Experimentally efficient methods for estimating the performance of quantum measurements

Easwar Magesan[1] and Paola Cappellaro[1,2]

[1] *Research Laboratory of Electronics, MIT, Cambridge MA, 02139, U.S.A.*
[2] *Nuclear Science and Engineering Department, MIT, Cambridge MA, 02139, U.S.A.*

Efficient methods for characterizing the performance of quantum measurements are important in the experimental quantum sciences. Ideally, one requires both a physically relevant distinguishability measure between measurement operations and a well-defined experimental procedure for estimating the distinguishability measure. Here, we propose the average measurement fidelity and error between quantum measurements as distinguishability measures. We provide protocols for obtaining bounds on these quantities that are both estimable using experimentally accessible quantities and scalable in the size of the quantum system. This gives an exponential reduction in the resources required for complete measurement tomography. We explain why the bounds should be valid in large generality and illustrate the method via numerical examples.

I. INTRODUCTION

Measurement plays a fundamental role in the quantum sciences as it allows one to evolve and extract information from a quantum system. From a practical perspective, measurements are imperative for the success of a wide variety of tasks from quantum information processing (QIP) and communication [1] to performing high-precision metrology [2]. If large-scale QIP is achieved experimentally, certain tasks that are conjectured to be classically inefficient will become realizable using a scalable number of resources [3, 4].

Various models of computation have been proposed for QIP, such as the standard circuit model [5], measurement based quantum computation (MBQC) [6], and topological quantum computation [7]. Each of these models relies greatly on the ability to perform accurate measurements. A large number of physical systems have been proposed as candidates for implementing QIP. A short, and certainly non-exhaustive, list of systems includes superconducting circuits [8, 9], Nitrogen-vacancy (NV) centers [10], trapped ions [11], NMR [12], quantum dots [13], and optical implementations [14]. Measurement schemes vary greatly across these systems and so it is important to have straightforward protocols for comparing measurements that are both independent of the particular type of implementation and scalable in the size of the system. Providing such protocols is the main goal of this paper.

There has been a significant amount of research towards *completely characterizing* the error on the operations used to process and measure quantum information via quantum process [15–17] and measurement tomography [18, 19]. In principle, these tomographic methods can be used to characterize the error affecting any quantum operation or measurement realized in an experimental setting. Unfortunately, there are various significant drawbacks to complete process and measurement tomography. Process tomography requires an exponential number of resources in the number n of quantum bits (qubits) that comprise the system (the number of parameters required to just describe the process scales as 16^n), is not robust against state-preparation and measurement errors [20], and requires intensive classical post-processing of measurement data [21]. Not surprisingly, since measurement tomography is realized by essentially inverting process tomography [18, 19], it suffers from similar drawbacks to quantum process tomography. In particular, complete measurement tomography requires the ability to prepare a complicated set of pure input states with extremely high precision (possibly by performing complex unitary gates), scales badly in the size of the system, and can require lengthy post-processing of the tomographic data.

In many cases, one may only be interested in a subset of parameters characterizing the noise, or determining the *strength* of the noise process rather than the entire process itself. For instance, in fault-tolerance [22–25], as long as the strength of the noise affecting state-preparation, gates, and measurements is bounded by some value, large-scale computation is possible. As a result, various methods for *partially characterizing* the error of quantum gates [26–35] have recently been proposed which attempt to circumvent many of the problems associated with full process tomography. The error is defined via the "quantum average gate fidelity", which is derived from the quantum channel fidelity between two quantum processes. Since quantum gates are unitary operations, the average gate fidelity (and higher order moments) take a simple form and can be calculated analytically [36–38].

Similarly, *partial* estimation of measurement errors by more efficient methods than full measurement tomography would be extremely valuable, but have yet be considered in significant detail. There are a variety of questions to consider in this endeavor, such as how to deal with abundance of different measurement schemes across the various implementations, and the ambiguity regarding what the single parameter characterizing the error should be. Since measurements are non-unitary processes, the quantum channel fidelity between the measurement and noise process is a very complex quantity and cannot be calculated analytically as is the case for quantum gates.

Here, we propose the average measurement fidelity and

error rate of a quantum measurement as natural measures of the error, and provide scalable, experimentally implementable protocols for estimating these quantities. The output of the protocol is an *upper (lower) bound* on the error rate (average fidelity) of the measurement. We focus mainly on the case where we attempt to implement a finite-dimensional projective measurement (also known as a projective-valued measure (PVM)), but instead implement a general quantum measurement in the form of a positive operator-valued measure (POVM). In particular, we analyze the following two scenarios:

1. The ideal measurement is a rank-1 PVM and we are only concerned with characterizing measurement probabilities.

2. The ideal measurement is a rank-1 PVM and we are concerned with simultaneously characterizing measurement probabilities and output states.

The case of rank-1 PVM's is significant in QIP since each computational model mentioned above can achieve universal quantum computation using rank-1 PVM's. We leave extending our results to arbitrary-rank PVM's as a direction of future research, however we anticipate the results presented here can be carried over to these measurements. Moreover, while the most general measurement model is a POVM, any POVM can be realized as a PVM on an extended Hilbert space by attaching ancilla systems to the Hilbert space of interest [39]. Thus, characterizing the quality of PVM's can provide direct information about one's ability to implement POVM's.

The methods we present here to obtain bounds on error rates of measurements are general and should be straightforward to implement in most systems. The protocol involves the ability to prepare a random state from the basis that constitutes the ideal measurement, and perform the noisy measurement (perhaps consecutively). Computing the upper bound from the measurement data is straightforward, and generic estimates of how many trials need to be performed can be obtained using statistical methods. We provide a direct comparison of the time-complexity between our protocol and that of completely reconstructing the noisy measurement POVM elements via tomography. Our method scales independently of d (where d is the dimension of the system, e.g. $d = 2^n$ for an n-qubit system) while a complete reconstruction scales as $O(d^3)$. Thus, while our method gives much less information than full tomography, it is scalable, straightforward to implement, and involves less post-processing of measurement data.

The structure of the presentation is as follows. First, in Sec. (II) we set notation, discuss the framework of PVM's and POVM's, and define the noisy measurement models. Next, in Sec.(III), we define the *quantum average measurement fidelity* and *quantum average measurement error*, which are the operationally relevant metrics we use to compare measurement operations.

In Sec. (IV) we present the main result for the first of the two scenarios listed above. We provide an expression for a lower (upper) bound on the fidelity (error rate), and present an experimental protocol for obtaining these bounds. In Sec. (V) we give a derivation of the bounds and show they satisfy certain necessary conditions to be useful in practice. Sec. (VI) provides a discussion of the general validity of the bounds and numerical examples for the case of a single-qubit system are analyzed. Our derived bounds are shown to be valid for *every* example analyzed, confirming the discussion regarding the validity of the method.

We discuss the extension of the method to rank-1 PVM's with output states in Sec. (VII) and provide the experimental protocol for this case in Sec. (VIII). The resource analysis for the protocols are provided in Sec. (IX) and concluding remarks are made in Sec. (X). The reader interested in the bounds and protocols without most of the technical details is referred to the text surrounding Eq. (4.2) for rank-1 PVM's and the text surrounding Eq. (7.14) for rank-1 PVM's with output states.

II. MEASUREMENT OF QUANTUM SYSTEMS AND THE ERROR MODEL

Denote the quantum system by $\mathcal{S}$ and suppose it is represented by a Hilbert space $\mathcal{H}$ of dimension $d < \infty$ ($d = 2^n$ for an n-qubit system). The most general type of measurement one can make is a "positive operator-valued measure" (POVM). A POVM consists of a set $\{E_k\}$ of linear operators on $\mathcal{H}$ that satisfy,

$$\begin{aligned} E_k &\geq 0, \\ \sum_k E_k &= \mathbb{1}. \end{aligned} \tag{2.1}$$

If the input state to the measurement is σ, the probability of obtaining outcome "k" is given by

$$r_k = \mathrm{tr}\left(E_k \sigma\right). \tag{2.2}$$

An important subset of POVM measurements are "projection-valued measures" (PVM's), which correspond to the case of each E_k being equal to a projection operator Π_k. Hence, in addition to the conditions listed in Eq. (2.1), $\Pi_k^2 = \Pi_k$ for each k.

PVM's are a very general class of quantum measurements and, as previously mentioned, are sufficient for performing universal quantum computation in various computational models [5, 6]. For instance, computational basis measurements allow for universality in the standard circuit model and single-qubit projective measurements provide universality in MBQC. Other important examples of PVM's in quantum information theory are parity measurements, which are used extensively in quantum error-correction and fault tolerance [40–43]. It is important to note that, by Naimark's theorem [39], any POVM can be performed via implementing a PVM on a larger Hilbert space. Hence, methods for determining error rates of PVM's can also give direct information

regarding the error associated with performing general POVM measurements.

PVM's are in a 1-1 correspondence with observables (Hermitian operators) $\mathcal{O}$, where $\mathcal{O}$ is non-degenerate if and only if the PVM consists of rank-1 projective elements . We write the observable $\mathcal{O}$ as

$$\mathcal{O} = \sum_{k=1}^{b} \lambda_k \Pi_k \tag{2.3}$$

where the eigenvalues λ_k correspond to the measurement values and $\Pi_k := \sum_{j=1}^{d_k} |\psi_j^k\rangle\langle\psi_j^k|$ is the rank-d_k projector onto the eigenspace spanned by the eigenvectors $|\psi_j^k\rangle$. Hence, $\sum_{k=1}^{b} d_k = d$ and the PVM is rank-1 if and only if $d_k = 1$ for every k. The ideal quantum operation that corresponds to the PVM (without post-selection), denoted $\mathcal{M}$, has action on input state σ,

$$\mathcal{M}(\sigma) = \sum_{k=1}^{b} \Pi_k \sigma \Pi_k = \sum_{k=1}^{b} p_k \left[\frac{\Pi_k \sigma \Pi_k}{\mathrm{tr}\left(\Pi_k \sigma\right)}\right] \tag{2.4}$$

where we define,

$$p_k = p_k(\sigma) = \mathrm{tr}\left(\Pi_k \sigma\right). \tag{2.5}$$

In the case of analyzing only measurement probabilities, the PVM is a mapping from the set of quantum states to the probability vectors $(p_1, ..., p_b)$,

$$\mathcal{M}(\sigma) = (p_1(\sigma), ..., p_b(\sigma)) = (\mathrm{tr}\left(\Pi_1 \sigma\right), ..., \mathrm{tr}\left(\Pi_b \sigma\right)). \tag{2.6}$$

This case is especially interesting since in many scenarios one is mainly interested in the output *value* of the measurement rather than the output state itself. For instance, the output of a quantum algorithm is usually the value of a measurement in the computational basis. Hence, as long as the measurement values are obtained with correct probabilities, the measurement is deemed successful.

