

All quantum measurements are asymptotically equivalent

Noah Linden¹ and Paul Skrzypczyk²

¹*School of Mathematics, Fry Building, University of Bristol, UK*

²*H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK.*

We consider the problem of reproducing one quantum measurement given the ability to perform another. In particular, given the availability of a – possibly imperfect – quantum measurement which can be performed multiple times, we study to what extent it can reproduce the measurement statistics and post-measurement state of a second target measurement, in the most general setting where both the available and target measurements are arbitrary generalised quantum measurements. We show that this general problem in fact reduces to the ability to reproduce the statistics of von Neumann measurements, and that in the asymptotic limit of infinitely many uses of the available measurement, a simple protocol based upon ‘classical cloning’ is able to perfectly achieve this task. This shows that asymptotically all (non-trivial) quantum measurements are equivalent. We also study optimal protocols for a fixed number of uses of the available measurement, and show that better protocols than classical cloning can be found in general. Our protocols show that the average error in reproducing a target measurement drops off exponentially fast, and that one can, for example, rapidly improve imperfect measurements by performing them a small number of times. This includes, but is not limited to, improving both noisy and lossy quantum measurements. Finally, we show that in a setting where we perform multiple measurements in parallel, we can moreover achieve finite-rate measurement reproduction, by using block-coding techniques from classical information theory.

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 generalised measurement – specified by a set of 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 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 generalised quantum measurements. The approach we will consider taking is to use the available measuring device *multiple times* – or equivalent to use multiple identical measuring devices – in order to realise a *single* use (or even potentially multiple uses) of the target measurement.

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,

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

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 behaviour 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 most general sense considered here in the literature before. The two key papers that have appeared were by Yuen [1] who proposed “purifying” photodetectors using pre-amplification, and by Dall’Arno D’Ariano and Sacchi [2], who considered much more generally the task of purifying noisy measurements. In a completely different direction, QISKIT describe a method of measurement error mitigation based upon linear algebra, post-processing a set of observed counts [3]. A key distinction here is that we are not interested specifically in purification of noisy 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.

The questions we study here are also interesting from a resource-theoretic perspective [4]. There has been interest lately in applying the resource-theory approach of quantum information to quantum measurements [5–8]. 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. The main insight of the present 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 [9]), given a single measuring device it can naturally be used many times, without being consumed or degraded. This shows that questions of measurement simulation [10, 11] – 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 non-trivial, and is not merely the classical cloning example given above. Finally, we consider the asymptotic scenario where we use the measurement many times, and want to approximate a large number of uses of a different measurement. We show that we can achieve *finite rate* transformations, using ideas from classical coding theory. 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 organised as follows:

- In Sec. II we consider a motivating example, using the ‘trine’ measurement to implement a von Neumann measurement. In Sec. II A we first consider the two-copy case, and a simple but non-optimal protocol based upon ‘classical cloning’. In Sec. II B we then present the general structure of deterministic protocol, before applying this insight back to find optimal (and non-trivial) 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. We end this section by briefly analysing noisy- Z measurements in Sec. II F, where we show that even here optimal protocols are non-trivial.
- In Sec. III we move onto the analysis of general measurements. In Sec. III A we outline the general set-up of quantum instruments (generalised measurements). In Sec. III B we then give provide general deterministic protocols in this setting. In Sec. III C we present our ‘post-measurement’ sub-routine, a key building block that implements a generalised measurement given access to a von Neumann measurement. In III D we present our ‘generalised classical cloning’ sub-routine, and present our main result, which combines the two sub-routines to show that any (non-trivial) measurement is able to implement any other measurement in the asymptotic limit of many uses.
- In Sec. IV we consider the problem of reproducing multiple uses of a target measurement in parallel, and present our second main result – that in this setting finite rate reproduction is possible, using block-coding protocols from classical information theory.

- Finally in Sec. V we conclude, and discuss future directions of research.

II. MOTIVATING EXAMPLE: REPRODUCING A 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 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 equatorial plane of the Bloch sphere.

Our goal in this example is to use this measurement N times to reproduce a single 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 formalise 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 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, then due to the linearity of quantum mechanics it follows immediately that the correct behaviour 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 normalised, 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.

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 von Neumann measurement, that demonstrates many of the key features we are interested in. This protocol is summarised 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

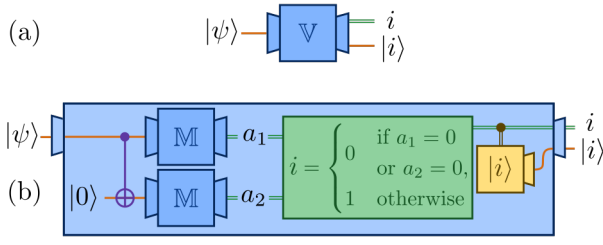


Figure 1. (a) A von Neumann measurement $\mathbb{V} = \{|i\rangle\langle i|\}$. The probability of the outcome i when measuring the state $|\psi\rangle$ is $\wp(i) = |\langle i|\psi\rangle|^2$. The state after measurement is $|i\rangle$. (b) A simple protocol using 2 copies of a trine measurement to approximately reproduce a 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.

unitary between this system and $|\psi\rangle$, with $|0\rangle$ as the target, leading to the (in general entangled) state

$$|\Psi\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle. \quad (6)$$

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

$$\begin{aligned} P(0,0) &= \frac{4}{9}|\alpha|^2, \\ P(0,1) &= P(0,2) \\ &= P(1,0) = P(2,0) = \frac{1}{9}|\alpha|^2, \\ P(1,1) &= P(2,2) = \frac{1}{36} + \frac{2}{9}|\beta|^2 + \frac{1}{12}(\alpha\beta^* + \alpha^*\beta), \\ P(1,2) &= P(2,1) = \frac{1}{36} + \frac{2}{9}|\beta|^2 - \frac{1}{12}(\alpha\beta^* + \alpha^*\beta), \end{aligned} \quad (7)$$

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 von Neumann measurement, we need to map these 9 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 particular choice, which exhaustive search confirms to be optimal among all deterministic protocols is

$$\begin{aligned} \{(0,0), (0,1), (0,2), (1,0), (2,0)\} &\mapsto 0, \\ \{(1,1), (1,2), (2,1), (2,2)\} &\mapsto 1, \end{aligned} \quad (8)$$

that is, in the first five cases from (7), where at least one measurement returns a zero outcome are mapped to 0, while the final four cases from (7), 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

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

These probabilities are intuitively ‘close’ to the probabilities of the von Neumann measurement. In particular, there is a small state-dependent probability of an ‘error’ in announcing the outcome 1 when in fact it should have 0. One way to quantify this closeness, that we will adopt here, is the root-mean-square (RMS) error, averaged over all states and outcomes,

$$\epsilon_{(2)} = \left(\frac{1}{2} \int d\psi \left([P^{(2)}(0) - \wp(0)]^2 + [P^{(2)}(1) - \wp(1)]^2 \right) \right)^{1/2}, \quad (10)$$

where, as a reminder, $\wp(0) = |\alpha|^2$ and $\wp(1) = |\beta|^2$ are the probabilities of the target von Neumann measurement (as given in (4) and (5)). In the case here of a dichotomic measurement, we have $\wp(1) = 1 - \wp(0)$ and $P^{(2)}(1) = 1 - P^{(2)}(0)$, and so $\epsilon_{(2)}$ simplifies to

$$\epsilon_{(2)} = \left(\int d\psi [P^{(2)}(0) - \wp(0)]^2 \right)^{1/2}. \quad (11)$$

This can be readily evaluated, and it is found that

$$\epsilon_{(2)} = \frac{1}{9\sqrt{3}} \approx 0.064. \quad (12)$$

Only in the case that the probabilities $P^{(2)}(i)$ were identical to the target probabilities $\wp(i)$ for all states $|\psi\rangle$ would ϵ 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

$$M'_0 = \frac{8}{9}|0\rangle\langle 0|, \quad M'_1 = \frac{1}{9}|0\rangle\langle 0| + |1\rangle\langle 1|. \quad (13)$$

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 von Neumann measurement, even though on the face of it, they are decidedly different measurements.

