Multiple-copy state discrimination of noisy qubits

Kieran Flatt,* Stephen M. Barnett, and Sarah Croke
School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK

Multiple-copy state discrimination is a fundamental task in quantum information processing. If there are two, pure, non-orthogonal states then both local and collective schemes are known to reach the Helstrom bound, the maximum probability of successful discrimination allowed by quantum theory. For mixed states, it is known that only collective schemes can perform optimally, so it might be expected that these schemes are more resilient to preparation noise. We calculate the probability of success for two schemes, a local scheme based on Bayesian updating and quantum data gathering, the simplest possible collective measurement, and consider imperfect preparation fidelity. We find two surprising results. Firstly, both schemes converge upon the same many-copy limit, which is less than unity. Secondly, the local scheme performs better in all cases. This highlights the point that one should take into account noise when designing state discrimination schemes.

I. INTRODUCTION

In many quantum information processing tasks, one needs to identify, by measurement, the state of a system given that the finite and discrete set of states from which it is taken is known. This task is called state discrimination [1–5]. Unless the set of possible states is an orthogonal basis for some space they cannot be perfectly discriminated and instead the user usually seeks to minimise one of two figures of merit, either the probability of incorrectly identifying or failing to identify the state. The measurement which minimises the former of these is the Helstrom, or minimum-error, measurement [6]. If there are two possible states, the optimal measurement has a simple analytic form. In more complex cases, such as three-or-more pure states [7] or mixed states [8, 9], only limited results are known.

Our above comments relate to single-copy state discrimination. Given a resource of multiple systems, all prepared in the same state, it might be expected that the correlations can be used to improve the probability of success. This intuition is correct and the Helstrom bound, the optimal value of this probability, is known for discriminating two states. However, in this case a physical implementation of the measurement is typically hard to find. Furthermore, the issue of locality versus collectivity arises: can the bound be achieved with local measurements, those on individual systems, only or must the discriminator use collective measurements, which are more difficult to perform? It is known that the best measurement to discriminate multi-partite states is often a collective measurement, even for product states. Famous examples of this are the double trine ensemble [10–13] and the domino states [14–16].

For two-pure-state discrimination, it is known that a local scheme can reach the Helstrom bound [17–19]. In other scenarios, very few analytic results have been acquired and most knowledge comes from numerical simulations [20, 21]. Here, some counterintuitive results emerge. One example is that the distinction between local and global optimality emerges. In some cases, among local schemes, the best overall measurement involves a fixed measurement on each qubit, which succeeds locally with a suboptimal probability [20]. For a small number of copies, adaptive schemes perform better than fixed schemes [22], but in the limit of large numbers of copies, this advantage disappears, even for mixed states [20, 22–25]. Further, it is for almost pure, but strictly speaking mixed, states that the gap in performance between collective stratgies and local strategies is most pronounced in the many-copy limit. Such unexpected results signal the need for further analytical work in this area. Whether collective measurements, which require a quantum memory, are required to reach the Helstrom bound is considered an important open question in state discrimination [26, 27].

The work presented in this paper investigates a separate, but related question. How resilient are multiple-copy state discrimination schemes to preparation noise? No real preparation is ever perfect, but for high enough fidelity we may consider the states to be pure. Further, decoherence properties of even state-of-the-art qubits can demonstrate significant variability [28, 29], resulting in a corresponding variability in the rate at which a preparation characterised as very high fidelity degrades over time. Finally, in a real-world physical communications system, instabilities in noise properties of a channel can lead to uncharacterised noise in the received states. How sensitive are schemes designed for pure states to a small amount of uncharacterized preparation noise? As the truly optimal scheme for noisy qubits will be collective, it might be expected that such schemes will be more resilient to preparation noise than the equivalent local scheme. Our approach is to compare two equivalent schemes, one local [17] and one collective [30], both of which reach the Helstrom bound for discriminating two pure states. We apply each scheme, optimised for a specific pair of pure states, to the corresponding mixed states and relate the probability of success to the preparation fidelity. Our results show that, surprisingly, the local scheme consistently performs better than the collective scheme. Neither, however, approaches unit success prob-

* k.flatt.1@research.gla.ac.uk

ability as the number of copies, N, grows. Rather, they approach the same fixed bound. We discuss how to use information which would otherwise be thrown away in the local adaptive scheme to improve on this bound. This recovers asymptotic behaviour which, as the number of qubits approaches infinity, tends towards perfect discrimination.

II. PRELIMINARIES

Two pure states of a qubit occupy a single great circle on the Bloch sphere. For this reason, they can be characterised in relation to each other by real numbers only and written in the form

$$|\psi_k\rangle = \cos(\theta)|0\rangle + (-1)^k \sin(\theta)|1\rangle \quad k = 0, 1. \tag{1}$$

The overlap of these two states is $\langle\psi_0|\psi_1\rangle = \cos(2\theta)$ and, without loss of generality, $0 \leq \theta \leq \pi/4$. If a single system is prepared in either of these states with probabilities p_k, the highest possible probability of successful discrimination is given by the Helstrom bound,

$$\mathrm{P}_1^H = \frac{1}{2}\left(1 + \sqrt{1 - 4p_0p_1\cos^2(2\theta)}\right). \tag{2}$$

If $\theta = \pi/4$ the two states are orthogonal. In such a case, $\mathrm{P}_1^H = 1$ and they can be perfectly discriminated. Otherwise, this quantity is less than one. The measurement which achieves this bound is a projective measurement onto the eigenvectors of $p_0|\psi_0\rangle\langle\psi_0| - p_1|\psi_1\rangle\langle\psi_1|$.

If instead there is a resource of N copies of the state, we are seeking to distinguish $|\psi_0\rangle^{\otimes N}$ from $|\psi_1\rangle^{\otimes N}$. As these can be considered as two single pure states on the total Hilbert space, the multiple-copy Helstrom bound is

$$\mathrm{P}_N^H = \frac{1}{2}\left(1 + \sqrt{1 - 4p_0p_1\cos^{2N}(2\theta)}\right). \tag{3}$$

In this case, the measurement which achieves this is again a von Neumann measurement, one that projects onto the eigenstates of $p_0|\psi_0\rangle\langle\psi_0|^{\otimes N} - p_1|\psi_1\rangle\langle\psi_1|^{\otimes N}$. To find these we must find the eigenvalues of a 2^N dimensional matrix, a task which is much simplified by the symmetry in the multiple-copy case. For pure states in particular, there are just two dimensions that are important, and a number of optimal schemes are known. We consider two.