In addition, in many current implementations of measurements, the action of performing the measurement operation to obtain the output value destroys or drastically alters the state of the system. For instance in photodetection, which forms the basis of various measurement schemes in optical, atomic, and superconducting systems, the measurement consists of recording the occurrence of a photon. If the state of the system is encoded into a degree of freedom (mode) of the photon, such as polarization or frequency, detection of the photon can record the output of a measurement of this degree of freedom. In common photodetectors, such as avalanche photodiodes, the photon is lost in the process of creating a current through the photoelectric effect. Hence, the system encoding the information to be measured is destroyed, but the measurement output can be accessed.

An example where the system is not destroyed yet the state is not preserved from the measurement is the Nitrogen-Vacancy (NV) center in diamond [10]. In this case, the processes of measurement and ground-state polarization are identical. Hence, while one can obtain the relevant measurement outcomes and statistics, the output state of the measurement is always the ground state.

In the general case of Eq. (2.4), the noisy measurement $\mathcal{E}$ is modeled by

$$\mathcal{E}(\sigma) = \sum_{k=1}^{b} r_k(\sigma) \rho_k(\sigma) \tag{2.7}$$

where we allow both the noisy measurement probabilities r_k and output states ρ_k to be functions of σ. By Eq. (2.4), $r_k(\sigma)$ and $\rho_k(\sigma)$ are ideally given by $p_k(\sigma) = \mathrm{tr}(\Pi_k \sigma)$ and $\frac{\Pi_k \sigma \Pi_k}{\mathrm{tr}(\Pi_k \sigma)}$ respectively. If we analyze only measurement probabilities, we have

$$\mathcal{E}(\sigma) = (r_1(\sigma), ..., r_b(\sigma)) = (\mathrm{tr}\left(E_1 \sigma\right), ..., \mathrm{tr}\left(E_b \sigma\right)) \tag{2.8}$$

where $\mathcal{E}$ is allowed to be of a completely general form by assuming it is modeled by a POVM $\{E_k\}_{k=1}^{b}$. The first scenario we are interested in is to compare the probability distributions in Eq. (2.6) and (2.8). We analyze the more general case of measurement probabilities and output states (ie. comparing Eq.'s (2.4) and (2.7)) in Sec. (VII). We first discuss the figure of merits we will use to compare the ideal measurement process $\mathcal{M}$ and noisy measurement process $\mathcal{E}$.

III. QUANTUM AVERAGE MEASUREMENT FIDELITY AND ERROR

A completely natural question is, how should we compare the ideal and actual measurements $\mathcal{M}$ and $\mathcal{E}$? A set of criteria that a distance measure, Δ, for comparing ideal and real quantum processes should satisfy has been given previously [44]. In practice, no Δ will satisfy all of these criteria simultaneously. Thus, one must settle for Δ to satisfy a subset of these criteria, in addition to other criteria that may be useful for the particular task at hand.

As mentioned previously, the average gate fidelity $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ is a useful method for comparing an intended unitary operation $\mathcal{U}$ and actual quantum process $\mathcal{E}$. There are various reasons for the utility of the average gate fidelity, for instance, it satisfies the following properties:

1. There is a straightforward method for evaluating $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ (given a description of $\mathcal{U}$ and $\mathcal{E}$),

2. $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ has a well-motivated physical interpretation,

3. All states are taken into account in an unbiased manner when calculating $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$,

4. $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$ is experimentally accessible via efficient protocols.

An important drawback of the average gate fidelity is that it is not a metric. We would like similar properties to hold for our method of comparing ideal and real measurements. Let us briefly outline how the average gate fidelity is derived from more general quantities, which will provide intuition for how to define our method for comparing measurements.

The average gate fidelity is derived from the state-dependent quantum channel fidelity, which is a standard method for comparing quantum operations. If $\mathcal{E}_1$ and $\mathcal{E}_2$ are quantum operations and σ is a quantum state, the quantum channel fidelity between $\mathcal{E}_1$ and $\mathcal{E}_2$, denoted $F_{\mathcal{E}_1,\mathcal{E}_2}$, is given by the standard state fidelity between $\mathcal{E}_1(\sigma)$ and $\mathcal{E}_2(\sigma)$,

$$\begin{aligned} F_{\mathcal{E}_1,\mathcal{E}_2}(\sigma) &= F(\mathcal{E}_1(\sigma),\mathcal{E}_2(\sigma)) \\ &= \left(\mathrm{tr}\sqrt{\sqrt{\mathcal{E}_1(\sigma)}\mathcal{E}_2(\sigma)\sqrt{\mathcal{E}_1(\sigma)}}\right)^2. \end{aligned} \tag{3.1}$$

When one of the operations in Eq. (3.1) is unitary (say $\mathcal{E}_2 = \mathcal{U}$), the channel fidelity is called the quantum gate fidelity, and when σ is pure ($\sigma = |\psi\rangle\langle\psi|$), the gate fidelity takes the extremely simple form,

$$\begin{aligned} F_{\mathcal{E}_1,\mathcal{U}}(|\psi\rangle\langle\psi|) &= \mathrm{tr}\left(\mathcal{E}_1(|\psi\rangle\langle\psi|)\,\mathcal{U}(|\psi\rangle\langle\psi|)\right) \\ &= \langle\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle \end{aligned} \tag{3.2}$$

where

$$\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}. \tag{3.3}$$

The *average quantum gate fidelity*, denoted $\overline{\mathcal{F}_{\mathcal{E}_1,\mathcal{U}}}$, is obtained by integrating over all pure input states. The integral is taken over the unitarily invariant Haar measure (also known as the Fubini-Study measure) on the set of pure states [45]. In this paper we denote the Fubini-Study measure by μ. This gives,

$$\begin{aligned} \overline{\mathcal{F}_{\mathcal{E}_1,\mathcal{U}}} &= \int \mathrm{tr}\left(\mathcal{E}_1(|\psi\rangle\langle\psi|)\,\mathcal{U}(|\psi\rangle\langle\psi|)\right) d\psi \\ &= \int \langle\psi|\Lambda(|\psi\rangle\langle\psi|)|\psi\rangle d\psi \\ &= \frac{\sum_j \mathrm{tr}(A_j)\mathrm{tr}(A_j) + d}{d^2 + d} \end{aligned} \tag{3.4}$$

where $\{A_j\}$ is any set of Kraus operators for Λ [36]. Thus, the average gate fidelity reduces to an extremely simple form because one of the operations is unitary.

Following this intuition, from Eq.'s (2.4), (2.7), and (3.1), we have that for the ideal ($\mathcal{M}$) and noisy ($\mathcal{E}$) measurements,

$$\begin{aligned} &F(\mathcal{E}(\sigma),\mathcal{M}(\sigma)) \\ &= \left(\mathrm{tr}\sqrt{\sqrt{\sum_{k=1}^b r_k\rho_k}\left(\sum_{k=1}^b p_k\left[\frac{\Pi_k\sigma\Pi_k}{\mathrm{tr}\left(\Pi_k\sigma\right)}\right]\right)\sqrt{\sum_{k=1}^b r_k\rho_k}}\right)^2 \end{aligned} \tag{3.5}$$

where the state-dependence in p_k, r_k, ρ_k is omitted for notational convenience. Unlike the unitary case (Eq. (3.2)), such an expression does not reduce to a simple form in general. However, ideally, the figure of merit we use to distinguish $\mathcal{M}$ and $\mathcal{E}$ will contain evenly weighted information from *all* possible input states and also have direct operational significance. As a result, we define the *quantum average measurement fidelity*, denoted $\overline{F_{\mathcal{E},\mathcal{M}}}$, to be the Haar integral of Eq. (3.5) over pure input states,

$$\overline{F_{\mathcal{E},\mathcal{M}}} = \int F(\mathcal{E}(|\psi\rangle),\mathcal{M}(\langle\psi|)). \tag{3.6}$$

We also define,

$$r_{\mathcal{E},\mathcal{M}} = 1 - \overline{F_{\mathcal{E},\mathcal{M}}} \tag{3.7}$$

to be the *quantum average measurement error*. In the case of only being concerned with measurement probabilities, that is $\mathcal{M}$ and $\mathcal{E}$ are given by Eq.'s (2.6) and (2.8) respectively, we have by Eq. (3.5),

$$\begin{aligned} \overline{F_{\mathcal{E},\mathcal{M}}} &= \int \left(\sum_{k=1}^b \sqrt{p_k r_k}\right)^2 d\psi \\ &= \int \left(\sum_{k=1}^b \sqrt{\mathrm{tr}(\Pi_k|\psi\rangle\langle\psi|)\mathrm{tr}(E_k|\psi\rangle\langle\psi|)}\right)^2 d\psi, \end{aligned} \tag{3.8}$$

$$r_{\mathcal{E},\mathcal{M}} = 1 - \int \left(\sum_{k=1}^b \sqrt{\mathrm{tr}(\Pi_k|\psi\rangle\langle\psi|)\mathrm{tr}(E_k|\psi\rangle\langle\psi|)}\right)^2 d\psi. \tag{3.9}$$

The direct relationship between $r_{\mathcal{E},\mathcal{M}}$ and $\overline{F_{\mathcal{E},\mathcal{M}}}$ implies statements and bounds proven about one directly applies to the other. Since $\overline{F_{\mathcal{E},\mathcal{M}}}$ and $r_{\mathcal{E},\mathcal{M}}$ will not have simple forms like the average gate fidelity in Eq. (3.4), our goal is to provide efficient methods for estimating $r_{\mathcal{E},\mathcal{M}}$ (equivalently $\overline{F_{\mathcal{E},\mathcal{M}}}$) that can be experimentally implemented in a simple manner.

An important reason for using the average measurement fidelity (error) as defined above to distinguish $\mathcal{M}$ and $\mathcal{E}$ is that this provides a state-independent distance measure to compare the operations which can be connected to the diamond norm distance [46] between quantum operations [47]. Computing the diamond norm is an exponentialy hard task since one needs a complete description of the quantum operations. Thus, the diamond norm is neither straightforward to calculate nor experimentally accessible. However, it is commonly used in fault-tolerant analyses for determining threshold error rates of physical operations. Thus, information about the diamond norm provided by the average measurement error defined above can potentially provide information regarding fault-tolerant computation.

IV. EXPERIMENTAL PROTOCOL: RANK-1 PVM'S

In this section, we look at obtaining an upper bound on $r_{\mathcal{E},\mathcal{M}}$ in the case of rank-1 PVM's where we are not concerned with output states. Again, we note that upper bounds on $r_{\mathcal{E},\mathcal{M}}$ are equivalent to lower bounds on $\overline{F_{\mathcal{E},\mathcal{M}}}$ and so for clarity we phrase the following discussion in terms of $r_{\mathcal{E},\mathcal{M}}$.

There are d PVM elements, $\{\Pi_1, ..., \Pi_d\}$, each of which is a rank-1 projection operator. From Eq. (2.8), the outcome of the measurement of a state $|\psi\rangle\langle\psi|$ can be associated to a member from $\{1, ..., d\}$ with frequency distribution

$$(r_1, ..., r_d) = (\mathrm{tr}(E_1|\psi\rangle\langle\psi|), ..., \mathrm{tr}(E_d|\psi\rangle\langle\psi|)). \tag{4.1}$$

Goal: Obtain an upper bound, ub, for $r_{\mathcal{E},\mathcal{M}}$ as defined in Eq. (3.9).

