

How to use arbitrary measuring devices to perform almost-perfect measurements

Noah Linden¹ and Paul Skrzypczyk^{2,3}¹*School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol BS8 1UG, United Kingdom*²*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, United Kingdom*³*CIFAR Azrieli Global Scholars program, CIFAR, Toronto, Canada*

(Received 8 August 2022; revised 1 November 2024; accepted 11 July 2025; published 1 August 2025)

We consider the problem of reproducing one quantum measurement given the ability to perform another. We give a general framework and specific protocols for this problem. For example, we show how to use available “imperfect” devices a small number of times to implement a target measurement with average error that drops off exponentially with the number of imperfect measurements used. We hope that could be useful in near-term applications as a type of lightweight error mitigation of the measuring devices. As well as the view to practical applications, we consider the question from a general theoretical perspective in the most general setting where both the available and target measurements are arbitrary generalized quantum measurements. We show that this general problem in fact reduces to the ability to reproduce the statistics of (complete) von Neumann measurements, and that in the asymptotic limit of infinitely many uses of the available measurement, a simple protocol based upon “classical cloning” can perfectly achieve this task. We show that asymptotically any nontrivial quantum measurement can reproduce any other. We also study optimal protocols for a fixed number of uses of the available measurement. This includes, but is not limited to, improving both noisy and lossy quantum measurements. Furthermore, we show that, in a setting where we perform multiple measurements in parallel, we can achieve finite-rate measurement reproduction, by using block-coding techniques from classical information theory. Finally, we show that advantages can also be gained by making use of probabilistic protocols.

DOI: [10.1103/8tph-mc2p](https://doi.org/10.1103/8tph-mc2p)

I. INTRODUCTION

The process of measurement plays a fundamental role in quantum mechanics, as it is only through measurement that classical information is obtained about the quantum state of a system. The most general measurement that can be performed in principle in quantum mechanics is a so-called generalized measurement (or instrument), specified by a set of subchannels (or usually Kraus operators). In practice, in any real experiment it is generally not possible to perform exactly the measurement one wants. For example, when measuring photons, detectors do not always click, leading to losses, or there will be coupling to additional degrees of freedom within the setup. Moreover, in different systems and architectures there may be fundamental or engineering constraints which make certain target measurements impossible or highly demanding to perform. Building high-quality measuring devices is thus a central and demanding task, which has seen significant progress in recent decades as quantum technologies have begun to emerge and mature.

Here we are interested in approaching the problem of performing a target measurement from the opposite direction: rather than directly engineering it, we are interested in how best to make use of available measuring devices, which may

either not be the ones sought after, or which may be imperfect or noisy. In particular, given an available measuring device or devices, we are interested in using it or them in an optimal way, such that at the end of the process, the effective measurement performed is as close as possible to the target measurement. In particular, we are interested in the most general situation, where the available and target measurements are both generalized quantum measurements. The approach we will consider taking is to use the available measuring device *multiple times*, or equivalently to use multiple identical measuring devices, in order to realize a *single* use (or even potentially multiple uses) of the target measurement.

We have both practical and theoretical aims in this paper.

In the context of the former we show how to use available “imperfect” devices a few times to implement a target measurement with average error that drops off exponentially with the number of imperfect measurements used. We hope that this could be useful in near-term applications as a type of lightweight error mitigation of the measuring devices.

At the more abstract theoretical level, we show that asymptotically any nontrivial quantum measurement can reproduce any other and that we can achieve finite-rate measurement reproduction, using block-coding techniques from classical information theory.

As an example of the potential types of processes we have in mind, consider that instead of using a measuring device to measure a given system, we first apply a “classical cloning” procedure on the incoming system,

$$|i\rangle \mapsto |i\rangle \dots |i\rangle, \quad (1)$$

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

e.g., we perform CNOTs with ancillary systems, “cloning” in the standard basis, and spreading information about the incoming state amongst the clones, and then measure each one of the clones using the available measuring device, before processing the string of outcomes into a single final outcome. What does this final measurement look like, and how does it compare to the available one? As we will show here, this cloning procedure is useful, and can provide lots of flexibility in changing the behavior of the measuring device, such that it can become close to a target measurement. Crucially, however, this procedure is provably not always optimal, and even better protocols can be found, as we will show.

Although this seems to be a rather natural question, it is one which does not appear to have been addressed in the literature before in the most general sense considered here. Key related papers are those of Yuen [1] who proposed “purifying” photodectors using preamplification, and by Dall’Arno D’Ariano and Sacchi [2], who considered the more specific task of purifying noisy measurements. More recently, [3,4] appeared, who propose methods to perform “active readout-error mitigation,” which are closely related to the prior works. These papers link to a modern and independent direction, known as “measurement error mitigation,” a method of based upon linear algebra, postprocessing a set of observed counts [5,6]. A key distinction with all of these results and ours is that we are not interested specifically in purification or improvement of noisy or imperfect measurements, but the general question of reproducing a target measurement with an available one, irrespective of how they relate to each other, e.g., even when the measurements have different numbers of outcomes, or are both extremal measurements. Nevertheless, our results can be seen as generalizing all of the above. On the experimental front, there is significant interest in performing high-quality measurements, and many results concerning improving measurements using a variety of techniques, e.g. [7–10]. Our theoretical results can be seen as providing a novel technique that is widely applicable for this purpose.

The questions we study here are also interesting from a resource-theoretic perspective [11]. There has been interest lately in applying the resource-theory approach of quantum information to quantum measurements [12–15]. In this approach, subsets of measurements are considered as the free objects, while any other measurement is considered a resource. In this approach, it has however been tacitly assumed that a measuring device is only used once. One main insight of this paper is that while it is natural in the context of resource theories of quantum states that each state can only be used once for some purpose (e.g., a single ebit which is consumed to perform quantum teleportation [16]), given a single measuring device it can normally be used many times, without being consumed or degraded.¹ This shows that questions of measurement simulation [17,18], which is another way of

phrasing the question considered here, can also naturally consider multiple uses of a measurement.

In what follows, we will show that, in fact, with the ability to perform any given measurement a sufficient number of times, it can approximate any other measurement as closely as desired. We will also show that the optimal procedure for this is nontrivial, in the sense that it is not merely the classical cloning example given above. Then, we consider the asymptotic scenario where we use the measurement many times, and want to approximate numerous uses of a different measurement. We show that we can achieve *finite-rate* transformations, using ideas from classical coding theory. Finally, we consider allowing ourselves to make use of probabilistic protocols, and demonstrate that these indeed provide further advantages. We will begin by considering a specific motivating example, which highlights many of the main ideas, before presenting the general results thereafter.

The paper is organized as follows:

(i) In Sec. II we consider a motivating example, using the trine measurement to implement a complete² von Neumann measurement. These two measurements are very different from each other and thus exhibit many key features of our results. In Sec. II A we first consider the two-copy case, and a simple but nonoptimal protocol based upon classical cloning. In Sec. II B we then present the general structure of deterministic protocols, before applying this insight back to find optimal (and nontrivial, i.e., not classical cloning) protocols in the case of two copies in Sec. II C, N copies in Sec. II D, and for single copies in Sec. II E.

(ii) In Sec. III we consider the practically important case of imperfect Z measurements (considering both the cases of noise and loss). An interesting feature is that even here optimal protocols can be nonclassical cloning.

(iii) In Sec. IV we move onto the analysis of general measurements. In Sec. IV A we outline the general setup of quantum instruments (generalized measurements). In Sec. IV B we then give general deterministic protocols in this setting. In Sec. IV C we present our postmeasurement subroutine, a key building block that implements a generalized measurement given access to a complete von Neumann measurement. In Sec. IV D we present our generalized classical cloning subroutine, and present our main result, which combines the two subroutines to show that any (nontrivial) measurement is able to implement any other measurement in the asymptotic limit of many uses.

(iv) In Sec. V we consider the problem of reproducing multiple uses of a target measurement in parallel, and show that in this setting finite rate reproduction is possible, using block-coding protocols from classical information theory.

(v) In Sec. VI we consider probabilistic protocols, and prove that they are able to outperform deterministic protocols.

¹There might be certain situations where this is not necessarily the case, e.g., a bipartite [local operations and classical communication (LOCC)] measurement that consumes entanglement. Here we will focus on situations where it is a reasonable assumption that measuring devices are degraded only very slowly.

²Here, by complete, we specifically refer to a von Neumann measurement where each projector is rank 1, and so the measurement has the same number of outcomes as the dimension of the Hilbert space. In contrast, a (noncomplete) von Neumann measurement may contain a smaller number of outcomes, with some higher-rank projectors.

(vi) Finally, in Sec. VII we conclude, and discuss future directions of research.

II. MOTIVATING EXAMPLE: REPRODUCING A COMPLETE VON NEUMANN MEASUREMENT USING THE TRINE

Let us assume that the only available measurement we can perform is the trine measurement $\mathbb{M} = \{M_a\}_a$, which has positive operator-valued measure (POVM) elements

$$M_a = \frac{2}{3}|\phi_a\rangle\langle\phi_a|, \quad (2)$$

where

$$|\phi_0\rangle = |0\rangle, \quad |\phi_1\rangle = \frac{|0\rangle + \sqrt{3}|1\rangle}{2}, \quad |\phi_2\rangle = \frac{|0\rangle - \sqrt{3}|1\rangle}{2}, \quad (3)$$

which are separated by 120° in the meridian plane of the Bloch sphere.³ Note that we only specify the POVM elements here, as we will assume the measurement is destructive (or alternatively that we are not interested in the postmeasurement state). In such cases, it is only the POVM elements that are relevant. We will come to the more general case in Sec. IV below.

Our goal in this example is to use this measurement N times to reproduce a single (complete) von Neumann measurement $\mathbb{V} = \{\Pi_0, \Pi_1\}$ on a qubit with elements $\Pi_0 = |0\rangle\langle 0|$ and $\Pi_1 = |1\rangle\langle 1|$. Since these two measurements are very different, at first sight this appears to be a challenging task.

In order to formalize our goal more precisely, let us imagine that we are given a single copy of an unknown state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, which we are to measure. Our goal then is to reproduce, as closely as possible the complete von Neumann measurement, that is to return outcome 0 with probability

$$\wp(0) = \langle\psi|\Pi_0|\psi\rangle = |\alpha|^2, \quad (4)$$

and leave the system in the state $|0\rangle$, and to return outcome 1 with probability

$$\wp(1) = \langle\psi|\Pi_1|\psi\rangle = |\beta|^2, \quad (5)$$

and leave the system in the state $|1\rangle$. We note that if we are able to achieve this goal perfectly, then due to the linearity of quantum mechanics it follows immediately that the correct behavior will also be reproduced when measuring either a mixed state, or part of a system. That is, if we measured the first particle of an entangled state $|\Phi\rangle = \gamma|0\rangle|\chi_0\rangle + \delta|1\rangle|\chi_1\rangle$, where $|\chi_0\rangle$ and $|\chi_1\rangle$ are normalized, but not necessarily orthogonal states, then we will find $\wp(0) = |\gamma|^2$, $\wp(1) = |\delta|^2$ and the states after measurement $|0\rangle|\chi_0\rangle$ and $|1\rangle|\chi_1\rangle$ respectively, as required.

In what follows, we will consider protocols which make use of trine measurements to approximate a complete von Neumann measurement. In a protocol that makes use of N trine measurements, we will denote by $P^{(N)}(a)$ the probability

that the result a is obtained. $P^{(N)}(a)$ depends upon $|\psi\rangle$, but for notational simplicity, we suppress this dependence. Our goal is that this should be similar to $\wp(a)$, but as we will see, there will be a small state-dependent error in announcing the wrong outcome.

In order to quantify how “close” our measurement is to the ideal one, we will adopt here the total variation distance (TVD), averaged over all states,

$$\langle\epsilon_{(N)}\rangle = \int d\psi (|P^{(N)}(0) - \wp(0)| + |P^{(N)}(1) - \wp(1)|). \quad (6)$$

We note that in the case here of a dichotomic measurement, we will always have $\wp(1) = 1 - \wp(0)$ and $P^{(N)}(1) = 1 - P^{(N)}(0)$, and so $\epsilon_{(N)}$ simplifies to

$$\langle\epsilon_{(N)}\rangle = 2 \int d\psi |P^{(N)}(0) - \wp(0)|. \quad (7)$$

Before presenting our simplest protocol, it will be useful to first consider what can be achieved by just performing the trine measurement, and guessing an outcome for a complete von Neumann measurement based upon the result. The probability distribution when measuring $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is

$$\begin{aligned} P(0) &= \frac{2}{3}|\alpha|^2, \\ P(1) &= \frac{1}{6} + \frac{1}{3}|\beta|^2 + \frac{\sqrt{3}}{6}(\alpha\beta^* + \alpha^*\beta), \\ P(2) &= \frac{1}{6} + \frac{1}{3}|\beta|^2 - \frac{\sqrt{3}}{6}(\alpha\beta^* + \alpha^*\beta). \end{aligned} \quad (8)$$

A simple way to approximate the von Neumann measurement is to associate the outcome 0 of the trine measurement with the outcome 0 of the von Neumann measurement, and to associate the other two outcomes of the trine measurement with the outcome 1. In this way, the implemented probabilities are

$$P^{(1)}(0) = \frac{2}{3}|\alpha|^2, \quad P^{(1)}(1) = \frac{1}{3} + \frac{2}{3}|\beta|^2. \quad (9)$$

Substituting these into (7), and evaluating the integral, we can calculate that this simple procedure has average TVD error of

$$\langle\epsilon_{(1)}\rangle = \frac{1}{3}. \quad (10)$$

Maybe unsurprisingly, this error is rather large, and demonstrates that, on average, the trine measurement is not able to approximately reproduce the von Neumann measurement under this simple procedure. In the following sections we will see that if we consider more interesting protocols, utilizing ancillary particles and making multiple measurements, the situation changes [and indeed these considerations lead us to the realization that even with one copy of the trine measurement, there are more subtle and better protocols (see Sec. II E)].

A. A simple protocol

We will exhibit first a simple protocol, which makes use of the trine measurement twice in order to implement a measurement which is not too dissimilar from the complete von Neumann measurement, that demonstrates many of the key features we are interested in. The basic intuition is to spread out the information contained in the state (the amplitudes α and β), into a larger Hilbert space, before attempting to

³One reason for treating this example in some detail is that, while it may not be particularly relevant in practical applications, many of the key features of the general protocols can be seen in this example.