B. More general protocols

We now want to 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

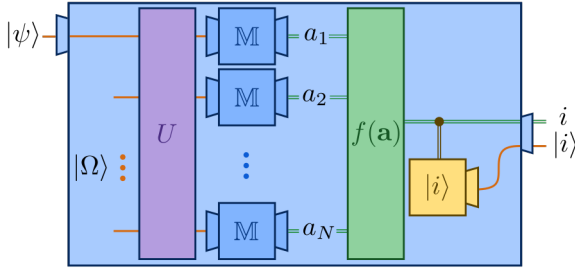


Figure 2. Protocols using N copies of the trine measurement in order to approximately reproduce a von Neumann measurement. $N - 1$ ancillary qubits are prepared in the state $|\Omega\rangle$ and a global unitary U is applied to these as well as the state $|\psi\rangle$. All N qubits are then measured 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 $\mathbb{M}^{(N)}$.

a single von Neumann measurement. The class of protocols we will consider are summarised pictorially in Fig. 2 and will comprise the following elements:

- (i) We prepare $N - 1$ ancillary qubits in a state $|\Omega\rangle$, independent of $|\psi\rangle$. Without loss of generality, this could be taken to be the tensor product of $N - 1$ copies of $|0\rangle$. As such, we are considering a collection of N qubits.
- (ii) We perform a global unitary transformation U on these N qubits, such that the state becomes

$$|\Psi\rangle = U|\psi\rangle|\Omega\rangle = \alpha|\Psi_0\rangle + \beta|\Psi_1\rangle, \quad (14)$$

where

$$|\Psi_0\rangle = U|0\rangle|\Omega\rangle, \quad |\Psi_1\rangle = U|1\rangle|\Omega\rangle, \quad (15)$$

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

- (iii) We measure all N systems, using the trine measurement.¹
- (iv) 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 a to the string \mathbf{a} .

In what follows we will restrict our attention to deterministic protocols, since these are more straightforward to analyse in the first instance, and since probabilistic protocols do not appear to provide any advantage.

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

$$Q_i^{(N)} = \sum_{a_1, \dots, a_N} \delta_{i, f(\mathbf{a})} M_{a_1} \otimes \dots \otimes M_{a_N}, \quad (16)$$

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

$$\begin{aligned} P^{(N)}(i) &= \langle \Psi | Q_i^{(N)} | \Psi \rangle, \\ &= \langle \psi | \langle \Omega | U^\dagger Q_i^{(N)} U | \psi \rangle | \Omega \rangle, \end{aligned} \quad (17)$$

from which it follows that the associated implemented POVM elements are

$$M_i^{(N)} = \text{tr}_{2\dots N}[(\mathbb{I} \otimes |\Omega\rangle\langle\Omega|) U^\dagger Q_i^{(N)} U], \quad (18)$$

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.

The figure of merit we consider will remain the same, and in this case is given by

$$\epsilon_{(N)} = \left(\int d\psi [P^{(N)}(0) - \wp(0)]^2 \right)^{1/2}. \quad (19)$$

In the Appendix we show that we can perform the integral over $|\psi\rangle$, and end up with an expression for $\epsilon_{(N)}$ which is the following simple quadratic problem

$$\begin{aligned} \epsilon_{(N)}^2 &= \min_{x, y} \frac{1}{3} [(1 - x - y)^2 + (1 - x)y] \\ \text{s.t.} \quad &\lambda_{\min} \leq x, y \leq \lambda_{\max}, \end{aligned} \quad (20)$$

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

$$x = \langle \Psi_0 | Q_0^{(N)} | \Psi_0 \rangle, \quad y = \langle \Psi_1 | Q_0^{(N)} | \Psi_1 \rangle, \quad (21)$$

are, respectively, the probabilities that the outcome 0 is returned when the state is $|\Psi_0\rangle$ and $|\Psi_1\rangle$ respectively.

We show in the Appendix, by using the fact that (20) is a quadratic form and the geometry of the problem, it is possible to analytically solve for the optimal x^* and y^* that achieve the minimum in (20) exactly. We find that there are three

¹ We 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.

regions, depending upon the relationship between λ_{\min} and λ_{\max} :

$$(x^*, y^*) = \begin{cases} (\lambda_{\max}, \frac{1-\lambda_{\max}}{2}) & \text{if } \lambda_{\max} \leq 1 - 2\lambda_{\min}, \\ (\lambda_{\max}, \lambda_{\min}) & \text{if } 1 - 2\lambda_{\min} \leq \lambda_{\max} \leq 1 - \frac{\lambda_{\min}}{2}, \\ (1 - \frac{\lambda_{\min}}{2}, \lambda_{\min}) & \text{if } 1 - \frac{\lambda_{\min}}{2} \leq \lambda_{\max}. \end{cases} \quad (22)$$

This shows that it is in fact straightforward to solve for the minimum RMS error $\epsilon_{(N)}$ 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 λ_{\max} and λ_{\min} . Moreover, once we have solved (19), we can finalise the protocol, since from (21) we see that we just need to find two orthogonal states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ that have the correct expectation values on $Q_0^{(N)}$, and we can always do this.² 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 $\epsilon_{(N)}$. 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 (19) it is straightforward to see, since both terms in the objective function are non-negative, that the only way that the average RMS error $\epsilon_{(N)}$ can vanish is if both $x = 0$ and $y = 1$. This is of course as we should have expected – it corresponds to the case where upon receiving $|i\rangle$ as the state to be measured, the measurement returns i with certainty. Crucially however, due to the constraints in (19) this in turn means that both $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$. This further means that $Q_0^{(0)}$ 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 $\epsilon_{(N)} > 0$, corresponding to an imperfect reproduction of the target von Neumann measurement. Therefore our problem has reduced to the study of the eigenvalues of the POVM $\mathbb{Q}^{(N)}$, and in particular how well we can arrange it such that each POVM element $Q_a^{(N)}$ has a unit eigenvalue.

