PAPER • OPEN ACCESS

Beating noise with abstention in state estimation

To cite this article: Bernat Gendra *et al* 2012 *New J. Phys.* **14** 105015

View the article online for updates and enhancements.

Beating noise with abstention in state estimation

Bernat Gendra, Elio Ronco-Bonvehi, John Calsamiglia[1], Ramon Muñoz-Tapia and Emilio Bagan

Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
E-mail: John.Calsamiglia@uab.cat

Abstract. We address the problem of estimating pure qubit states with non-ideal (noisy) measurements in the multiple-copy scenario, where the data consist of a number N of identically prepared qubits. We show that the average fidelity of the estimates can increase significantly if the estimation protocol allows for inconclusive answers, or abstentions. We present the optimal protocol and compute its fidelity for a given probability of abstention. The improvement over standard estimation, without abstention, can be viewed as an effective noise reduction. These and other results are illustrated for small values of N. For asymptotically large N, we derive analytical expressions of the fidelity and the probability of abstention and show that for a fixed fidelity gain the latter decreases with N at an exponential rate given by a Kulback–Leibler (relative) entropy. As a byproduct, we give an asymptotic expression in terms of this very entropy of the probability that a system of N qubits, all prepared in the same state, has a given total angular momentum. We also discuss an extreme situation where noise increases with N and where estimation with abstention provides the most significant improvement as compared to the standard approach.

[1] Author to whom any correspondence should be addressed.

IOP Institute of Physics ⬤ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

Knowing the state of a system is a key task in quantum information processing. An unknown quantum state can only be unveiled by means of measurements. These, however, provide only partial knowledge about the system and, furthermore, this information gain comes always at the expense of destroying the state. Only when a reasonably large number N of identically prepared copies of the system are available, is an accurate estimation of the state possible. For a given N, the aim then is to find the measurement protocol that yields the best estimate of the input state.

The standard estimation optimization problem is suited for a situation where, say, an experimentalist is confronted with an unknown state of a system of which she is asked to provide an estimate, based, of course, on the results of a measurement of her choice. A quantitative assessment of her performance is usually given by the expected value of the fidelity (or some other distinguishability measure) between the unknown input and her guess (see below). Hence, it is implicitly assumed that the experimentalist is obliged to provide such a guess regardless of the measurement outcome she obtains. In this context, many results have been obtained in a large variety of scenarios over the last few years [1–12].

Here, we will study a variation of this setting suited for a situation where the experimentalist is allowed to decide whether to provide a guess or abstain from doing so. Of course, this decision cannot be based on the actual state of the system (which is unknown by definition) but rather on the result of a measurement. This relaxation of the original setting is very useful because it enables the experimentalist to post-select her measurement outcomes in order to provide a more accurate guess. That is, the possibility of abstaining enables her to discard instances where the measurement outcome turns out to be not informative enough. We will find that abstention can provide an important advantage, especially in noisy scenarios. The problem of 'state estimation with abstention'[2] is especially relevant in situations where the experimentalist can afford to re-run the experiment, i.e. she can easily prepare a new instance of the problem, or where she prioritizes having high-quality estimates.

Post-selection is a widely used tool in quantum information, particularly in experimental scenarios, where one has special demands or constraints. A form of abstention has already been explored in state discrimination [13], another important quantum statistical inference

_____

[2] We thank G Chiribella for suggesting this word.

primitive. Discrimination aims at identifying in which one, out of a set of known quantum states, a system has been prepared. Two fundamental approaches are usually considered: the so-called 'minimum-error', where the experimentalist always has to provide a conclusive answer, at the expense, of course, of being wrong with a certain probability [14], and 'unambiguous discrimination', where no errors are permitted, but instead an inconclusive answer (abstention) may be given with some probability [15]. By varying the allowed rate of inconclusive answers [16–19], we may go from one approach to the other [20–22]. The possibility of abstaining has been studied in [23] for phase estimation, and in [24] for direction estimation with arbitrary pure input signals. In both cases the results show a significant improvement over the standard (without abstention) approach.