The experimental protocol to obtain ub is as follows:

Protocol:

Step 1: Choose a pair of indices (l, m) uniformly at random from the set $\mathcal{D} = \{0, ..., d-1\} \times \{0, ..., d-1\}$.

Step 2: For each $j \in \{l, m\}$,

a): Prepare the quantum state $|\psi_j\rangle$, perform the noisy measurement $\mathcal{E}$ on $\Pi_j = |\psi_j\rangle$, and record whether outcome "j" is obtained,

b): Repeat a) many times and denote the frequency of obtaining "j" by u_j, that is, $u_j = \mathrm{tr}(\Pi_j E_j)$,

(see Sec. IX A for a discussion of the number of repetitions required to estimate u_j to a desired accuracy and confidence).

Step 3: Repeat Steps 1 and 2 K times, where K is dictated by the desired accuracy and confidence in estimating ub

(see Sec. IX B for a discussion of the size of K).

Step 4: Compute the upper bound of $r_{\mathcal{E},\mathcal{M}}$, ub, via the formula (derived in Sec. V),

$$\begin{aligned} ub &= 1 - \frac{1 + d\overline{X}}{1 + d} \\ &= \frac{d}{1 + d}(1 - \overline{X}) \end{aligned} \tag{4.2}$$

where

$$\overline{X} := \frac{1}{d^2} \sum_{(l,m)\in\mathcal{D}} \sqrt{u_l u_m} \sim \frac{1}{K} \sum_{(k_1,k_2)} \sqrt{u_{k_1} u_{k_2}}, \tag{4.3}$$

and $\{(1_1, 1_2), ..., (K_1, K_2)\}$ are the K trials dictated by Step 3.

This concludes the protocol.

There are various important points about the protocol that should be emphasized. First, the number of trials required in Steps 2b) and 3 are independent of d. Thus, the time-complexity of the entire protocol is independent of d, and depends only on the desired accuracy and confidence of the estimate of ub (see Sec. IX). Second, ub can be computed from the above protocol using only:

1. Applications of the noisy measurement and,

2. The ability to prepare the d pure input states $|\psi_j\rangle$.

Lastly, it is straightforward to show the following two properties of ub, (see Sec. V D)

1. In the limit of $r_{\mathcal{E},\mathcal{M}} \downarrow 0$,

$$ub \downarrow 0, \tag{4.4}$$

2. ub scales well in d.

These are clearly necessary conditions for ub to be a good upper bound on $r_{\mathcal{E},\mathcal{M}}$. Moreover, property 1 implies that, loosely speaking, in the small error limit ub can be taken as an estimate of $r_{\mathcal{E},\mathcal{M}}$.

V. DERIVATION OF THE UPPER BOUND

From Eq.'s (3.7), (4.2), and (4.3), showing ub is an upper bound for $r_{\mathcal{E},\mathcal{M}}$ is equivalent to showing

$$lb := \frac{d + \sum_{k=1}^{d} u_k + \sum_{l\neq m} \sqrt{u_l u_m}}{d(d+1)} \tag{5.1}$$

is a lower bound for $\overline{F_{\mathcal{E},\mathcal{M}}}$. First, we have from Eq. (3.8),

$$F(\mathcal{E}(\sigma), \mathcal{M}(\sigma)) = \left(\sum_k \sqrt{r_k p_k} \right)^2 \tag{5.2}$$

where the state-dependence is implicit in the p_k and r_k. Taking the integral over all pure states gives,

$$\begin{aligned} &\int F(\mathcal{E}(|\psi\rangle\langle\psi|), \mathcal{M}(|\psi\rangle\langle\psi|)) d\psi \\ &= \sum_{k=1}^{d} \left[\int r_k p_k d\psi \right] + \sum_{l\neq m} \left[\int \sqrt{r_l r_m p_l p_m} d\psi \right]. \end{aligned} \tag{5.3}$$

To proceed, by Schur's Lemma, we have that for any positive integer t,

$$\int |\psi\rangle\langle\psi|^{\otimes t} d\psi = \frac{\Pi_{\mathrm{sym}}(t, d)}{\mathrm{tr}\left[\Pi_{\mathrm{sym}}(t, d)\right]} \tag{5.4}$$

where $\Pi_{\rm sym}(t,d)$ is the projector onto the symmetric subspace of the t-partite Hilbert space whose factor spaces are each d-dimensional Hilbert spaces [38, 48]. It is straightforward to show that $\Pi_{\rm sym}(t,d)$ is equal to the normalized sum of the $t!$ elements in the group of permutation operators on the t-partite Hilbert space. We look at the sums in Eq. (5.3) separately.

A. $\sum_{k=1}^{d}\int r_k p_k d\psi$

We have from Eq. (5.4)

$$\begin{aligned}\int r_k p_k d\psi &= \int \mathrm{tr}(E_k|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_k|\psi\rangle\langle\psi|)d\psi \\ &= \int \mathrm{tr}\left([E_k\otimes\Pi_k]\,|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right)d\psi \\ &= \frac{2\mathrm{tr}\left([E_k\otimes\Pi_k]\,\Pi_{\rm sym}(2,d)\right)}{d(d+1)}. \end{aligned} \tag{5.5}$$

Since the group of permutations of two objects consists of the identity and SWAP operations, we have

$$\Pi_{\rm sym}(2,d) = \frac{\mathbb{1}\otimes\mathbb{1}+\mathrm{SWAP}}{2}. \tag{5.6}$$

Therefore,

$$\int r_k p_k d\psi = \frac{\mathrm{tr}(E_k\Pi_k)+d\mathrm{tr}(E_k\frac{\mathbb{1}}{d})}{d(d+1)} \tag{5.7}$$

and so, since $\sum_{k=1}^{d}E_k=\mathbb{1}$,

$$\begin{aligned}\sum_{k=1}^{d}\int r_k p_k d\psi &= \sum_{k=1}^{d}\frac{\mathrm{tr}(E_k\Pi_k)+d\mathrm{tr}(E_k\frac{\mathbb{1}}{d})}{d(d+1)} \\ &= \frac{\sum_{k=1}^{d}u_k+d}{d(d+1)}\end{aligned} \tag{5.8}$$

which gives the first two terms in the numerator in Eq. (5.1).

B. $\sum_{l\neq m}\int\sqrt{r_l r_m p_l p_m}d\psi$

We have,

$$\begin{aligned}&\int\sqrt{r_l r_m p_l p_m}d\psi \\ &= \int\sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(E_l|\psi\rangle\langle\psi|)}\sqrt{\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)\mathrm{tr}(E_m|\psi\rangle\langle\psi|)}d\psi.\end{aligned} \tag{5.9}$$

Computing the above integral is analytically is not possible because of the square root in the argument. Now we assume the following inequality holds (see Sec. VI),

$$\begin{aligned}&\int\sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(E_l|\psi\rangle\langle\psi|)}\sqrt{\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)\mathrm{tr}(E_m|\psi\rangle\langle\psi|)}d\psi \\ &\geq \int\sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_l E_l\Pi_l(|\psi\rangle\langle\psi|))}\sqrt{\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_m E_m\Pi_m(|\psi\rangle\langle\psi|))}d\psi \\ &= \int\sqrt{\langle\psi_l|E_l|\psi_l\rangle}\mathrm{tr}(\Pi_l(|\psi\rangle\langle\psi|))\sqrt{\langle\psi_m|E_m|\psi_m\rangle}\mathrm{tr}(\Pi_m(|\psi\rangle\langle\psi|))d\psi \\ &= \sqrt{u_l u_m}\int\mathrm{tr}(\Pi_l(|\psi\rangle\langle\psi|))\mathrm{tr}(\Pi_m(|\psi\rangle\langle\psi|))d\psi.\end{aligned} \tag{5.10}$$

Note that for each j it is generally not true that $E_j-u_j\Pi_j$ is positive semidefinite. If this were the case then the above inequality would always hold.

From Eq. (5.4) we have,

$$\begin{aligned}&\int\mathrm{tr}(\Pi_l(|\psi\rangle\langle\psi|))\mathrm{tr}(\Pi_m(|\psi\rangle\langle\psi|))d\psi \\ &= \int\mathrm{tr}\left[(\Pi_l\otimes\Pi_m)\,|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|\right]d\psi \\ &= \frac{\mathrm{tr}\left(\Pi_l\Pi_m\right)+\mathrm{tr}\left(\Pi_l\right)\mathrm{tr}\left(\Pi_m\right)}{d(d+1)} \\ &= \frac{1}{d(d+1)}\end{aligned} \tag{5.11}$$

since Π_l and Π_m are projectors onto orthogonal subspaces. Thus,

$$\int \sqrt{r_l r_m p_l p_m} d\psi \geq \frac{\sqrt{\mathrm{tr}\left(E_l \Pi_l\right) \mathrm{tr}\left(E_m \Pi_m\right)}}{d(d+1)} = \frac{\sqrt{u_l u_m}}{d(d+1)} \tag{5.12}$$

which gives the last term in the numerator of Eq. (5.1).

In total, assuming the inequality in Eq. (5.10) holds,

$$\frac{1+d\overline{X}}{1+d} = \frac{d + \sum_{k=1}^{d} u_k + \sum_{l \neq m} \sqrt{u_l u_m}}{d(d+1)} \tag{5.13}$$

is a lower bound for $\overline{F_{\mathcal{E},\mathcal{M}}}$. Before analyzing the validity of the lower bound, we briefly compare its expression with that of the average gate fidelity and show that it satisfies certain necessary conditions for it to be useful.

C. Comparison With Quantum Gate Fidelity

There is a nice parallel between the lower bound on the average measurement fidelity and the exact expression for the average quantum gate fidelity. As previously mentioned, the quantum average gate fidelity takes a particularly nice form because one of the operations is unitary [36]. Indeed, if one compares the unitary $\mathcal{U}$ and quantum operation $\mathcal{E}$, the average gate fidelity between $\mathcal{U}$ and $\mathcal{E}$, $\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}}$, is given by,

$$\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}} = \frac{\overline{A}d+1}{d+1} \tag{5.14}$$

where

$$\overline{A} = \frac{1}{d^2} \sum_k \mathrm{tr}(A_k)\mathrm{tr}(A_k^\dagger) \tag{5.15}$$

and $\{A_k\}$ is any set of Kraus operators for the quantum operation $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$.

Our expression for a lower bound on the average measurement fidelity $\overline{F_{\mathcal{E},\mathcal{M}}}$ takes a similar form,

$$lb = \frac{\overline{X}d+1}{d+1} \tag{5.16}$$

where

$$\overline{X} = \frac{1}{d^2} \sum_{l,m} \sqrt{\mathrm{tr}(\Pi_l E_l)\mathrm{tr}(\Pi_m E_m)}. \tag{5.17}$$

In some sense this is not surprising since each quantity is a Haar integral over functions of two copies of a quantum state $|\psi\rangle\langle\psi|$. A direction of further research is to understand properties of the average measurement fidelity in more detail, and draw more parallels with well-known distinguishability measures such as the average gate fidelity.