In this article we are concerned with systems in which the resource qubits are prepared imperfectly. This is represented by a parameter $\delta\theta_i$ which characterises the displacement of the ith qubit's state from the ideal case such that

$$\begin{aligned}|\tilde{\psi}_k^i\rangle &= \cos(\theta + \delta\theta_i)|0\rangle + (-1)^k \sin(\theta + \delta\theta_i)|1\rangle \\ &= \cos(\delta\theta_i)|\psi_k\rangle - \sin(\delta\theta_i)|\psi_{k\perp}\rangle.\end{aligned} \tag{4}$$

In the second line we relate the noisy form of the state to the ideal case, Eq. 1 and introduce $|\psi_{k\perp}\rangle$ to indicate the state orthogonal to $|\psi_k\rangle$. The fidelity F is the standard way to parameterise the noise on a system. It can be understood operationally as the probability that a measurement of the prepared state will identify it as the ideal state [4]. For pure states, it is defined as the overlap of the prepared and ideal states, averaged over the noise's probability distribution which we assume is symmetric, i.e., $\mathrm{P}(\delta\theta_i) = \mathrm{P}(-\delta\theta_i)$. One can consider this a Gaussian distribution however that level of detail is not required in what follows. The two noise parameters are then related by

$$\langle\cos^2(\delta\theta_i)\rangle = \int |\langle\tilde{\psi}_k^i|\psi_k\rangle|^2 \mathrm{P}(\delta\theta_i) = F, \tag{5}$$

where $1/2 \leq F \leq 1$. From this we also have

$$\langle\sin^2(\delta\theta_i)\rangle = 1 - F \tag{6}$$

$$\langle\cos(2\delta\theta_i)\rangle = 2F - 1 \tag{7}$$

$$\langle\sin(2\delta\theta_i)\rangle = 0. \tag{8}$$

The first two of these follows from the definition of the fidelity while the third uses the symmetry of the probability distribution. These are the only functions which are averaged in what follows. We assume that the noise on each qubit is independent of the others and average at each stage.

Using these results we express the noisy form of the state, Eq. 4, as a mixed state. We obtain

$$\rho_k = F|\psi_k\rangle\langle\psi_k| + (1 - F)|\psi_{k\perp}\rangle\langle\psi_{k\perp}|, \tag{9}$$

where we have averaged over the probabilty distribution of $\delta\theta_i$. If $F = 1$ it is the relevant pure state. If instead $F = 1/2$, which is the smallest possible value of the fidelity, it is a maximally mixed state, so that maximum noise erases all information about the state. For other values of F, the state varies monotically between these two points. Our interest throughout this paper will be in systems which are close to perfect fidelity.

III. LOCAL-ADAPTIVE MEASUREMENT

An important result in multiple-copy state discrimination is that it is possible to reach the Helstrom bound, Eq. 3, using local measurements only. We follow here the scheme of Acín *et al* [17] but similar results have been found by others [18, 19]. They examine a local and adaptive scheme in which the measurement of the nth copy can depend upon the outcome of measurements on the previous $(n-1)$ copies. We first need to introduce some notation. The sequence of measurement outcomes is represented by a bit string x as long as the number N of qubits, with the nth result labelled i_n. The measurement onto the nth qubit is a projector onto the basis

$$\begin{aligned}|\omega_0^n\rangle &= \cos(\phi_x)|0\rangle + \sin(\phi_x)|1\rangle \\ |\omega_1^n\rangle &= \sin(\phi_x)|0\rangle - \cos(\phi_x)|1\rangle.\end{aligned} \tag{10}$$

Here, we use x for the bit string of the first $(n-1)$ results and adopt a different notation when it is required. In the local-adaptive measurement scheme, the measurement at each point depends on the previous outcomes in the scheme however the overall result is determined by the final measurement outcome alone. The optimal scheme of this kind turns out to be Bayesian updating [17]. On the first qubit, one projects onto the eigenvectors of $p_0|\psi_0\rangle\langle\psi_0| - p_1|\psi_1\rangle\langle\psi_1|$. On the rest, the relevant eigenbasis is instead $\mathrm{P}(0|x)|\psi_0\rangle\langle\psi_0| - \mathrm{P}(1|x)|\psi_1\rangle\langle\psi_1|$, in which $\mathrm{P}(k|x)$ is the probability, calculated from Bayes' theorem, that the state $|\psi_k\rangle$ was prepared given that bit string x is the measurement record. The ϕ_x for which this measurement satisifies the Helstrom bound is found to be

$$\cos(2\phi_x) = (-1)^{i_{N-1}}\sqrt{\frac{1-4p_0p_1\cos^{2N-2}(2\theta)}{1-4p_0p_1\cos^{2N}(2\theta)}}. \quad (11)$$

The only appearance of the bit string x here is in the single index i_{N-1}, which is the value of the prior measurement. Thus the scheme does not use the entire measurement and is in this sense Markovian as well as Bayesian.

Here we apply the local-adaptive scheme, in the form optimised for pure states, to the mixed states relevant to imperfect preparation. A true Bayesian scheme, one that uses the entire measurement record, would be the best way to generalise the scheme to mixed states. We return to this point later. For now, we are interested in a direct comparison of the pure state schemes and so proceed with the Markovian form.

We begin by showing that this scheme reaches the Helstrom bound in the case of perfect preparation. We use a different approach to that in Ref. [17] as it does not straightforwardly generalise to include noise. This calculation gives a form for the success probability with N qubits in terms of that for $(N-1)$ qubits, an inductive formula which is solved by the Helstrom bound. We then modify the calculation to include noise. This leads to a different inductive formula, which is then solved to give the overall success probability. In thse calculations, we make repeated use of the result

$$\mathrm{P}(i_N|x,k) = \frac{1}{2}\left[1 + (-1)^{i_N}\cos(2\theta)\cos(2\phi_x) + (-1)^{i_N+k}\sin(2\theta)\sin(2\phi_x)\right] \quad (12)$$

for the probability that the Nth outcome is i_N given that the state $|\psi_k\rangle$ was sent and that the initial $(N-1)$ results were x. This is calculated using Eqs. 1 and 10.

In the local-adaptive scheme, the identification of the prepared state is made with the final outcome. For this reason, the probability of success is

$$\mathrm{P}_N^{ad} = \sum_{x,k} p_k \mathrm{P}(k|x,k)\mathrm{P}(x|k). \quad (13)$$

This is a sum over both signal states $k = 0, 1$ and over all bit strings x of length $(N-1)$, none of which contribute directly to the state identification. We first substitute Eq. 12, with $i_N = k$, into this result to give

$$\mathrm{P}_N^{ad} = \frac{1}{2}\left[1 + \sum_{x,k}\left(\sin(2\theta)\sin(2\phi_x)p_k\mathrm{P}(x|k) + \cos(2\theta)\cos(2\phi_x)(-1)^{i_{N-1}+k}p_k\mathrm{P}(x|k)\right)\right]. \quad (14)$$

Then next step is to use Eq. 11 for the optimal value of $2\phi_x$ in this equation:

$$\mathrm{P}_N^{ad} = \frac{1}{2}\left[1 + \frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\sum_{x,k} p_k\mathrm{P}(x|k) + \cos^2(2\theta)\sqrt{\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^{2N}(2\theta)}}\sum_{x,k}(-1)^{i_{N-1}+k}p_k\mathrm{P}(x|k)\right]. \quad (15)$$