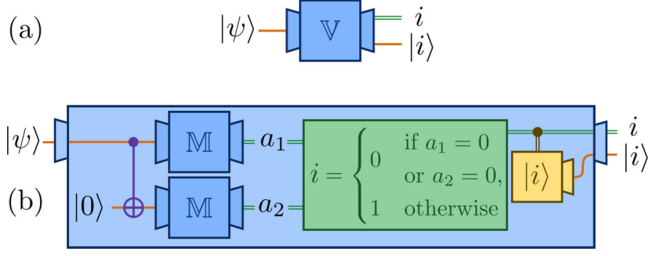


FIG. 1. (a) A complete von Neumann measurement $\mathbb{V} = \{|i\rangle\langle i|$. The probability of the outcome i when measuring the state $|\psi\rangle$ is $\rho(i) = |\langle i|\psi\rangle|^2$. The state after measurement is $|i\rangle$. (b) A simple protocol using two copies of a trine measurement to approximately reproduce a complete von Neumann measurement. An ancillary qubit is prepared in the state $|0\rangle$, and the incoming state $|\psi\rangle$ is classically cloned by applying a CNOT. Both particles are then measured using the trine measurement, and the outcomes are mapped into the outcome i of the measurement. Conditional on the outcome i , the corresponding state $|i\rangle$ is prepared.

retrieve the information using multiple measurements. This protocol is summarized pictorially in Fig. 1(b). We will see shortly, however, that this simple protocol is not optimal.

Let us consider that we prepare an ancillary system in the state $|0\rangle$, and classically clone, i.e., apply a controlled-not (CNOT) unitary between this system and $|\psi\rangle$, with $|0\rangle$ as the target, leading to the (in general entangled) state

$$|\Psi\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle. \quad (11)$$

We now measure both of the qubits, each one with the trine measurement. The probability distribution for the pairs of outcomes is found to be

$$\begin{aligned} P(0, 0) &= \frac{4}{9}|\alpha|^2, \\ P(0, 1) &= P(0, 2) = P(1, 0) = P(2, 0) = \frac{1}{9}|\alpha|^2, \\ P(1, 1) &= P(2, 2) = \frac{1}{36} + \frac{2}{9}|\beta|^2 + \frac{1}{12}(\alpha\beta^* + \alpha^*\beta), \\ P(1, 2) &= P(2, 1) = \frac{1}{36} + \frac{2}{9}|\beta|^2 - \frac{1}{12}(\alpha\beta^* + \alpha^*\beta), \end{aligned} \quad (12)$$

where $P(a_1, a_2) = \langle\Psi|M_{a_1} \otimes M_{a_2}|\Psi\rangle$ is the joint probability for the pair of outcomes (a_1, a_2) . In order to try and implement the complete von Neumann measurement, we need to map these nine outcomes pairs onto the two outcomes 0 and 1 of the target measurement, which is what will be declared as the result of the measurement. A choice which works well is

$$\begin{aligned} \{(0, 0), (0, 1), (0, 2), (1, 0), (2, 0)\} &\mapsto 0, \\ \{(1, 1), (1, 2), (2, 1), (2, 2)\} &\mapsto 1, \end{aligned} \quad (13)$$

that is, in the first five cases from (12), where at least one measurement returns a zero outcome are mapped to 0, while the final four cases from (12), where neither measurement returns a zero outcome are mapped to 1. Under this mapping, the implemented probabilities $P^{(2)}(i)$ [where the superscript (2) reminds us that we are considering two uses of the trine] are seen to be

$$\begin{aligned} P^{(2)}(0) &= \frac{8}{9}|\alpha|^2, \\ P^{(2)}(1) &= \frac{1}{9} + \frac{8}{9}|\beta|^2. \end{aligned} \quad (14)$$

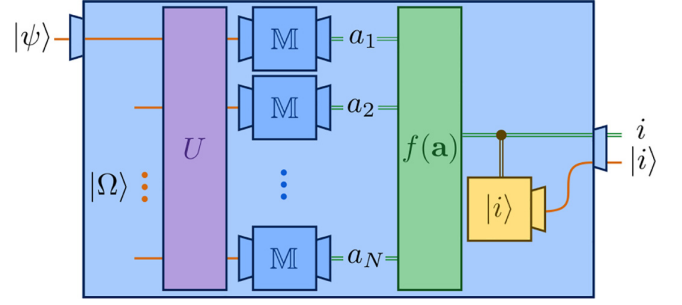


FIG. 2. Protocols using N copies of the trine measurement in order to approximately reproduce a complete von Neumann measurement. $N - 1$ ancillary qubits are prepared in the state $|\Omega\rangle$ and a global unitary U is applied to these as well as the state $|\psi\rangle$. All N qubits are then measured using the trine, and the string of outcomes $\mathbf{a} = (a_1, \dots, a_N)$ is mapped into an outcome $i = f(\mathbf{a})$. Conditional on i , the state $|i\rangle$ is prepared. This implements a measurement $\mathbb{M}^{(N)}$.

Given these probabilities, it is straightforward to calculate the average TVD error $\epsilon_{(2)}$, which is found to be

$$\langle\epsilon_{(2)}\rangle = \frac{1}{9}. \quad (15)$$

Only in the case that the probabilities $P^{(2)}(i)$ were identical to the target probabilities $\rho(i)$ for all states $|\psi\rangle$ would $\langle\epsilon_{(2)}\rangle$ equal zero. The small value found here shows that, on average, the error in the probabilities is indeed small, as expected.

As we explain in more detail below, we can also calculate the POVM $\mathbb{M}' = \{M'_i\}_i$ that the above procedure implements. We find

$$M'_0 = \frac{8}{9}|0\rangle\langle 0|, \quad M'_1 = \frac{1}{9}|0\rangle\langle 0| + |1\rangle\langle 1|. \quad (16)$$

We also want to leave the system in the correct state after the measurement. This step is, however, straightforward to achieve. Whenever we announce the result i , we simply prepare the system in the state $|i\rangle$.

Altogether, this simple example, making use of the trine measurement only twice, demonstrates that we can in fact implement a measurement that is not far from the complete von Neumann measurement, even though on the face of it, they are decidedly different measurements.

B. More general protocols

We will now show that using the available measurement multiple times allows an exponential decrease of the average error in the number of uses.

We consider a general class of protocols of the above type, where we are now able to perform trine measurements N times, and want to reproduce as closely as possible a single von Neumann measurement. The class of protocols we will consider are summarized pictorially in Fig. 2 and will comprise the following elements:

(1) We prepare $N - 1$ ancillary qubits in a state $|\Omega\rangle$, independent of $|\psi\rangle$. Without loss of generality (because of the next step), this can be taken to be the tensor product of $N - 1$ copies of $|0\rangle$. As such, we are considering a collection of N qubits.

(2) We perform a global unitary transformation U on these N qubits, such that the state becomes

$$|\Psi\rangle = U|\psi\rangle|\Omega\rangle = \alpha|\Psi_0\rangle + \beta|\Psi_1\rangle, \quad (17)$$

where

$$|\Psi_0\rangle = U|0\rangle|\Omega\rangle, \quad |\Psi_1\rangle = U|1\rangle|\Omega\rangle, \quad (18)$$

are two orthogonal and normalized, but otherwise arbitrary states of the N qubits.

(3) We measure all N systems, using the trine measurement.⁴

(4) Dependent on the string $\mathbf{a} = (a_1, a_2, \dots, a_N)$ of outcomes of the N measurements, we return as outcome of the target measurement either 0 or 1, and prepare the corresponding state $|0\rangle$ or $|1\rangle$.

There are two approaches that can be taken in the final step, either a deterministic or a probabilistic assignment. In the former case, to each string of outcomes we associate a definite outcome, either 0 or 1. We can do this, for example, by specifying a function $f(\mathbf{a})$, such that we assign the value $i = f(\mathbf{a})$ to the string of outcomes $\mathbf{a}$, as we did earlier in the two-copy case. In the latter case, on the other hand, this assignment is allowed to be probabilistic. This is then specified instead by a conditional probability distribution $q(i|\mathbf{a})$, which gives the probability of assigning the outcome i to the string $\mathbf{a}$.

In what follows we will restrict our attention to deterministic protocols. We will return to probabilistic protocols later on in Sec. VI, where we will show that, surprisingly, they can offer an advantage over deterministic protocols, at the expense of being much harder to analyze in general.

In either of these two cases, the above procedure defines a POVM $\mathbb{Q}^{(N)}$ with elements $Q_0^{(N)}$ and $Q_1^{(N)} = (\mathbb{I}_N - Q_0^{(N)})$ on the N systems, associated to the outcomes 0 and 1, respectively, where $\mathbb{I}_n$ denotes the identity on n qubits. In particular, in the deterministic case, these are given by

$$Q_i^{(N)} = \sum_{a_1, \dots, a_N} \delta_{i, f(\mathbf{a})} M_{a_1} \otimes \dots \otimes M_{a_N}. \quad (19)$$

That is, $Q_0^{(N)}$ and $Q_1^{(N)}$ are sums of tensor products of trine POVM elements. The implemented probabilities are then given by

$$\begin{aligned} P^{(N)}(i) &= \langle \Psi | Q_i^{(N)} | \Psi \rangle, \\ &= \langle \psi | \langle \Omega | U^\dagger Q_i^{(N)} U | \psi \rangle | \Omega \rangle, \end{aligned} \quad (20)$$

from which it follows that the associated implemented POVM elements are

$$M_i^{(N)} = \text{tr}_{2\dots N}[(\mathbb{I} \otimes |\Omega\rangle\langle\Omega|) U^\dagger Q_i^{(N)} U], \quad (21)$$

where $\text{tr}_{2\dots N}$ denotes the partial trace over all but the first system. In what follows, we will find it useful to think about properties of both the POVM $\mathbb{Q}^{(N)}$ acting on all N systems and in the effective POVM $\mathbb{M}^{(N)}$ acting on the system alone.

⁴We will see below that in fact this is not quite as general as we need; we will also need in special cases to only measure a *subset* of the systems, and trace out the remaining, unmeasured systems.

The figure of merit we consider will remain the same, as given in (7). In the Appendix we show that the average (integral) over all states can be performed, under the natural assumption that the implemented POVM elements are diagonal in the basis of the target von Neumann measurement.⁵ Under this assumption, the implemented POVM element $M_0^{(N)}$ has the form

$$M_0^{(N)} = (1 - \epsilon_0)|0\rangle\langle 0| + \epsilon_1|1\rangle\langle 1|, \quad (22)$$

where

$$\epsilon_0 = \langle \Psi_0 | Q_1^{(N)} | \Psi_0 \rangle = 1 - \langle \Psi_0 | Q_0^{(N)} | \Psi_0 \rangle, \quad (23)$$

$$\epsilon_1 = \langle \Psi_1 | Q_0^{(N)} | \Psi_1 \rangle, \quad (24)$$

are the errors, respectively, that the outcome 1 is returned when the state is $|0\rangle$ and that 0 is returned when the state is $|1\rangle$, and we used the fact that $Q_1^{(N)} = \mathbb{I} - Q_0^{(N)}$. Thus, using the standard parametrization for the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle$, it follows that $P^{(N)}(0) = (1 - \epsilon_0)\cos^2\frac{\theta}{2} + \epsilon_1\sin^2\frac{\theta}{2}$. Substituting this into (7), we are then able to explicitly integrate, after identifying the two regions where the $P^{(N)}(0) - \wp(0)$ has differing sign. A straightforward but lengthy calculation then leads to the simplified form

$$\langle \epsilon_{(N)} \rangle = \frac{\epsilon_0^2 + \epsilon_1^2}{\epsilon_0 + \epsilon_1}. \quad (25)$$

Given (23), $1 - \epsilon_0$ and ϵ_1 must lie between the minimal and maximal eigenvalues of $Q_0^{(N)}$, but are otherwise unconstrained. This now provides the possibility of finding the *best* possible average TVD error [for a fixed assignment $f(\mathbf{a})$, and under the assumption of diagonal POVM elements], by performing the simple optimization

$$\begin{aligned} \langle \epsilon_{(N)} \rangle^* &= \min_{\epsilon_0, \epsilon_1} \frac{\epsilon_0^2 + \epsilon_1^2}{\epsilon_0 + \epsilon_1} \\ \text{s.t.} \quad &1 - \lambda_{\max} \leq \epsilon_0 \leq 1 - \lambda_{\min}, \\ &\lambda_{\min} \leq \epsilon_1 \leq \lambda_{\max}, \end{aligned} \quad (26)$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $Q_0^{(N)}$, which must each lie in the interval $[0, 1]$.

In the Appendixes, using the structure of (26) and identifying the geometry of the problem, we show that it is possible to analytically solve for the optimal ϵ_0^* and ϵ_1^* that achieve the minimum in (26). The results are summarized in Table I.

This shows that it is in fact straightforward to solve for the minimum average TVD error $\langle \epsilon_{(N)} \rangle^*$ once a partition of the outcome strings has been made. Indeed, once we have chosen a partition, this defines the POVM $\mathbb{Q}^{(N)}$, which determines $\lambda_{\max}$ and $\lambda_{\min}$. Moreover, once we have solved (26), we can finalize the protocol, since from (23) we see that we just need to find two orthogonal states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ that have

⁵This can in fact be achieved without any loss of generality; if we have a protocol which does not have this property, we can always dephase it (making it diagonal in the target basis) by interacting with one further ancillary system. We have not found any advantage in considering protocols which produce coherence in the target basis.

TABLE I. Regions and behavior of $\langle \epsilon_{(N)} \rangle^*$. As shown in the Appendix, there are five regions, depending upon the relation between $\lambda_{\min}$ and $\lambda_{\max}$. In each of these regions, it is possible to analytically find the minimizing pair $(\epsilon_0^*, \epsilon_1^*)$ of (26), and the corresponding optimal average TVD error $\langle \epsilon_{(N)} \rangle^*$ [for a fixed assignment $f(\mathbf{a})$, and under the assumption of diagonal POVM elements].