D. Necessary Conditions For ub To Be a Useful Upper Bound on $r_{\mathcal{E},\mathcal{M}}$

In this subsection, we discuss two necessary conditions lb must satisfy in order to be a useful lower bound on $\overline{F_{\mathcal{E},\mathcal{M}}}$. By the equivalence between ub being an upper bound for $r_{\mathcal{E},\mathcal{M}}$ and lb being a lower bound for $\overline{F_{\mathcal{E},\mathcal{M}}}$, these conditions are equivalent to corresponding conditions for ub.

Two necessary conditions that lb must satisfy in order to be a useful lower bound on $\overline{F_{\mathcal{E},\mathcal{M}}}$ are,

1. Limit of No Error: As the measurement error goes to 0, $lb \uparrow 0$, and,
2. Scaling in d: lb scales well in the dimension d of the system.

Here, we show lb satisfies both of these criteria.

1. *Limit of No Error*

Suppose the measurement error goes to 0 in that the POVM elements converge to the ideal PVM elements,

$$E_j \to \Pi_j. \tag{5.18}$$

This implies for each $j \in \{1, ..., d\}$,

$$u_j = \mathrm{tr}(\pi_j E_j) \uparrow 1. \tag{5.19}$$

Hence, from the set of equations in Eq. (6.1), for each (l, m),

$$g_{l,m}(|\psi\rangle) \uparrow f_{l,m}(|\psi\rangle) \tag{5.20}$$

which implies,

$$lb \uparrow 1. \tag{5.21}$$

Thus lb satisfies the necessary condition of converging to 1 in the limit of no errors.

2. *Scaling in d*

Let us now show that lb scales well in d. By Eq. (6.2), suppose that for each (l, m), $\epsilon_{l,m} \geq 0$ is such that,

$$\int f_{l,m}(|\psi\rangle)d\psi - \int g_{l,m}(|\psi\rangle)d\psi = \epsilon_{l,m}. \tag{5.22}$$

This is equivalent to,

$$\int f_{l,m}(|\psi\rangle)d\psi = \frac{\sqrt{u_l u_m}}{d(d+1)} + \epsilon_{l,m}. \tag{5.23}$$

Since $\epsilon_{l,m} = 0$ if $l = m$, summing over all $l \neq m$ gives,

$$\sum_{l\neq m} \int f_{l,m}(|\psi\rangle) d\psi = \sum_{l\neq m} \frac{\sqrt{u_l u_m}}{d(d+1)} + \sum_{l\neq m} \epsilon_{l,m}. \quad (5.24)$$

Assume the error in the POVM elements remains roughly independent of d in that $\mathrm{tr}(\Pi_j E_j)$ is independent of d. Let us suppose for the sake of clarity of the argument that

$$\mathrm{tr}(\Pi_j E_j) = 1 - \delta \quad (5.25)$$

for some small $\delta > 0$ independent of j and d. Then,

$$\begin{aligned} \sum_{l\neq m} \int f_{l,m}(|\psi\rangle) d\psi = & \sum_{l\neq m} \frac{\sqrt{(1-\delta)^2}}{d(d+1)} + \sum_{l\neq m} \epsilon_{l,m} \\ = & \frac{(1-\delta)d(d-1)}{d(d+1)} + \sum_{l\neq m} \epsilon_{l,m} \\ = & \frac{(1-\delta)(d-1)}{(d+1)} + \sum_{l\neq m} \epsilon_{l,m}. \end{aligned} \quad (5.26)$$

Hence as $d \to \infty$

$$\sum_{l\neq m} \int f_{l,m}(|\psi\rangle) d\psi \to (1-\delta) + \sum_{l\neq m} \epsilon_{l,m}. \quad (5.27)$$

Define

$$\epsilon = \sum_{(l,m)} \epsilon_{l,m}. \quad (5.28)$$

Then,

$$\overline{F_{\mathcal{E},\mathcal{M}}} = \sum_{(l,m)} \int f_{l,m}(|\psi\rangle) d\psi \to (1-\delta) + \epsilon \quad (5.29)$$

and so, since $\overline{F_{\mathcal{E},\mathcal{M}}} \leq 1$, each value of $\epsilon_{l,m}$ must scale as $\frac{1}{d^2}$. Hence, lb scales well in large dimensions since it essentially depends on the size of the u_k (ie. $1-\delta$).

VI. VALIDITY OF THE UPPER BOUND

The upper bound in Eq. (4.2) is valid provided the inequality in Eq. (5.10) holds. The goal of this section is to show that this inequality holds in very general situations. To set notation, we define

$$\begin{aligned} f_{l,m}(|\psi\rangle) &= \sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(E_l|\psi\rangle\langle\psi|)}\sqrt{\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)\mathrm{tr}(E_m|\psi\rangle\langle\psi|)}, \\ g_{l,m}(|\psi\rangle) &= \sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_l E_l \Pi_l|\psi\rangle\langle\psi|)}\sqrt{\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_m E_m \Pi_m|\psi\rangle\langle\psi|)} \\ &= \sqrt{u_l u_m}\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|) \end{aligned} \quad (6.1)$$

so that we want to show,

$$\int [f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)] d\psi \geq 0. \quad (6.2)$$

We have,

$$\int [f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)] d\psi = \int k_{l,m}(|\psi\rangle) h_{l,m}(|\psi\rangle) d\psi \quad (6.3)$$

where

$$k_{l,m}(|\psi\rangle) := \sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)} \quad (6.4)$$

$$h_{l,m}(|\psi\rangle) := \sqrt{\mathrm{tr}(E_l|\psi\rangle\langle\psi|)\mathrm{tr}(E_m|\psi\rangle\langle\psi|)} \quad (6.5)$$

$$- \sqrt{u_l u_m}\sqrt{\mathrm{tr}(\Pi_l|\psi\rangle\langle\psi|)\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)}. \quad (6.6)$$

Since $k_{l,m}(|\psi\rangle) \geq 0$ for all $|\psi\rangle$, we have,

$$f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle) < 0 \Leftrightarrow h_{l,m}(|\psi\rangle) < 0. \quad (6.7)$$

However, $h_{l,m}(|\psi\rangle) < 0$ if and only if one of

$$\langle\psi|E_l - u_l\Pi_l|\psi\rangle < 0 \quad (6.8)$$

or

$$\langle\psi|E_m - u_m\Pi_m|\psi\rangle < 0 \quad (6.9)$$

is true. Without loss of generality, suppose $\langle\psi|E_l - u_l\Pi_l|\psi\rangle < 0$ holds.

As a matrix written in the $\{|\psi_k\rangle\}$ basis, we have $E_l - u_l\Pi_l$ is equal to E_l except in the (l,l)'th entry which is 0. If $|\psi\rangle = |\psi_l\rangle$ then $\langle\psi|E_l - u_l\Pi_l|\psi\rangle = 0$, and if $|\psi\rangle$ is orthogonal to $|\psi_l\rangle$, $\langle\psi|E_l - u_l\Pi_l|\psi\rangle \geq 0$. Hence, Eq. (6.8) can only be satisfied if $|\psi\rangle$ has large enough amplitude in the subspace defined by $|\psi_l\rangle$. The amount of amplitude that is required depends on the size of the off-diagonal elements (coherence) in the l'th column (row) of E_l. If the coherence is not too large then $|\psi\rangle$ will need to have large amplitude in $|\psi_l\rangle$, and as the coherence grows larger, $|\psi\rangle$ can potentially satisfy Eq. (6.8) without being close to $|\psi_l\rangle$.

The key point is that, if $|\psi\rangle$ has large amplitude in $|\psi_l\rangle$, and so there is a chance of $\langle\psi|E_l - u_l\Pi_l|\psi\rangle < 0$, then $\mathrm{tr}(\Pi_m|\psi\rangle\langle\psi|)$ is small. Thus, $k_{l,m}(|\psi\rangle)$ will be small and so $k_{l,m}(|\psi\rangle)$ acts like a *modulating factor* in the expression for $f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)$ to ensure

$f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)$ is small when $|\psi\rangle$ is close to $|\psi_l\rangle$ (or $|\psi_m\rangle$ by symmetry). Moreover, the measure of the set of states for which $|\psi\rangle$ is close to $|\psi_l\rangle$ (or $|\psi_m\rangle$) is small, and by Levy's Lemma [49] decreases exponentially in the dimension d. Thus, we have that the measure of states for which $|\psi\rangle$ is close to $|\psi_l\rangle$ (or $|\psi_m\rangle$) is small and, for such a state, $f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)$ is small in magnitude.

In total, if the POVM element E_l does not have strong coherence then, since $|\psi\rangle$ must be close to $|\psi_l\rangle$ to make $f_{l,m} - g_{l,m} < 0$, we expect

$$\int [f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle)] d\psi \geq 0. \tag{6.10}$$

Hence, our bounds should be valid for most practical cases since, realistically, the coherence will not be overwhelmingly large. Note that the bounds are guaranteed to be valid if there is no coherence at all, that is, if the POVM elements are diagonal in the $\{|\psi_k\rangle\}$ basis. In the next section we perform the above analysis for a single qubit and in Sec. VI B we numerically investigate the single-qubit case. Interestingly, we find that for *all* values of coherence magnitude, the bounds are valid. This provides evidence that the bounds will be valid in large generality. Clearly, a more in-depth investigation of sufficient conditions for the bounds to be valid is desirable, however the above argument and numerical results of Sec. (VI B) indicate that the bounds and protocol will be valid in most practical cases.