The first sum in this expression is straightforward to evaluate. It is simply a sum over a complete set of possible scenarios and we have $\sum_{x,k} p_k\mathrm{P}(x|k) = 1$. The other series is a little more complicated. We use the usual rules of conditional probability to write

$$\mathrm{P}(x|k) = \mathrm{P}(i_{N-1}\dot{x}|k) = \mathrm{P}(i_{N-1}|\dot{x},k)\mathrm{P}(\dot{x}|k), \quad (16)$$

where we introduce the notation $\dot{x}$ for the bit string of the first $(N-2)$ results. We use also Eq. 12, with x replaced by $\dot{x}$ and i_N replaced by i_{N-1}, for the probabilities $\mathrm{P}(i_{N-1}|\dot{x},k)$ in this equation. Bringing together all of these results, a short calculation reveals

$$\sum_{x,k}(-1)^{i_{N-1}+k}p_k\mathrm{P}(x|k) = \sum_{\dot{x},k}\left(\sin(2\theta)\sin(2\phi_{\dot{x}})p_k\mathrm{P}(\dot{x}|k) + \cos(2\theta)\cos(2\phi_{\dot{x}})(-1)^{i_{N-2}+k}p_k\mathrm{P}(\dot{x}|k)\right). \quad (17)$$

This should be compared with Eq. 14, in which the same expression occurs but over the final rather than penultimate outcome. This can be used to write the expression as

$$\sum_{x,k}(-1)^{i_{N-1}+k}p_k\mathrm{P}(x|k) = 2P_{N-1}^{ad} - 1. \quad (18)$$

After subsituting this into Eq. 15, we are left with the inductive expression

$$\mathrm{P}_N^{ad} = \frac{1}{2}\left[1 + \frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}} + \cos^2(2\theta)\sqrt{\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^{2N}(2\theta)}}(2\mathrm{P}_{N-1}^{ad} - 1)\right]. \quad (19)$$

The general solution to this equation is the multiple-copy Helstrom bound, Eq. 3, which can be verified by direct substitution. The $N = 1$ case corresponds to single-copy state discrimination and that bound is derived in the

usual manner. That the probability expression has this inductive form follows as the measurement strategy is Markovian. We have followed others in showing that the Helstrom bound can be reached with local measurements only [17–19]. Our main result in this section is a generalisation of this expression to the regime of imperfect preparation fidelity.

The calculation proceeds in the same manner as that without noise. The difference is in the probability of a specific result i_N given that the state $|\psi_k\rangle$ was prepared, which changes when the latter is replaced with a noisy state. To take this into account, Eq. 12 is replaced by an equivalent expression calculated using Eqs. 9 and 10. The new probability is

$$\begin{aligned} &\mathrm{P}(i_N|x,k) \\ &= \frac{1}{2}\left[1+(2F-1)(-1)^{i_N}\cos(2\theta)\cos(2\phi_x)\right. \\ &\qquad \left.+(2F-1)(-1)^{i_N+k}\sin(2\theta)\sin(2\phi_x)\right] \end{aligned} \tag{20}$$

so that the only change in the noisy case is the appearance of the factor $(2F-1)$ here. We use this to derive, in exactly the same manner as in the perfect-fidelity case, the probability of success. The result of this, as might be expected based on the change to the individual probabilities, is simply

$$\begin{aligned} \mathrm{P}_N^{ad} = \frac{1}{2}\left[1+(2F-1)\frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\right. \\ \left.+(2F-1)\cos^2(2\theta)\sqrt{\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^{2N}(2\theta)}}(2\mathrm{P}_{N-1}^{ad}-1)\right]. \end{aligned} \tag{21}$$

This relation is hardly more complicated than the noisless case, Eq. 19, but its solution is much more complicated. By recursive application of this formula using the $N=1$ case, which can be evaluated analytically, we establish that the solution is

$$\begin{aligned} \mathrm{P}_N^{ad} = \frac{1}{2}\left(1+(2F-1)^N\sqrt{1-\cos^{2N}(2\theta)}\right. \\ \left.+\frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\mathcal{S}_N\right), \end{aligned} \tag{22}$$

where we introduce the notation

$$\mathcal{S}_N = \sum_{i=1}^{N}(2F-1)^{N+1-i}(1-(2F-1)^{i-1})\cos^{2N-2i}(2\theta).$$

This solution can be verified by substitution into the inductive relationship. The series $\mathcal{S}_N$ can be evaluated using the usual formulae for geometric progressions. After some algebraic manipulation we find

$$\begin{aligned} \mathcal{S}_N =&(2F-1)\frac{1-(2F-1)^N\cos^{2N}(2\theta)}{1-(2F-1)\cos^2(2\theta)} \\ &-(2F-1)^N\frac{1-\cos^{2N}(2\theta)}{1-\cos^2(2\theta)}. \end{aligned} \tag{23}$$

Between Eqs. 22 and 23, the probability that the local-adpative scheme successfully identifies the state is defined in terms of the preparation fidelity F. In the perfect-fidelity case $F=1$, substitution shows that $\mathcal{S}_N = 0$, and we have that the usual Helstrom bound is achieved. If instead $F = 1/2$, the prepared state is by definition a completely mixed state for both $|\psi_0\rangle$ and $|\psi_1\rangle$, so that the states are indistinguishable. For this value of the fidelity, the probability becomes one-half which corresponds to guessing. The other interesting limit is the behaviour of the scheme if there are many copies of the state. We look at this in a later section where we also plot the success probability.

IV. QUANTUM DATA GATHERING

The previous measurement scheme is purely local. It produces a classical bit value for each of the resource qubits. It is known that schemes of this type are in general not able to perform optimal state discrimination when the possible states are mixed. One requires collective measurements. Here, we are interested in the ability of these schemes to function in the presence of preparation noise. As an example of one scheme which measures collectively, we consider quantum data gathering [30]. A collective measurement does not need to explore the entirety of $\mathcal{H}^{\otimes N}$ as the two product states $|\psi_k\rangle$ occupy a subspace with two dimensions. A single qubit, therefore, has enough capacity to store the information which is required for state discrimination. Quantum data gathering is an algorithm which transfers the information from the multiple-copy product state to a further qubit. The latter is then individually measured, which identifies the signal state with a probability equal to the Helstrom bound.

This scheme requires a quantum memory, a qubit which does not decohere between interactions. This probe is initialised in the state $|0\rangle$. When required, we label this space $\mathcal{H}_A$. The interaction with the first qubit is a SWAP gate. The remaining interactions leave the resource qubits, labelled S_i (where $i = 1, 2, ..., N$), each in the state $|0\rangle$ and, if there is no preparation noise, leave the probe in one of two states

$$|\psi_k^{(n)}\rangle = \cos(\theta_n)|0\rangle + (-1)^k\sin(\theta_n)|1\rangle, \quad k=0,1 \tag{24}$$

in which

$$\cos(\theta_n) = \sqrt{\frac{1}{2}\left(1+\cos^n(2\theta)\right)}. \tag{25}$$

These two states have an overlap $\langle\psi_0^{(n)}|\psi_1^{(n)}\rangle = \cos^n(2\theta)$, where n is the number of qubits which the probe has interacted with until that point in the scheme. Thus, the probability of success is the Helstrom bound. The protocol works as the product state of the N systems exists in a two-dimensional subspace of the overall Hilbert space. This, of course, no longer holds for mixed states,