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 amounted to choosing

$$\begin{aligned} Q_0^{(2)} &= M_0 \otimes M_0 + M_0 \otimes M_1 + M_0 \otimes M_2 \\ &\quad + M_1 \otimes M_0 + M_2 \otimes M_0 \\ &= \mathbb{I} \otimes \mathbb{I} - (\mathbb{I} - M_0) \otimes (\mathbb{I} - M_0), \\ Q_1^{(2)} &= M_1 \otimes M_1 + M_1 \otimes M_2 + M_2 \otimes M_1 + M_2 \otimes M_2 \\ &= (\mathbb{I} - M_0) \otimes (\mathbb{I} - M_0). \end{aligned} \quad (23)$$

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

$$|\Psi_0\rangle = |0\rangle|0\rangle, \quad |\Psi_1\rangle = |1\rangle|1\rangle, \quad (24)$$

and

$$\begin{aligned} P^{(2)}(0) &= \frac{8}{9} \text{ when } |\psi\rangle = |0\rangle, \\ P^{(2)}(0) &= 0 \text{ when } |\psi\rangle = |1\rangle, \end{aligned} \quad (25)$$

which led, as a reminder, to $\epsilon_{(2)} = \frac{1}{9\sqrt{3}} \approx 0.064$.

Now, $Q_0^{(2)}$ has minimum and maximum eigenvalues equal to $\lambda_{\min} = 0$ and $\lambda_{\max} = \frac{8}{9}$ respectively. Solving (19) in this case leads to the smaller RMS error

$$\epsilon_{(2)}^* = \frac{1}{18}, \quad (26)$$

which is achieved at

$$(x^*, y^*) = (\frac{8}{9}, \frac{1}{18}). \quad (27)$$

and is strictly better than what we found before, which we see is $\frac{2}{\sqrt{3}} \approx 1.15$ times larger than the optimal RMS error. We thus see that the classical cloning protocol we gave in Sec. II A, although intuitive, isn't 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 $x^* = \lambda_{\max}$, on the contrary the optimal value $y^* > \lambda_{\min} = 0$. That is, in contrast to the procedure from before, which led to the probabilities given in (25), the optimal probabilities are in fact

$$\begin{aligned} P^{(2)}(0) &= \frac{8}{9} \text{ when } |\psi\rangle = |0\rangle \\ P^{(2)}(0) &= \frac{1}{18} \text{ when } |\psi\rangle = |1\rangle, \end{aligned} \quad (28)$$

and the optimal implemented POVM is

$$\begin{aligned} M_0^{(2)} &= \frac{8}{9}|0\rangle\langle 0| + \frac{1}{18}|1\rangle\langle 1|, \\ M_1^{(2)} &= \frac{17}{18}|1\rangle\langle 1| + \frac{1}{9}|0\rangle\langle 0|. \end{aligned} \quad (29)$$

² As 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.

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 y^* .

Since $x^* = \lambda_{\max}$, 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

$$|\Psi_1\rangle = |1\rangle \left(\sqrt{\frac{1}{12}}|0\rangle + \sqrt{\frac{11}{12}}|1\rangle \right), \quad (30)$$

which is still 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 may look at though 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 $\frac{1}{18}$. 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 $P^{(2)}(0) = \frac{4}{9}$, while the optimal protocol gives $P^{(2)}(0) = \frac{17}{36}$, 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 (20).

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 $\mathbb{Q}^{(N)}$. This raises the possibility that there is in fact a better choice of $\mathbb{Q}^{(N)}$, i.e. a better partition of the strings a into two subsets. A simple argument shows that for deterministic protocols making use of the trine measurement twice, this is in fact the best possible choice. In particular, since we already have $\lambda_{\min} = 0$, this is already optimal, and cannot be made any smaller. Therefore any other strategy we can consider must have a larger λ_{\max} , and must compensate for any increase in λ_{\min} . We have already achieved $\lambda_{\max} = \frac{8}{9}$. It is easy to see that we cannot arrange for any eigenvalue between $\frac{8}{9}$ and 1, hence the only possibility is $\lambda_{\max} = 1$, in which case we find $\lambda_{\min} = \frac{1}{9}$, and we achieve an equally optimal solution, where we have essentially relabelled 0 and 1, which then has to be compensated for in U .

D. N -copy protocols

We now briefly outline how the above generalises to the case of N uses of the trine. Here, we find that the optimal

POVM is

$$\begin{aligned} Q_1^{(N)} &= (M_1 + M_2)^{\otimes N} \equiv (\mathbb{I} - M_0)^{\otimes N}, \\ Q_0^{(N)} &= \mathbb{I}_N - (M_1 + M_2)^{\otimes N} \equiv \mathbb{I}_N - (\mathbb{I} - M_0)^{\otimes N}. \end{aligned} \quad (31)$$

We see that this 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

$$\lambda_{\min} = 0, \quad \lambda_{\max} = 1 - \frac{1}{3^N}. \quad (32)$$

Solving (19) in this case leads to

$$\epsilon_{(N)}^* = \frac{1}{2} \frac{1}{3^N}, \quad (33)$$

which is achieved at

$$(x^*, y^*) = \left(1 - \frac{1}{3^N}, \frac{1}{2} \frac{1}{3^N} \right). \quad (34)$$

Crucially, this shows that the RMS 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 von Neumann measurement, if we allow ourselves the possibility of performing the measurement many times, it in fact becomes equivalent. The measurement we implement through this N -use protocol, as given by (18) is found to be

$$\begin{aligned} M_0^{(N)} &= \left(1 - \frac{1}{3^N} \right) |0\rangle\langle 0| + \frac{1}{2} \frac{1}{3^N} |1\rangle\langle 1|, \\ M_1^{(N)} &= \left(1 - \frac{1}{2} \frac{1}{3^N} \right) |1\rangle\langle 1| + \frac{1}{3^N} |0\rangle\langle 0|. \end{aligned} \quad (35)$$

Finally, an optimal unitary U achieving this protocol³ produces

$$\begin{aligned} |\Psi_0\rangle &= U|0\rangle^{\otimes N} = |0\rangle^{\otimes N}, \\ |\Psi_1\rangle &= U|1\rangle|0\rangle^{\otimes(N-1)} = |1\rangle^{\otimes(N-1)}|\phi^{(N)}\rangle, \end{aligned} \quad (36)$$

with

$$|\phi^{(N)}\rangle = \frac{1}{2} \frac{1}{3^{(N-1)/2}} |0\rangle + \sqrt{1 - \frac{1}{4} \frac{1}{3^{(N-1)}}} |1\rangle. \quad (37)$$

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,

³ As 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 lab.

by an amount which decreases with N . A simple calculation shows that if we instead did use the classical cloning protocol, this would amount to picking the sub-optimal pair $(x, y) = (1 - \frac{1}{3^N}, 0) = (\lambda_{\max}, \lambda_{\min})$, which would achieve $\epsilon_{(N)} = \frac{1}{\sqrt{3}} \frac{1}{3^N}$. Notably this is $\frac{2}{\sqrt{3}}$ times worse in RMS error compared to the optimal choice, independent of N , and therefore still has an error which drops off exponentially fast to zero.

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 non-trivial protocol to follow. In particular, substituting $N = 1$ into (35) indicates we should be able to achieve an implementation of the noisy von Neumann measurement

$$M_0^{(1)} = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{6}|1\rangle\langle 1|, \quad M_1^{(1)} = \frac{5}{6}|1\rangle\langle 1| + \frac{1}{3}|0\rangle\langle 0|, \quad (38)$$

which achieves an RMS error of $\epsilon_{(1)}^* = \frac{1}{6}$, which outperforms the ‘naive’ protocol of just measuring using the trine and outputting 0 when the outcome is 0, which achieves $\epsilon_{(1)} = \frac{1}{3\sqrt{3}}$.