A. Single-Qubit

Our goal in this section is to obtain a lower bound for

$$\int f_{l,m}(|\psi\rangle) - g_{l,m}(|\psi\rangle) d\psi \tag{6.11}$$

and provide sufficient conditions for when Eq. (6.2) holds in the case of a single qubit. We use the Bloch sphere representation of a single qubit:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{\mathrm{i}\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle \tag{6.12}$$

where $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$. As the results will depend on the magnitude of the coherence in the POVM elements, we assume without loss of generality that $\gamma \in \mathbb{R}$ and $\gamma > 0$.

$$E_0 = \begin{pmatrix} u_0 & \gamma \\ \gamma & \mathrm{tr}(E_0\Pi_1) \end{pmatrix}, \tag{6.13}$$

$$E_1 = \begin{pmatrix} \mathrm{tr}(E_1\Pi_0) & -\gamma \\ -\gamma & u_1 \end{pmatrix}, \tag{6.14}$$

Since our goal is obtaining a lower bound on $\int (f_{0,1} - g_{0,1})(|\psi\rangle) d\psi$, let us attempt to understand when one of Eq.'s (6.8) or (6.9) is true (ie. $\langle\psi|E_0 - u_0\Pi_0|\psi\rangle < 0$ or $\langle\psi|E_1 - u_1\Pi_1|\psi\rangle < 0$) since this is a necessary condition for $h_{l,m}(|\psi\rangle) < 0$. Without loss of generality, let us assume

$$\langle\psi|E_0 - u_0\Pi_0|\psi\rangle < 0. \tag{6.15}$$

We have,

$$\begin{aligned}\langle\psi|E_0 - u_0\Pi_0|\psi\rangle = 2\mathrm{Re}\left(\gamma\cos\left(\frac{\theta}{2}\right)e^{i\phi}\sin\left(\frac{\theta}{2}\right)\right) \\ + \mathrm{tr}(E_0\Pi_1)\sin^2\left(\frac{\theta}{2}\right)\end{aligned} \tag{6.16}$$

and it is straightforward to show $\langle\psi|E_0 - u_0\Pi_0|\psi\rangle < 0$ is satisfied if

$$\begin{aligned}\phi &\in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \\ \theta &\in \left[0, 2\,\mathrm{arccot}\left(-\frac{\mathrm{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)\right].\end{aligned} \tag{6.17}$$

Hence, denoting the set of states that satisfy $\langle\psi|E_0 - u_0\Pi_0|\psi\rangle < 0$ by $A_{0,\gamma}$, we have that $A_{0,\gamma}$ is contained in the set of all states with ϕ and θ given by Eq. (6.17). As expected, the measure of $A_{0,\gamma}$, $\mu(A_{0,\gamma})$, is extremely small for weak coherence and grows larger as the coherence increases in magnitude.

One can obtain an upper bound for $\mu(A_{0,\gamma})$ by noting that the uniform measure on the Bloch sphere has density function

$$\frac{1}{4\pi}\sin(\theta). \tag{6.18}$$

Hence, $\mu(A_{0,\gamma})$ is at most

$$\frac{1}{4} - \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\left(2\,\mathrm{arccot}\left(-\frac{\mathrm{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)\right) d\phi. \tag{6.19}$$

Now, if $|\psi\rangle \in A_{0,\gamma}$, $\theta < 2\,\mathrm{arccot}\left(-\frac{\mathrm{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)$, and so,

$$\begin{aligned}\mathrm{tr}(\Pi_1|\psi\rangle\langle\psi|) &\leq \sin^2\left(\mathrm{arccot}\left(-\frac{\mathrm{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)\right) \\ &= \frac{\cos^2(\phi)}{\cos^2(\phi) + \left(\frac{\mathrm{tr}(E_0\Pi_1)}{2\gamma}\right)^2}.\end{aligned} \tag{6.20}$$

Hence, we have that for weak coherence γ, the set of states which satisfy Eq. (6.15) will have small measure (bounded by Eq. (6.19)) and by Eq. (6.20), the value of $g_{0,1}$ on these states is small. Using a symmetric argument for the state $|1\rangle$ gives the following set of equations,

$$\mu(A_{0,\gamma}) \leq \frac{1}{4} - \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\left(2\text{arccot}\left(-\frac{\text{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)\right),$$
$$\mu(A_{1,\gamma}) \leq \frac{1}{4} - \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\left(2\text{arccot}\left(-\frac{\text{tr}(E_1\Pi_0)}{2\gamma\cos(\phi)}\right)\right),$$
$$\text{If } |\psi\rangle \in A_{0,\gamma}\ ,\ \text{tr}(\Pi_1|\psi\rangle\langle\psi|) \leq \frac{\cos^2(\phi)}{\cos^2(\phi) + \left(\frac{\text{tr}(E_0\Pi_1)}{2\gamma}\right)^2} \text{ with } \phi \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right),$$
$$\text{If } |\psi\rangle \in A_{1,\gamma}\ ,\ \text{tr}(\Pi_0|\psi\rangle\langle\psi|) \leq \frac{\cos^2(\phi)}{\cos^2(\phi) + \left(\frac{\text{tr}(E_1\Pi_0)}{2\gamma}\right)^2} \text{ with } \phi \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right). \tag{6.21}$$

Using these results we can obtain a lower bound for $\int h_{0,1}(|\psi\rangle)d\psi = \int (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi$. We have,

$$\int (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi = \sum_{j=0}^{1}\int_{A_{j,\gamma}} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi + \int_{(A_{0,\gamma}\cup A_{1,\gamma})^c} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi. \tag{6.22}$$

Using the set of equations in Eq. (6.21) one can show

$$\begin{aligned}&\int_{A_{0,\gamma}} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi \\ &\geq \int_{A_{0,\gamma}} 0 - \sqrt{u_0 u_1}\text{tr}(\Pi_0|\psi\rangle\langle\psi|)\text{tr}(\Pi_1|\psi\rangle\langle\psi|)d\psi \\ &= -\frac{\sqrt{u_0 u_1}}{2\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \sin^4\left(\text{arccot}\left(-\frac{\text{tr}(E_1\Pi_0)}{2\gamma\cos(\phi)}\right)\right)d\phi.\end{aligned} \tag{6.23}$$

By Eq. (6.20),

$$\begin{aligned}&\int_{A_{0,\gamma}} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi \\ &\geq -\frac{\sqrt{u_0 u_1}}{2\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \left[\frac{\cos^2(\phi)}{\cos^2(\phi) + \left(\frac{\text{tr}(E_0\Pi_1)}{2\gamma}\right)^2}\right]^2 d\phi \\ &=: -\delta_{0,\gamma}\end{aligned} \tag{6.24}$$

and similarly,

$$\begin{aligned}&\int_{A_{1,\gamma}} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi \\ &\geq -\frac{\sqrt{u_0 u_1}}{2\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \left[\frac{\cos^2(\phi)}{\cos^2(\phi) + \left(\frac{\text{tr}(E_1\Pi_0)}{2\gamma}\right)^2}\right]^2 d\phi \\ &=: -\delta_{1,\gamma}.\end{aligned} \tag{6.25}$$

Hence, in total,

$$\begin{aligned}&\int (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi \\ &\geq -(\delta_{0,\gamma} + \delta_{1,\gamma}) + \int_{(A_{0,\gamma}\cup A_{1,\gamma})^c} (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi\end{aligned} \tag{6.26}$$

where

$$\begin{aligned}&\mu\left((A_{0,\gamma}\cup A_{1,\gamma})^c\right) \\ &\geq \frac{1}{2} + \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\left(2\text{arccot}\left(-\frac{\text{tr}(E_0\Pi_1)}{2\gamma\cos(\phi)}\right)\right) \\ &\quad + \frac{1}{4\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\left(2\text{arccot}\left(-\frac{\text{tr}(E_1\Pi_0)}{2\gamma\cos(\phi)}\right)\right)\end{aligned} \tag{6.27}$$

and if $|\psi\rangle \in (A_{0,\gamma}\cup A_{1,\gamma})^c$,

$$[f_{0,1} - g_{0,1}](|\psi\rangle) \geq 0. \tag{6.28}$$

Eq. (6.26) provides an explicit lower bound on $\int h_{0,1}(|\psi\rangle)d\psi = \int (f_{0,1} - g_{0,1})(|\psi\rangle)d\psi$. Next, we analyze numerical results for the single-qubit case

B. Numerical Examples

We now look at some simple single-qubit examples to help clarify the technical details presented thus far. In order to perform integration over the Fubini-Study measure on single-qubit states, we use the Bloch sphere representation of a single qubit state,

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\phi}|1\rangle, \tag{6.29}$$

where $\theta \in [0,\pi]$, $\phi \in [0, 2\pi]$. Fubini-Study integration reduces to uniform integration over the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$. For any function $f(\theta,\phi)$ we have that the integral

of f with respect to this uniform probability measure is the double integral,

$$\frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} f(\theta,\phi)\sin(\theta)d\theta d\phi. \tag{6.30}$$

The normalization $\frac{1}{4\pi}$ ensures we have a probability measure. We have the following quantities,

$$\begin{aligned} p_0 &= \cos^2\left(\frac{\theta}{2}\right), \\ p_1 &= 1-p_0, \\ r_0 &= E_0^{0,0}\cos^2\left(\frac{\theta}{2}\right) + E_0^{1,1}\sin^2\left(\frac{\theta}{2}\right) + \mathrm{Re}\left(E_0^{0,1}e^{i\phi}\right)\sin(\theta), \\ r_1 &= 1-r_0. \end{aligned} \tag{6.31}$$

It is straightforward to show,

$$\begin{aligned} &\int \sqrt{p_0 r_0 p_1 r_1}\,d\psi \\ &= \frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \sqrt{\left[E_0^{0,0}\cos^4\left(\frac{\theta}{2}\right)\sin^2\left(\frac{\theta}{2}\right) + E_0^{1,1}\cos^2\left(\frac{\theta}{2}\right)\sin^4\left(\frac{\theta}{2}\right) + 2\mathrm{Re}\left(E_0^{0,1}e^{i\phi}\right)\cos^3\left(\frac{\theta}{2}\right)\sin^3\left(\frac{\theta}{2}\right)\right]} \\ &\quad \sqrt{\left[1 - E_0^{0,0}\cos^2\left(\frac{\theta}{2}\right) - E_0^{1,1}\sin^2\left(\frac{\theta}{2}\right) - 2\mathrm{Re}\left(E_0^{0,1}e^{i\phi}\right)\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right)\right]}\, d\phi\sin(\theta)d\theta. \end{aligned} \tag{6.32}$$

With Eq. (6.32) in hand, we look at the following examples.

Suppose our goal is to characterize a single-qubit projective measurement in the computational basis. The POVM elements $\{E_0, E_1\}$ that model the noise must each be positive semidefinite and $E_0+E_1 = \mathbb{1}$. Hence, we write

$$E_0 = \begin{pmatrix} u_0 & \gamma \\ \gamma^* & 1-u_0 \end{pmatrix}, \tag{6.33}$$

$$E_1 = \begin{pmatrix} 1-u_0 & -\gamma \\ -\gamma^* & u_0 \end{pmatrix}, \tag{6.34}$$

where $\gamma \in \mathbb{C}$ is given by

$$\gamma = Re^{i\phi} \tag{6.35}$$

and $\phi \in [0, 2\pi)$. We have the linear operator E_0 (E_1) is positive semidefinite if and only if its leading principal minors are non-negative. Since $d = 2$, this reduces to the condition,

$$|\gamma| = R \leq \sqrt{u_0(1-u_0)}. \tag{6.36}$$

For various cases, we observe how the exact measurement fidelity compares to ub as R approaches its maximum value. Since the results will only depend on R and not ϕ, without loss of generality, one could assume $\gamma = R$. As a verification of this, the numerics given below were performed for various $\phi \in [0, 2\pi)$ and, as expected, the results were independent of ϕ.

Three of the cases we analyzed are $u_0 = 0.99$, 0.995, and 0.999. The values of lb and ub for each case are contained in Table I. Plots of the average measurement error $r_{\mathcal{E},\mathcal{M}}$ as a function of $|\gamma|$ for each case are given in Fig. 1.