for which some information about the states will be retained in the resource qubits. The interactions between the probe and resource qubits are unitary operations which map this subspace onto the two dimensions of the probe's space through the index k, which is the only piece of information needed to characterise each state. The unitary operator U_n that performs such an operation has the property $U_n|\psi_k\rangle_{S_n}|\psi_k^{(n-1)}\rangle_A = |0\rangle_{S_n}|\psi_k^{(n)}\rangle_A$. Alone, this does not span the Hilbert space and we need to include also the state's components which appear only if the preparation is imperfect. The choice we make is $U_n|\psi_{k\perp}\rangle_{S_n}|\psi_{k\perp}^{(n-1)}\rangle_A = |1\rangle_{S_n}|\psi_{k\perp}^{(n)}\rangle$. The unitary operator, written in the computational basis for both qubits, is

$$\begin{aligned}
U_n|0_{S_n}0_A\rangle &= \frac{\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)}|0_{S_n}0_A\rangle \\
&\quad + \frac{\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)}|1_{S_n}0_A\rangle \\
U_n|1_{S_n}1_A\rangle &= \frac{\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)}|0_{S_n}0_A\rangle \\
&\quad - \frac{\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)}|1_{S_n}0_A\rangle \\
U_n|1_{S_n}0_A\rangle &= \frac{\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)}|0_{S_n}1_A\rangle \\
&\quad + \frac{\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)}|1_{S_n}1_A\rangle \\
U_n|0_{S_n}1_A\rangle &= \frac{\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)}|0_{S_n}1_A\rangle \\
&\quad - \frac{\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)}|1_{S_n}0_A\rangle. \qquad (26)
\end{aligned}$$

After all of the resource qubits have been processed, the qubit is measured with a Helstrom measurement which corresponds to distinguishing $|\psi_0^{(N)}\rangle$ from $|\psi_1^{(N)}\rangle$. Again, the quantity we calculate is the probability that this measurement is successful if the prepared qubits are instead mixed states.

The strategy that we use to calculate this probability is to find, by representing the interactions as Kraus operators acting on $\mathcal{H}_A$, the probe's state at each stage of the protocol. These Kraus operators are derived by considering that the resource qubits are subsequently measured in the computational basis though we sum over both outcomes. This strategy gives us the possibility of considering that such a measurement, which could be used as a diagnostic for the protocol's behaviour, does occur. The Kraus operators are calculated as $M_{i,k}^{(n)} = \langle i|_{S_n}U_n|\tilde{\psi}_k^n\rangle_{S_n}$, where $i = 0, 1$ and we use the noisy form of the state. As the calculation involves pairs of Kraus operators, at this point we do not average over the noise. The Kraus operators are best expressed in the form

$$\begin{aligned}
M_{0,k}^{(n)}|\psi_k^{(n-1)}\rangle &= \cos(\delta\theta_n)|\psi_k^{(n)}\rangle \qquad (27) \\
M_{0,k}^{(n)}|\psi_{k\perp}^{(n-1)}\rangle &= -\frac{\sin(2\theta)\sin(\delta\theta_n)\cos(2\theta_{n-1})}{\sin(2\theta_n)}|\psi_k^{(n)}\rangle \\
&\quad + \frac{\cos(2\theta+\delta\theta_n)\sin(2\theta_{n-1})}{\sin(2\theta_n)}|\psi_{k\perp}^{(n)}\rangle \\
M_{1,k}^{(n)}|\psi_k^{(n-1)}\rangle &= \sin(2\theta+\delta\theta_n)|\psi_{k\perp}^{(n)}\rangle \\
M_{1,k}^{(n)}|\psi_{k\perp}^{(n-1)}\rangle &= -\frac{\sin(\delta\theta_n)\sin(2\theta_{n-1})}{\sin(2\theta_n)}|\psi_k^{(n)}\rangle \\
&\quad + \frac{\sin(\delta\theta_n)\cos(2\theta_{n-1}) - \sin(2\theta+\delta\theta_n)\cos(2\theta_n)}{\sin(2\theta_n)}|\psi_{k\perp}^{(n)}\rangle.
\end{aligned}$$

One way to think about these is objects is that the outcome $M_{0,k}^{(n)}$ indicates that the protocol is running well and conversely for $M_{1,k}^{(n)}$. This is because the former is the only outcome if the fidelity is perfect. This is seen in the Kraus representation as the action of $M_{0,k}^{(n)}$ is to map the state $|\psi_k^{(n-1)}\rangle$ onto $|\psi_k^{(n)}\rangle$, thus preserving the information which is encoded in that basis, whereas the operator $M_{1,k}^{(n)}$ has the opposite effect: by mapping the state $|\psi_k^{(n-1)}\rangle$ onto $|\psi_{k\perp}^{(n)}\rangle$ it deletes all the information which has been acquired up to that point. This is the origin of the claim that a subsequent measurement of the prepared qubit can act as a diagnostic. This point is later considered in more detail.

We are now in a position to calculate the density matrix of the probe after N interactions. We assume that all noise is in the state preparation and that the operations are implemented perfectly. At the first step, the sample is swapped with the probe, so that the probe is left in the state

$$\rho_1 = F|\psi_k\rangle\langle\psi_k| + (1-F)|\psi_{k\perp}\rangle\langle\psi_{k\perp}|, \qquad (28)$$

as was shown earlier (to simplify the notation, we drop the index k from the density operator ρ). We evaluate the next step in full detail and the result allows us to find, by inspection, the form of the density matrix in general. The full calculation is discussed in Appendix A. It is a straightforward though rather involved process, which results in a density matrix

$$\begin{aligned}
\rho_N &= A_N|\psi_k^{(N)}\rangle\langle\psi_k^{(N)}| + (1-A_N)|\psi_{k\perp}^{(N)}\rangle\langle\psi_{k\perp}^{(N)}| \\
&\quad + B_N\sigma_x^{(N)}, \qquad (29)
\end{aligned}$$

in which

$$A_N = 1 - (1-F)\frac{1-\cos^{2N}(2\theta)(2F-1)^N}{1-\cos^2(2\theta)(2F-1)}, \qquad (30)$$

and

$$B_N = (1-F)\frac{\sin^2(2\theta)\cos^{N-1}(2\theta)}{\sin(2\theta_N)} \times \left[\frac{1-(2F-1)^{N-1}}{1-(2F-1)} - \cos^{N+1}(2\theta)(2F-1)^{N-1}\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^2(2\theta)}\right]. \quad (31)$$

Between Eqs. 29, 30 and 31, we have characterised the probe's density matrix in terms of the fidelity and state parameters only. If $F = 1$, which corresponds to the perfect fidelity case, $A_N = 1$ and $B_N = 0$. This corresponds to the probe being in the pure state $|\psi_k^{(N)}\rangle$, as one would expect. If instead $F = 1/2$, which corresponds to maximum infidelity, then again $B_N = 0$ however here $A_N = 1/2$. This means that the probe is in a maximally mixed state so that it carries no information about the prepared state. This corroborates the analysis of the similar cases in the local-adaptive scheme.