We do however appear to run into a difficulty when we consider the unitary U that we need to perform from (36), since this is in fact no longer unitary, taking orthogonal states to non-orthogonal 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.

$$U|0\rangle|0\rangle = |0\rangle|0\rangle, \quad U|1\rangle|0\rangle = |1\rangle \left(\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \right). \quad (39)$$

We can furthermore notice that the state we rotate to when the qubit to be measured is $|1\rangle$ is in fact $|\phi_1\rangle$, one of the three directions of the trine measurement. It is straightforward to see that this protocol achieves the optimum. 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.

F. Moving from the trine to noisy- Z measurements

Finally, it is also interesting to question why is it that protocols become non-trivial, 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 non-orthogonal rank-one elements, which do not commute

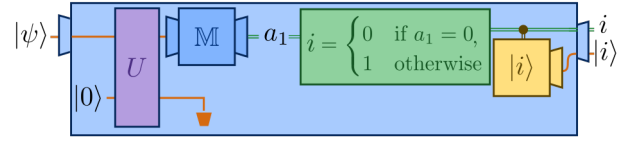


Figure 3. A modified protocol in order to optimally make use of a single trine measurement. This protocol makes use of an ancillary qubit, so that the unitary U from (39) can be applied, which effectively allows the states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ to be non-orthogonal. Since only a single trine measurement is performed, the ancillary particle is left unmeasured / is thrown away at the end of the protocol.

with each other. Maybe it is this non-commutativity which leads to the interesting structure to the problem we identified above. In what follows, we will show that this is not the case, and that this is not the origin of the non-trivial aspects of our measurement implementation problem.

Consider a simple class of measurements with POVM elements

$$\begin{aligned} N_0(p, q) &= p|0\rangle\langle 0| + (1 - q)|1\rangle\langle 1|, \\ N_1(p, q) &= (1 - p)|0\rangle\langle 0| + q|1\rangle\langle 1|. \end{aligned} \quad (40)$$

where we can view $p, q \in (\frac{1}{2}, 1]$ as parametrising the asymmetric noise⁴ of the measurement. 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 $\lambda_{\min} = 1 - q < \frac{1}{2}$ and $\lambda_{\max} = p > \frac{1}{2}$.

From (22), we know that there are three regions. In the middle region, where

$$2q - 1 \leq p \leq \frac{1 + q}{2}, \quad (41)$$

the optimal solution is obtained at $(x^*, y^*) = (\lambda_{\max}, \lambda_{\min})$. This regime includes the standard noisy Z measurement. In this regime the optimal protocol is just the trivial one, and no implementation in fact occurs.

Surprisingly, the remaining two regions, where there is a non-trivial protocol to follow, are in fact non-empty. In particular, when either

$$p > \frac{1 + q}{2} \quad \text{or} \quad q > \frac{1 + p}{2}, \quad (42)$$

i.e. when we consider a sufficiently asymmetric situation, this is the case. Focusing on the former case, we find the optimal solution is at

$$(x^*, y^*) = \left(\frac{1 + q}{2}, 1 - q \right), \quad (43)$$

⁴ Note that we restrict the interval for later convenience, as it will be sufficient for our purposes here.

which achieves $\epsilon_{(1)}^* = \frac{1-q}{2}$, and for which an optimal protocol, which requires an ancillary qubit as in the case of a single use of the trine, is to perform the transformation

$$U|0\rangle|0\rangle = (\gamma|0\rangle + \delta|1\rangle)|0\rangle, \quad U|1\rangle|1\rangle = |1\rangle|1\rangle, \quad (44)$$

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

$$\gamma = \sqrt{\frac{3q-1}{2(p+q-1)}}. \quad (45)$$

One can confirm that $0 \leq \gamma \leq 1$ for all p and q in the first regime. 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 utilise a non-trivial protocol in order to use it to optimally implement a von Neumann measurement. Thus, the need for non-trivial protocols does not derive from the non-commutativity of the POVM elements, and is a general feature of optimal protocols.

III. REPRODUCING GENERAL MEASUREMENTS

Up until this point we focused on a particularly insightful problem – that of reproducing a single use of a qubit von Neumann measurement, given the ability to perform the trine measurement N times. 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 non-trivial post-measurement states (in contrast to the von Neumann measurement, as we will explain more carefully below). A key insight is that in fact the ability to reproduce a von Neumann measurement can be viewed as an important sub-routine, which although may not be optimal, gives a concrete protocol which allows any measurement to reproduce any other measurement.

A. The general set-up

In the above we described our set-up 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 post-measurement 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 post-measurement 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

sub-channels), 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

$$P(a) = \text{tr}[\Lambda_a(\rho)], \quad (46)$$

and the state after the measurement is

$$\rho'_a = \frac{\Lambda_a(\rho)}{P(a)}. \quad (47)$$

In order for the measurement to be properly normalised, 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 sub-channel of the instrument is completely positive and trace non-increasing, it can be written as a sum of *Kraus operators*,

$$\Lambda_a(\cdot) = \sum_i K_i^a(\cdot) K_i^{a\dagger}, \quad (48)$$

where $\sum_i K_i^{a\dagger} K_i^a \leq \mathbb{I}$. The normalisation condition then reads $\sum_{i,a} K_i^{a\dagger} K_i^a = \mathbb{I}$. 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.⁵ Nevertheless, to each outcome we can still associate a POVM element M_a , which determines the probability of a given outcome (but not the post-measurement state). We see that

$$M_a = \sum_i K_i^{a\dagger} K_i^a, \quad (49)$$

where $P(a) = \text{tr}[M_a \rho]$. For a general instrument, we find that the post-measurement state ρ'_a will depend both upon the measurement outcome *and* the state measured. This is in contrast to the case of a von Neumann measurement, where the post-measurement 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 post-measurement states of the von Neumann measurement are ‘trivial’, and we see that a potentially new aspect of the reproduction problem arises when we want to reproduce the post-measurement states of a measurement where these are non-trivial, 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 *non-trivial*, 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 don’t require any measurement to be performed. They amount simply to producing a random (fictitious) measurement result, and

⁵ This 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.

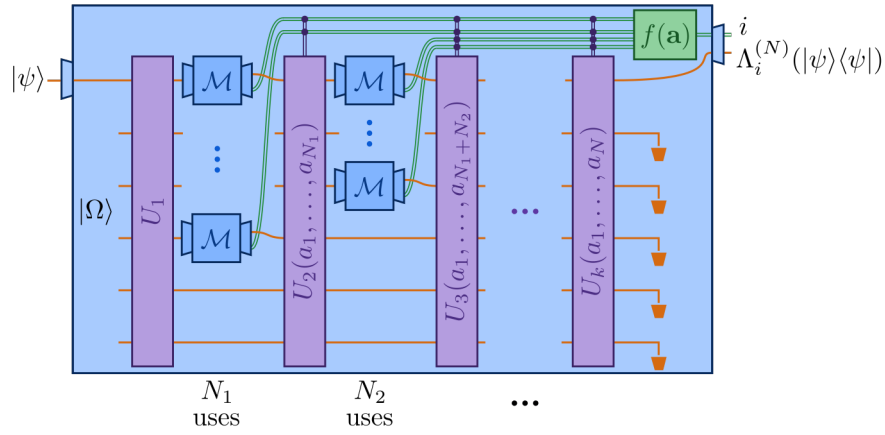


Figure 4. The general class of protocols we consider here. N_A ancillary particles are prepared in the state $|\Omega\rangle$, and an initial unitary U_1 is applied to all $N_A + 1$ particles. In stage one, the first N_1 particles are measured, and then a conditional unitary $U_2(a_1, \dots, a_{N_1})$ 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 $\mathcal{M}^{(N)} = \{\Lambda_i^{(N)}\}$.

