma \in \mathbb{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \sigma \rangle\rangle^2 = \sum_{\sigma \in \mathbb{P}^*} \langle\langle \sigma | \Lambda_u | \hat{\sigma} \rangle\rangle^2 \langle\langle \sigma | \Lambda_u | \hat{\sigma} \rangle\rangle^2 = \sum_{\sigma \in \mathbb{P}^*} \langle\langle \hat{\sigma} | \Lambda_u^\dagger \Lambda_u | \hat{\sigma} \rangle\rangle^2, \quad (\text{C34})$$

$$\sum_{\sigma, \tau \in \mathbb{P}^*} \langle\langle \sigma | \Lambda_u \Lambda_u^\dagger | \tau \rangle\rangle^2 = \sum_{\sigma, \tau, \hat{\sigma}, \hat{\tau} \in \mathbb{P}^*} \langle\langle \sigma | \Lambda_u | \hat{\sigma} \rangle\rangle \langle\langle \tau | \Lambda_u | \hat{\sigma} \rangle\rangle \langle\langle \sigma | \Lambda_u | \hat{\tau} \rangle\rangle \langle\langle \tau | \Lambda_u | \hat{\tau} \rangle\rangle = \sum_{\hat{\sigma}, \hat{\tau} \in \mathbb{P}^*} \langle\langle \hat{\sigma} | \Lambda_u^\dagger \Lambda_u | \hat{\tau} \rangle\rangle^2. \quad (\text{C35})$$

Now Eq. (C34) implies that  $a_{1,2}(\Lambda) = a_{1,2}(\Lambda^\dagger) = b_{1,2}(\Lambda)$ . Subtracting Eq. (C34) from Eq. (C35) implies that  $a_S(\Lambda) = a_S(\Lambda^\dagger) = b_S(\Lambda)$ . This shows the second and third claim of this proposition [Eq. (C30) and Eq. (C31)], using the bounds and expressions for  $a_{1,2}$  and  $a_S$  from Proposition 15 and 16

For  $b_{[\text{adj}]}$  it is not clear that  $b_{[\text{adj}]}$  equals  $a_{[\text{adj}]}$ . However, by copying the technique of the proof of Proposition 17 we show that the same bounds hold. Since  $b_{[\text{adj}]}(\Lambda) = a_{[\text{adj}]}(\Lambda^\dagger)$ , Proposition 17 implies that

$$b_{[\text{adj}]} = \frac{2}{(d^2 - 4)(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathbb{P}^*} \left( \sum_{\sigma \in C_\tau} \langle\langle \sigma \cdot \tau | \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle \right)^2. \quad (\text{C36})$$

The bound is proven in exactly the same spirit as Proposition 17. We first bound the square of the mean by the mean of the squares (Lemma 10) and then use that  $\langle\langle \sigma \cdot \tau | \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle^2 = \langle\langle \sigma \cdot \tau | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle^2$  (since  $\sigma \cdot \tau \neq \pm \sigma$ ). The matrix  $\mathbf{I} - \Lambda_u^\dagger \Lambda_u$  is then shown to be positive semidefinite using  $\|\Lambda_u^\dagger \Lambda_u\|_\infty \leq 1$  (by the assumptions on  $\Lambda$  and Lemma 7) together with the fact that  $\Lambda_u^\dagger \Lambda_u \geq 0$  is positive semidefinite. Thus Sylvester's criterion can be applied (Lemma 8) Therefore

$$\begin{aligned} b_{[\text{adj}]} &= \frac{2}{(d^2 - 4)(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathbb{P}^*} \left( \sum_{\sigma \in C_\tau} \langle\langle \sigma \cdot \tau | \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle \right)^2 \\ &\leq \frac{1}{(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathbb{P}^*} \sum_{\sigma \in C_\tau} \langle\langle \sigma \cdot \tau | \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle^2 \\ &= \frac{1}{(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathbb{P}^*} \sum_{\sigma \in C_\tau} \langle\langle \sigma \cdot \tau | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle^2 \\ &\leq \frac{1}{(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau \in \mathbb{P}^*} \sum_{\sigma \in C_\tau} \langle\langle \sigma | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle \langle\langle \sigma \cdot \tau | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \cdot \tau \rangle\rangle \\ &\leq \frac{1}{(d^2 - 1)^{\frac{3}{2}}} \sum_{\tau, \sigma \in \mathbb{P}^*} \langle\langle \sigma | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle \langle\langle \tau | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \tau \rangle\rangle \\ &= \sqrt{d^2 - 1} (1 - u)^2, \end{aligned} \quad (\text{C37})$$