In the quantum data gathering routine, following the unitary interactions between the probe and all resource qubits, the probe is left in the density matrix that we have calculated. If the fidelity is perfect, this will be one of two possible states, either $|\psi_0^{(N)}\rangle$ or $|\psi_1^{(N)}\rangle$. At this stage in the protocol, the probe is then measured with the Helstrom measurement which best distinguishes these states. This is the final piece of the calculation, which gives us the probability P_N^{qdg} that the scheme succeeds. Helstrom's conditions tell us that the best measurement is a projector onto the eigenvalues of $p_0|\psi_0^{(N)}\rangle\langle\psi_0^{(N)}| - p_1|\psi_1^{(N)}\rangle\langle\psi_1^{(N)}|$. The case $p_0 \neq p_1$ is significantly more involved without adding further understanding. For this reason we restrict our attention to equiprobable preparation $p_0 = p_1 = 1/2$ for this scheme. The relevant eigenvectors are

$$|\psi_+^{(N)}\rangle = \sqrt{\frac{1+\sin(2\theta_N)}{2}}|\psi_k^{(N)}\rangle + \sqrt{\frac{1-\sin(2\theta_N)}{2}}|\psi_{k\perp}^{(N)}\rangle$$
$$|\psi_-^{(N)}\rangle = \sqrt{\frac{1-\sin(2\theta_N)}{2}}|\psi_k^{(N)}\rangle - \sqrt{\frac{1+\sin(2\theta_N)}{2}}|\psi_{k\perp}^{(N)}\rangle, \quad (32)$$

where the subscript $\pm$ indicates an associated eigenvalue of $\lambda = \pm 1$. It is the positive eigenvalue which corresponds to the correct outcome. The success probability derived from this

$$\mathrm{P}_N^{qdg} = \langle\psi_+^{(N)}|\rho_N|\psi_+^{(N)}\rangle = \frac{1-\sin(2\theta_N)}{2} + A_N\sin(2\theta_N) - B_N\cos(2\theta_N). \quad (33)$$

The Helstrom bound is written in a form useful here as $\mathrm{P}_N^H = (1+\sin(2\theta_N))/2$. We see that P_N^{qdg} is leading order in the Helstrom bound (once A_N and B_N are entered), followed by terms which are linearly and inversely proportional to that object. This structure is similar to the equivalent expression for the local-adaptive measurement scheme. Eqs. 30, 31 and 33 together define the probability of success for the quantum data gathering.

V. DISCUSSION

In Fig. 1 we plot, as a function of the number N of resource qubits, the probability of failure for both local-adaptive measurements and quantum data gathering, alongside a majority voting fixed measurement scheme, for three values of the angle θ and two values of the fidelity F. For now we focus on the former two schemes. In both cases, we have used equiprobable preparation $p_0 = p_1 = 1/2$. Despite the range of parameters, some broad features emerge. We comment on the many-copy limit, in which both quantitites converge upon the same value, below. What is relevant at this point is that, in all cases, the local scheme approaches this limit with fewer qubits than the collective scheme. This improvement is small enough, in the fourth or fifth decimal place for some cases, that it is probably not experimentally significant. Nonetheless, we have shown that the local adaptive scheme is more resilient to noise than the quantum data gathering scheme.

The third scheme plotted in Fig. 1 is a majority voting scheme in which the Helstrom measurement is performed on each qubit and the most common outcome in the measurement record is the overall outcome. There is no simple analytic expression for the success probability but it is straightforward to find numerically [22]. The Helstrom measurement is that for discriminating the two mixed states, rather than the original pure states, though this will be the same for equal priors. Thus, this measurement scheme takes into account both the whole measurement record and the noise in the preparation. In general, we find that this simple generalisation is enough to outperform the other two schemes. In particular, it is not limited by the same asymptotic behaviour as those schemes. If there is only a small amount of noise, as can be seen in the graph with $F = 0.999$ and $\theta = \pi/12$, majority voting no longer outperforms the local adapative and quantum data gathering schemes. As the fidelity becomes closer to one, the two previously analysed schemes will become closer to the genuine optimal scheme, hence they perform better for moderate N in the high-fidelity case.

Asymptotic regime

Special attention should be paid to the many-copy limit of both quantum data gathering and the local adaptive scheme. Interestingly, one finds the same value in