preparing some quantum state. Clearly such trivial measurements will never be able to reproduce any non-trivial 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 $\wp(i) = \text{tr}[\Lambda_i(|\psi\rangle\langle\psi|)]$, and to leave the system in the state $\rho'_i = \Lambda_i(|\psi\rangle\langle\psi|)/\wp(i)$. 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 post-measurement 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, and is the following:

General Protocol:

- (i) Prepare N_A ancillary particles in a state $|\Omega\rangle$, independent of $|\psi\rangle$. Without loss of generality, this can be taken to be $|\Omega\rangle = |0\rangle^{\otimes N_A}$.

- (ii) Perform a global unitary U_1 on the $N_A + 1$ particles $|\psi\rangle|\Omega\rangle$.
- (iii) 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).
- (iv) Depending upon the string of outcomes \mathbf{a} , apply a unitary transformation $U_2(\mathbf{a})$ on all of the particles.
- (v) Measure the first N_2 particles using the measurement \mathcal{M} . Append the results of these measurements onto the existing string, so that $\mathbf{a} = (a_1, \dots, a_{N_1}, a_{N_1+1}, \dots, a_{N_1+N_2})$.
- (vi) 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.
- (vii) 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 $p(i|\mathbf{a})$.
- (viii) Apply a final unitary transformation $U_k(\mathbf{a})$, and return as the post-measurement state the first system (without loss of generality).

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 generalised form of RMS error from before, taking

into account now also the post-measurement state, namely

$$\epsilon_{(N)} = \left(\int d\psi \frac{1}{m} \sum_i \left\| \Lambda_i^{(N)}(|\psi\rangle\langle\psi|) - \Gamma_i(|\psi\rangle\langle\psi|) \right\|_F^2 \right)^{\frac{1}{2}}, \quad (50)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The above class of protocols are hard to analyse in general, due to their generality. In 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 sub-routine, which will reduce the complexity of our problem substantially.

C. Post-measurement sub-routine

When considering general measurement reproduction, with non-trivial post-measurement states, at first sight it appears that the problem is inherently more difficult than reproducing a von Neumann measurement, with its trivial post-measurement 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 sub-routine*, which shows that we can reduce the problem of reproducing a given measurement to the problem of reproducing a perfect von Neumann measurements. In particular, if we are given access to a von Neumann measurement, the following sub-routine is able to *perfectly* reproduce any other measurement. Thus, in order to use one measurement to reproduce another, we simply need only show that we can use any measurement to reproduce a von Neumann measurement, and then use this sub-routine, 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 von Neumann measurement $\mathbb{V} = \{|i\rangle\langle i|\}$. Consider the following protocol, which we will refer to as the *post-measurement sub-routine*:

Post-measurement sub-routine:

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

$$V|\psi\rangle|0\rangle|0\rangle = \sum_{i,j} K_j^i |\psi\rangle|i\rangle|j\rangle \quad (51)$$

- (iii) Ignore the third particle, and measure the second using a perfect von Neumann measurement.
- (iv) Return as outcome the result of the von Neumann measurement, and return as the post-measurement state the first particle.

It is straightforward to see that the probability of returning

the outcome a is

$$P(a) = \sum_j \text{tr}[K_j^a |\psi\rangle\langle\psi| K_j^{a\dagger}] = \text{tr}[M_a |\psi\rangle\langle\psi|], \quad (52)$$

and that the state after measurement is

$$\rho'_a = \frac{1}{P(a)} \sum_i K_i^a |\psi\rangle\langle\psi| K_i^{a\dagger} = \frac{\Lambda_a(|\psi\rangle\langle\psi|)}{P(a)}. \quad (53)$$

That is, this shows that a 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 (51) we would need to encode $|a\rangle$ into an appropriate collection of qudits and perform the required number of von Neumann measurements (e.g. if the target measurement had 7 outcomes, and we had access to a qubit von Neumann measurement, then we would encode into 3 qubits, and measure all 3 of them, with only 7 of the 8 possible outcomes ever occurring).

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 non-trivial* measurement, we can reproduce arbitrarily well *any other* (generalised) measurement, by making sufficient many uses of the first measurement.

In order to prove the main result, we can first note that, given the above post-measurement sub-routine, all that is left to be shown is that any measurement is able to reproduce a perfect von Neumann measurement arbitrarily well in the asymptotic limit. We will now show this, focusing on a simple (albeit non-optimal) protocol, based upon generalised 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 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 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 von Neumann measurement can indeed distinguish between the d basis states $\{|i\rangle\}$. In the backward direction, a perfect von Neumann measurement is the only measurement that is able to perfectly distinguish the basis $\{|i\rangle\}$, and hence the equivalence follows.

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 ‘generalised classical cloning protocol’ achieves this goal:

Generalised classical cloning sub-routine:

- (i) Prepare $N - 1$ ancillary particles in the state $|\Omega\rangle = |0\rangle^{\otimes(N-1)}$.
- (ii) Perform the unitary U such that
$$|\Psi_i\rangle = U|i\rangle|\Omega\rangle = |e_i\rangle^{\otimes N}, \quad (54)$$

where $|e_i\rangle$ form an orthonormal set of d states, which could be chosen at random.
- (iii) Measure the basis state to be identified, and all $N - 1$ ancillary particles using the available measurement \mathcal{M} , keeping only the string of measurement results $\mathbf{a} = (a_1, \dots, a_N)$, and discarding all post-measurement states.
- (iv) Use maximum-likelihood estimation to guess which basis state $|i\rangle$ was measured.

We will now analyse this sub-routine, 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

$$P(a|i) = \langle e_i | M_a | e_i \rangle, \quad (55)$$

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

$$P^{(N)}(\mathbf{a}|i) = \prod_{k=1}^N P(a_k|i). \quad (56)$$

Now, the key insight is to view this as N samples from one of d different probability distributions, with each distribution being labelled 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.

Crucially, in the asymptotic limit we know that all non-identical random variables can be perfectly distinguished, and that the probability of making an error drops off exponentially in the number of samples N when using maximum-likelihood estimation, which follows immediately from the Chernoff-Stein Lemma.