	Region	$(\epsilon_0^*, \epsilon_1^*)$	$\langle \epsilon_{(N)} \rangle^*$
I	$\frac{1}{\sqrt{2}} < (1 - \lambda_{\max})$	$(1 - \lambda_{\max}, \lambda_{\max})$	$\frac{1}{2} + 2(\lambda_{\max} - \frac{1}{2})^2$
II	$(\sqrt{2} + 1)\lambda_{\min} < (1 - \lambda_{\max}) \leq \frac{1}{\sqrt{2}}$	$(1 - \lambda_{\max}, (\sqrt{2} - 1)(1 - \lambda_{\max}))$	$2(\sqrt{2} - 1)(1 - \lambda_{\max})$
III	$(\sqrt{2} - 1)(1 - \lambda_{\max}) \leq \lambda_{\min} \leq (\sqrt{2} + 1)(1 - \lambda_{\max})$	$(1 - \lambda_{\max}, \lambda_{\min})$	$\frac{(1 - \lambda_{\max})^2 + \lambda_{\min}^2}{1 - \lambda_{\max} + \lambda_{\min}}$
IV	$(\sqrt{2} + 1)(1 - \lambda_{\max}) < \lambda_{\min} \leq \frac{1}{\sqrt{2}}$	$((\sqrt{2} - 1)\lambda_{\min}, \lambda_{\min})$	$2(\sqrt{2} - 1)\lambda_{\min}$
V	$\frac{1}{\sqrt{2}} < \lambda_{\min}$	$(1 - \lambda_{\min}, \lambda_{\min})$	$\frac{1}{2} + 2(\lambda_{\min} - \frac{1}{2})^2$

the correct expectation values on $Q_0^{(N)}$, and we can always do this.⁶ Thus, we see that we have reduced our initial problem to a much simpler one, of solving a simple quadratic problem. The final stage is then to find the optimal partitions, which needs to be done either by appealing to a careful argument, or by simply checking the (large) number of possibilities.

There are furthermore a number of interesting lessons we can draw from this reformulation of $\langle \epsilon_{(N)} \rangle$. First, we can notice that it only depends upon two specific states, $|\psi\rangle = |0\rangle$ and $|\psi\rangle = |1\rangle$, through $|\Psi_0\rangle$ and $|\Psi_1\rangle$. These are the two orthogonal states that the von Neumann measurement is able to perfectly distinguish. As such, we see that the question of how well our implemented measurement performs on the entire Bloch sphere is reduced to a modified figure of merit on only two natural states.

Second, we note that from the form (25) it is straightforward to see since both terms in the numerator are non-negative, and since the denominator cannot become negative when $0 \leq \epsilon_0, \epsilon_1 \leq 1$, that the only way that the average TVD error $\langle \epsilon_{(N)} \rangle$ can vanish is if $\epsilon_0 = 0$ and $\epsilon_1 = 0$. This is of course as we should have expected: it corresponds to the case whereupon receiving $|i\rangle$ as the state to be measured, the measurement returns i with certainty. Crucially, however, when considering finding the optimal error, due to the constraints in (26) this in turn means that $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$. This further means that $Q_0^{(N)}$ must have an eigenvalue 0 and an eigenvalue 1. The former condition, due to the fact that $Q_1^{(N)} = \mathbb{I} - Q_0^{(N)}$, means that $Q_1^{(N)}$ must have an eigenvalue equal to 1. Altogether, therefore, this shows that in order to have zero error, the POVM $\mathbb{Q}^{(N)}$ must be such that each of its elements has a unit eigenvalue, and then it is clear that we must arrange things such that $|\Psi_0\rangle = U|0\rangle|\Omega\rangle$ and $|\Psi_1\rangle = U|1\rangle|\Omega\rangle$ are the corresponding eigenvectors.

Conversely, the above shows that if $\lambda_{\min} > 0$ or $\lambda_{\max} < 1$, we will for certain have $\langle \epsilon_{(N)} \rangle^* > 0$, corresponding to an imperfect reproduction of the target complete von Neumann measurement. Therefore, our problem has reduced to the study of the eigenvalues of the POVM $\mathbb{Q}^{(N)}$, and in particular how close to 1 we can arrange the largest eigenvalue of $Q_0^{(N)}$ to be.

⁶As we will see later, while it is simple to achieve the correct expectation values, it is not so obvious at first sight how to ensure that the states are orthogonal. This can, however, always be achieved by making use of extra ancillary particles, which are not measured.

C. Returning to two copies

Let us return to our example from Sec. II A, using two copies of the trine, $N = 2$, and rephrase everything in the above terms. The strategy that we outlined previously amounts to choosing

$$\begin{aligned}
 Q_0^{(2)} &= M_0 \otimes M_0 + M_0 \otimes M_1 + M_0 \otimes M_2 \\
 &\quad + M_1 \otimes M_0 + M_2 \otimes M_0 \\
 &= \mathbb{I} \otimes \mathbb{I} - (\mathbb{I} - M_0) \otimes (\mathbb{I} - M_0), \\
 Q_1^{(2)} &= M_1 \otimes M_1 + M_1 \otimes M_2 + M_2 \otimes M_1 + M_2 \otimes M_2 \\
 &= (\mathbb{I} - M_0) \otimes (\mathbb{I} - M_0)
 \end{aligned} \tag{27}$$

for the POVM $\mathbb{Q}^{(2)}$, and U equal to controlled-not such that

$$|\Psi_0\rangle = |0\rangle|0\rangle, \quad |\Psi_1\rangle = |1\rangle|1\rangle, \tag{28}$$

and

$$\begin{aligned}
 P^{(2)}(0) &= \frac{8}{9} \quad \text{when } |\psi\rangle = |0\rangle, \\
 P^{(2)}(0) &= 0 \quad \text{when } |\psi\rangle = |1\rangle,
 \end{aligned} \tag{29}$$

which led, as a reminder, to $\langle \epsilon_{(2)} \rangle = \frac{1}{9}$.

Now, $Q_0^{(2)}$ has minimum and maximum eigenvalues equal to $\lambda_{\min} = 0$ and $\lambda_{\max} = \frac{8}{9}$, respectively. From Table I we can verify we are in region II, and that the average TVD error is thus

$$\langle \epsilon_{(2)} \rangle^* = \frac{2(\sqrt{2} - 1)}{9} \approx 0.092 \tag{30}$$

which is achieved at

$$(\epsilon_0^*, \epsilon_1^*) = \left(\frac{1}{9}, \frac{\sqrt{2} - 1}{9} \right) \tag{31}$$

and is strictly better than what we found before, which we see is $\frac{1}{2(\sqrt{2}-1)} \approx 1.21$ times larger than the optimal average TVD error. We thus see that the classical cloning protocol we gave in Sec. II A, although intuitive, is not in fact the optimal strategy. It will be very insightful now to understand why there is a better strategy, and how it differs from our initial guess.

To that end, we can notice, crucially, that while $\epsilon_0^* = 1 - \lambda_{\max}$, on the contrary the optimal value $\epsilon_1^* > \lambda_{\min} = 0$. That is, in contrast to the procedure from before, which led to the probabilities given in (29), the optimal probabilities are in

fact

$$\begin{aligned} P^{(2)*}(0) &= \frac{8}{9} \quad \text{when } |\psi\rangle = |0\rangle, \\ P^{(2)*}(0) &= \frac{\sqrt{2}-1}{9} \quad \text{when } |\psi\rangle = |1\rangle, \end{aligned} \quad (32)$$

and the optimal implemented POVM is

$$\begin{aligned} M_0^{(2)*} &= \frac{8}{9}|0\rangle\langle 0| + \frac{\sqrt{2}-1}{9}|1\rangle\langle 1|, \\ M_1^{(2)*} &= \frac{10-\sqrt{2}}{9}|1\rangle\langle 1| + \frac{1}{9}|0\rangle\langle 0|. \end{aligned} \quad (33)$$

This shows that it is not optimal to take $\langle \Psi_1 | Q_0^{(2)} | \Psi_1 \rangle$ to be equal to the smallest eigenvalue of $Q_0^{(2)}$, and furthermore that in general $|\Psi_1\rangle$ will not be an eigenstate of $Q_0^{(2)}$. This can be seen explicitly in this example: the eigenvalues of $Q_0^{(2)}$ are $\{\frac{8}{9}, \frac{2}{3}, \frac{2}{3}, 0\}$ with corresponding eigenvectors $\{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\}$, and so we cannot take $|\Psi_1^*\rangle$ to be an eigenstate of $Q_0^{(2)}$ and achieve ϵ_1^* .

Since $\epsilon_0^* = 1 - \lambda_{\max}$, we must take $|\Psi_0^*\rangle = |0\rangle|0\rangle$, the corresponding eigenvector, just as before. For $|\Psi_1^*\rangle$ we see that we have some freedom, showing that there are infinitely many optimal solutions. One particularly simple choice is

$$|\Psi_1^*\rangle = |1\rangle \left(\sqrt{\frac{\sqrt{2}-1}{6}}|0\rangle + \sqrt{\frac{7-\sqrt{2}}{6}}|1\rangle \right), \quad (34)$$

which is still achieved by U being a controlled unitary transformation, except now when the control is $|1\rangle$ the unitary applied is no longer a full X rotation, but a rotation by a slightly smaller angle.

At first sight, it might appear that this protocol should be worse than the previous protocol since before if we were given the state $|\psi\rangle = |1\rangle$, then there was zero probability of erroneously returning the result 0, whereas with this new strategy the probability is $\frac{\sqrt{2}-1}{9}$. However, if we consider a state on the equator of the Bloch sphere $|\psi\rangle = (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$, then the previous protocol gave $P^{(2)}(0) = \frac{4}{9} \approx 0.44$, while the optimal protocol gives $P^{(2)*}(0) = \frac{7+\sqrt{2}}{18} \approx 0.47$, independent of ϕ , which is closer to the target statistics of $\wp(0) = \frac{1}{2}$. The lesson here is that looking only at the statistics of the basis states can be misleading compared to the statistics on average over the entire Bloch sphere, and this averaging is precisely what is taken into account in order to arrive at the form (26).

Finally, we should note that in the above we have found the optimal unitary U , and hence optimal states $|\Psi_0^*\rangle$ and $|\Psi_1^*\rangle$ after having fixed a choice of $Q^{(2)}$ (under the assumption of diagonal POVM elements). This raises the possibility that there is in fact a better choice of $Q^{(2)}$, i.e., a better partition of the strings **a** into two subsets. An exhaustive search finds this not to be the case; the protocol presented is in fact an optimal choice of $Q^{(2)}$ (other optimal choices also exist, which relate to this choice by suitable symmetries).

D. N -copy protocols

We now briefly outline how the above generalizes to the case of N uses of the trine. The natural generalization is

$$\begin{aligned} Q_1^{(N)} &= (M_1 + M_2)^{\otimes N} \equiv (\mathbb{I} - M_0)^{\otimes N}, \\ Q_0^{(N)} &= \mathbb{I}_N - (M_1 + M_2)^{\otimes N} \equiv \mathbb{I}_N - (\mathbb{I} - M_0)^{\otimes N}. \end{aligned} \quad (35)$$

Exhaustive search confirms this to be an optimal choice amongst all deterministic protocols for two- and three-copy protocols.⁷ The generalization to N copies corresponds to outputting 1 if none of the trine measurements have as outcome 0, and returning 0 in the other case, when one or more trine measurement returns 0.

$Q_0^{(N)}$ has minimum and maximum eigenvalues equal to

$$\lambda_{\min} = 0, \quad \lambda_{\max} = 1 - \frac{1}{3^N}. \quad (36)$$

Making use of Table I, we remain in region II, and have

$$\langle \epsilon_{(N)} \rangle^* = \frac{2(\sqrt{2}-1)}{3^N}, \quad (37)$$

which is achieved at

$$(\epsilon_0^*, \epsilon_1^*) = \left(\frac{1}{3^N}, \frac{\sqrt{2}-1}{3^N} \right). \quad (38)$$

Crucially, this shows that the average TVD error drops off exponentially fast in the number of uses of the trine and, surprisingly, tends to a perfect reproduction in the asymptotic limit of infinitely many uses. Thus, even though the trine measurement is fundamentally distinct from the complete von Neumann measurement, if we allow ourselves the possibility of performing the measurement many times, it in fact becomes equivalent.

An optimal unitary U achieving this protocol⁸ produces

$$\begin{aligned} |\Psi_0\rangle &= U|0\rangle^{\otimes N} = |0\rangle^{\otimes N}, \\ |\Psi_1\rangle &= U|1\rangle|0\rangle^{\otimes(N-1)} = |1\rangle^{\otimes(N-1)}|\phi^{(N)}\rangle, \end{aligned} \quad (39)$$

with

$$|\phi^{(N)}\rangle = \sqrt{\frac{\sqrt{2}-1}{2 \times 3^{N-1}}} |0\rangle + \sqrt{1 - \frac{\sqrt{2}-1}{2 \times 3^{N-1}}} |1\rangle. \quad (40)$$

That is, the optimal protocol is very close to a classical cloning protocol, whereby we copy $|0\rangle$ to N copies of $|0\rangle$, and $|1\rangle$ to N copies of $|1\rangle$. The key difference is that the final qubit should not be copied exactly, but must be rotated toward $|1\rangle$, with distance decreasing exponentially with N . A simple calculation shows that if we instead did use the classical cloning protocol, this would amount to picking the suboptimal pair $(\epsilon_0, \epsilon_1) = (\frac{1}{3^N}, 0) = (1 - \lambda_{\max}, \lambda_{\min})$, which would achieve $\langle \epsilon_{(N)} \rangle = \frac{1}{3^N}$. Notably this is always $\frac{1}{2(\sqrt{2}-1)} \approx 1.21$ times worse in average TVD error compared to the optimal choice, independent of N ,

⁷We note that this is not the unique optimal protocol; other optimal protocols are related to this one by symmetry.

⁸As in the previous example, there are infinitely many unitaries we can choose between; here we focus on what we believe is the simplest possible protocol to achieve in the laboratory.

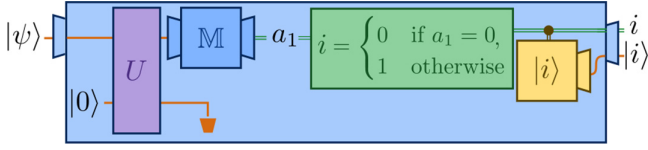


FIG. 3. A modified protocol in order to optimally make use of a single trine measurement. This protocol makes use of an ancillary qubit, so that the unitary U from (43) can be applied, which effectively allows the states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ to be nonorthogonal. Since only a single trine measurement is performed, the ancillary particle is left unmeasured or is thrown away at the end of the protocol.

and therefore still has an error which drops off exponentially fast to zero.