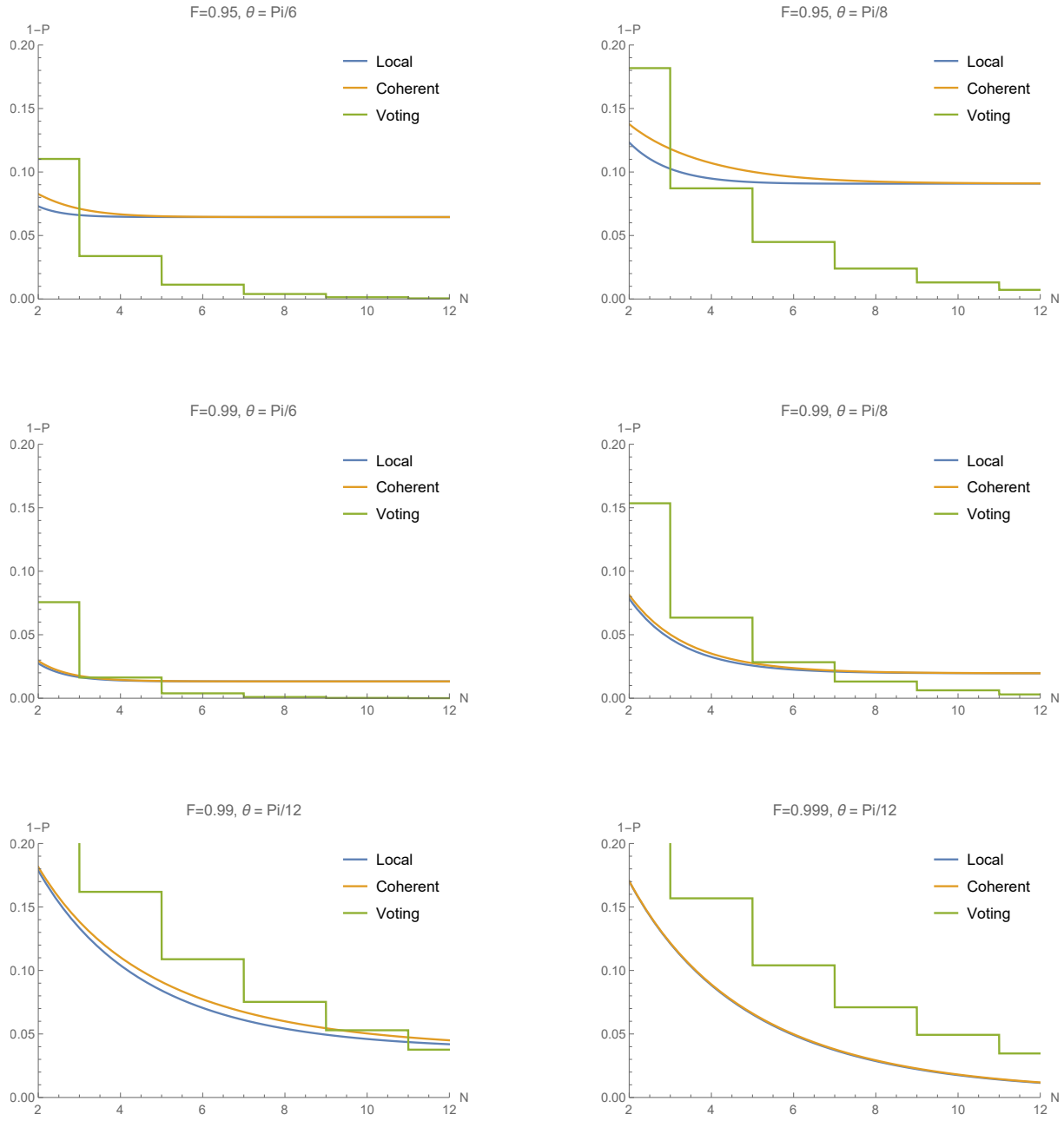

FIG. 1. The probability of failure for three different multiple-copy state discrimination schemes with imperfect preparation: local-adaptive measurements (Local, blue), quantum data gathering (Coherent, orange) and voting based on fixed measurements (Voting, green). The former two schemes do not take into account the preparation noise however the latter scheme does. A range of parameters for the angle θ and fidelity F are used.

both cases:

$$\lim_{N\to\infty} \mathrm{P}_N^{qdg} = \lim_{N\to\infty} \mathrm{P}_N^{ad} = 1 - \frac{1-F}{1-(2F-1)\cos^2(2\theta)}. \quad (34)$$

In the limit $F = 1$, this equation reaches unity and so the states can be perfectly discriminated given an infinite number of copies. If instead the two states are the same $\theta = 0$, then we find a probability of one-half. This makes sense as it should be impossible to distinguish two equal states and all that can be done is to guess. These two limits are non-commuting. This occurs because the measurement schemes are ill-defined when discriminating equal states, i.e., the unitary operation for quantum data gathering would need to map two orthogonal states onto the same state, which is clearly not possible.

It is intriguing that both the local adaptive and quantum data gathering scheme approach the same asymp-

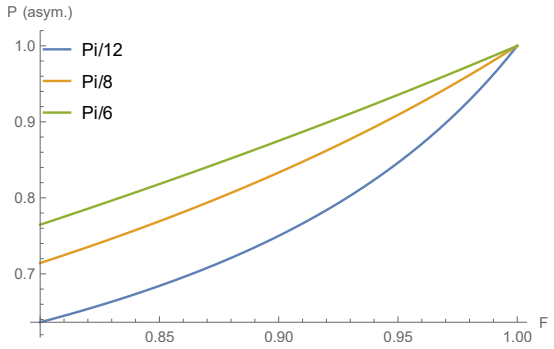

FIG. 2. The asymptote of the probability of success varying with preparation fidelity F for a variety of different angles.

tote, which in general is less than unity, reaching that value only if $F = 1$. This behaviour suggests that the asymptote could be identified with a systematic error which arises when a state discrimination scheme, one which is optimised for a particular pair of states, is applied to a different pair of states. As the preparation fidelity in some cases may be further from $F = 1$ than those we considered above, we plot in Fig. 2 Eq. 34 as a function of F. This gives an idea of the kind of values which will be found experimentally.

The specific form of the many-copy limit can be calculated in a different manner, by understanding the behaviour of the local-adaptive measurement scheme in such a regime. Analysing this behaviour also helps in improving intuition of that scheme. Inspection of Eq. 11 reveals that the scheme in this case can be understood as hypothesis checking. If the outcome on one qubit suggests that $|\psi_0\rangle$ was the prepared state, the next measurement will be onto the basis $|\psi_0\rangle, |\psi_{0\perp}\rangle$, with the latter outcome associated with a preparation of $|\psi_1\rangle$. This explains why the strategy cannot perfectly discriminate. When applied to mixed states, neither measurement outcome is impossible. This hypothesis-checking scheme can be used to calculate the probability of success. We assume that the strategy of hypothesis checking is used for an infinite number of qubits. We find agreement with the original calculation. Two probabilities are required. Firstly,

$$\mathrm{P}(a|i_{N-1} = a, a) = F \tag{35}$$

is the probability of finding outcome a, given that the previous outcome was a (so that the measurement at this stage is $|\psi_a\rangle, |\psi_{a\perp}\rangle$), given that $|\psi_a\rangle$ was sent. We require also

$$\mathrm{P}(a|i_{N-1} = \overline{a}, a) = F - \cos^2(2\theta)(2F-1), \tag{36}$$

which is the probability of outcome a (i.e., the state $|\psi_{\overline{a}\perp}\rangle$) given that the previous measurement gave the outcome $\overline{a}$, the other possible state, and that $|\psi_a\rangle$ was sent. In terms of these objects, the probability of success on the $(N+1)$th qubit is written in terms of the probability of success on the Nth qubit:

$$\begin{aligned}\mathrm{P}^{ad}_{N+1} &= \mathrm{P}(a|i_{N-1} = a, a)\mathrm{P}^{ad}_N \\ &\quad + \mathrm{P}(a|i_{N-1} = \overline{a}, a)(1 - \mathrm{P}^{ad}_N) \\ &= F - \cos^2(2\theta)(2F-1) + \cos^2(2\theta)(2F-1)\mathrm{P}^{ad}_N.\end{aligned} \tag{37}$$

This result is then used iteratively to find an expression for the probability of success after N' more measurements:

$$\begin{aligned}\mathrm{P}^{ad}_{N+N'} &= \left(\cos^2(2\theta)(2F-1)\right)^{N'} \mathrm{P}^{ad}_N \\ &\quad + \left(F - \cos^2(2\theta)(2F-1)\right) \sum_{i=0}^{N'-1} \cos^{2i}(2\theta)(2F-1)^i.\end{aligned} \tag{38}$$

Finally, as N' is increased the first term will be suppressed, and in the limit of an infinite number of copies, the probability of success at a given point makes no contribution to the overall expression. All that remains is to evaluate the geometric summation and to rearrange for

$$\lim_{N'\to\infty} \mathrm{P}^{ad}_{N+N'} = 1 - \frac{1-F}{1-(2F-1)\cos^2(2\theta)}, \tag{39}$$

the same value which was found previously. Here, it has been found with a different method to the more general case. Unfortunately, a similar method for confirming the calculation does not exist for quantum data gathering, however inspection of the unitary Eq. 26 reveals similar behaviour. In the many-copy limit there, the probe states become the diagonal basis states $|+\rangle, |-\rangle$. One example of the behaviour of the unitary in this regime is $U_N|\psi_0\rangle_{S_N}|+\rangle_A = |0\rangle_{S_N}|+\rangle_A$, so that all the unitary has done is to delete the resource qubit's information conditioned upon it matching what is already known.