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

$$\begin{aligned} P(0|0) &= \frac{2}{3}, & P(1|0) &= \frac{1}{6}, & P(2|0) &= \frac{1}{6}, \\ P(0|1) &= 0, & P(1|1) &= \frac{1}{2}, & P(2|1) &= \frac{1}{2}. \end{aligned} \quad (57)$$

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

$$M_0 = |0\rangle\langle 0|, \quad M_1 = |1\rangle\langle 1| + |2\rangle\langle 2|. \quad (58)$$

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 $P(1|1) = P(1|2) = 1$. Consider therefore generalised classical cloning, with

$$\begin{aligned} |e_0\rangle &= \cos\theta|0\rangle + \sin\theta|1\rangle, \\ |e_1\rangle &= -\sin\theta|0\rangle + \cos\theta|1\rangle, \\ |e_2\rangle &= |2\rangle, \end{aligned} \quad (59)$$

and $\theta \in (0, \pi/4)$. In this case, the probabilities obtained are

$$\begin{aligned} P(0|0) &= \cos^2\theta, & P(1|0) &= \sin^2\theta, \\ P(0|1) &= \sin^2\theta, & P(1|1) &= \cos^2\theta, \\ P(0|2) &= 0, & P(1|2) &= 1. \end{aligned} \quad (60)$$

Now all three distributions are distinct, and so the basis states can be distinguished perfectly asymptotically.

As we show more concretely in the Appendix, we can always find a basis $\{|e_i\rangle\}$ whereby, just as in the above example, we can ensure that the distributions $p(a|i)$ for different values of i are genuinely distinct.⁶ In particular, as in the above example, if in the standard basis $|i\rangle$, two basis states $|j\rangle$ and $|j'\rangle$ lead to identical probability distributions, $P(a|j) = P(a|j')$, then by considering a change of basis generated by rotations in a two-dimensional subspace, then we can ensure that $P(a|j) \neq P(a|j')$, and therefore the states can be distinguished.

Although this protocol will in general definitely not be the optimal one – in terms of minimising the RMS 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

⁶ The only exception to this is the case we already excluded, when the available measurement is *trivial*, with all elements proportional to the identity operator, $M_a = q(a)\mathbb{I}$.

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 5 outcomes, and the measurement we have access to acts on qutrits, then we will consider 2 uses of the measurement, which we treat (conceptually) as a single measurement on a 9 dimensional space, and the states $|e_i\rangle$ forming a 5 dimensional subspace of this 9 dimensional space.⁷ The above analysis then applies, without any additional difficulties introduced.

IV. 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 analogue 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 sub-channels

$$\Gamma_{\mathbf{i}} = \Gamma_{i_1} \otimes \Gamma_{i_2} \otimes \dots \otimes \Gamma_{i_k}. \quad (61)$$

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

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

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 realise that the previous analysis, which showed that the problem of reproducing an arbitrary measurement reduces to the problem of reproducing just a 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 von Neumann measurement. This follows, since sub-routine 1 only makes use of a 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 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

$$p(a|x) = \langle x|M_a|x\rangle, \quad (62)$$

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

From the noisy-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*,

$$C = \max_{p(x)} I(A : X), \quad (63)$$

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.

⁷ If 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 don't need to treat the incoming system as a 9 dimensional system, but simply map its state into a 9-dimensional Hilbert space, and 'reset' the incoming particle to a fixed state $|0\rangle$.

Now, the noisy coding theorem guarantees the existence of encodings and decodings, namely functions from $\mathbf{x} = x_1 \cdots x_k$ (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 von Neumann measurements in parallel, the input will be a state of the form

$$|\psi\rangle = \sum_{\mathbf{i}} \alpha_{\mathbf{i}} |i_1\rangle \cdots |i_k\rangle, \quad (64)$$

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

Finite-rate classical-coding-based protocol:

- (i) Prepare $N - k = (R_k - 1)k$ ancillary particles in the state $|\Omega\rangle$.
- (ii) Apply a unitary U which sends the $N = R_k k$ particles into states encoding the classical codewords,

$$U|i_1\rangle \cdots |i_k\rangle |\Omega\rangle = |c_{\mathbf{i}}\rangle, \\ = |(c_{\mathbf{i}})_1\rangle \cdots |(c_{\mathbf{i}})_N\rangle, \quad (65)$$
 where $(c_{\mathbf{i}})_j$ denotes the j^{th} element of the codeword $c_{\mathbf{i}}$.
- (iii) Measure each of the N particles using the available measurement \mathbb{M} , producing a string of outcomes $\mathbf{a} = a_1 \cdots a_N$.
- (iv) 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.
- (v) Prepare the state $|j_1\rangle \cdots |j_k\rangle$, 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 $|\mathbf{i}\rangle = |i_1\rangle \cdots |i_k\rangle$, 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 (62). 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 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-coding schemes, in general would not be expected to be optimal. Nevertheless, their existence highlights how classical 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 post measurement 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, (62) is only one particular classical channel that we can associate with a quantum measurement. In improved rate can be achieved by optimising over associated classical channels, by varying the states measured, that is by considering the class of channels $\mathcal{C}_{\mathbb{M}}$,

$$\mathcal{C}_{\mathbb{M}} = \{p(a|x) | p(a|x) = \text{tr}(M_a \rho_x)\}, \quad (66)$$

where ρ_x are quantum states for all x .

V. 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 measurement are asymptotically equivalent to each other. In order to show this general result, we demonstrated two key sub-routines, which are interesting in their own right - a subroutine for implementing arbitrarily well a von Neumann measurement, and a subroutine making use of a von Neumann measurement to implement an arbitrary measurement - including the post measurement state. Finally, 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.

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 a noisy measurement is used multiple times in order to implement a less noisy measurement. We have shown that even with a small number of 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. 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 among multiple systems, and in the process completely overcome noise or other limitations which arise from specific fixed measurements. Our results also extend the resource-theory approach to measurements, to enlarge the class of allowed operations to incorporate multiple uses of the available resourceful measurement.

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 root-mean-squared error as our figure of merit to optimise. 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 which 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 of the unitary operations should be realisable 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 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 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 characterise the ultimate optimal rate of measurement reproduction.

ACKNOWLEDGMENTS

We thank Nicolas Brunner, Sandu Popescu and Andrew Daley for insightful discussions. PS gratefully acknowledges support from a Royal Society University Research Fellowship (UHQT/NFQI). NL 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.

-
- [1] H. P. Yuen, *Opt. Lett.*, **OL 12**, 789 (1987).
 - [2] M. Dall’Arno, G. M. D’Ariano, and M. F. Sacchi, *Phys. Rev. A* **82**, 042315 (2010).
 - [3] 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, P. Jayasinha, H. Kang, A. h. Karamlou, R. Lored, D. McKay, A. Maldonado, A. Macaluso, A. Mezzacapo, Z. Minev, R. Movassagh, G. Nannicini, P. Nation, A. Phan, M. Pistoia, A. Rattew, J. Schaefer, J. Shabani, J. Smolin, J. Stenger, K. Temme, M. Tod, E. Wanzambi, S. Wood, and J. Wootton., “Learn quantum computation using qiskit,” (2020).
 - [4] E. Chitambar and G. Gour, *Rev. Mod. Phys.* **91**, 025001 (2019).
 - [5] P. Skrzypczyk and N. Linden, *Phys. Rev. Lett.* **122**, 140403 (2019).
 - [6] M. Oszmaniec and T. Biswas, *Quantum* **3**, 133 (2019).
 - [7] R. Takagi and B. Regula, *Phys. Rev. X* **9**, 031053 (2019).
 - [8] T. Guff, N. A. McMahon, Y. R. Sanders, and A. Gilchrist, arXiv:1902.08490 [quant-ph] (2019), arXiv:1902.08490 [quant-ph].