The measurement we implement through this N -use protocol, as given by (21), is found to be

$$\begin{aligned} M_0^{(N)} &= \left(1 - \frac{1}{3^N}\right) |0\rangle\langle 0| + \frac{\sqrt{2}-1}{3^N} |1\rangle\langle 1|, \\ M_1^{(N)} &= \left(1 - \frac{\sqrt{2}-1}{3^N}\right) |1\rangle\langle 1| + \frac{1}{3^N} |0\rangle\langle 0|. \end{aligned} \quad (41)$$

Before moving onto the case of more general measurement implementations, there are two final points worth discussing.

E. Single-use protocol

The first point worth noting is that our procedure shows that already in the scenario where the trine measurement is performed only once there is a nontrivial protocol to follow. In particular, substituting $N = 1$ into (41) indicates we should be able to achieve an implementation of the noisy complete von Neumann measurement

$$\begin{aligned} M_0^{(1)} &= \frac{2}{3} |0\rangle\langle 0| + \frac{\sqrt{2}-1}{3} |1\rangle\langle 1|, \\ M_1^{(1)} &= \frac{4-\sqrt{2}}{3} |1\rangle\langle 1| + \frac{1}{3} |0\rangle\langle 0|, \end{aligned} \quad (42)$$

which achieves an average TVD error of $\langle \epsilon_{(1)} \rangle^* = \frac{2}{3}(\sqrt{2}-1) \approx 0.276$. Notably, this outperforms the naive procedure of just measuring using the trine and outputting 0 when the outcome is 0: the procedure that we started with at the beginning of Sec. II. We saw previously that the naive procedure achieves $\langle \epsilon_{(1)} \rangle = \frac{1}{3}$.

We do, however, appear to run into a difficulty when we consider the unitary U that we need to perform from (39) since this is in fact no longer unitary, taking orthogonal states to nonorthogonal states. There is, however, an avenue to overcome this small difficulty, as depicted in Fig. 3. Consider preparing an ancillary qubit in the state $|0\rangle$, which is *not* measured. This ancillary qubit allows us to enact the transformation we need on the first qubit, and flip the second so as to satisfy the orthogonality requirements, i.e.,

$$\begin{aligned} U|0\rangle|0\rangle &= |0\rangle|0\rangle, \\ U|1\rangle|0\rangle &= \left(\sqrt{\frac{\sqrt{2}-1}{2}} |0\rangle + \sqrt{\frac{3-\sqrt{2}}{2}} |1\rangle \right) |1\rangle. \end{aligned} \quad (43)$$

It is interesting that it is necessary to go beyond the use of a single qubit, which effectively allows us to move from unitary transformations to more general CPTP maps at the level of the first qubit, and that this is useful.

III. IMPERFECT Z MEASUREMENTS

We now turn to the case that our available physical devices perform an imperfect versions of a (complete) von Neumann measurement. These cases are similar to one that occurs in many physically important applications. We will consider the two primary forms of imperfection: (i) noise, i.e., the possibility of declaring the wrong outcome, and (ii) loss, i.e., the possibility of declaring no outcome.

Interestingly, these examples also shed light on the question of why is it that protocols become nontrivial, requiring us to move beyond simple classical cloning. A natural conjecture would be that this derives from the fact that the trine measurement contains three nonorthogonal rank-one elements, which do not commute with each other. Maybe it is this noncommutativity which leads to the interesting structure to the problem we identified above. In what follows, when considering noisy Z measurement, we will show that this is not the case, and that this is not the origin of the nontrivial aspects of our measurement implementation problem.

A. Noisy Z measurements

Consider a simple class of measurements with POVM elements

$$\begin{aligned} N_0(p, q) &= p|0\rangle\langle 0| + (1-q)|1\rangle\langle 1|, \\ N_1(p, q) &= (1-p)|0\rangle\langle 0| + q|1\rangle\langle 1|. \end{aligned} \quad (44)$$

Without loss of generality we take $p \geq q$ and $p \geq (1-q)$, so that $p \in [\frac{1}{2}, 1]$.⁹ We can think of this class of measurements as being asymmetric noisy Z measurements, which include the standard noisy Z measurement in the special symmetric case $p = q$.

We will consider the situation of a single-copy ($N = 1$) approximate implementation of the von Neumann measurement using these noisy measurements. In this case, we have $\lambda_{\min} = 1 - q$ and $\lambda_{\max} = p$.

From Table I, we know that there are five regions to consider. Given our constraints, it is straightforward to check that regions I and II are excluded. In the region III, where

$$(\sqrt{2}+1)q - \sqrt{2} \leq p \leq 2 - \sqrt{2} + (\sqrt{2}-1)q, \quad (45)$$

the optimal solution is obtained at $(\epsilon_0^*, \epsilon_1^*) = (1 - \lambda_{\max}, \lambda_{\min})$. This region includes the standard noisy Z measurement. In this region an optimal protocol is just the trivial one (the unitary is the identity).

Surprisingly, the remaining two regions are in fact nonempty, and in these regions it is necessary to employ a nontrivial protocol (i.e., one which is not simply classical

⁹Note that we restrict the interval for later convenience, as it will be sufficient for our purposes here, and all other cases can be mapped into this one by suitable relabeling.

cloning). In particular, focusing on region IV (similar calculations hold for region V), then we are in this region whenever

$$1 - \frac{1}{\sqrt{2}} \leq q < (\sqrt{2} + 1)p - \sqrt{2}. \quad (46)$$

From Table I the optimal solution is at

$$(\epsilon_0^*, \epsilon_1^*) = ((\sqrt{2} - 1)(1 - q), 1 - q), \quad (47)$$

which achieves $(\epsilon_{(1)})^* = 2(\sqrt{2} - 1)(1 - q)$, and for which an optimal protocol, which requires an ancillary qubit (which will not be measured, as in the case of a single use of the trine), is to perform the transformation

$$U|0\rangle|0\rangle = (\gamma|0\rangle + \delta|1\rangle)|0\rangle, \quad U|1\rangle|1\rangle = |1\rangle|1\rangle, \quad (48)$$

where $\delta = \sqrt{1 - \gamma^2}$, and γ depends upon p and q , and is equal to

$$\gamma = \sqrt{\frac{1 - \sqrt{2} + \sqrt{2}q}{p + q - 1}}. \quad (49)$$

One can confirm that $0 \leq \gamma \leq 1$ for all p and q satisfying (46). The protocol is to measure the first of the two qubits, and return as the outcome the observed outcome (and prepare the corresponding state).

What this shows is that even though we have a very simple measurement, with two commuting POVM elements, there is a regime where it is nevertheless necessary to utilize a nontrivial (not classical-cloning) protocol in order to use it to optimally implement a complete von Neumann measurement. Thus, the need for nontrivial protocols does not derive from the noncommutativity of the POVM elements, and is a general feature of optimal protocols.

B. Lossy Z measurements

We now consider instead the case of lossy Z measurements, i.e., the model of a measurement where the system can be lost as part of the measuring process, returning a null result. We will refer to this null result (without loss of generality) as 2. The POVM elements of a lossy Z measurement are

$$\begin{aligned} L_0(\eta) &= (1 - \eta)|0\rangle\langle 0|, \\ L_1(\eta) &= (1 - \eta)|1\rangle\langle 1|, \\ L_2(\eta) &= \eta\mathbb{I}. \end{aligned} \quad (50)$$

Here η is the *loss parameter*, the probability that the system is lost. The ideal measurement is recovered when $\eta = 0$, corresponding to zero loss (such that the outcome 2 is never seen).

When considering multiple lossy measurements, rather than taking the target measurement to be an ideal Z measurement, it is instructive instead to take the target to be another lossy measurement, i.e., one with three outcomes. In particular, for a given string out outcomes $\mathbf{a} = (a_1, \dots, a_N)$, when using N measuring devices and classical cloning, consider the following deterministic assignment:

$$i = \begin{cases} 0 & \text{if any } a_k = 0 \text{ for } k = 1, \dots, N, \\ 1 & \text{if any } a_k = 1 \text{ for } k = 1, \dots, N, \\ 2 & \text{if all } a_k = 2 \text{ for } k = 1, \dots, N. \end{cases} \quad (51)$$

Note that this covers all cases that can occur, due to the structure of the measurement (50) (in combination with classical cloning). In particular, if the qubit to be measured is in the state $|0\rangle$, then after cloning, the N -qubit state that is measured is $|0\rangle^{\otimes N}$. It is therefore impossible that *any* measurement returns result $a_k = 1$; the only possible results are strings of 0 and 2 (and all can occur). A similar reasoning shows holds when the system starts in the state $|1\rangle$. The assignment (51) is therefore seen to capture all possible strings.

It is straightforward to calculate the POVM elements of the reproduction. They are given by

$$\begin{aligned} L_0^{(N)} &= (1 - \eta^N)|0\rangle\langle 0|, \\ L_1^{(N)} &= (1 - \eta^N)|1\rangle\langle 1|, \\ L_2^{(N)} &= \eta^N\mathbb{I}, \end{aligned} \quad (52)$$

i.e., they are of the same form as (50), but with η replaced by η^N . This thus corresponds to exponential suppression of the losses. It is intuitive why this is the case: in a lossy measurement, the only imperfection is when the system is lost, in which case no information is learned about its identity. In this protocol, the basis states are cloned, and the only time that no information is gained is if *all* clones are lost, which occurs with probability η^N , an exponential suppression.

Finally, in this section we note that lossy Z measurements can also be used to reproduce noisy Z measurements, by binning the outcomes. One possible strategy (which is sufficient to exhibit the scaling with N) is as follows: when the outcome 2 is obtained, the result 0 can be returned. In this case, the POVM elements become of the form (44) with

$$p = 1, \quad q = 1 - \eta. \quad (53)$$

We thus have $\lambda_{\max} = 1$ and $\lambda_{\min} = \eta$, and from Table I we will be in region IV if $\eta \leq 1/\sqrt{2}$ and region V otherwise.

Given access to multiple uses of the measurement, we can first employ the protocol from the previous paragraph, before carrying out the binning. The combined effect leads to a measurement with $p = 1$ and $q = 1 - \eta^N$, hence with $\lambda_{\max} = 1$ and $\lambda_{\min} = \eta^N$. This thus always ends up in the region IV for sufficiently large N , with the final average TVD error of $(\epsilon_{(N)})^* = 2(\sqrt{2} - 1)\eta^N$, exponentially small, as we would expect.

IV. REPRODUCING GENERAL MEASUREMENTS

Up until this point we focused on insightful examples, reproducing a single use of a qubit von Neumann measurement given the ability to perform the trine measurements, imperfect Z measurements, or lossy (inefficient) Z measurements. In this section we will now study the more general problem of reproducing the single use of one measurement given the ability to perform another measurement N times. The main result of this section will be to show that in fact any measurement can reproduce any other measurement, showing that all measurements are in fact equivalent in the setting we consider. Crucially, this is true even when the measurement we are trying to reproduce has nontrivial postmeasurement states (in contrast to the complete von Neumann measurement, as we will explain more carefully below). A key insight is that in fact the ability to reproduce a complete von Neumann

measurement can be viewed as an important subroutine, which although not always optimal, gives a concrete protocol which allows any measurement to reproduce any other measurement.

A. The general setup

In the above we described our setup in terms of the POVM elements of the trine. This was sufficient for our needs previously, as we in fact did not make use of the postmeasurement states of the measurement, but only the result of the measurement. In general, after a system is measured, it will leave the system in a postmeasurement state, dependent upon the outcome observed. From a modern perspective, the most general measurement that can be performed is an *instrument* $\mathcal{M}$, which is a collection of completely positive and trace non-increasing maps $\mathcal{M} = \{\Lambda_a\}_a$ (also commonly referred to as *subchannels*), one associated to each outcome a . We will denote by m the number of outcomes of the measurement, which can be arbitrary. When a state ρ is measured, the probability of the outcome a is

$$P(a) = \text{tr}[\Lambda_a(\rho)], \quad (54)$$

and the state after the measurement is

$$\rho'_a = \frac{\Lambda_a(\rho)}{P(a)}. \quad (55)$$

In order for the measurement to be properly normalized, we must have $\sum_a P(a) = 1$ for all states, which is satisfied as long as $\Lambda = \sum_a \Lambda_a$ is a trace-preserving map.

Note that since each subchannel of the instrument is completely positive and trace nonincreasing, it can be written as a sum of *Kraus operators*,

$$\Lambda_a(\cdot) = \sum_i K_i^a(\cdot) K_i^{a\dagger}, \quad (56)$$

where $\sum_i K_i^{a\dagger} K_i^a \leq \mathbb{I}$. The normalization condition then reads as $\sum_{i,a} K_i^{a\dagger} K_i^a = \mathbb{I}$. This highlights why instruments can be viewed as more general than Kraus operators since they allow each individual outcome of the measurement to be associated with a collection of Kraus operators, rather than a single Kraus operator.¹⁰ Nevertheless, to each outcome we can still associate a POVM element M_a , which determines the probability of a given outcome (but not the postmeasurement state). We see that

$$M_a = \sum_i K_i^{a\dagger} K_i^a, \quad (57)$$

where $P(a) = \text{tr}[M_a \rho]$. For a general instrument, we find that the postmeasurement state ρ'_a will depend both upon the measurement outcome *and* the state measured. This is in contrast to the case of a complete von Neumann measurement, where the postmeasurement states depend only upon the measurement outcome (and are independent of the state measured). It is in this sense that we can say that the postmeasurement states

of the complete von Neumann measurement are trivial, and we see that a potentially new aspect of the reproduction problem arises when we want to reproduce the postmeasurement states of a measurement where these are nontrivial, and retain a dependence upon the measured state.

The general scenario we wish to consider there is that we want to reproduce a target measurement $\mathcal{T} = \{\Gamma_i\}$, given the ability to perform the measurement $\mathcal{M} = \{\Lambda_a\}$ N times.