Gate noise

We have considered here only preparation noise. In the quantum data gathering scheme, there will also be noise in the gates needed to implement the unitary Eq. 26. This operation takes the form of a rotation controlled upon binary addition of the register of each individual qubit, which can be implemented by two CNOT gates alongside single-qubit gates. Thus, $2N$ two-qubit gates are needed to perform quantum data gathering on a resource of N qubits. We assume that the contribution to the noise from single qubit gates is negligible. Because the diamond norm [31–33], the standard measure of gate noise, satisfies the triangle inequality, that there are $2N$ gates required means that the total gate noise scales linearly with N. This will appear as a noisy channel acting upon the probe's state, and decrease further the probability of success. To make further comments, we would need to understand the form of the noise in more detail [34].

Improving the scheme

In the large N limit, any two states are in principle distinguishable, as with an infinite-qubit resource one could simply perform tomography to reconstruct the state. It is clear that the bound Eq. 34 can be improved upon. With finite N, it is desirable to tailor the scheme to the states used.

Some obvious methods exist to improve the performance of each scheme. The local scheme which we have considered up to this point is Markovian, so that $N-1$ of the previous results are discarded. In the pure state case, there are symmetries between the probabilities of different bit strings such that this does not lessen the overall probability of success [17] however when the preparation is imperfect this no longer holds, and information is wasted. The whole measurement record could be taken into account when updating the prior probabilities of the state, and this would improve the performance of the scheme.

A full analysis of the modified scheme would be beyond the scope of this article. What can be discussed is the asymptotic behaviour of the scheme, which was shown earlier to consist of measuring projectively with $|\psi_0\rangle$ or $|\psi_1\rangle$ as one of the outcomes, depending on the previous measurement result, so that the measurement record in the many-copy limit can be broken down into two fixed measurements. Post-selection of either measurement is equivalent to classical sampling of two probability distributions, a problem which is known to asymptotically decay with an exponent determined by the Chernoff bound [20, 35], and thus outperforms the Markovian scheme. This exponent will be determined by the number of fixed measurements in each set and will give worse performance than the quantum Chernoff bound which is found as the optimal bound in asymptotic multiple-copy state discrimination [36–38].

In quantum data gathering, the scheme can be modified by measuring the resource qubits at each step and acting based upon the outcome. As noted, an outcome of $|1\rangle$ indicates that all quantum information gathered up until that point has been lost. So, one way to modify the protocol is to restart whenever such an outcome occurs. Some care must be taken as only a finite number of consecutive $|0\rangle$ outcomes can occur before a bad outcome. In Fig. 1 it is seen that only small number, three or four, of interactions are required to get very close to the best-possible probability. However, numerical evaluation of the relevant probabilities reveals that even this small number is unlikely enough (while still being highly probably, $p \geq 0.97$ typically) to bring the overall probability of state discrimination below that which occurs if the qubits are not measured. This is played off against two things. Firstly, success here is heralded at the expense of increasing ambiguity in some cases, similar to unambiguous state discrimination. Secondly, if there are many resource qubits available, a small run of successes becomes likely to occur at some point. Thus, in some scenarios it may be advantageous to post-select based on the measurement outcomes. A hybrid scheme in which subsets of systems are measured collectively, followed by majority voting on the measurement output would give an improved probability of success, but still less than the local scheme. Here we chose to evaluate the performance of a scheme requiring a single qubit of memory as the quantum probe. A fully general scheme, which would achieve the optimal Helstrom measurement for arbitrary many-copy states, would require a processor of size $\log N$ [30]. Our results show that how a collective measurement is implemented has a considerable effect on its robustness to noise.

VI. CONCLUSION

We have considered the ability of two multiple-copy state discrimination schemes to perform when the state preparation is imperfect. We find two surprising results. Firstly, that for small amounts of uncharacterised noise, the optimal local adaptive measurement is more robust than the simple, single qubit collective scheme. We also find that both schemes have the same many-copy limit, which is less than unity. Despite the different physical mechanisms used in each scheme, they have precisely the same behaviour in this regime. This suggests that the quantity found is a generic property of applying an incorrect scheme, and should be investigated further. With a modification of the local adaptive scheme, we are able to surpass this limit and recover the desirable exponential decay with increasing N.

It would be useful to know an optimal state discrimination scheme for mixed states of the type considered here. A natural starting point would be to generalise the local-adaptive scheme to use the entire measurement record when updating the prior probabilities in a Bayesian manner and to calculate the range, if any, in which this strategy is optimal. In general, more analytic work is required in multiple-copy state discrimination. Some of the techniques used here may be found to be useful in that task.

Appendix A: State of the probe qubit in quantum data gathering

In this appendix we evaluate the density matrix of the probe qubit which is used in the quantum data gathering scheme.

The denisty matrix of the probe in this scheme after the first interaction, a SWAP, is given by

$$\rho_1 = F|\psi_k\rangle\langle\psi_k| + (1-F)|\psi_{k\perp}\rangle\langle\psi_{k\perp}|. \qquad \text{(A1)}$$

From this point on the probe and resource qubit interact according the relevant unitary, defined above. In order to find the former object's density matrix after the second interaction, we multiply by the Kraus operators in Eq.

27 and then average over $\delta\theta_i$ in one step here. It is a straightforward (though longwinded) calculation to find

$$\begin{aligned}\rho_2 &= M_{0,k}^{(2)}\rho_1 M_{0,k}^{(2)\dagger} + M_{1,k}^{(2)}\rho_1 M_{1,k}^{(2)\dagger} \\ &= \big(F-(1-F)(2F-1)\cos^2(2\theta)\big)\,|\psi_k^{(2)}\rangle\langle\psi_k^{(2)}| \\ &+ \big(1-F+(1-F)(2F-1)\cos^2(2\theta)\big)\,|\psi_{k\perp}^{(2)}\rangle\langle\psi_{k\perp}^{(2)}| \\ &+ (1-F)\frac{\cos^2(2\theta)\sin^2(2\theta)}{\sin^2(2\theta_2)}\sigma_x^{(2)}, \end{aligned} \tag{A2}$$

where $\sigma_x^{(2)} = |\psi_k^{(2)}\rangle\langle\psi_{k\perp}^{(2)}| + |\psi_{k\perp}^{(2)}\rangle\langle\psi_k^{(2)}|$. We keep the convention of using a superscript on the Pauli matrix to indicate the basis in which it is written. This density matrix can be understood as two pieces: a trace-one, diagonal piece consisting of the first two terms and another consiting of only the σ_x matrix. We can expect, based on this, that the same is true of the general density matrix, which we expect can be written

$$\begin{aligned}\rho_N &= A_N|\psi_k^{(N)}\rangle\langle\psi_k^{(N)}| + (1-A_N)|\psi_{k\perp}^{(N)}\rangle\langle\psi_{k\perp}^{(N)}| \\ &+ B_N\sigma_x^{(N)}. \end{aligned} \tag{A3}$$