where in the last inequality the sum is completed by adding the nonnegative terms  $\langle\langle \sigma | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \sigma \rangle\rangle \langle\langle \tau | \mathbf{I} - \Lambda_u^\dagger \Lambda_u | \tau \rangle\rangle$  for all  $\tau \in \mathbb{P}^*$  and  $\sigma \in \mathbb{P}^* \setminus C_\tau$ . Note that this is the same bound as on  $a_{[\text{adj}]}$ .  $\blacksquare$

Finally two more propositions are needed to bound the inner products in the expanded variance expression. The tool for this is the following. This proposition is formulated for any general CPTP map  $\mathcal{E}$  and Hermitian operators  $X, Y \in \mathcal{L}(\mathcal{H})$ . This theorem is applicable to inner products involving the map  $\mathcal{N}^{m-s-1}$ , since this is a CPTP map.

*Proposition 19.* Let  $\mathcal{E}$  be a CPTP map on a general Hilbert space  $\mathcal{H}$ . Then for any pair of Hermitian operators  $X, Y \in \mathcal{L}(\mathcal{H})$

$$\langle\langle X | \mathcal{E} | Y \rangle\rangle \leq \|X\|_\infty \|Y\|_1. \quad (\text{C38})$$

*Proof.* By Von Neumann's trace inequality and Hölder's inequality (Lemma 9) it follows that

$$\langle\langle X|\mathcal{E}|Y\rangle\rangle = \text{Tr}[X\mathcal{E}(Y)] \leq \|X\|_\infty \|\mathcal{E}(Y)\|_1, \quad (\text{C39})$$

using that  $X$  and  $\mathcal{E}(Y)$  are Hermitian. We then use the induced trace norm (the  $1 \rightarrow 1$  norm) and the fact that the map  $\mathcal{E}$  is a CPTP map so that  $\|\mathcal{E}\|_{1 \rightarrow 1} \leq 1$  (Lemma 5). Therefore

$$\|X\|_\infty \|\mathcal{E}(Y)\|_1 \leq \|X\|_\infty \|\mathcal{E}\|_{1 \rightarrow 1} \|Y\|_1 \leq \|X\|_\infty \|Y\|_1. \quad (\text{C40})$$

Putting this together proves the bound.  $\blacksquare$

In order to apply the above proposition to the inner products occurring in the variance proof, a bound on the norms of the operators  $A_i$  with  $i \in \mathcal{Z}_{TS}$  is needed.

*Proposition 20 (Norm bounds on  $A_i$ ).* Let  $\{A_i : i \in \mathcal{Z}_{TS}\}$  be defined as in Eq. (B6). Then for  $d \geq 4$  the following bounds hold:

$$\|A_i\|_1 \leq d^2 \quad \text{and} \quad \|A_i\|_\infty \leq \sqrt{\frac{6}{(d-2)(d-1)}}, \quad \forall i \in \mathcal{Z}_{TS}. \quad (\text{C41})$$

If  $d = 2$ , then  $\mathcal{Z}_{TS} = \{S; 1, 2\}$ , and

$$\|A_S\|_1 = \frac{5}{\sqrt{3}}, \quad \|A_S\|_\infty = \frac{1}{\sqrt{3}}, \quad (\text{C42})$$

$$\|A_{1,2}\|_1 = 2\sqrt{2}, \quad \|A_{1,2}\|_\infty = \frac{\sqrt{2}}{3}. \quad (\text{C43})$$

*Proof.* For the  $d = 2$  case, the norms can be computed directly, since  $A_S$  and  $A_{1,2}$  are explicitly defined in Eq. (B9)–Eq. (B10). By direct computation the result follows. For  $d \geq 4$ , the trace norm bound is trivial, since

$$\|A_i\|_1 \leq \sqrt{d^4} \|A_i\|_2 = d^2, \quad (\text{C44})$$

by Hölder's inequality. The last equality uses the fact that  $A_i$  are Hilbert-Schmidt normalized ( $\|A_i\|_2 = 1$ ). The effort of the proof is in the bound on  $\|A_i\|_\infty$ .