The only requirement we place on $\mathcal{M}$ is that it is *nontrivial*, meaning that its POVM elements are not all proportional to the identity. Such measurements are not measurements at all from a physical perspective since they do not require any measurement to be performed. They amount simply to producing a random (fictitious) measurement result, and preparing some quantum state. Clearly, such trivial measurements will never be able to reproduce any nontrivial measurement, and hence we exclude them from investigations from here on.

Similarly to before, we consider that we are given a single copy of an unknown pure state $|\psi\rangle$, and at the end of the protocol, in an ideal reproduction we wish to produce an outcome i with probability $\wp(i) = \text{tr}[\Gamma_i(|\psi\rangle\langle\psi|)]$, and to leave the system in the state $\rho'_i = \Gamma_i(|\psi\rangle\langle\psi|)/\wp(i)$. This will not be possible in general, and so we seek to approximate this as closely as possible using an optimal protocol involving N uses of the available measurement.

B. General protocols

We now outline the general class of protocols that we consider, as well as the figure of merit we will focus on. The primary difference from before arises due to the fact that we now consider instruments, with postmeasurement states. This means that the most general protocols will have a *sequential* aspect to them. In particular, we may consider protocols involving a number of stages, where at one stage we make N_1 measurements on a subset of the particles. Depending upon the outcome of these measurements (and previous measurements), we may then apply a unitary on the whole system, before measuring a new subset of N_2 particles, and so on. The general structure of protocols we consider, depicted in Fig. 4, is the following:

General protocol

- (1) Prepare N_A ancillary particles in a state $|\Omega\rangle$, independent of $|\psi\rangle$. Without loss of generality, this can be taken to be $|\Omega\rangle = |0\rangle^{\otimes N_A}$.
- (2) Perform a global unitary U_1 on the $N_A + 1$ particles $|\psi\rangle|\Omega\rangle$.
- (3) Measure a subset $\mathcal{S}_1$ of the particles using the available measurement $\mathcal{M}$ N_1 times, leading to a string of outcomes $\mathbf{a} = (a_1, \dots, a_{N_1})$. Due to the freedom in specifying the unitary U_1 in the previous step, the measurements can always be taken to be on the first N_1 particles (since systems can be permuted before the measurement).
- (4) Depending upon the string of outcomes $\mathbf{a}$, apply a unitary transformation $U_2(\mathbf{a})$ on all of the particles.
- (5) Measure the first N_2 particles using the measurement $\mathcal{M}$. Append the results of these

¹⁰This is also sometimes referred to as non-fine-grained measurements since it can be viewed as a measurement with a pair of outcomes a and i , such that i is ignored.

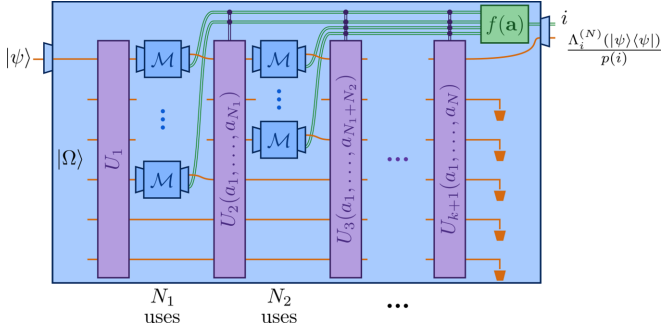


FIG. 4. The general class of protocols we consider here. N_A ancillary particles are prepared in the state $|\Omega\rangle$, and an initial unitary U_1 is applied to all $N_A + 1$ particles. In stage one, the first N_1 particles are measured, and then a conditional unitary $U_2(a_1, \dots, a_{N_1})$ is applied. This is iterated k times, such that in stage j the first N_j particles are measured, and then a unitary $U_j(a_1, \dots, a_{N_1+\dots+N_j})$ is applied, which depends upon all previous measurement outcomes. After N measurements have been performed, the final unitary is used to prepare the output particle (taken to be the first without loss of generality), and the measurement outcome is computed according to the function $i = f(\mathbf{a})$. This implements a measurement $\mathcal{M}^{(N)} = \{\Lambda_i^{(N)}\}$.

measurements onto the existing string, so that $\mathbf{a} = (a_1, \dots, a_{N_1}, a_{N_1+1}, \dots, a_{N_1+N_2})$.

(6) Repeat steps (iv) and (v) until $N = N_1 + N_2 + \dots + N_k$ measurements have been made, i.e., so that in total there are k measurement stages, and in total N uses of $\mathcal{M}$ are made.

(7) Dependent on the full string of outcomes $\mathbf{a} = (a_1, \dots, a_N)$, return as outcome the result i . As before, this can be done deterministically, where $i = f(\mathbf{a})$, or probabilistically, according to $p(i|\mathbf{a})$.

(8) Apply a final unitary transformation $U_{k+1}(\mathbf{a}, i)$, and return as the postmeasurement state the first system (without loss of generality).¹¹

This protocol will enact a measurement $\mathcal{M}^{(N)} = \{\Lambda_i^{(N)}\}$ on the system. We can quantify the performance of the protocol using a generalized form of TVD error from before, taking into account now also the postmeasurement state, namely,

$$\langle \epsilon_{(N)} \rangle = \int d\psi \sum_i \|\Lambda_i^{(N)}(|\psi\rangle\langle\psi|) - \Gamma_i(|\psi\rangle\langle\psi|)\|_1, \quad (58)$$

where $\|\cdot\|_1$ denotes the trace norm. The above class of protocols are hard to analyze in general, due to their generality. In

¹¹Note that $U_{k+1}(\mathbf{a}, i)$ will only depend explicitly upon i in the case of probabilistic protocols; for deterministic ones, since $i = f(\mathbf{a})$, $U_{k+1}(\mathbf{a})$ will only depend upon i implicitly through $\mathbf{a}$. Note also that since the first system could contain multiple subsystems, this controlled unitary allows rather general postmeasurement state reproduction; as an example, it could include a controlled-swap, thus allowing for effective “measure and prepare.”

what follows we will restrict our attention to specific families of protocols which are simpler and less general. Nevertheless, we will see that they are powerful enough to show that any measurement can reproduce arbitrarily well any other measurement. Before doing so, we will first exhibit an important subroutine, which will reduce the complexity of our problem substantially.

C. Postmeasurement subroutine

When considering general measurement reproduction, with nontrivial postmeasurement states, at first sight it appears that the problem is inherently more difficult than reproducing a complete von Neumann measurement, with its trivial postmeasurement state, that depends solely upon the measurement outcome, and not upon the system being measured. In this subsection we will show that there is in fact a *universal subroutine*, which shows that we can reduce the problem of reproducing a given measurement to the problem of reproducing destructive perfect complete von Neumann measurements. “Destructive” means that we only need to get the correct classical outcome of the measurement; we do not need access to the postmeasurement state. In particular, if we are given access to a destructive complete von Neumann measurement, the following subroutine is able to *perfectly* reproduce any other measurement. Thus, in order to use one measurement to reproduce another, we simply only need to show that we can use any measurement to reproduce a (destructive) complete von Neumann measurement, and then use this subroutine, in a bootstrap fashion, to reproduce the desired measurement.

Let us assume that we wish to reproduce a measurement $\mathcal{M}$ as above, and we have the ability to perform a destructive complete von Neumann measurement $\mathbb{V} = \{|i\rangle\langle i|\}$. Consider the following protocol, which we will refer to as the *postmeasurement subroutine*:

Postmeasurement subroutine

- (1) Prepare two ancillary particles in the state $|0\rangle|0\rangle$.
- (2) Perform a global unitary V such that

$$V|\psi\rangle|0\rangle|0\rangle = \sum_{a,j} K_j^a |\psi\rangle|a\rangle|j\rangle. \quad (59)$$

- (3) Ignore the third particle, and measure the second using a destructive perfect complete von Neumann measurement.
- (4) Return as outcome the result of the von Neumann measurement, and return as the postmeasurement state the first particle.

It is straightforward to see that the probability of returning the outcome a is

$$P(a) = \sum_j \text{tr}[K_j^a |\psi\rangle\langle\psi| K_j^{a\dagger}] = \text{tr}[M_a |\psi\rangle\langle\psi|], \quad (60)$$

and that the state (of the first particle) after the measurement is

$$\rho'_a = \frac{1}{P(a)} \sum_j K_j^a |\psi\rangle\langle\psi| K_j^{a\dagger} = \frac{\Lambda_a(|\psi\rangle\langle\psi|)}{P(a)}. \quad (61)$$

That is, this shows that a destructive complete von Neumann measurement is able to perfectly reproduce any other measurement. Interestingly, in this protocol the dimension of the system the von Neumann measurement needs to act upon is determined by the number of outcomes of the target measurement. Note that this does not introduce any difficulties. If, for example, we can only perform von Neumann measurements of dimension d , then in (59) we would need to encode $|a\rangle$ into an appropriate collection of qudits and perform the required number of complete von Neumann measurements (e.g., if the target measurement had seven outcomes, and we had access to a qubit von Neumann measurement, then we would encode into three qubits, and measure all three of them, with only seven of the eight possible outcomes ever occurring; equivalently, we could take the final outcome to have Kraus operator 0).

D. Asymptotic reproduction of any measurement by any other measurement

We are now in a position to give our first main result:

Main result 1: Given the ability to perform *any nontrivial* measurement, we can reproduce arbitrarily well *any other* (generalized) measurement, by making sufficiently many uses of the first measurement.

In order to prove the main result, we can first note that, given the above postmeasurement subroutine, all that is left to be shown is that any measurement is able to reproduce a perfect destructive von Neumann measurement arbitrarily well in the asymptotic limit (that is, we do not even need to worry about getting the postmeasurement state correct, even though this is in fact completely straightforward, as we do not make any use of it). We will now show this, focusing on a simple (albeit nonoptimal) protocol, based upon generalized classical cloning.

We will begin by considering the simplest situation, where we assume that the measurement we have access to is on systems of the same dimension as the target complete von Neumann measurement, which we will denote by d . At the end of this section we will explain how to adapt the analysis to take into account the more general situation, where the dimensions differ.

The first key observation is that the ability to perfectly perform a complete von Neumann measurement is equivalent to being able to perfectly distinguish between the d basis states $\{|i\rangle\}$. In the forward direction this is immediate since a perfect complete von Neumann measurement can indeed distinguish between the d basis states $\{|i\rangle\}$. In the backward direction, a perfect complete von Neumann measurement is

the only measurement that is able to perfectly distinguish the basis $\{|i\rangle\}$, and hence the equivalence follows.¹²

Thus, we can focus on the task of perfectly distinguishing a basis of states, and if we show that we can do this asymptotically, then we know this implies that we have the ability to perform an ideal measurement. The following generalized classical cloning protocol achieves this goal:

Generalized classical cloning subroutine

(1) Prepare $N - 1$ ancillary particles in the state $|\Omega\rangle = |0\rangle^{\otimes(N-1)}$.

(2) Perform the unitary U such that

$$|\Psi_i\rangle = U|i\rangle|\Omega\rangle = |e_i\rangle^{\otimes N}, \quad (62)$$

where $|e_i\rangle$ form an orthonormal set of d states (and will be specified more fully in the main text).

(3) Measure all N particles using the available measurement $\mathcal{M}$, keeping only the string of measurement results $\mathbf{a} = (a_1, \dots, a_N)$, (and discarding any and all post-measurement states).

(4) Use maximum-likelihood estimation to guess which basis state $|i\rangle$ was initially prepared.

We will now analyze this subroutine to show that in the asymptotic limit of large N , it is able to correctly identify the basis state $|i\rangle$ perfectly, independent of the measurement $\mathcal{M}$ used.

The probability of obtaining the outcome a when measuring the incoming basis state $|i\rangle$ is

$$P(a|i) = \langle e_i | M_a | e_i \rangle, \quad (63)$$

where $\{M_a\}_a$ are the POVM elements of the instrument $\mathcal{M}$. The probabilities of the string of outcomes $\mathbf{a} = (a_1, \dots, a_N)$ will be

$$P^{(N)}(\mathbf{a}|i) = \prod_{k=1}^N P(a_k|i). \quad (64)$$

Now, the key insight is to view this as N samples from one of d different probability distributions, with each distribution being labeled by i . That is, we can map the problem of distinguishing basis states onto the equivalent problem of having to distinguish between d different classical random variables, given N samples of one of the variables. This is thus an instance of d -ary hypothesis testing.

Crucially, in the asymptotic limit, all nonidentical random variables can be perfectly distinguished. Moreover, the probability of making an error drops off exponentially in the number

¹²Note that, strictly speaking, we also ignore trivial counterexamples to the above, such as the three outcome measurements with outcomes $M_0 = |0\rangle\langle 0|$, $M_1 = \frac{1}{2}|1\rangle\langle 1|$ and $M_2 = \frac{1}{2}|1\rangle\langle 1|$. That is, examples where the measurement is obtained by a trivial postprocessing of the von Neumann measurement. In this example, $M_1 = M_2$, and this measurement is equivalent to performing a complete von Neumann measurement, except one outcome involves a fair coin flip before announcing the result.

of samples N when using maximum-likelihood estimation [19], with the rate given by the minimum pairwise Chernoff distance between the d probability distributions.

Finally, since the probability of making an error decreases exponentially in the asymptotic limit, so too does the average TVD error $\langle \epsilon_{(N)} \rangle$. In particular, we find that $\epsilon_{a|i}$ approaches 0 exponentially fast for $a \neq i$, and similarly $\epsilon_{a|a}$ approach 1 exponentially fast, in both cases as N tends towards infinity. As we show in the Appendix, we can find an upper bound on the TVD error that depends only upon $\epsilon_{a|i}$. Given that this upper bound approaches 0 exponentially fast, a direct calculation shows that $\langle \epsilon_{(N)} \rangle$ must too approach 0 exponentially.