- [9] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [10] L. Guerini, J. Bavaresco, M. Terra Cunha, and A. Acín, *Journal of Mathematical Physics* **58**, 092102 (2017).
- [11] M. Ozmaniec, L. Guerini, P. Wittek, and A. Acín, *Phys. Rev. Lett.* **119**, 190501 (2017).

APPENDICES

Appendix A: Evaluating the root-mean-square (RMS) error

In this appendix we will show how the integral over all states can be performed, in order to obtain an explicit optimisation problem for the root-mean-square error. For concreteness, we will focus on (19), before giving the general result, which follows the same technique, thereafter.

Our starting point is to notice that we can re-express $\epsilon_{(N)}^2$ from (19) as

$$\begin{aligned}\epsilon_{(N)}^2 &= \int d\psi [P^{(N)}(0) - \wp(0)]^2, \\ &= \int d\psi [\langle \psi | M_0^{(N)} | \psi \rangle - \langle \psi | \Pi_0 | \psi \rangle]^2,\end{aligned}\quad (\text{A1})$$

where we have used (17) and (18) to write $P^{(N)}(0) = \langle \psi | M_0^{(N)} | \psi \rangle$, and used the fact that $\wp(0) = |\alpha|^2 = |\langle 0 | \psi \rangle|^2 = \langle \psi | \Pi_0 | \psi \rangle$. We can now further re-express this as

$$\epsilon_{(N)}^2 = \int d\psi \text{tr} \left[|\psi\rangle\langle\psi|^{\otimes 2} \left(M_0^{(N)} - \Pi_0 \right)^{\otimes 2} \right]. \quad (\text{A2})$$

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

$$\int d\psi |\psi\rangle\langle\psi|^{\otimes 2} = \frac{1}{3} \Pi_{\text{sym}}, \quad (\text{A3})$$

that is, that the integral is proportional to the projector onto the symmetric subspace (here of two qubits).⁸ Using further the fact that $\Pi_{\text{sym}} = \frac{1}{2}(\mathbb{1} + S)$, with S the SWAP operator, we arrive at

$$\begin{aligned}\epsilon_{(N)}^2 &= \frac{1}{6} \left[\left(\text{tr} \left(M_0^{(N)} \right) \right)^2 + \text{tr} \left(M_0^{(N)2} \right) \right. \\ &\quad \left. - 2 \text{tr} \left(M_0^{(N)} \right) - 2 \langle 0 | M_0^{(N)} | 0 \rangle + 2 \right].\end{aligned}\quad (\text{A4})$$

Let us define now

$$x = \langle 0 | M_0^{(N)} | 0 \rangle = \langle \Psi_0 | Q_0^{(N)} | \Psi_0 \rangle, \quad (\text{A5})$$

$$y = \langle 1 | M_0^{(N)} | 1 \rangle = \langle \Psi_1 | Q_0^{(N)} | \Psi_1 \rangle \quad (\text{A6})$$

$$z = \langle 0 | M_0^{(N)} | 1 \rangle = \langle \Psi_0 | Q_0^{(N)} | \Psi_1 \rangle, \quad (\text{A7})$$

where the latter equalities follow from (18). With these definitions in place, we have

$$\text{tr} \left(M_0^{(N)} \right) = x + y, \quad (\text{A8})$$

$$\text{tr} \left(M_0^{(N)2} \right) = x^2 + y^2 + 2|z|^2, \quad (\text{A9})$$

⁸ For qudits, the right-hand side would instead be $\frac{1}{d_{\text{sym}}} \Pi_{\text{sym}}$, where $d_{\text{sym}} = \binom{d+1}{2}$ is the dimension of the symmetric subspace of two qudits.

and so, after simple manipulations we arrive at

$$\epsilon_{(N)}^2 = \frac{1}{3} [(1-x-y)^2 + (1-x)y + |z|^2]. \quad (\text{A10})$$

We are interested in minimising this root-mean-square error over all implementation protocols. From (A5) and (A6), we see that x and y , as expectation values of $Q_0^{(N)}$, are constrained by its minimum and maximum eigenvalues. Moreover, since $|z|^2$ is a non-negative quantity, it is immediate that the optimal choice of z , which can always be achieved, is to take $z = 0$, which corresponds to $M_0^{(N)}$ being diagonal in the $\{|0\rangle, |1\rangle\}$ basis. We thus arrive at (20), namely

$$\begin{aligned} \epsilon_{(N)}^2 = \min_{x,y} & \quad \frac{1}{3} [(1-x-y)^2 + (1-x)y] \\ \text{s.t.} & \quad \lambda_{\min} \leq x, y \leq \lambda_{\max}. \end{aligned} \quad (\text{A11})$$

The above generalises to qudits in a straightforward way. In this case, we find that the generalised expression for $\epsilon_{(N)}^2$ becomes

$$\begin{aligned} \epsilon_{(N)}^2 = \frac{1}{d(d+1)} \sum_a & \left[\left(\sum_i x_{ai} - 1 \right)^2 \right. \\ & \left. + \sum_i x_{ai}^2 - 2x_{aa} + 1 \right], \end{aligned} \quad (\text{A12})$$

where

$$x_{ai} = \langle i | M_a^{(N)} | i \rangle = \langle \Psi_i | Q_a^{(N)} | \Psi_i \rangle. \quad (\text{A13})$$

The constraints that the x_{ai} need to satisfy are now seen to be

$$\lambda_{\min}^a \leq x_{ai} \leq \lambda_{\max}^a, \quad (\text{A14})$$

where λ_{\min}^a and λ_{\max}^a are, respectively, the smallest and largest eigenvalues of the operator $Q_a^{(N)}$, and

$$\sum_a x_{ai} = 1, \quad (\text{A15})$$

which follows from the fact that $\sum_a \langle i | M_a^{(N)} | i \rangle = \langle i | \mathbb{1} | i \rangle = 1$. Altogether, the minimum RMS error is therefore found by minimising (A12), subject to the constraints (A14) and (A15). Since the constraints are linear, this problem is an instance of a quadratic program, which can be solved efficiently numerically.

Appendix B: Solving analytically for the minimal root-mean-square error

In this appendix we will show that it is possible to analytically solve for the minimal root-mean-square error that can be achieved in the case of approximately implementing a 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 analysing the constraints in (20). Geometrically, we see that we are optimising over a rectangle in the $x-y$ plane. Denoting the co-ordinate by (x, y) , the vertices of this rectangle are seen to be at $(\lambda_{\min}, \lambda_{\min})$, $(\lambda_{\min}, \lambda_{\max})$, $(\lambda_{\max}, \lambda_{\min})$ and $(\lambda_{\max}, \lambda_{\max})$.

The function we are optimising is a quadratic form, and hence the contours of this function are ellipses, centred at the point $(1, 0)$. This point will only be within the rectangle only if $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$, which implies that $M_0^{(N)} = |0\rangle\langle 0|$, as we saw previously. In this case, the ellipse has zero width and height, and $\epsilon_{(N)} = 0$.

When $\lambda_{\min} > 0$ and/or $\lambda_{\max} < 1$, then the centre of the ellipse is outside of the rectangle, and our optimisation problem can be understood geometrically as finding the size of the smallest ellipse which *just touches* the rectangle. The point at which the ellipse and rectangle touch will be the optimal solution (x^*, y^*) .