The proof of this statement uses the description of the tensor-2 Liouville representation of Ref. [34] over Ref. [36], since their description is basis-free. Reference [34] considers the action of the Clifford group  $\mathbf{C}(d)$  on  $\mathcal{H}^{\otimes 4}$ . The representation  $\mathcal{H}^{\otimes 4}$  of the Clifford group  $\mathbf{C}(d)$  decomposes as

$$\mathcal{H}^{\otimes 4} = \bigoplus_k W_k \otimes \mathbb{C}^{d_k}, \quad (\text{C45})$$

where  $W_k$  are irreducible, pairwise inequivalent representations of the Clifford group that occur with multiplicity  $d_k$ . Here  $k$  is just an index for the irreducible, inequivalent representations. Descriptions of these spaces and explicit expressions for their dimensions are given in [34] (there the index  $k$  runs over Young Diagrams  $\lambda$  and signs  $s$ ). We will show that

$$\|A_i\|_\infty \leq \max_k \frac{1}{\sqrt{|W_k|}}. \quad (\text{C46})$$

Since the dimensions of all  $W_k$  are given, the maximization can easily be done.

Using the intertwining isomorphism  $\mathbf{L}(\mathcal{H}) \simeq \mathcal{H} \otimes \mathcal{H}^*$  the tensor-4 Liouville representation on  $\mathbf{L}(\mathcal{H}^{\otimes 4})$  can be written in terms of the decomposition Eq. (C45):

$$\mathbf{L}(\mathcal{H}^{\otimes 4}) = \bigoplus_{k,l} \mathbf{L}(W_l, W_k) \otimes \mathbf{L}(\mathbb{C}^{d_l}, \mathbb{C}^{d_k}). \quad (\text{C47})$$

In principle  $\mathbf{L}(W_l, W_k)$  are not irreducible representations. However, only the trivial subrepresentations of  $\mathbf{L}(W_l, W_k)$  [denoted  $(\mathbf{L}(W_l, W_k))^{\mathbb{C}(d)}$ ] are relevant, since

$$(\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)} = \bigoplus_{k,l} (\mathbf{L}(W_l, W_k))^{\mathbb{C}(d)} \otimes \mathbf{L}(\mathbb{C}^{d_l}, \mathbb{C}^{d_k}). \quad (\text{C48})$$

The key point is that every element  $\varphi \in (\mathbf{L}(W_l, W_k))^{\mathbb{C}(d)}$  is an intertwining operator between the representations  $W_k$  and  $W_l$  [37]. By Schur's Lemma [37] and the fact that  $W_k$  are mutually inequivalent irreducible representations it follows that  $\varphi \propto \delta_{k,l} I_{W_k}$ . Therefore

$$(\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)} = \bigoplus_k \text{Span}\{I_{W_k}\} \otimes \mathbf{L}(\mathbb{C}^{d_k}). \quad (\text{C49})$$

This description provides a simple orthogonal basis for the space  $(\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)}$ , namely,

$$\mathcal{A} = \{P_{W_k} \otimes E_{m,n} | k; m, n = 1, \dots, d_k\}, \quad (\text{C50})$$

where  $P_{W_k}$  is the orthogonal projection onto  $W_k$  and  $\{E_{m,n} | m, n = 1, \dots, d_k\}$  is the canonical (or any other) orthonormal basis of  $\mathbf{L}(\mathbb{C}^{d_k})$ . Normalizing with respect to the Hilbert-Schmidt norm yields the orthonormal basis operators

$$A_{k,m,n} = \frac{P_{W_k}}{\sqrt{|W_k|}} \otimes E_{m,n}. \quad (\text{C51})$$

Note that our basis operators  $\{A_i : i \in \mathcal{Z}_{TS}\}$  might be different than these  $A_{k,m,n}$ . However, these  $A_i$  also span trivial subrepresentations of  $\mathbf{L}(\mathcal{H}^{\otimes 4})$ , so  $A_i \in (\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)}$ . We now show that  $\|A_i\| \leq \max_k |W_k|^{-\frac{1}{2}}$  for all  $A \in (\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)}$  such that  $\|A\|_2 = 1$ . Therefore this bound holds in particular for our  $A_i$  of interest. To do so,  $A$  is written in the basis  $\mathcal{A}$  as

$$A = \sum_k \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n}, \quad \text{s.t.} \quad \sum_k \sum_{m,n=1}^{d_k} |\alpha_{k,m,n}|^2 = 1. \quad (\text{C52})$$