The only subtlety that arises then is whether the distributions $P(a|i)$ are different for all states $|i\rangle$, which is necessary in order to distinguish the states. In order to understand this issue, let us consider two simple examples. First, returning to the trine measurement from Sec. II, and consider standard classical cloning, with $|e_i\rangle = |i\rangle$, then the probabilities obtained are

$$\begin{aligned} P(0|0) &= \frac{2}{3}, & P(1|0) &= \frac{1}{6}, & P(2|0) &= \frac{1}{6}, \\ P(0|1) &= 0, & P(1|1) &= \frac{1}{2}, & P(2|1) &= \frac{1}{2}. \end{aligned} \quad (65)$$

Since the two distributions $P(a|0)$ and $P(a|1)$ are distinct, asymptotically the states $|0\rangle$ and $|1\rangle$ can be perfectly distinguished. As a second example, consider the degenerate qutrit measurement with POVM elements

$$M_0 = |0\rangle\langle 0|, \quad M_1 = |1\rangle\langle 1| + |2\rangle\langle 2|. \quad (66)$$

Using standard classical cloning, this measurement is not able to distinguish between the basis states $|1\rangle$ and $|2\rangle$ since they both lead to the same probability distribution $P(0|1) = P(0|2) = 0$ and $P(1|1) = P(1|2) = 1$. However, it is able to distinguish the basis state $|0\rangle$ from $|1\rangle$ and $|2\rangle$. If we were therefore to permute the states $|0\rangle$ and $|1\rangle$ we would be able to distinguish instead $|1\rangle$ from the other two. Using this idea, we can therefore consider a generalized cloning protocol, which consists of performing different hypothesis tests in parallel, by dividing the N clones into two groups. That is, we can consider the following protocol:

$$\begin{aligned} U|0\rangle|0\rangle^{\otimes(N-1)} &= |0\rangle^{\otimes N/2}|1\rangle^{\otimes N/2}, \\ U|1\rangle|0\rangle^{\otimes(N-1)} &= |1\rangle^{\otimes N/2}|0\rangle^{\otimes N/2}, \\ U|2\rangle|0\rangle^{\otimes(N-1)} &= |2\rangle^{\otimes N}. \end{aligned} \quad (67)$$

We then use the first $N/2$ measurement results to decide if the state is $|0\rangle$ or not, and the second $N/2$ results to decide if it is $|1\rangle$ or not. At the end of the procedure, we therefore identify the state, with error which still drops exponentially in the asymptotic limit.

In the Appendixes we show more generally that this procedure allows us to identify any basis $|i\rangle$, with error that drops exponentially in the asymptotic limit, using any measurement that is not trivial.

Although this protocol will in general definitely not be the optimal one, in terms of minimizing the average TVD error, we find it nevertheless fascinating that such a simple protocol exists, and can be used as a subroutine to show that any measurement can reproduce any other measurement, with arbitrarily small error, in the asymptotic regime.

Finally, we must also address the case where the dimension of the measurement we have access to differs from the number of outcomes of the target measurement. If the dimension is larger, then the above analysis goes through, as we can just take the states $|e_i\rangle$ to span a subspace smaller than the dimension of the measurement, without any difficulty. The problematic case arises when the number of outcomes is larger than the dimension of the measurement. In this case, we have to consider a sufficient number of uses of the measurement at a time. As a concrete example, if we want to reproduce a measurement with five outcomes, and the measurement we have access to acts on qutrits, then we will consider two uses of the measurement, which we treat (conceptually) as a single measurement on a nine-dimensional space, and the states $|e_i\rangle$ forming a five-dimensional subspace of this nine-dimensional space.¹³ The above analysis then applies, without any additional difficulties introduced.

V. BLOCK-CODING AND FINITE-RATE PROTOCOLS

Up until now, we have considered the problem of reproducing a single use of a target measurement device, and shown that it can be perfectly reproduced, in the limit where we use the available measuring device an unbounded number of times. If we therefore consider the *rate* of the protocol, defined as the number of measurements performed in the protocol per measurement reproduced, we see that the protocols we have considered are *zero rate*. An interesting and important question is whether it is possible to have *finite-rate* measurement reproduction, with protocols which make use of only a finite number of measurements in the protocol per measurement reproduced.

In this section, we will see that this is possible, if we allow ourselves the analog of *block coding* from classical and quantum information theory. In this context, this means that we will not try to reproduce measurements one by one, but instead will attempt to reproduce only *multiple measurements in parallel*. In particular, if the target measurement is $\mathcal{T}$, when we will aim to reproduce $\mathcal{T}^{\otimes k}$, that is, the product measurement on k systems, with outcomes $\mathbf{i} = i_1 i_2 \dots i_k$, and associated subchannels

$$\Gamma_{\mathbf{i}} = \Gamma_{i_1} \otimes \Gamma_{i_2} \otimes \dots \otimes \Gamma_{i_k}. \quad (68)$$

We then want to consider the question of reproducing this parallel measurement, and are specifically interested in the asymptotic scenario, where k becomes large, and ask for how many measurements N are necessary for a perfect reproduction, and whether a vanishing error can be achieved even if the ratio $\frac{k}{N}$ is kept constant, i.e., such that the number of measurements performed by the protocol is proportional to the number of target measurements. The motivation for studying this limit comes from classical (and quantum) information

¹³If this is viewed as problematic per se, then we can consider a modified protocol, where $|\Omega\rangle$ is a state on $(\mathbb{C}^9)^{\otimes N/2}$, and a unitary U which acts on $\mathbb{C}^5 \otimes (\mathbb{C}^9)^{\otimes N/2}$, such that $U|i\rangle|\Omega\rangle = |0\rangle|i\rangle^{\otimes N/2}$. In this way, we do not need to treat the incoming system as a nine-dimensional system, but simply map its state into a nine-dimensional Hilbert space, and "reset" the incoming particle to a fixed state $|0\rangle$.

theory, where we know that by using block-coding schemes, finite-rate protocols for communicating over noisy channels are possible.

Our main result in this section is that we can directly make use of finite-rate classical coding schemes in order to show that measurement reproduction can also be achieved at finite rate. In particular, we have the following:

Main result 2. Measurement reproduction can be achieved at finite rate when reproducing multiple measurements in parallel. The asymptotic rate R is lower bounded by the classical capacity

$$C = \max_{p(x)} I(A : X),$$

of the classical channel $p(a|x) = \langle x|M_a|x \rangle$ associated to the measurement, and where $I(A : X)$ is the mutual information of the joint probability distribution $p(a, x) = p(x)p(a|x)$ between the input and output of the channel $I(A : X) = H(X) - H(X|A)$, with $H(X)$ and $H(X|A)$ the Shannon entropy and conditional entropy, respectively.

In order to prove this result, the first step is to realize that the previous analysis, which showed that the problem of reproducing an arbitrary measurement reduces to the problem of reproducing just a complete von Neumann measurement, carries over into the present setting. In particular, in order to show that it is possible to find a finite-rate measurement reproduction protocol, we only need to show that there exists a finite-rate protocol for reproducing a complete von Neumann measurement. This follows since subroutine 1 only makes use of the von Neumann measurement in order to perform an arbitrary measurement and does not involve any further measurement. It therefore does not alter any conclusion about the rate of the combined protocol.

As stated in the main result 2, to every measurement $\mathbb{M} = \{M_a\}_a$ (specified only in terms of its POVM elements), we can naturally associate a classical channel, that is, a conditional probability distribution $p(a|x)$, via

$$p(a|x) = \langle x|M_a|x \rangle, \quad (69)$$

where the input to the channel is x , which labels the basis state, and the output of the channel is the measurement result a .

From the noisy-channel coding theorem of classical information theory, we know that there exists a sequence of codes (of increasing length), with the error in transmission decreasing with the size of the code, at a rate which approaches the *capacity*

$$C = \max_{p(x)} I(A : X), \quad (70)$$

where $I(A : X)$ is the mutual information of the joint probability distribution $p(a, x) = p(x)p(a|x)$ between the input and output of the channel $I(A : X) = H(X) - H(X|A)$, with $H(X)$ and $H(X|A)$ the Shannon entropy and conditional entropy, respectively.

Now, the noisy-channel coding theorem guarantees the existence of encodings and decodings, namely, functions from

$\mathbf{x} = x_1 \dots x_k$ (strings of length k), to codewords $c_{\mathbf{x}}$ (strings of length $N = R_k k$), and a decoding function $g(\cdot)$, which maps back from length- N strings into length- k strings, with R_k , the rate at block length k , approaching the capacity of the channel C as k tends to infinity.

We can make use of this in the below single-round protocol, making use of $N = R_k k$ imperfect measurements. Before stating the protocol, we note that since we are now aiming to perform k complete von Neumann measurements in parallel, the input will be a state of the form

$$|\psi\rangle = \sum_{\mathbf{i}} \alpha_{\mathbf{i}} |i_1\rangle \dots |i_k\rangle, \quad (71)$$

where $\alpha_{\mathbf{i}}$ are arbitrary amplitudes.

Finite-rate classical-coding-based protocol

(1) Prepare $N - k = (R_k - 1)k$ ancillary particles in the state $|\Omega\rangle$.

(2) Apply a unitary U which sends the $N = R_k k$ particles into states encoding the classical codewords

$$\begin{aligned} U|i_1\rangle \dots |i_k\rangle |\Omega\rangle &= |c_{\mathbf{i}}\rangle, \\ &= |(c_{\mathbf{i}})_1\rangle \dots |(c_{\mathbf{i}})_N\rangle, \end{aligned} \quad (72)$$

where $(c_{\mathbf{i}})_j$ denotes the j th element of the codeword $c_{\mathbf{i}}$.

(3) Measure each of the N particles using the available measurement $\mathbb{M}$, producing a string of outcomes $\mathbf{a} = a_1 \dots a_N$.

(4) Apply the decoding function $g(\cdot)$ of the classical code to the string of outcomes, to produce $\mathbf{j} = g(\mathbf{a})$, a length- k string, which is returned as the outcome of the k measurements.

(5) Prepare the state $|j_1\rangle \dots |j_k\rangle$, and return as the post-measurement state.

By construction, with error which vanishes as k becomes large, the string $\mathbf{j}$ will be perfectly correlated with the computational basis state $|\mathbf{i}\rangle = |i_1\rangle \dots |i_k\rangle$ since the unitary U turned this basis state into states encoding the classical codewords. The measurements performed on these codeword states produced measurement outcomes, which can be identified with the outcomes of the associated channel (69). However, these codewords were precisely the encodings which could then be decoded after the application of the channel, which in this case corresponds to identifying the original basis state.

Putting everything together, we therefore see that this (approximately) performs a complete von Neumann measurement on k particles since the probability of having a given computational basis state is precisely $|\alpha_{\mathbf{i}}|^2$, with error which vanishes as k tends to infinity. Crucially, unlike our previous analysis, the number of additional particles required per particle being measured is now finite, given by the rate R_k of the code used, which we know approaches the capacity C in the limit of large k . This shows that we can achieve measurement reproduction at finite rate.

The above protocol, based upon classical channel coding schemes, in general would not be expected to be optimal. Nevertheless, their existence highlights how classical channel

coding theory can be used directly for measurement reproduction, even if not at the optimal possible rate. A highly relevant open question is to understand the optimal rates that can be achieved, and how more general protocols might achieve them. In particular, we can see immediately that the above protocol is highly constrained in a number of ways. First, it only uses a single round of measurements; second, it does not rely on the postmeasurement states at all; and lastly the encodings are into product states, not general entangled states. It is reasonable to believe that relaxing all of these could provide significant improvements in the ultimate rates achievable.

It is also worth noting that even a straightforward modification of the above idea can in principle improve the rate. In particular, (69) is only one particular classical channel that we can associate with a quantum measurement. An improved rate can be achieved by optimizing over associated classical channels, by varying the states measured, that is by considering the class of channels $\mathcal{C}_{\mathbb{M}}$,

$$\mathcal{C}_{\mathbb{M}} = \{p(a|x) \mid p(a|x) = \text{tr}(M_a \rho_x)\}, \quad (73)$$

where ρ_x are quantum states for all x , and using the associated encoding and decoding scheme.

VI. PROBABILISTIC PROTOCOLS

In this final section we return to the possibility of using slightly generalized protocols, which *probabilistically* assign outcomes for the target measurement. We will see that, for the average TVD error considered here, such probabilistic protocols can in fact outperform any deterministic protocol.

We will present the main idea by returning to the example from Sec. II which made multiple uses of a trine measurement in order to approximately reproduce a complete von Neumann measurement. For the case of two uses of the trine, we gave an optimal (deterministic) protocol in (13): that is, we gave an assignment of an outcome i , given the pair of trine measurement outcomes (a_1, a_2) . A more general possibility is to decide which outcome to assign according to a probabilistic protocol. That is, given the outcomes of the trine measurements, we now need to specify the probabilities $q(i|a_1, a_2)$. From this perspective, we see that the class of protocols considered up until now allowed only for deterministic assignments, such that $q(i|a_1, a_2) = \delta_{i,f(a_1, a_2)}$, for some function $f(a_1, a_2)$.

We now demonstrate, surprisingly, that probabilistic protocols are more powerful than deterministic ones, by presenting a probabilistic protocol which outperforms the optimal deterministic protocol in (13). The probabilistic protocol we consider is closely related to the deterministic protocol (13). The difference is that now, when we obtain one of the four strings (a_1, a_2) which were previously mapped to 1 all the time, we now map them to 1 with probability z , and flip them to 0 with probability $(1 - z)$. The five strings which were previously always mapped to 0 remain so. Mathematically,

$$\begin{aligned} q(1|1, 1) &= q(1|1, 2) = q(1|2, 1) = q(1|2, 2) = z, \\ q(0|1, 1) &= q(0|1, 2) = q(0|2, 1) = q(0|2, 2) = (1 - z), \\ q(0|0, 0) &= q(0|0, 1) = q(0|0, 2) = q(0|1, 0) \\ &= q(0|2, 0) = 1. \end{aligned} \quad (74)$$

(The previous deterministic protocol corresponds to taking $z = 1$.) In terms of POVM elements, this has the effect of changing $Q_0^{(2)}$ and $Q_1^{(2)}$ from (27) into

$$\tilde{Q}_0^{(2)} = zQ_0^{(2)} + (1 - z)\mathbb{I}, \quad \tilde{Q}_1^{(2)} = zQ_1^{(2)}. \quad (75)$$

A direct calculation allows us to analytically solve for the optimal choice of z . We find that

$$z^* = \frac{9}{8} - \frac{9}{8\sqrt{41}} \approx 0.95 \quad (76)$$

is the optimal value for this family of protocols parametrized by z , and leads to a slightly improved average TVD error of

$$\langle \tilde{\epsilon}_{(2)} \rangle^* = \frac{\sqrt{41} - 5}{16} \approx 0.088, \quad (77)$$

which should be compared to the deterministic case ($z = 1$) that has error $\langle \epsilon_{(2)} \rangle^* = \frac{2(\sqrt{2}-1)}{9} \approx 0.092$.