TABLE I. Values of lb and ub for 3 Cases of u_0.

u_0	0.99	0.995	0.999
lb	0.9933	0.9967	0.9993
ub	0.0067	0.0033	0.0007

Various important features are evident from Fig. 1. First, ub is *always* an upper bound for $r_{\mathcal{E},\mathcal{M}}$ (the curves in the plot do not intersect). This agrees with the discussion in Sec. (V) where we noted that ub should be a valid upper bound in large generality (even for large coherence) due to both the anti-correlation of the random variables $\mathrm{tr}(\Pi_0|\psi\rangle\langle\psi|)$ and $\mathrm{tr}(\Pi_1|\psi\rangle\langle\psi|)$, and the measure of the set of states for which $g_{0,1} > f_{0,1}$ being small. Second, as expected, the difference between ub and $r_{\mathcal{E},\mathcal{M}}$ grows smaller as the magnitude of the coherence increases. Thus, the bounds become more tight as the coherence in the POVM elements increases.

These results suggest that the bounds are always valid regardless of how large $|\gamma|$ is relative to u_0. We tested whether there were any violations when u_0 was varied from 0.9 to 0.9999 in increments of 10^{-4}, and the coherence was varied from its minimum to maximum in each case. We found *no* violations of the upper bound.

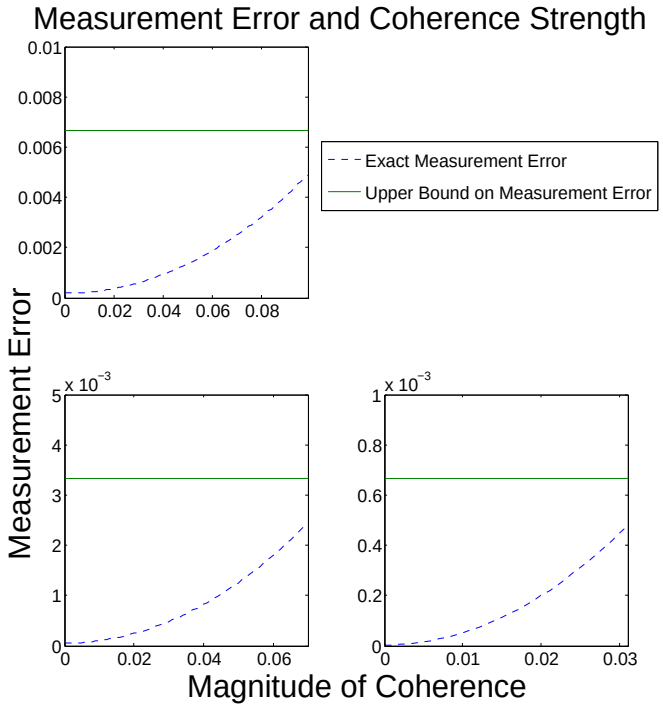

FIG. 1. Plot of exact average measurement error $r_{\mathcal{E},\mathcal{M}}$ (blue-dashed) against coherence strength $|\gamma|$ for three different values of u_0: 0.99 (upper left), 0.995 (bottom left), and 0.999 (bottom right). The upper bound ub for $r_{\mathcal{E},\mathcal{M}}$ in each case is also given (green).

VII. CHARACTERIZING MEASUREMENT PROBABILITIES AND OUTPUT STATES FOR RANK-1 PVM'S

Ideally, a quantum measurement produces an output state that can be utilized for further purposes. In these cases one would like to characterize the error on both the measurement statistics and output states of the measurement. In this section we briefly outline such a method for rank-1 PVM's and discuss conditions for the upper bound on the measurement error $r_{\mathcal{E},\mathcal{M}}$ to be valid.

From Eq. (2.4) we have that $\mathcal{M}$ is given by,

$$\mathcal{M}(\sigma) = \sum_{k=1}^{d} \Pi_k \sigma \Pi_k = \sum_{k=1}^{b} p_k \Pi_k \tag{7.1}$$

where

$$p_k = \mathrm{tr}\left(\Pi_k \sigma\right) \tag{7.2}$$

and the output state Π_k is *independent* of the input state σ. Note that this assumption is only valid because the ideal measurement is a rank-1 PVM. We assume the noisy measurement process $\mathcal{E}$ is of the form,

$$\mathcal{E}(\sigma) = \sum_{k=1}^{b} r_k \rho_k \tag{7.3}$$

where, as before,

$$r_k = \mathrm{tr}(E_k \sigma) \tag{7.4}$$

is modeled via a POVM $\{E_k\}_{k=1}^{d}$. We assume the ρ_k are independent of σ since, ideally, the output states Π_k are independent of σ. The average measurement fidelity is still given by

$$\begin{aligned} \overline{F_{\mathcal{E},\mathcal{M}}} &= \int F_{\mathcal{E},\mathcal{M}}(|\psi\rangle\langle\psi|) d\psi \\ &= \int F\left(\sum_k r_k \rho_k, \sum_k p_k \Pi_k\right) d\psi \end{aligned} \tag{7.5}$$

and, as before,

$$r_{\mathcal{E},\mathcal{M}} = 1 - \int F_{\mathcal{E},\mathcal{M}}(|\psi\rangle\langle\psi|) d\psi. \tag{7.6}$$

In this case where we are also interested in output states, the expression for $\overline{F_{\mathcal{E},\mathcal{M}}}$ is much more complicated. However, we can use strong concavity of the fidelity [50] to obtain,

$$\begin{aligned} \overline{F_{\mathcal{E},\mathcal{M}}} &= \int F(\mathcal{E}(|\psi\rangle), \mathcal{M}(\langle\psi|)) \\ &\geq \int \left(\sum_k \sqrt{r_k p_k} \sqrt{\mathrm{tr}\left(\rho_k \Pi_k\right)}\right)^2 d\psi. \end{aligned} \tag{7.7}$$

Since $\mathrm{tr}\left(\rho_k \Pi_k\right)$ is just the fidelity between the k'th ideal and noisy output states of the measurement, we define,

$$F_k := F(\rho_k, \Pi_k) = \mathrm{tr}\left(\rho_k \Pi_k\right). \tag{7.8}$$

Moreover, since there is only state-dependence in p_k and r_k, we have,

$$\overline{F_{\mathcal{E},\mathcal{M}}} \geq \sum_{(l,m)\in\mathcal{D}} \left[\int \sqrt{r_l r_m p_l p_m} d\psi\right] \sqrt{F_l F_m}. \tag{7.9}$$

From Sec. V this gives,

$$\overline{F_{\mathcal{E},\mathcal{M}}} \geq \frac{1 + d\overline{Y}}{1 + d} \tag{7.10}$$

where

$$\overline{Y} := \frac{1}{d^2} \sum_{(l,m)\in\mathcal{D}} \sqrt{u_l u_m} \sqrt{F_l F_m}. \tag{7.11}$$

At this point, we could define $lb = \frac{1+d\overline{Y}}{1+d}$ (and thus $ub = 1 - \frac{1+d\overline{Y}}{1+d}$), except the F_k cannot be obtained directly from the noisy measurement $\mathcal{E}$. As our goal is an experimentally implementable protocol to obtain bounds on $\overline{F_{\mathcal{E},\mathcal{M}}}$ and $r_{\mathcal{E},\mathcal{M}}$, we need to find a lower bound for this quantity in terms of accessible quantities. Note that the values $\mathrm{tr}\,(\rho_k E_k)$ are accessible in that, if the output of the measurement is "k" and one feeds the output state ρ_k into a subsequent measurement, then the probability of obtaining output "k" in the second measurement is $\mathrm{tr}\,(\rho_k E_k)$. Let us denote Q_k to be this probability,

$$\begin{aligned} Q_k &= \mathbb{P}\left[\text{obtaining outcome } k \text{ from } \mathcal{E} \text{ if the input is } \rho_k\right] \\ &= \mathrm{tr}(\rho_k E_k) \end{aligned} \tag{7.12}$$

where the notation "$\mathbb{P}(\cdot)$" means "probability of". Thus, *if* it is true that for each k,

$$F_k = \mathrm{tr}\,(\rho_k \Pi_k) \geq Q_k = \mathrm{tr}\,(\rho_k E_k)\,, \tag{7.13}$$

then we have,

$$\begin{aligned} \overline{F_{\mathcal{E},\mathcal{M}}} &\geq \frac{1+d\overline{Z}}{1+d}, \\ r_{\mathcal{E},\mathcal{M}} &\leq \frac{d}{1+d}(1-\overline{Z}), \end{aligned} \tag{7.14}$$

where

$$\overline{Z} := \frac{1}{d^2} \sum_{(l,m)\in\mathcal{D}} \sqrt{u_l u_m}\sqrt{Q_l Q_m}. \tag{7.15}$$

Motivation for why the assumption in Eq. (7.13) will often be true is that ideally, for each k, $E_k = \Pi_k$. Hence, $\mathrm{tr}\,(\rho_k \Pi_k)$ is just the k'th diagonal element of ρ_k, $\rho_k^{k,k}$. Since $E_k \leq \mathbb{1}$, if E_k is diagonal in the $\{|\psi_k\rangle\}$ basis, then Eq. (7.13) is guaranteed to hold. This is because the probability amplitude of Π_k is spread across the diagonal elements of E_k. Even when there is coherence in the E_k (non-zero off-diagonal elements) we expect the inequality should still hold in realistic situations, however a more detailed analysis of conditions for this to occur are required. Note also that, realistically, one does not need Eq. (7.13) to hold for every k, just a large enough number to ensure $\overline{Y} \geq \overline{Z}$. We now provide the experimental protocol for determining ub.

VIII. EXPERIMENTAL PROTOCOL: RANK-1 PVM'S WITH OUTPUT STATES

Goal: Obtain ub in the case of imperfect measurement probabilities *and* output states.

Protocol:

Step 1: Choose a pair of indices (l, m) uniformly at random from $\mathcal{D} = \{0, ..., d-1\} \times \{0, ..., d-1\}$.

Step 2: For each $j \in \{l, m\}$,

a) Prepare the quantum state Π_j, perform the noisy measurement $\mathcal{E}$ on Π_j, and record whether outcome "j" is obtained,

b) If outcome "j" is obtained in a), repeat the measurement on the current state of the system, and record whether outcome "j" is obtained again,

c) Repeat Steps a) and b) many times and denote the frequency of obtaining "j" in each Step by u_j and Q_j respectively.

(see below for a discussion of the number of repetitions required to estimate u_j and Q_j to desired accuracies).

Step 3: Repeat Steps 1 and 2 K times, where K is dictated by the desired accuracy and confidence in estimating ub

(see below for a discussion of the size of K).

Step 4: Compute the upper bound ub of $r_{\mathcal{E},\mathcal{M}}$ (Eq. (7.14)) to accuracy and confidence dictated by the K trials $\{(1_1, 1_2), ..., (K_1, K_2)\}$ from Step 3,

$$\begin{aligned} ub &= \frac{d}{1+d}(1-\overline{Z}) \\ &\sim \frac{d}{1+d}\left(1 - \frac{1}{K}\sum_{(k_1,k_2)} \sqrt{u_{k_1}u_{k_2}}\sqrt{Q_{k_1}Q_{k_2}}\right). \end{aligned} \tag{8.1}$$

This concludes the protocol.