Now we use that the operator  $A \in (\mathbf{L}(\mathcal{H}^{\otimes 4}))^{\mathbb{C}(d)}$  is block diagonal with respect to the spaces  $\text{Span}\{I_{W_k}\} \otimes \mathbf{L}(\mathbb{C}^{d_k})$  [see Eq. (C49)]. Therefore the infinity norm can be computed as the maximum over  $k$  of the infinity norm of  $A$  restricted to  $\text{Span}\{I_{W_k}\} \otimes \mathbf{L}(\mathbb{C}^{d_k})$ , yielding

$$\begin{aligned} \|A\|_\infty &= \left\| \sum_k \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n} \right\|_\infty \\ &= \max_k \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} A_{k,m,n} \right\|_\infty \\ &= \max_k \left\| \frac{P_{W_k}}{\sqrt{|W_k|}} \otimes \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty. \end{aligned} \quad (\text{C53})$$

Using some basic properties of the Schatten  $p$ -norms, this is bounded as follows:

$$\begin{aligned} \|A\|_\infty &= \max_k \left\| \frac{P_{W_k}}{\sqrt{|W_k|}} \right\|_\infty \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty \\ &= \max_k \frac{\|P_{W_k}\|_\infty}{\sqrt{|W_k|}} \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty \\ &\leq \max_k \frac{1}{\sqrt{|W_k|}}, \end{aligned} \quad (\text{C54})$$

using that  $\|P_{W_k}\|_\infty = 1$  and

$$\begin{aligned} \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_\infty &\leq \left\| \sum_{m,n=1}^{d_k} \alpha_{k,m,n} E_{m,n} \right\|_2 \\ &\leq \sum_{m,n=1}^{d_k} |\alpha_{k,m,n}|^2 \|E_{m,n}\|_2 = 1. \end{aligned} \quad (\text{C55})$$

By Lemma 1 of Ref. [34], which gives all dimensions  $|W_k|$ , it follows that

$$\|A\|_\infty \leq \max_k \frac{1}{\sqrt{|W_k|}} = \sqrt{\frac{6}{(d-1)(d-2)}}, \quad (\text{C56})$$

provided that  $d = 2^q \geq 4$ ,  $q \in \mathbb{N}$ . This proves the last bound.  $\blacksquare$

Finally, there is one inner product in the proof of Theorem 1 for which a sharper bound can be found than using Proposition 19 and 20. This sharper bound is given in the following proposition.

*Proposition 21.* Let  $\mathcal{N}$  be defined as in Eq. (B14), with  $\Lambda$  a single-qubit or unital quantum channel. Then for any  $m \in \mathbb{N}$  the following bound holds:

$$\langle\langle A_i | \mathcal{N}^m | B_2 B_2 \rangle\rangle \leq \frac{1}{\sqrt{|V_i|}}, \quad \forall i \in \mathcal{Z}_{TS}. \quad (\text{C57})$$

*Proof.* Slightly rewriting the inner product yields

$$\langle\langle A_i | \mathcal{N}^m | B_2 B_2 \rangle\rangle = \langle\langle A_i | \mathcal{N}^m (B_2 B_2) \rangle\rangle. \quad (\text{C58})$$

From the definition of  $\mathcal{N}$  Eq. (B14) it follows that

$$\begin{aligned} \mathcal{N}^m (B_2 B_2) &= \frac{1}{|\mathcal{C}(d)|^m} \sum_{\mathbf{j}} \mathcal{G}_{\mathbf{j}}^{\otimes 4} (B_2 B_2) \\ &= \frac{1}{|\mathcal{C}(d)|^m} \sum_{\mathbf{j}} [\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)]^{\otimes 2}, \end{aligned} \quad (\text{C59})$$

where the sum is over all noisy sequences of length  $m$  indexed by  $\mathbf{j}$  (i.e.,  $\mathbf{j}$  is a multi-index of length  $m$ ). We will show that  $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq 1$ . We treat the multiqubit and single-qubit case separately. In the multiqubit case, we have

$$\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}\|_{2 \rightarrow 2} \|B_2\|_2 = \|\mathcal{G}_{\mathbf{j}}\|_{2 \rightarrow 2}^2. \quad (\text{C60})$$

The inequality follows from the definition of the induced Schatten norms [see Eq. (A7)]. The equality is due to the fact that  $\|B_2\|_2 = 1$  is normalized. Under the assumption that  $\Lambda$  is unital, the entire sequence  $\mathcal{G}_{\mathbf{j}}$  is unital. Therefore by Lemma 5 (Pérez-García),  $\|\mathcal{G}_{\mathbf{j}}\|_{2 \rightarrow 2}^2 \leq 1$ . This shows that  $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq 1$ .