Numerical investigation indicates this is not the optimal probabilistic protocol, and even lower average TVD errors can be achieved. Nevertheless, it demonstrates that (simple) probabilistic protocols can provably outperform the best deterministic ones.

One comment is in order. The advantage offered by probabilistic protocols can only arise when the figure of merit considered is nonlinear. If the figure of merit would be linear, then a simple argument shows that a deterministic protocol will always perform at least as well as any probabilistic one. Since our figure of merit here is well motivated, quantifying the average discrepancy between the target and implemented measurement, it is genuinely of interest to realize that injecting randomness into the process can be beneficial. We leave it as future research to fully investigate and understand the nature of this probabilistic advantage.

VII. CONCLUSIONS

In this paper we have considered the problem of using one measurement to implement another. In particular, we have been interested in the situation where the available measurement is used multiple times, a type of implementation (or simulation) which has received limited attention to date. Surprisingly, we have shown that a sufficient number of uses of any measurement allows for the implementation of any other measurement, with error dropping off exponentially fast in the number of measurement uses. As a concise summary, this shows that all measurements are asymptotically equivalent to each other. In order to show this general result, we demonstrated two key subroutines, which are interesting in their own right: a subroutine for implementing arbitrarily well a complete von Neumann measurement, and a subroutine making use of a complete von Neumann measurement to implement an arbitrary measurement, including the postmeasurement state. Furthermore, we showed that measurement implementation can be achieved at finite rate in a block-coding setting; in order to implement multiple uses of a measurement in parallel, it is sufficient to make use of a finite number of

measurements per measurement performed. To show this, we demonstrated how results from classical Shannon theory can be directly harnessed to build measurement implementation protocols. Finally, we showed that allowing protocols to further be probabilistic can be beneficial by exhibiting a simple probabilistic protocol using the trine which outperforms the best deterministic protocol.

Our results are important from a number of different perspectives. First, as a special case of our general results, we can consider the task of measurement purification, where an imperfect measurement is used multiple times in order to implement a high-quality complete von Neumann measurement. We have shown that even with a few uses, improvements can be made. This may be directly relevant for current quantum technologies, including noisy intermediate scale (NISQ) devices, and suggests an alternative approach to directly engineering less noisy measurements, instead to overcome noise through a measurement implementation protocol instead. This possibility will be relevant in the situation that entangling unitaries can be performed with relatively high fidelity, whereas measurements are comparatively imperfect. In real applications, of course, neither process will be perfect, and it is an important future question to probe this tradeoff.

Ultimately, we could envisage that measurement implementation becomes a standard technique for improving measurements. Similarly, if a given measurement is hard to engineer, our methods show how a simpler-to-engineer measurement could in principle be used instead, that is, without viewing the available measurement as noisy. This is precisely what is shown in our detailed analysis of the trine measurement: it can implement the von Neumann measurement, even though it is not a noisy measurement, and even though at first sight it appears to be a very different measurement, one which is itself extremal in the space of measurements.

From a fundamental perspective, our results raise interesting questions concerning how the act of measurement leads to classical information about a quantum system. Through our protocols, we see how information can be spread out amongst multiple systems, and in the process completely overcome noise or other limitations which arise from specific fixed measurements. Our results can also be viewed from a resource-theoretic perspective, whereby it can be viewed as an enlargement of the class of allowed operations to incorporate multiple uses of the available resourceful measurement. From this perspective, it shows that if no cost is associated with making multiple uses, then this trivializes the theory (making all measurements equivalent). However, if only finite numbers of uses are allowed, our results indicate that the theory has a rich structure, allowing for an interesting set of interconversions between measuring devices in general.

There are many fascinating open directions to explore. First, it would be interesting to understand optimal protocols, and the properties they exhibit. We have seen that classical-cloning-based protocols are not always optimal (as demonstrated by a detailed analysis of the trine), but sometimes they are. We do not have a full understanding of what determines the properties of an optimal protocol. Furthermore, we considered here the average TVD error as our

figure of merit to optimize. It is important to understand how robust our findings are to other figures of merit, e.g., when we look at the worst case error, instead of the average case error.

The protocols described here made use of detailed structure and properties of the available measurement. An important extension is to devise protocols that work for measurements about which we only have partial information. Preliminary investigation shows that some of our protocols indeed work for a range of available measurements (e.g., a single classical cloning protocol will work for a whole range of noisy measurements), however, these are only preliminary observations, and a complete understanding of how to optimally make use of partial information is missing.

In the paradigm here, we placed no restrictions on the types of unitary interactions we are allowed to perform in the protocol. This is a valid choice to make when interested in the fundamental limitations which arise from the available measurements themselves. It would nevertheless be interesting to further place restrictions on the unitary operations allowed, and to understand what effect, if any, this has on the results. For example, we may demand that all the unitary operations should be realizable by a polynomial-sized quantum circuit. This would be particularly interesting to study from a practical perspective, where the complexity of the unitary circuit could become a limiting factor in making the protocols realistic and feasible.

In quantum computers, one often only needs to measure a small subset of the qubits at the end of the computation. This raises the possibility that the remaining qubits could be reinitialized, and used as the ancillary qubits in our measurement scheme. A particularly interesting possibility is to make use of some form of block-coding scheme. Then we could use all the qubits to form codewords before measuring them in order to reproduce a more accurate measurement on the small subset of qubits.

In the case of block coding, we have only shown that finite-rate reproduction is possible by relying on results from classical Shannon theory, in particular from the noisy-channel coding theorem. Our belief is that significantly better rates could be achieved by considering more general protocols, not based upon classical information theory. In particular, we leave it as an open question whether we can make use of quantum codes for classical communication to improve our results, and whether we can characterize the ultimate optimal rate of measurement reproduction.

ACKNOWLEDGMENTS

We thank N. Brunner, S. Popescu, I. Čepaitė, A. Daley, G. Pelegrí, J. Pritchard, and C. Corlett for insightful discussions. We thank A. Essery for bringing to our attention that Table I in the previous version of this manuscript was missing two rows. P.S. acknowledges support as a CIFAR Azrieli Global Scholar in the Quantum Information Science Program. P.S. gratefully acknowledges support from a Royal Society University Research Fellowship (UHQT/NFQI). N.L. gratefully acknowledges support from the UK Engineering and Physical Sciences Research Council through Grants No. EP/R043957/1, No. EP/S005021/1, and No. EP/T001062/1.

APPENDIX A: EVALUATING OR BOUNDING THE AVERAGE TOTAL VARIATION DISTANCE (TVD) ERROR

In this Appendix we will show how the integral over all states can be performed for qubits, in order to obtain an explicit optimization problem for the TVD error.

For higher-dimensional system, while this average does not appear to be easily computable, we show how it can be upper bounded, using ideas from [20]. This upper bound can then be evaluated, and proves useful for bounding the error of protocols in general situations.

1. Two outcome qubit measurements

Our starting point is to express (7) more explicitly. In particular, we will write $|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$ (i.e., in Bloch sphere notation), such that $d\psi = \frac{1}{4\pi} \sin \theta d\theta d\phi$.

We will assume that $M_i^{(N)}$ is diagonal in the $\{|0\rangle, |1\rangle\}$ basis, with $M_0^{(N)}$ explicit given as

$$M_0^{(N)} = (1 - \epsilon_0)|0\rangle\langle 0| + \epsilon_1|1\rangle\langle 1|. \quad (\text{A1})$$

Two comments are in order here. First, given any protocol that leads to a measurement with nonvanishing off-diagonal elements, we can always find another protocol with vanishing off-diagonal elements. This just says that we can always dephase any protocol. One way to see this is that we can always begin the protocol by dephasing the state to be measured by entangling it with another ancillary system. If we instead include this dephasing of the state in the measurement (which essentially amounts to consider the Heisenberg picture of this process), it will have the effect of dephasing the POVM

elements. Hence, there is no difficulty in obtaining diagonal POVMs.

On the other hand, there *is* a lack of generality in the class of protocols that we consider by restricting to protocols which dephase the incoming state, and one may wonder if it could in fact be optimal to consider such nondiagonal protocols. We do not have a proof that this cannot happen, but we do note that this would be a rather surprising result if it were true, and would be attributable to our figure of merit. This is because if we were to instead consider a trace-distance-based measure, which would focus on the *worst-case* behavior of the protocol, then it follows straight away that an optimal protocol cannot necessarily have off-diagonal terms. This is because we can view the error instead from the perspective of distinguishability; the fact that there is an error says that the target measurement and the simulated measurement can be distinguished. We know that under the trace distance, two measurements cannot become more distinguishable under preprocessing (which is a statement of the data processing inequality). However, dephasing is a preprocessing, which leaves the target measurement unaltered.

This argument shows that as far as the worst-case error is concerned, we can restrict to diagonal measurements without loss of generality. Furthermore, in Appendix A 2 below, we will show how to bound the average TVD error from above. This bound was also shown to satisfy the data processing inequality in [20]. Hence, it seems a reasonable assumption that the average TVD error should have this behavior too, and we will therefore restrict our attention to diagonal measurements from here on.

With this in place, we now have

$$P^{(N)}(0) = (1 - \epsilon_0) \cos^2 \frac{\theta}{2} + \epsilon_1 \sin^2 \frac{\theta}{2}, \quad (\text{A2})$$

and so

$$\begin{aligned} \langle \epsilon_{(N)} \rangle &= \frac{1}{2\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta |P^{(N)}(0) - \wp(0)| \\ &= \int_0^\pi \sin \theta d\theta \left| (1 - \epsilon_0) \cos^2 \frac{\theta}{2} + \epsilon_1 \sin^2 \frac{\theta}{2} - \cos^2 \frac{\theta}{2} \right| \\ &= \int_0^\pi \sin \theta d\theta \left| \epsilon_1 \sin^2 \frac{\theta}{2} - \epsilon_0 \cos^2 \frac{\theta}{2} \right|. \end{aligned} \quad (\text{A3})$$

The term $\epsilon_1 \sin^2 \frac{\theta}{2} - \epsilon_0 \cos^2 \frac{\theta}{2}$ changes sign at $\theta = \theta_c$, where θ_c satisfies

$$\tan^2 \frac{\theta_c}{2} = \frac{\epsilon_0}{\epsilon_1}. \quad (\text{A4})$$

We can therefore express the integral as

$$\langle \epsilon_{(N)} \rangle = \int_{\theta_c}^\pi \sin \theta d\theta \left(\epsilon_1 \sin^2 \frac{\theta}{2} - \epsilon_0 \cos^2 \frac{\theta}{2} \right) - \int_0^{\theta_c} \sin \theta d\theta \left(\epsilon_1 \sin^2 \frac{\theta}{2} - \epsilon_0 \cos^2 \frac{\theta}{2} \right). \quad (\text{A5})$$

This can be now be readily integrated. We find

$$\langle \epsilon_{(N)} \rangle = \frac{1}{4}(\epsilon_0 + \epsilon_1)(1 - 4 \cos \theta_c + \cos 2\theta_c) + 2\epsilon_1 \cos \theta_c. \quad (\text{A6})$$

Substituting for θ_c , and simplifying, we finally arrive at

$$\langle \epsilon_{(N)} \rangle = \frac{\epsilon_0^2 + \epsilon_1^2}{\epsilon_0 + \epsilon_1}. \quad (\text{A7})$$

2. General upper bound

Going beyond qubits, it appears challenging to explicitly evaluate the average TVD error. Nevertheless, we can make progress by finding an upper bound, based upon the results in [20]. For completeness, we reproduce the required results here.

The quantity of interest in the general case of a qudit von Neumann measurement is

$$\langle \epsilon_{(N)} \rangle = \sum_i \int d\psi |P^{(N)}(i|\psi) - \wp(i|\psi)|, \quad (\text{A8})$$

where the average is over all pure states $|\psi\rangle$ in $\mathbb{C}^d$, and we use a notation that makes it explicit that $P^{(N)}(i|\psi) = \langle \psi | M_i^{(N)} | \psi \rangle$ and $\wp(i|\psi) = |\langle i | \psi \rangle|^2$ depend upon the state $|\psi\rangle$.

We start by making use of Jensen's inequality, which states that for a concave function f , and a random variable X , $\langle f(X) \rangle \leq f(\langle X \rangle)$. In particular, for each i we define a random variable $X_i = [P^{(N)}(i|\psi) - \wp(i|\psi)]^2$ (with the average over ψ), and take $f(x) = \sqrt{x}$, such that $f(x^2) = |x|$. Applying Jensen's inequality, we therefore see that

$$\langle \epsilon_{(N)} \rangle = \sum_i \int d\psi |P^{(N)}(i|\psi) - \wp(i|\psi)| \leq \sum_i \sqrt{\int d\psi [P^{(N)}(i|\psi) - \wp(i|\psi)]^2} = \sum_i \sqrt{\int d\psi [\langle \psi | M_i^{(N)} | \psi \rangle - \langle \psi | \Pi_i | \psi \rangle]^2}. \quad (\text{A9})$$

We can reexpress this by noting that

$$[\langle \psi | M_i^{(N)} | \psi \rangle - \langle \psi | \Pi_i | \psi \rangle]^2 = \text{tr}[\psi \langle \psi |^{\otimes 2} (M_i^{(N)} - \Pi_i)^{\otimes 2}]. \quad (\text{A10})$$

The benefit of this is that we can now use the fact that

$$\int d\psi |\psi\rangle \langle \psi|^{\otimes 2} = \frac{1}{d_{\text{sym}}} \Pi_{\text{sym}}, \quad (\text{A11})$$

where $d_{\text{sym}} = \frac{d(d+1)}{2}$ is the dimension of the symmetric subspace of two qudits. That is, the integral is proportional to the projector onto the symmetric subspace. Finally, we can use the fact that $\Pi_{\text{sym}} = \frac{1}{2}(\mathbb{I} + S)$, with S the SWAP operator. Putting everything together, after a lengthy but straightforward calculation, we arrive at the bound

$$\langle \epsilon_{(N)} \rangle \leq \frac{1}{\sqrt{2d_{\text{sym}}}} \sum_i \sqrt{\left(\sum_{j \neq i} \epsilon_{j|i} - \sum_{k \neq i} \epsilon_{i|k} \right)^2 + \left(\sum_{j \neq i} \epsilon_{j|i} \right)^2 + \sum_{k \neq i} \epsilon_{i|k}^2}, \quad (\text{A12})$$

where, as above, we assume that $M_i^{(N)}$ is diagonal in the basis of the von Neumann measurement $\{|i\rangle\}$, and can be expressed as

$$M_i^{(N)} = \left(1 - \sum_{j \neq i} \epsilon_{j|i} \right) |i\rangle \langle i| + \sum_{k \neq i} \epsilon_{i|k} |k\rangle \langle k|, \quad (\text{A13})$$

which is in terms of the *errors* $\epsilon_{i|j}$ to return the outcome i when the incoming basis state was j . This is a convenient way to represent each POVM element, and focuses on the errors that it makes. Notably, due to the normalization of measurements, the probability of correctly identifying the state $|i\rangle$ is completely determined by the probabilities of incorrectly announcing the outcome i when the state is any other state $|j\rangle \neq |i\rangle$.