We should emphasize various points about the protocol. First, similar to the protocol with no output states (see Sec IV), the number of trials required to implement the above protocol is independent of d, and only depends on the desired accuracy and confidence of the estimates in each of Steps 2 c) and 3. See Sec.'s IX A and IX B for respective discussions about the time-complexity of each of these steps.

Second, ub can be computed using only:

1. Applications of the noisy measurement (sequentially on output states as well) and,
2. The ability to prepare each of the d pure input states Π_j.

Lastly, it is straightforward to show that the two necessary conditions given previously in Sec. V D for ub (lb) to be a useful upper (lower) bound also hold here. Indeed,

1. In the limit of $r_{\mathcal{E},\mathcal{M}} \downarrow 0$,

$$ub \downarrow 0 \tag{8.2}$$

and,

2. *ub* scales well in d.

The first condition holds since as $r_{\mathcal{E},\mathcal{M}} \downarrow 0$ it must be the case that $E_k \to \Pi_k$ and $\rho_k \to \Pi_k$. Therefore, for every k, $u_k \to 1$ and $Q_k \to 1$. The second condition holds using an analogous argument as that given in Sec. V D 2

IX. RESOURCE ANALYSIS

In this section we discuss the time-complexity and resources required for the protocols in detail, and also compare the time-complexity of the protocol with a full reconstruction of the noisy measurement (ie. a full reconstruction of the POVM elements E_k). First, we analyze the number of trials required in Step 2b) of the protocol in Sec. IV and Step 2c) of the protocol in Sec. VIII. Afterwards, we analyze the number of trials required in Step 3 of both protocols.

A. Number of Trials Required in Step 2b) of Sec. IV and Step 2c) of Sec. VIII

We first explicitly analyze the number of trials for Step 2b) of Sec. IV. The discussion carries over in a straightforward manner to Step 2c) of Sec. VIII.

1. *Step 2b) of Sec.IV*

For each $k \in \{0, 1, ..., d-1\}$, we would like to understand how many samples are required to obtain an estimate of u_k. Let us fix k and define $z_k = \text{tr}(\Pi_k(\mathbb{1} - E_k))$ where, since $\{E_k\}_{k=1}^d$ is a POVM,

$$\mathbb{1} - E_k = \sum_{j \neq k}^{d} E_j. \tag{9.1}$$

Then,

$$\begin{aligned} z_k &\geq 0, \\ u_k + z_k &= \text{tr}(\Pi_k) = 1, \end{aligned} \tag{9.2}$$

so $\{u_k, z_k\}$ forms a probability distribution on the binary measurement outcome space $\{$"k", "not k"$\}$.

Let $\hat{u}_k$ be the estimator of u_k obtained in Sub-step 2b of the protocol. Since we are trying to find a parameter that defines an underlying probability distribution, we can use maximum likelihood estimation (MLE) techniques [51]. As the trials are independent and identically distributed, standard MLE gives the estimate of u_k to be equal to the frequency of obtaining outcome "k", as in Step 4. Thus, if N is the number of trials and n_k is the number of times outcome "k" is observed, standard MLE gives the obvious estimate,

$$\hat{u}_k = \frac{n_k}{N}. \tag{9.3}$$

Since, in practice, u_k will be close to 1, one could utilize additive smoothing methods for MLE, which assign higher weight to low probability outcomes. In our case there is one low probability outcome "not k". The formula for $\hat{u}_k$ under a λ-additive smoothing method is,

$$\hat{u}_k = \frac{n_k + \lambda}{N + 2\lambda}, \tag{9.4}$$

where the factor of 2 multiplying λ in the denominator arises from there being two possible outcomes of the experiment, and is required to ensure $\{\hat{u}_k, 1 - \hat{u}_k\}$ is a probability distribution. Note that when this smoothing technique is used, it will fairly bias the estimate of u_k to be smaller than the actual value of u_k, so the estimation will be honest.

The key point in terms of time complexity is that the number of trials N required to estimate u_k to accuracy ϵ with confidence $1 - \delta$ is *independent* of the size of the system. Since the set of possible probability distributions $\{u_k, z_k\}$ satisfies certain consistency conditions, the estimator $\hat{u}_k$ converges in distribution to u_k,

$$\hat{u}_k \xrightarrow{D} u_k. \tag{9.5}$$

Moreover, the mean of $\hat{u}_k$ is equal to u_k and the variance of $\hat{u}_k$ scales as

$$\text{Var}\left(\hat{u}_k\right) \sim \frac{1}{NI(\{u_k, z_k\})} \tag{9.6}$$

where $I(\{u_k, z_k\})$ is the Fisher information [52] of one observation of the true probability distribution. The Fisher information contained in one observation is a measure of how much information about u_k is gained on average from observing $\{$"k", "not k"$\}$ (with distribution $\{u_k, z_k\}$). If $u_k \sim 1$ then the amount of information gained on average is large.

Since $\{$"k", "not k"$\}$ is a Bernoulli random variable, it is possible to explicitly compute the Fisher information of one observation,

$$I(\{u_k, z_k\}) = \frac{1}{u_k z_k} = \frac{1}{u_k(1 - u_k)}. \tag{9.7}$$

Hence, from Eq.'s (9.6) and (9.7), we have that

$$\text{Var}\left(\hat{u}_k\right) \sim \frac{u_k(1 - u_k)}{N}. \tag{9.8}$$

Now, suppose we want to estimate u_k to accuracy ϵ with confidence $1 - \delta$, ie.

$$\mathbb{P}\left(|\hat{u}_k - u_k| \geq \epsilon\right) \leq \delta. \tag{9.9}$$

By Chebyshev's theorem we have for any integer $j \geq 1$,

$$\mathbb{P}\left(|\hat{u}_k - u_k| \geq j\sigma(\hat{u}_k)\right) \leq \frac{1}{j^2}. \tag{9.10}$$

Choose j_δ to be the smallest j such that

$$\frac{1}{j_\delta^2} \leq \delta. \tag{9.11}$$

Then we have

$$\mathbb{P}\left(|\hat{u}_k - u_k| \geq \frac{j_\delta \sqrt{u_k(1-u_k)}}{\sqrt{N}}\right) \leq \frac{1}{j_\delta^2} \tag{9.12}$$

and so we set

$$\frac{j_\delta \sqrt{u_k(1-u_k)}}{\sqrt{N}} \leq \epsilon. \tag{9.13}$$

This gives

$$N \geq \frac{j_\delta^2 u_k(1-u_k)}{\epsilon^2}. \tag{9.14}$$

In total, if $N \geq \frac{j_\delta^2 u_k(1-u_k)}{\epsilon^2}$ with $j_\delta \geq \frac{1}{\delta}$ then,

$$\mathbb{P}\left(|\hat{u}_k - u_k| \geq \epsilon\right) \leq \delta. \tag{9.15}$$

Since N is independent of d, the number of trials required to estimate u_k to accuracy ϵ and confidence $1-\delta$ is independent of d.

2. *Step 2c) of Sec. VIII*

In the protocol of Sec. VIII we have to estimate both u_k and Q_k where, u_k is the probability of obtaining outcome "k" on the first measurement and Q_k is the probability of obtaining outcome "k" on a second measurement *given* the result of the first measurement is "k". Note that observing "k" in the second measurement is also a Bernoulli random variable, so the discussion from above carries over in an analogous manner to estimating Q_k.

Let N_1 and N_2 be the number of trials required to estimate u_k and Q_k to each of their respective accuracies and confidences. Suppose we perform the first measurement N_1 times (so we have estimated u_k to its desired accuracy and confidence). If the number of times "k" is observed over these trials is greater than N_2 then we have estimated both u_k and Q_k to their desired accuracies. If the number of times "k" is observed in the first measurement is less than N_2 then we keep repeating until N_2 outcomes of "k" are recorded in the first measurement. Let M_2 be the number of times the first measurement has to be performed before N_2 values of "k" are recorded. The total number of trials is no more than

$$N = \min\{N_1, M_2\}. \tag{9.16}$$

As N only depends on the accuracies, confidences, and values of u_k and Q_k, it is independent of d.

B. Number of Trials Required in Step 3 of Sec. (IV) and Step 3 of Sec. (VIII)

Let us now discuss how many trials are required in Step 3 of each of the protocols. The argument is the same for each protocol so, without loss of generality, we use the notation from Sec. IV (the discussion for Sec. (VIII) follows by replacing X with Z and making appropriate changes). We have that $\overline{X}$ is equal to the expectation value of the random variable $X : \mathcal{D} \to [0,1]$ by

$$X(l,m) = \sqrt{u_l u_m} \tag{9.17}$$

(for Sec. VIII this will be $Z(l,m) = \sqrt{u_l u_m}\sqrt{Q_l Q_m}$). Suppose one wants to estimate $\overline{X}$ to accuracy ϵ and confidence $1-\delta$ using K independent samples of X. Let S_K denote the random variable that represents the average of K independent samples $\{X_1, ..., X_K\}$ of X. By Hoeffding's inequality,

$$\begin{aligned}\mathbb{P}\left(\left|S_K - \overline{X}\right| \geq \epsilon\right) &\leq 2e^{\frac{-2(k\epsilon)^2}{K(b-a)^2}} \\ &= 2e^{\frac{-2K\epsilon^2}{(b-a)^2}}\end{aligned} \tag{9.18}$$

where $[a,b]$ is the range of X (here $[a,b] \subseteq [0,1]$). Hence, setting

$$\delta = 2e^{\frac{-2K\epsilon^2}{(b-a)^2}} \tag{9.19}$$

gives,

$$K = \frac{\ln\left(\frac{2}{\delta}\right)(b-a)^2}{2\epsilon^2} \leq \frac{\ln\left(\frac{2}{\delta}\right)}{2\epsilon^2} \tag{9.20}$$

which is independent of d. In practice, $b - a << 1$ which will improve this bound on K.

C. Complete Reconstruction of POVM Elements

For completeness, we provide an explicit protocol and determination of the time-complexity of complete tomography of the noisy POVM elements $\{E_k\}_{k=1}^d$. The idea is to input various pure states into the noisy measurement and analyzing the frequency of obtaining particular outcomes. If the matrix representation of E_k in the $|\psi_m\rangle$ basis is written as $E_k^{i,j}$, then there are $\frac{d^2+d}{2}$ elements on the main diagonal and upper triangular section of E_k that need to be estimated (since each $E_k \geq 0$, the lower triangular part of E_k is completely determined by the upper triangular part). One can estimate the elements of E_k by first defining the pure states,

$$|\psi_{i,j}^+\rangle = \frac{|\psi_i\rangle + |\psi_j\rangle}{\sqrt{2}} \quad \text{and} \quad |\psi_{i,j}^-\rangle = \frac{|\psi_i\rangle + i|\psi_j\rangle}{\sqrt{2}}.$$

Then, since,

$$|\psi_i\rangle\langle\psi_j| = |\psi^+_{i,j}\rangle\langle\psi^+_{i,j}| + i|\psi^-_{i,j}\rangle\langle\psi^-_{i,j}| - \left(\frac{1+i}{2}\right)|\psi_i\rangle\langle\psi_i| - \left(\frac{1+i}{2}\right)|\psi_j\rangle\langle\psi_j|, \tag{9.21}$$

we have,

$$E_k^{i,j} = \mathrm{tr}\left(E_k\Pi^+_{i,j}\right) + i\mathrm{tr}\left(E_k\Pi^-_{i,j}\right) - \left(\frac{1+i}{2}\right)\mathrm{tr}(E_k\Pi_i) - \left(\frac{1+i}{2}\right)\mathrm{tr}(E_k\Pi_j). \tag{9.22}$$

The algorithm to determine the set $\{E_k\}_{k=1}^d$ is as follows.