In case of a single-qubit, nonunital error channels  $\Lambda$ , some extra care must be taken. Let us denote  $\mathcal{L}(\mathcal{H})^H := \{A \in \mathcal{L}(\mathcal{H}) : \text{Tr}[A] = 0, A = A^\dagger\} = \text{Span}_{\mathbb{R}}\{\sigma : \sigma \in \mathcal{P}^*\}$  as the traceless Hermitian subspace of  $\mathcal{L}(\mathcal{H})$ . This space is a vector space over  $\mathbb{R}$ , with an orthonormal basis  $\mathcal{P}^*$ . Since  $\mathcal{G}_{\mathbf{j}}$  is positive (and thus maps Hermitian operators to Hermitian operators) and trace-preserving, it maps the traceless Hermitian subspace  $\mathcal{L}(\mathcal{H})^H$  to itself. Observe that  $B_2 \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2}$ . Therefore restrict  $\mathcal{G}_{\mathbf{j}}^{\otimes 2}$  to  $(\mathcal{L}(\mathcal{H})^H)^{\otimes 2}$ . This results in

$$\begin{aligned} \|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 &= \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|_{(\mathcal{L}(\mathcal{H})^H)^{\otimes 2}} (B_2)\|_2 \\ &\leq \|\mathcal{G}_{\mathbf{j}}^{\otimes 2}|_{(\mathcal{L}(\mathcal{H})^H)^{\otimes 2}}\|_{2 \rightarrow 2} \|B_2\|_2 \\ &= \|\mathcal{G}_{\mathbf{j}}|_{\mathcal{L}(\mathcal{H})^H}\|_{2 \rightarrow 2}^2. \end{aligned} \quad (\text{C61})$$

The first equality is the restriction of  $\mathcal{G}_{\mathbf{j}}$  to the traceless Hermitian subspace. The inequality follows from the definition of the induced Schatten norm Eq. (A7). The final equality is due to the fact that  $\|B_2\|_2 = 1$ . The key point of restricting to the traceless Hermitian subspace  $\|\mathcal{G}_{\mathbf{j}}|_{\mathcal{L}(\mathcal{H})^H}\|_{2 \rightarrow 2}$  allows for the application of statement Eq. (A25) of Lemma 5 (Pérez-García). By the lemma (where  $\|\mathcal{G}_{\mathbf{j}}|_{\mathcal{L}(\mathcal{H})^H}\|_{2 \rightarrow 2}$  is denoted  $\|\mathcal{G}_{\mathbf{j}}\|_{2 \rightarrow 2}^H$ ), we have

$$\|\mathcal{G}_{\mathbf{j}}|_{\mathcal{L}(\mathcal{H})^H}\|_{2 \rightarrow 2} \leq \sqrt{\frac{d}{2}}, \quad (\text{C62})$$

which in the single-qubit case means  $\|\mathcal{G}_{\mathbf{j}}|_{\mathcal{L}(\mathcal{H})^H}\|_{2 \rightarrow 2} \leq 1$ . Therefore, we also have  $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq 1$  in the single-qubit, nonunital case.

We have thus established that  $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq 1$  for single-qubit or unital noise maps  $\Lambda$ . Therefore, the following upper bound is valid:

$$\begin{aligned} \langle\langle A_i | \mathcal{N}^m | B_2 B_2 \rangle\rangle &= \frac{1}{|\mathcal{C}(d)|^m} \sum_{\mathbf{j}} \langle\langle A_i | [\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)]^{\otimes 2} \rangle\rangle \\ &\leq \frac{1}{|\mathcal{C}(d)|^m} \sum_{\mathbf{j}} \max_{\substack{Q \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2} \\ \|Q\|_2 \leq 1}} \langle\langle A_i | Q^{\otimes 2} \rangle\rangle \\ &= \max_{\substack{Q \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2} \\ \|Q\|_2 \leq 1}} \langle\langle A_i | Q^{\otimes 2} \rangle\rangle. \end{aligned} \quad (\text{C63})$$