This is confirmed by the following analysis, in which we evaluate A_N and B_N by calculating how each piece (diagonal and Pauli) is updated. We again multiply by the Kraus operators and average over $\delta\theta_i$ in a single step. The first result is

$$\begin{aligned}&A_{N-1}\sum_i M_{i,k}^{(n)}|\psi_k^{(N-1)}\rangle\langle\psi_k^{(N-1)}|M_{i,k}^{(n)\dagger} \\ &+ (1-A_{N-1})\sum_i M_{i,k}^{(n)}|\psi_{k\perp}^{(N-1)}\rangle\langle\psi_{k\perp}^{(N-1)}|M_{i,k}^{(n)\dagger} \\ &= \big(F-(1-A_{N-1})(2F-1)\cos^2(2\theta)\big)\,|\psi_k^{(N)}\rangle\langle\psi_k^{(N)}| \\ &+ \big(1-F+(1-A_{N-1})(2F-1)\cos^2(2\theta)\big)\,|\psi_{k\perp}^{(N)}\rangle\langle\psi_{k\perp}^{(N)}| \\ &- (1-A_{N-1})\,\frac{(2F-1)\sin^2(2\theta)\cos(2\theta_N)}{\sin(2\theta_N)}\sigma_x^{(N)}. \end{aligned} \tag{A4}$$

Notice that again we find the same structure, that of a diagonal piece and a Pauli matrix. The other update is

$$\begin{aligned}&\sum_i M_{i,k}^{(n)}\sigma_x^{(N-1)}M_{i,k}^{(n)\dagger} \\ &\qquad = \frac{(2F-1)\cos(2\theta)\sin(2\theta_{N-1})}{\sin(2\theta_N)}\sigma_x^{(N)}. \end{aligned} \tag{A5}$$

It is seen that both terms contribute in the form, written in the natural basis of the next step, that we have predicted and the density matrix will always take the form of Eq. A3. Repeated application of the above two results allow us to evaluate A_N and B_N, which are both written in terms of geometric progressions. We find

$$\begin{aligned}A_N &= F\cos^{2N-2}(2\theta)(2F-1)^{N-1} \\ &+ \big(F-(2F-1)\cos^2(2\theta)\big)\sum_{i=0}^{N-2}\cos^{2i}(2\theta)(2F-1)^i \end{aligned} \tag{A6}$$

$$\begin{aligned}B_N &= (2F-1)\sin^2(2\theta) \\ &\times\sum_{i=1}^{N-1}\frac{\cos(2\theta_{i+1})}{\sin(2\theta_{i+1})}\prod_{j=i+2}^{N}(2F-1)\cos(2\theta)\frac{\sin(2\theta_{j-1})}{\sin(2\theta_j)}. \end{aligned} \tag{A7}$$

It is straightforward to evaluate the summation to give

$$A_N = 1-(1-F)\frac{1-\cos^{2N}(2\theta)(2F-1)^N}{1-\cos^2(2\theta)(2F-1)}, \tag{A8}$$

which is then used alongside Eq. 25 to evaluate

$$\begin{aligned}B_N &= (1-F)\frac{\sin^2(2\theta)\cos^{N-1}(2\theta)}{\sin(2\theta_N)} \\ &\times\left[\frac{1-(2F-1)^{N-1}}{1-(2F-1)}\right. \\ &\qquad\left.-\cos^{N+1}(2\theta)(2F-1)^{N-1}\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^2(2\theta)}\right]. \end{aligned} \tag{A9}$$

The denominators of the two fractions inside the braces could each be simplified however we leave them in this form so that it is clear that there are no convergence issues in the limit $F \to 1/2$ or $\theta \to 0$.

ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council and by the Royal Society (RP150122). We thank both referees for their suggestions on the content of this article.

[1] S. M. Barnett and S. Croke, Adv. Opt. Photon. **1**, 238 (2009).
[2] J. Bae and L.-C. Kwek, J. Phys. A: Math. Theor **48**, 083001 (2015).
[3] J. Bergou, Mod. Opt. **57**, 160 (2010).
[4] S. M. Barnett, *Quantum Information* (Oxford University Press, Oxford, 2009).
[5] A. Chefles, Contemp. Phys. **41**, 401 (2010).
[6] C. W. Helstrom, *Quantum detection and estimation theory* (Academic, 1976).
[7] D. Ha and Y. Kwon, Phys. Rev. A **87**, 062302 (2013).
[8] G. Weir, S. M. Barnett, and S. Croke, Phys. Rev. A **96**, 022312 (2017).
[9] J. Bae and W.-Y. Hwang, Phys. Rev. A **87**, 012334 (2013).

[10] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
[11] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
[12] E. Chitambar and M.-H. Hsieh, Phys. Rev. A **88**, 020302 (2013).
[13] S. Croke, G. Weir, and S. M. Barnett, Phys. Rev. A **95**, 052308 (2017).
[14] C. H. Bennett *et al.*, Phys. Rev. A **59** (1999).
[15] A. M. Childs *et al.*, Comms. Math. Phys. **323**, 1121 (2013).
[16] S. Croke and S. M. Barnett, Phys. Rev. A **95**, 012337 (2017).
[17] A. Acín, E. Bagan, *et al.*, Phys. Rev. A **71**, 032338 (2005).
[18] D. Brody and B. Meister, Phys. Rev. Lett. **76** (1996).
[19] M. Ban, K. Yamazaki, and O. Hirota, Phys. Rev. A **55**, 22 (1997).
[20] B. L. Higgins, A. C. Doherty, *et al.*, Phys. Rev. A **83**, 052314 (2011).
[21] S. Slussarenko, M. M. Weston, *et al.*, Phys. Rev. Lett. **118**, 030502 (2017).
[22] B. Higgins *et al.*, Phys. Rev. Lett. **103**, 220503 (2009).
[23] J. Calsamiglia *et al.*, Phys. Rev. A **77**, 032311 (2008).
[24] J. Calsamiglia *et al.*, Phys. Rev. Lett. **105**, 080504 (2010).
[25] M. Hayashi, IEEE Trans. Inf. Theory **55**, 3807 (2009).
[26] R. Koenig, U. Maurer, and R. Renner, IEEE Trans. Inf. Theory **51**, 2391 (2005).
[27] "Open problems in quantum information theory, 31," `http://qig.itp.uni-hannover.de/qiproblems/31`, accessed 6th Aug. 2019.
[28] J. Burnett *et al.*, "Decoherence benchmarking of superconducting qubits," (2019), arxiv:1901.04417.
[29] S. Schlör *et al.*, "Correlating decoherence in transmon qubits: Low frequency noise by single fluctuators," (2019), arxiv:1901.05352.
[30] R. Blume-Kohout, S. Croke, and M. Zwolak, Scientific Reports **3** (2013).
[31] D. Aharanov, A. Kitaev, and N. Nisan, in *Proceedings of the Thirteenth Annual ACM Symposium on theory of Computation* (STOC, 1997) pp. 20–30.
[32] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Info. Comput. **6**, 97 (2006).
[33] D. Aharanov and M. Ben-Or, SIAM J. Comput. **38**, 1207 (2008).
[34] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, New J. Phys. **18**, 012002 (2016).
[35] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, 2006).
[36] K. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, Commun. Math. Phys. **279**, 251 (2008).
[37] M. Nussbaum and A. Szkoła, Ann. Stat. **39**, 3211 (2011).
[38] M. Nussbaum and A. Szkoła, Ann. Stat. **37**, 1040 (2009).