Step 1: For each state $|\psi_{j,j}\rangle \in \{|\psi_{k,k}\rangle\}_{k=1}^d$, input $|\psi_{j,j}\rangle$ into the noisy measurement $\mathcal{E}$ many times and record the frequency of obtaining *each* of the d different possible outcomes "k". This gives, for each k, $\mathrm{tr}(E_k|\psi_{j,j}\rangle\langle\psi_{j,j}|)$.

Step 2: For each state $|\psi^+_{i,j}\rangle \in \{|\psi^+_{k,l}\rangle\}_{k,l=1}^d$, input $|\psi^+_{i,j}\rangle$ into the noisy measurement $\mathcal{E}$ many times and record the frequency of obtaining *each* of the d different possible outcomes "k". This gives, for each k, $\mathrm{tr}(E_k|\psi^+_{i,j}\rangle\langle\psi^+_{i,j}|) = \mathrm{tr}\left(E_k\Pi^+_{i,j}\right)$.

Step 3: For each state $|\psi^-_{i,j}\rangle \in \{|\psi^-_{k,l}\rangle\}_{k,l=1}^d$, input $|\psi^-_{i,j}\rangle$ into the noisy measurement $\mathcal{E}$ many times and record the frequency of obtaining *each* of the d different possible outcomes "k". This gives, for each k, $\mathrm{tr}(E_k|\psi^-_{i,j}\rangle\langle\psi^-_{i,j}|) = \mathrm{tr}\left(E_k\Pi^-_{i,j}\right)$.

Step 4: Combine all of the elements estimated in Steps 1 through 3 to construct the E_k. Step 1 gives the diagonal elements of the E_k since

$$E_k^{j,j} = \mathrm{tr}(E_k|\psi_{j,j}\rangle\langle\psi_{j,j}|). \tag{9.23}$$

Eq. (9.22) and Steps 1 through 3 give all of the off-diagonal elements $E_k^{i,j}$. This concludes the protocol.

The number of trials one will have to perform is again dictated by MLE. It is important to note that, as opposed to the simple Bernoulli procedure described in Sec. IX A for estimating the u_k, the MLE procedure in this case is more involved. This is because when the noisy measurement is performed, one must keep track of which value of j is obtained (not just whether the outcome was j or not). Thus the number of trials in each step will be greater than that required to estimate the u_k because events with small probability may rarely be seen (if at all). Smoothing techniques will likely have to be employed to ensure rare events are not assigned zero probability.

This discussion shows that the determination of each $\mathrm{tr}(E_k|\phi\rangle\langle\phi|)$, where $|\phi\rangle$ is one of $|\psi_{j,j}\rangle$, $|\psi^+_{i,j}\rangle$, or $|\psi^-_{i,j}\rangle$, requires greater time-complexity than that of estimating each u_k. Since there are $d + 2\frac{d(d-1)}{2} = d^2$ such $|\phi\rangle$, one will have to estimate d^3 different probabilities over Steps 1 through 3 using MLE. Hence, the full reconstruction requires the estimation of d^3 probabilities with more complicated post-processing of the measurement data (as well as a larger number of trials to estimate each probability). In addition, there is an added complexity in preparing d^3 different input states since more complex rotations may be required.

X. DISCUSSION

We have provided simple experimentally scalable protocols for obtaining upper (lower) bounds on the average measurement error (fidelity) of quantum measurements. We discussed conditions for the validity of the bounds and explained why they should be valid in extremely general situations. The bounds could also be useful as actual estimates of the average measurement error (fidelity) in the small-error regime. We have presented a set of numerical examples for a single-qubit system. In every instance the bounds were shown to be valid (the bounds were not violated for any magnitude of coherence parameters in the noisy POVM operators). In addition, as expected, the bounds became better approximations as the magnitude of the coherence increased. This provides further evidence that the bounds should hold in extremely general situations and should be useful in practice.

The protocols are scalable and only require the ability to prepare states from a basis set, and perform the noisy measurement (sequentially when there are output states) to estimate d probabilities. In addition, post-processing of the data is completely straightforward and avoids the difficulties in associating large sets of tomographic data to valid mathematical objects. This can be compared with a full reconstruction of the POVM elements of the noisy measurement, which requires the preparation of d^2 input states, the estimation of d^3 probabilities, and more involved post-processing. In addition, the d^2 input states can be highly complex. In many situations, these input states will have to be prepared using complicated unitary rotations.

There are a number of different questions and avenues for future research. First, we have focused on the case of rank-1 PVM's, however we expect that our results can be extended to higher-rank PVM's, especially low-rank PVM's in large Hilbert spaces. As well, since any POVM can be implemented via a PVM on an extended Hilbert space, our protocol can potentially give information regarding the quality of POVM measurements. It will also be interesting to analyze the extent to which the ideas presented here can be used to characterize non-ideal POVM measurements (as opposed to PVM's). While we have shown the bounds derived here should hold in large generality, a deeper understanding of the validity of the bounds will clearly be useful.

Lastly, the algorithms given here require the ability to prepare the basis states $\{|\psi_k\rangle\}$ which constitute the

measurement. In practice, these states typically have errors. We expect that running the protocol with noisy states will still provide valid bounds, however they will not be as tight as the upper bounds obtained when the protocol is run with perfect states. For instance, if the noisy input states are a convex combination of elements of $\{|\psi_k\rangle\}$, then the obtained bounds are still valid.

The scalable protocols presented here can be useful for determining the quality of experimental quantum measurements. There is still much to investigate with regards to useful metrics for comparing measurements and proposing experimentally efficient methods for characterizing measurement devices. As experimental quantum systems scale to larger sizes, such methods will be indispensable for characterizing and controlling multi-qubit systems.

ACKNOWLEDGMENTS

E.M. acknowledges financial support from the National Science Foundation through grant NSF PHY-1125846. The authors are grateful for helpful discussions with Alexandre Cooper-Roy, Joseph Emerson, Jay Gambetta, Masoud Mohseni, and Marcus Silva.

[1] R. Feynman, International Journal of Theoretical Physics **21** (1982).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
[3] P. Shor, in *Proceedings of the 35'th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Press, Los Alamitos, CA, 1994).
[4] S. Lloyd, Science **273**, 1073 (1996).
[5] D. Deutsch, Proc. Roy. Soc. Lond. A **400**, 97 (1985).
[6] R. Raussendorf and H. Briegel, Phys. Rev. Lett. **86**, 5188 (2001).
[7] A. Kitaev, Annals of Physics **303**, 2 (1997).
[8] V. Bouchiat et al., Phys. Scr. A **T76**, 165 (1998).
[9] J. Koch et al., Phys. Rev. A **76**, 042319 (2007).
[10] J. Wrachtrup and F. Jelezko, Journal of Physics: Condensed Matter **18**, S807 (2006).
[11] I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).
[12] D. Cory, A. Fahmy, and T. Havel, in *Proceedings of the 4th Workshop on Physics and Computation* (Boston, MA, 1996).
[13] D. Loss and D. Divincenzo, Phys. Rev. A **57**, 120 (1998).
[14] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46 (2001).
[15] I. Chuang and M. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
[16] J. B. Altepeter et al., Phys. Rev. Lett. **90**, 193601 (2003).
[17] M. Mohseni and D. A. Lidar, Phys. Rev. Lett. **97**, 170501 (2006).
[18] J. S. Lundeen et al., Nature Physics **5**, 27 (2009).
[19] L. Zhang et al., Nature Photonics **6**, 364 (2012).
[20] S. Merkel et al. (2012), arXiv:quant-ph/1211.0322.
[21] T. Monz et al., Phys. Rev. Lett. **106**, 130506 (2011).
[22] P. Shor, in *Proceedings of the 37'th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Press, Burlington, VT, 1996).
[23] D. Aharonov and M. Ben-Or, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)* (1997).
[24] E. Knill, R. Laflamme, and W. Zurek, Proc. R. Soc. Lond. A **454**, 365 (1997).
[25] J. Preskill (1997), arXiv:quant-ph/9712048.
[26] J. Emerson et al., Science **317**, 1893 (2007).
[27] M. Silva et al., Phys. Rev. A **78**, 012347 (2008).
[28] E. Knill et al., Physical Review A **77**, 012307 (2008).
[29] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).
[30] E. Magesan et al., Phys. Rev. Lett. **109**, 080505 (2012).
[31] O. Moussa et al., Phys. Rev. Lett. **109**, 070504 (2012).
[32] A. Bendersky, F. Pastawski, and J. Paz, Phys. Rev. Lett. **100**, 190403 (2008).
[33] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).
[34] C. Schmiegelow et al., Phys. Rev. Lett. **107**, 100502 (2011).
[35] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).
[36] M. Nielsen, Physics Letters A **303**, 249 (2002).
[37] J. Emerson, R. Alicki, and K. Zyczkowski, Journal of Optics B: Quantum and Semiclassical Optics **7**, S347 (2005).
[38] E. Magesan, R. Blume-Kohout, and J. Emerson, Phys. Rev. A **84**, 012309 (2011).
[39] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, UK, 2002).
[40] D. Gottesman (1997), ph.D. Thesis, arXiv:quant-ph/9705052.
[41] D. Gottesman, Chaos, Solitons and Fractals **10**, 1749 (1999).
[42] M. D. Reed et al., Nature **482**, 382 (2012).
[43] D. DiVincenzo and F. Solgun (2012), arXiv:1205.1910.
[44] A. Gilchrist, N. Langford, and M. Nielsen, Phys. Rev. A **71**, 062310 (2005).
[45] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, UK, 2006).
[46] A. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[47] B. Salman and R. Koenig, New J. Phys. **13** (2011).
[48] J. Renes et al., J. Math. Phys. **45**, 2171 (2004).
[49] M. Ledoux, *The Concentration of Measure Phenomenon* (American Mathematical Society, 2001).
[50] C. A. Fuchs, Ph.D. thesis, University of New Mexico (2006), preprint quant-ph 9601020.
[51] H. Jeffreys, *Theory of Probability* (Oxford University Press, USA, 1998), 3rd ed.
[52] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, USA, 1991).