In the second line, we have replaced the particular operator  $\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2) \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2}$  which satisfies  $\|\mathcal{G}_{\mathbf{j}}^{\otimes 2} (B_2)\|_2 \leq 1$  with the maximization over all operators  $Q \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2}$  that satisfy  $\|Q\|_2 \leq 1$ . To continue, we use the definition of  $A_i$  Eq. (B6), which is given by

$$A_i = \frac{1}{\sqrt{|V_i|}} \sum_{s=1}^{|V_i|} v_s^{(i)} v_s^{(i)}, \quad \forall i \in \mathcal{Z}_{TS}, \quad (\text{C64})$$

where  $\{v_s^{(i)}\}$  is an orthonormal basis of  $V_i \subset (\mathcal{L}(\mathcal{H})^H)^{\otimes 2}$ . Let us expand  $Q$  in this basis,

$$\begin{aligned} Q &= q_{\perp} v_{\perp}^{(i)} + \sum_{s=1}^{|V_i|} q_s v_s^{(i)} \quad \text{s.t.} \quad |q_{\perp}|^2 + \sum_{s=1}^{|V_i|} |q_s|^2 \leq 1, \\ q_{\perp}, q_s &\in \mathbb{C}, \quad \forall s = 1, \dots, |V_i|. \end{aligned} \quad (\text{C65})$$

Here  $q_{\perp} v_{\perp}^{(i)}$  is the component of  $Q$  in the space orthogonal to  $V_i$ , i.e.,  $q_{\perp} v_{\perp}^{(i)} \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2} \setminus V_i$ . The condition on  $q_{\perp}$  and the  $q_s$  follow from the requirement that  $\|Q\|_2 \leq 1$ . Actually, there are additional constraints on  $q_{\perp}$  and the  $q_s$  needed to ensure that  $Q$  is traceless and Hermitian, but these constraints are not necessary to prove the result. Using the expansion Eq. (C65) it follows that

$$\begin{aligned} & \max_{\substack{Q \in (\mathcal{L}(\mathcal{H})^H)^{\otimes 2} \\ \|Q\|_2 \leq 1}} \langle\langle A_i | Q^{\otimes 2} \rangle\rangle \\ & \leq \max_{\substack{\{q_s\} \\ \sum_s |q_s|^2 \leq 1}} \frac{1}{\sqrt{|V_i|}} \sum_{s,t,k=1}^{|V_i|} |q_s q_t| |\langle\langle v_k^{(i)} v_k^{(i)} | v_s^{(i)} v_t^{(i)} \rangle\rangle| \\ & = \max_{\substack{\{q_s\} \\ \sum_s |q_s|^2 \leq 1}} \frac{1}{\sqrt{|V_i|}} \sum_{k=1}^{|V_i|} |q_k|^2 \leq \frac{1}{\sqrt{|V_i|}}, \end{aligned} \quad (\text{C66})$$

using the fact that  $\langle\langle v_k^{(i)} v_k^{(i)} | v_s^{(i)} v_t^{(i)} \rangle\rangle = \delta_{sk} \delta_{tk}$  by orthonormality of the basis. This completes the proof.  $\blacksquare$