In (A12) we can see immediately that if all the individual errors vanish, then on average there is no error in the measurement, as must be the case. Furthermore, we see that if the individual errors (on basis states) can be shown to decrease exponentially with N , then the average error must also decrease exponentially with N .

APPENDIX B: SOLVING ANALYTICALLY FOR THE MINIMAL AVERAGE TVD ERROR

In this Appendix we will show that it is possible to analytically solve for the minimal average TVD error that can be achieved in the case of approximately implementing an ideal complete qubit von Neumann measurement. The method we will apply is a geometrical one, which shows that there are three regimes, which we can explicitly identify.

Let us begin analyzing the constraints in (26). Geometrically, we see that we are optimizing over a rectangle in the ϵ_0 - ϵ_1 plane. Denoting the coordinate by (ϵ_0, ϵ_1) , the vertices of this rectangle are seen to be at $(1 - \lambda_{\text{max}}, \lambda_{\text{min}})$, $(1 - \lambda_{\text{max}}, \lambda_{\text{max}})$, $(1 - \lambda_{\text{min}}, \lambda_{\text{min}})$, and $(1 - \lambda_{\text{min}}, \lambda_{\text{max}})$.

We can also straightforwardly identify the contours of the objective function. In particular, points which satisfy

$$\epsilon_0^2 + \epsilon_1^2 = c(\epsilon_0 + \epsilon_1) \quad (\text{B1})$$

are those where the function takes value c . This can be reexpressed as

$$\left(\epsilon_0 - \frac{1}{2}c \right)^2 + \left(\epsilon_1 - \frac{1}{2}c \right)^2 = \left(\frac{c}{\sqrt{2}} \right)^2, \quad (\text{B2})$$

that is, the contours are circles of radius $\frac{c}{\sqrt{2}}$, centered at $(\epsilon_0^c, \epsilon_1^c) = (\frac{1}{2}c, \frac{1}{2}c)$. Clearly, $c = 0 = \langle \epsilon_{(N)} \rangle^*$ corresponds to

$\epsilon_0 = \epsilon_1 = 0$, and this will only be in the rectangle if $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$. From (A1), this implies that $M_0^{(N)} = |0\rangle\langle 0|$, as expected.

When $\lambda_{\min} > 0$ and/or $\lambda_{\max} < 1$, then there will be a range of c such that the center of the circle is outside the rectangle. Our optimization problem can be understood geometrically as finding the size of the smallest circle which *just touches* the rectangle. The point at which the circle and rectangle touch will be the optimal solution $(\epsilon_0^*, \epsilon_1^*)$.

We can now make two key observations. First, since the center of the circle always moves from the point $(0, 0)$ along the line $\epsilon_0 = \epsilon_1$, the point where the circle touches will either be along the bottom edge, when $\epsilon_1 = \lambda_{\min}$, along the left-hand edge, when $\epsilon_0 = 1 - \lambda_{\max}$, or at the bottom left-hand vertex, when $(\epsilon_0, \epsilon_1) = (1 - \lambda_{\max}, \lambda_{\min})$.

Second, if the smallest circle touching one of these edges (and not the corner), the edge will necessarily be tangent to the circle.

Let us therefore consider first the case where the ellipse touches along the bottom edge of the rectangle. In this case, the point at which the two touch is where the tangent to the circle is horizontal. This will be at the point $(\epsilon_0^c, \epsilon_1^c + \frac{c}{\sqrt{2}}) = (\frac{c}{2}, \frac{c}{2} + \frac{c}{\sqrt{2}})$. That is, these points lie on the line

$$\epsilon_1 = (1 + \sqrt{2})\epsilon_0. \quad (\text{B3})$$

This line must intersect the bottom of the rectangle. This will only happen if the intersection point is between the corners $(1 - \lambda_{\max}, \lambda_{\min})$ and $(1 - \lambda_{\min}, \lambda_{\min})$. The intersection occurs at $\epsilon_1 = \lambda_{\min}$, and so we obtain the region

$$(1 + \sqrt{2})(1 - \lambda_{\max}) \leq \lambda_{\min} \leq \frac{1}{\sqrt{2}}. \quad (\text{B4})$$

The case where the circle touches along the left-hand edge is similar. We find that points with vertical tangent lie on the line $\epsilon_0 = (1 + \sqrt{2})\epsilon_1$, the intersection occurs at $\epsilon_0 = 1 - \lambda_{\max}$, and this will be between the corners when

$$(1 + \sqrt{2})\lambda_{\min} < 1 - \lambda_{\max} \leq \frac{1}{\sqrt{2}}. \quad (\text{B5})$$

Finally, when neither of these cases hold, we see that the intersection will occur at one of corner of the rectangle (either the bottom corner, or one of the top two corners, respectively). Putting everything together leads to Table I.

APPENDIX C: GENERALIZED CLASSICAL CLONING CAN ALWAYS DISTINGUISH A BASIS OF STATES

In this Appendix we will show that it is always possible to use a generalized classical cloning protocol to distinguish between a basis of states. As shown in the example in the main body in Sec. IV D, all we in fact need to be able to do is distinguish a single basis state $|i\rangle$ (which we will take to be $|0\rangle$ without loss of generality) from the remaining basis states. That is, if we have $P(a|0) \neq P(a|i)$ for $i \neq 0$, even if all other $P(a|i)$ are equal (and hence do not allow us to distinguish the states $|i\rangle$, for $i \neq 0$, then we can still distinguish the entire basis by a generalized protocol. In particular, the fact that $P(a|0) \neq P(a|i)$ means that with error which drops exponentially in the asymptotic limit of N uses of the measurement, we can decide whether the state was $|0\rangle$ or not.

Now, just as in the example in the main text, we could now consider simple pairwise permutations of the basis states, i.e., permuting $|0\rangle$ and $|1\rangle$, before applying the cloning procedure, and this would allow us then to determine whether the state was $|1\rangle$ or not. Using this basic idea, we can consider carrying out $d - 1$ hypothesis tests in parallel by dividing the N particles ($N - 1$ ancillary particles and the incoming system) into $(d - 1)$ groups (we choose N to be divisible by $d - 1$), and apply the following unitary:

$$\begin{aligned} U|0\rangle|0\rangle^{\otimes(N-1)} &= |e_0\rangle^{\otimes N/(d-1)}|e_1\rangle^{\otimes N/(d-1)} \dots |e_{d-2}\rangle^{\otimes N/(d-1)}, \\ U|1\rangle|0\rangle^{\otimes(N-1)} &= |e_1\rangle^{\otimes N/(d-1)}|e_0\rangle^{\otimes N/(d-1)}|e_1\rangle^{\otimes (d-3)N/(d-1)}, \\ &\vdots \\ U|d-2\rangle|0\rangle^{\otimes(N-1)} &= |e_{d-2}\rangle^{\otimes (d-2)N/(d-1)}|e_0\rangle^{\otimes N/(d-1)}, \\ U|d-1\rangle|0\rangle^{\otimes(N-1)} &= |e_{d-1}\rangle^{\otimes N}. \end{aligned} \quad (\text{C1})$$

Considering the first group of $N/(d - 1)$ particles, this can be used to determine whether the state is $|0\rangle$ or not. Considering the second group of $N/(d - 1)$ particles, this can be used to determine whether the state is $|1\rangle$ or not. This continues until the last group of $N(d - 1)$ particles, which is used to determine if the state is $|d - 2\rangle$. If in each case we determine the state not to be one of the $d - 1$ states checked for, then we can conclude that the state must have been the remaining state $|d - 1\rangle$. This procedure thus allows us to determine the state.

Concerning the errors made, each hypothesis test performed now uses $N/(d - 1)$ particles, rather than the N particles considered altogether previously. However, since d is a constant, the probability of making an error still drops off exponentially in N , and hence it follows that altogether the error drops off exponentially in the number of uses N .

The only case in which the above fails to work is if it is impossible to distinguish a single state $|e_0\rangle$ from the rest of a basis of states $|e_i\rangle$. That is, it must be the case that for all bases, the probability distributions $P(a|i) = \langle e_i|M_a|e_i\rangle$ are independent of i .¹⁴ To see what this implies about the POVM elements M_a , let us consider that we use the eigenbasis $|v_i^a\rangle$ of M_a as the basis. It follows that in order to fail, we have $\langle v_i^a|M_a|v_i^a\rangle = \lambda_i^a = \lambda^a$ is constant for all i . That is, $M_a = \lambda^a \mathbb{I}$, each POVM element must be proportional to the identity operator. These are precisely the trivial measurements, which we have ruled out from the start, since it is clear that such measurements provide no information about the quantum state being measured, and therefore cannot be used to simulate any measurement.

Altogether, this shows that the above procedure will always work for a nontrivial measurement; it must be possible to find a basis of states where at least one state is distinguishable from the remainder, and once this can be accomplished, the above shows that we can then in fact distinguish the entire basis of states by carefully applying permutations.

We note finally that the above procedure is by no means optimal in general. In the general case, we would expect

¹⁴Note that this does not mean that they have to be the same for all choices of basis $|e_i\rangle$, but just that for all bases, the probabilities are independent of the basis state.

multiple states to be distinguished, and in this case, we can adapt the generalized cloning scheme to better make use of the information available. Nevertheless, it is important to

realize that even with the weakest possible requirement (of only one state distinguishable), we can prove the desired result.

-
- [1] H. P. Yuen, Design of transparent optical networks by using novel quantum amplifiers and sources, *Opt. Lett.* **12**, 789 (1987).
 - [2] M. Dall’Arno, G. M. D’Ariano, and M. F. Sacchi, Purification of noisy quantum measurements, *Phys. Rev. A* **82**, 042315 (2010).
 - [3] R. Hicks, B. Kobrin, C. W. Bauer, and B. Nachman, Active readout-error mitigation, *Phys. Rev. A* **105**, 012419 (2022).
 - [4] J. M. Günther, F. Tacchino, J. R. Wootton, I. Tavernelli, and P. K. Barkoutsos, Improving readout in quantum simulations with repetition codes, *Quantum Sci. Technol.* **7**, 015009 (2022).
 - [5] A. Abbas, S. Andersson, A. Asfaw, A. Corcoles, L. Bello, Y. Ben-Haim, M. Bozzo-Rey, S. Bravyi, N. Bronn, L. Capelluto, A. C. Vazquez, J. Ceroni, R. Chen, A. Frisch, J. Gambetta, S. Garion, L. Gil, S. D. L. P. Gonzalez, F. Harkins, T. Imamichi *et al.*, Learn quantum computation using qiskit (2020), <https://github.com/Qiskit/textbook>.
 - [6] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
 - [7] D. B. Hume, T. Rosenband, and D. J. Wineland, High-fidelity adaptive qubit detection through repetitive quantum nondemolition measurements, *Phys. Rev. Lett.* **99**, 120502 (2007).
 - [8] M. Sinhal, Z. Meir, K. Najafian, G. Hegi, and S. Willitsch, Quantum-nondemolition state detection and spectroscopy of single trapped molecules, *Science* **367**, 1213 (2020).
 - [9] M. Raha, S. Chen, C. M. Phenicie, S. Ourari, A. M. Dibos, and J. D. Thompson, Optical quantum nondemolition measurement of a single rare earth ion qubit, *Nat. Commun.* **11**, 1605 (2020).
 - [10] S. Motlakunta, N. Kotibhaskar, C.-Y. Shih, A. Vogliano, D. McLaren, L. Hahn, J. Zhu, R. Häblützel, and R. Islam, Preserving a qubit during state-destroying operations on an adjacent qubit at a few micrometers distance, *Nat. Commun.* **15**, 6575 (2024).
 - [11] E. Chitambar and G. Gour, Quantum resource theories, *Rev. Mod. Phys.* **91**, 025001 (2019).
 - [12] P. Skrzypczyk and N. Linden, Robustness of measurement, discrimination games, and accessible information, *Phys. Rev. Lett.* **122**, 140403 (2019).
 - [13] M. Oszmaniec and T. Biswas, Operational relevance of resource theories of quantum measurements, *Quantum* **3**, 133 (2019).
 - [14] R. Takagi and B. Regula, General resource theories in quantum mechanics and beyond: operational characterization via discrimination tasks, *Phys. Rev. X* **9**, 031053 (2019).
 - [15] T. Guff, N. A. McMahon, Y. R. Sanders, and A. Gilchrist, A Resource theory of quantum measurements, *J. Phys. A: Math. Theor.* **54**, 225301 (2021).
 - [16] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, *Phys. Rev. Lett.* **70**, 1895 (1993).
 - [17] L. Guerini, J. Bavaresco, M. Terra Cunha, and A. Acín, Operational framework for quantum measurement simulability, *J. Math. Phys.* **58**, 092102 (2017).
 - [18] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Simulating positive-operator-valued measures with projective measurements, *Phys. Rev. Lett.* **119**, 190501 (2017).
 - [19] C. Leang and D. Johnson, On the asymptotics of M-hypothesis Bayesian detection, *IEEE Trans. Inf. Theory* **43**, 280 (1997).
 - [20] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, Exploring quantum average-case distances: Proofs, properties, and examples, *IEEE Trans. Inf. Theory* **69**, 4600 (2023); Operational quantum average-case distances, *Quantum* **7**, 1106 (2023).