We can now make two key observations. First, since the centre of the ellipse is always below and to the right of the rectangle, the point where the ellipse touches will either be along the bottom edge, when $y = \lambda_{\min}$, along the right-hand edge, when $x = \lambda_{\max}$, or at the bottom right-hand vertex, when $(x, y) = (\lambda_{\max}, \lambda_{\min})$.

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

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 ellipse is horizontal. A straightforward calculation shows that the tangent to the ellipse at a general point satisfies

$$(-1 + x + 2y) \frac{dy}{dx} = 2 - 2x - y. \quad (\text{B1})$$

Therefore all points with horizontal tangent lie on the line

$$y = 2(1 - x). \quad (\text{B2})$$

This line must intersect the bottom of the rectangle. This will only happen if the intersection point is before the corner. The intersection occurs at $x = 1 - \frac{\lambda_{\min}}{2}$, and so obtain the region

$$1 - \frac{\lambda_{\min}}{2} \leq \lambda_{\max}. \quad (\text{B3})$$

The case where the ellipse touches along the right-hand edge is similar. We find that points with vertical tangent lie on the line $y = \frac{1-x}{2}$, the intersection occurs at $y = \frac{1-\lambda_{\max}}{2}$, and this will be above the corner when

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

Finally, when neither of these cases hold, we see that the intersection will occur at the corner of the rectangle. Put together, leads to (22), namely

$$(x^*, y^*) = \begin{cases} (\lambda_{\max}, \frac{1-\lambda_{\max}}{2}) & \text{if } \lambda_{\max} \leq 1 - 2\lambda_{\min}, \\ (\lambda_{\max}, \lambda_{\min}) & \text{if } 1 - 2\lambda_{\min} \leq \lambda_{\max} \leq 1 - \frac{\lambda_{\min}}{2}, \\ (1 - \frac{\lambda_{\min}}{2}, \lambda_{\min}) & \text{if } 1 - \frac{\lambda_{\min}}{2} \leq \lambda_{\max}. \end{cases} \quad (\text{B5})$$

Appendix C: Generalised classical cloning can always distinguish a basis of states

In this Appendix we will show that it is always possible to use a generalised classical cloning protocol to distinguish between a basis of states. As explained in the main body, all we need to ensure is that for each basis state $|i\rangle$, there is a state $|e_i\rangle$, such that the corresponding probability distribution $P(a|i) = \langle e_i|M_a|e_i\rangle$ is distinct for all i .

The simplest case, corresponding to (standard) classical cloning, is when $|e_i\rangle = |i\rangle$, and all of the distributions are distinct. In this case, there is no difficulty. Therefore, let us assume the converse, that there are at least two states which, using standard classical cloning, lead to the same probability distribution. In this case, we can first calculate, one by one, the POVM elements in the subspace spanned by the states that lead to identical probability distributions. That is, if is the states $|j_0\rangle$ and $|j_1\rangle$ that lead to identical probabilities, we compute the operator

$$|j_0\rangle\langle j_0|M_0|j_0\rangle\langle j_0| + |j_0\rangle\langle j_0|M_0|j_1\rangle\langle j_1| + |j_1\rangle\langle j_1|M_0|j_0\rangle\langle j_0| + |j_1\rangle\langle j_1|M_0|j_1\rangle\langle j_1|. \quad (C1)$$

If M_0 restricted to this subspace is not proportional to the identity, then it will have two distinct eigenvalues, $\lambda_0 \neq \lambda_1$, with two corresponding eigenvectors, $|\nu_0\rangle$ and $|\nu_1\rangle$. We can now use these eigenvectors, instead of $|j_0\rangle$ and $|j_1\rangle$ in a generalised classical cloning protocol, that is, we take $|e_0\rangle = |\nu_0\rangle$ and $|e_1\rangle = |\nu_1\rangle$. With this choice, we will have $P(0|0) = \lambda_0 \neq \lambda_1 = P(0|1)$. Thus, it is no longer the case that the two states lead to identical probability distributions.

The above will fail if M_0 is proportional to the identity operator in the subspace (in which case, since both eigenvalues are equal, the above will not allow us to generate two distinct probability distributions). We can nevertheless consider other POVM elements, M_a , until we find one that is not proportional to the identity, and then use its eigenvectors in a generalised classical cloning protocol.

The only case in which the above fails to work is if all POVM elements happen to be proportional to the identity in the subspace spanned by $|j_0\rangle$ and $|j_1\rangle$. If on the one hand we are considering qubits, then $|j_0\rangle$ and $|j_1\rangle$ span the Hilbert space, and so the fact that all POVM elements are proportional to the identity mean that we have in fact a *trivial measurement*, which we already excluded from our analysis, as it is clear that such measurements can never used to perform non-trivial measurements.

Therefore, let us consider a qutrit or higher dimensional

system, such that all POVM elements are proportional to the identity in a two-dimensional subspace. We can now proceed by searching for any other basis state that leads to a non-identical probability distribution. Let us assume that $|j_2\rangle$ is such a basis vector, with $P(a|2) \neq P(a|0)$. We then consider a basis similar to (59),

$$\begin{aligned} |e_0\rangle &= \cos\theta|j_0\rangle + \sin\theta|j_2\rangle, \\ |e_1\rangle &= |j_1\rangle, \\ |e_2\rangle &= -\sin\theta|j_0\rangle + \cos\theta|j_2\rangle, \end{aligned} \quad (C2)$$

This will leave $P(a|1)$ unchanged, but will transform $P(a|0)$ and $P(a|2)$ into

$$\begin{aligned} P'(a|0) &= \cos^2\theta P(a|0) + \sin^2\theta P(a|2) + \sin 2\theta z(a), \\ P'(a|2) &= \sin^2\theta P(a|0) + \cos^2\theta P(a|2) - \sin 2\theta z(a), \end{aligned} \quad (C3)$$

where $z(a) = \Re[\langle j_0|M_a|j_2\rangle]$. The difference is therefore

$$P'(a|0) - P'(a|2) = \cos 2\theta [P(a|0) - P(a|2)] + 2 \sin 2\theta z(a). \quad (C4)$$

This difference cannot be zero for all values of θ , since this would require $P(a|0) = P(a|2)$, which by assumption is not the case. This means that we can definitely find a value of θ such that the all three states lead to distinct probability distributions.

The only case in which the above does not work is if $P(a|2) = P(a|0) = P(a|1)$. In this case, we proceed in the same manner as above, except now we calculate the POVM elements in the three-dimensional subspace spanned by the states $|j_0\rangle$, $|j_1\rangle$ and $|j_2\rangle$. We then have the same type of possibilities. If all eigenvalues of the operator in this subspace are distinct, then we can use the eigenvectors as the basis for classical cloning. If one of the eigenvalues is degenerate, we can consider a rotation between one of the degenerate eigenvalues and the non-degenerate one to obtain three vectors all of which lead to distinct probability distributions. If all three eigenvalues are degenerate (so that the operator is proportional to the identity), then if we are considering qutrits then we have a trivial measurement, and hence this case is uninteresting. If we have ququarts or higher, we consider a fourth state $|j_3\rangle$, and repeat the analysis once again, now considering a subspace of one higher dimension. We thus see that we can proceed inductively, always increasing the dimension, unless we have identified a set of states all of which lead to distinct probability distributions.