- 
- [1] J. Emerson, R. Alicki, and K. Życzkowski, *J. Opt. B* **7**, 347 (2005).
- [2] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Phys. Rev. A* **77**, 012307 (2008).
- [3] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [4] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. A* **85**, 042311 (2012).
- [5] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **82**, 040305 (2010).
- [6] S. Olmschenk, R. Chicireanu, K. D. Nelson, and J. V. Porto, *New J. Phys.* **12**, 113007 (2010).
- [7] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, *Phys. Rev. Lett.* **108**, 260503 (2012).
- [8] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, *Nature (London)* **508**, 500 (2014).
- [9] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, *J. Phys.: Condens. Matter* **27**, 154205 (2015).
- [10] T. Xia, M. Lichtman, K. Maller, A. W. Carr, M. J. Piotrowicz, L. Isenhower, and M. Saffman, *Phys. Rev. Lett.* **114**, 100503 (2015).
- [11] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [12] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *New J. Phys.* **17**, 113020 (2015).
- [13] J. J. Wallman, M. Barnhill, and J. Emerson, *Phys. Rev. Lett.* **115**, 060501 (2015).
- [14] J. J. Wallman, M. Barnhill, and J. Emerson, *New J. Phys.* **18**, 043021 (2016).
- [15] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, [arXiv:1702.03688](https://arxiv.org/abs/1702.03688) (2017).
- [16] G. Feng, J. J. Wallman, B. Buonacorsi, F. H. Cho, D. K. Park, T. Xin, D. Lu, J. Baugh, and R. Laflamme, *Phys. Rev. Lett.* **117**, 260501 (2016).
- [17] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, *Phys. Rev. A* **93**, 012301 (2016).
- [18] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, [arXiv:1610.05296](https://arxiv.org/abs/1610.05296) (2016).
- [19] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, *New J. Phys.* **18**, 012002 (2016).
- [20] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, *Phys. Rev. Lett.* **117**, 170502 (2016).
- [21] J. J. Wallman, [arXiv:1511.00727](https://arxiv.org/abs/1511.00727) (2015).
- [22] L. P. Thinh, P. Faist, J. Helsen, D. Elkouss, and S. Wehner, [arXiv:1808.00358](https://arxiv.org/abs/1808.00358) (2018).
- [23] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, *Phys. Rev. A* **89**, 062321 (2014).
- [24] C. Granade, C. Ferrie, and D. G. Cory, *New J. Phys.* **17**, 013042 (2015).
- [25] I. Hincks, J. J. Wallman, C. Ferrie, C. Granade, and D. G. Cory, [arXiv:1802.00401](https://arxiv.org/abs/1802.00401) (2018).
- [26] J. J. Wallman and S. T. Flammia, *New J. Phys.* **16**, 103032 (2014).
- [27] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, [arXiv:1701.04299](https://arxiv.org/abs/1701.04299) (2017).
- [28] D. Gross, K. Audenaert, and J. Eisert, *J. Math. Phys.* **48**, 052104 (2007).
- [29] W. Hoeffding, *J. Am. Stat. Assoc.* **58**, 301 (1963).
- [30] N. Johnston, QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9 (2016).
- [31] C. A. Fuchs and J. van de Graaf, *IEEE Trans. Inf. Theory* **45**, 1216 (1999).
- [32] N. A. Weiss, *A Course in Probability* (Addison-Wesley, Boston, MA, 2005), pp. 380–383.
- [33] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, MA, 2018).
- [34] H. Zhu, R. Kueng, M. Grassl, and D. Gross, [arXiv:1609.08172](https://arxiv.org/abs/1609.08172) (2016).

- [35] H. Zhu, *Phys. Rev. A* **96**, 062336 (2017).
- [36] J. Helsen, J. J. Wallman, and S. Wehner, *J. Math. Phys.* **59**, 072201 (2018).
- [37] W. Fulton and J. Harris, *Representation Theory* (Springer, New York, NY, 2004).
- [38] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, *J. Math. Phys.* **47**, 083506 (2006).
- [39] T. Chasseur and F. K. Wilhelm, *Phys. Rev. A* **92**, 042333 (2015).
- [40] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *Phys. Rev. Lett.* **119**, 130502 (2017).
- [41] J. J. Wallman, *Quantum* **2**, 47 (2018).
- [42] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Phys. Rev. A* **80**, 012304 (2009).
- [43] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, *Phys. Rev. A* **92**, 060302 (2015).
- [44] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, *npj Quantum Inf.* **2**, 16012 (2016).
- [45] D. S. França and A. K. Hashagen, *J. Phys. A: Math. Theor.* **51**, 395302 (2018).
- [46] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, *Quantum* **2**, 85 (2018).
- [47] W. G. Brown and B. Eastin, *Phys. Rev. A* **97**, 062323 (2018).
- [48] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, [arXiv:1806.02048](https://arxiv.org/abs/1806.02048) (2018).
- [49] J. M. Farinholt, *J. Phys. A* **47**, 305303 (2014).
- [50] M. Ozols, Clifford group (2008).
- [51] M. A. Nielsen, *Phys. Lett. A* **303**, 249 (2002).
- [52] P. Etingof, O. Golberg, S. Hensel, T. Liu, A. Schwendner, D. Vaintrob, and E. Yudovina, *Introduction to Representation Theory* (American Mathematical Society, Providence, RI, 2009).
- [53] D. Braun, O. Giraud, I. Nechita, C. Pellegrini, and M. Žnidarič, *J. Phys. A* **47**, 135302 (2014).
- [54] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, Cambridge, MA, 2013).
- [55] C. D. Aliprantis and O. Burkinshaw, *Principles of Real Analysis*, 3rd ed. (Academic Press, San Diego, CA, 1998).