In this work we consider the optimal estimation of a completely unknown pure qubit state when N copies of it are available for measurement and when a certain rate (probability and rate will be used interchangeably throughout the paper) of abstention Q is permitted. Our approach can be viewed as a probabilistic purification that succeeds with probability Q (similar probabilistic maps for broadcasting and cloning have been employed, for instance, in [25–27]). We will show that in an ideal noise-free scenario, abstention does not improve the estimation accuracy. However, it does in a realistic noisy scenario, as we claimed above. Here we will consider a simplified model where noisy measurements will be replaced by local depolarizing channels followed by ideal measurements.

The paper is organized as follows. In section 2, we consider estimation without abstention. More precisely, we obtain the protocol that gives the best estimate of the state of a qubit based on non-ideal measurements on N independent and identically prepared systems. In section 3, estimation with abstention is introduced, and the optimal protocol for a fixed value of the abstention rate Q is obtained. We study the asymptotic regime of large N and derive the corresponding maximum fidelity and probability of abstention. As an example, we also consider a scenario where abstention gives a drastic improvement. This is the case when noise increases with N in such a way that the fidelity of the estimation approaches a finite value less than 1 as N becomes large. We close the paper with some brief conclusions and present an outlook for future work.

2. No abstention

Let us consider N copies of a completely unknown pure qubit state $|\vec{n}\rangle$ (throughout the paper $\vec{n}$ will denote a unit Bloch vector) that we wish to estimate by performing a realistic, and therefore noisy, quantum measurement. We model it as an ideal measurement preceded by the single-qubit depolarizing channel acting on every copy:

$$\mathcal{E}(\rho) = (1 - \eta)\rho + \frac{\eta}{3}(\sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z), \tag{1}$$

where with probability $1 - \eta$ no error occurs, while with probability η the state is affected by either a bit-flip, a phase-flip or both. This error probability η is assumed to be known by the experimentalist; therefore, for the purpose of analyzing the effects of noise in the estimation process, we will transfer its effect to the input states and optimize the estimation protocol over ideal measurements. Hence, we will consider input states of the form

$$\rho(\vec{n}) = r |\vec{n}\rangle\langle\vec{n}| + (1 - r)\frac{\mathbb{1}}{2} = \frac{\mathbb{1} + r\,\vec{n} \cdot \vec{\sigma}}{2}, \tag{2}$$

with $r = 1 - (4/3)\eta$. In other words, we will assume that the input states either do not change with probability r or they become completely randomized with probability $1 - r = (4/3)\eta$. The original problem is thus equivalent to the estimation of a pure state $|\vec{n}\rangle$ (or of a uniformly distributed Boch vector $\vec{n}$) based on the outcomes of an appropriate ideal measurement on N copies of the mixed state $\rho(\vec{n})$ in equation (2), i.e. on the state $\rho(\vec{n})^{\otimes N} = \tau(\vec{n})$.

For each measurement outcome χ an estimate $|\vec{n}_\chi\rangle$ is provided according to some guessing rule $\chi \to |\vec{n}_\chi\rangle$. We choose to quantify the quality of the estimate by means of the squared overlap

$$f(\vec{n}, \vec{n}_\chi) = |\langle \vec{n}|\vec{n}_\chi\rangle|^2, \tag{3}$$

also known as the fidelity. The overall quality of the estimation protocol is then given by the average fidelity

$$F = \sum_\chi \int \mathrm{d}n f(\vec{n}, \vec{n}_\chi)\, p(\chi|\vec{n}), \tag{4}$$

where $\mathrm{d}n = \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi/(4\pi)$ is the uniform probability distribution on the two-sphere and $p(\chi|\vec{n})$ is the conditional probability of obtaining the outcome χ if the input state is $\tau(\vec{n})$. This probability is given by the Born rule $p(\chi|\vec{n}) = \mathrm{tr}[\Pi_\chi \tau(\vec{n})]$, where $\Pi_\chi \geqslant 0$ are the elements of a positive operator valued measure (POVM). They satisfy the completeness relation $\sum_\chi \Pi_\chi = \mathbb{1}$, where $\mathbb{1}$ denotes the identity operator in the space spanned by the input states $\{\tau(\vec{n})\}$. The index χ may be discrete, continuous or both.

A protocol (i.e. a measurement $\{\Pi_\chi\}$ and a guessing rule $\chi \to |\vec{n}_\chi\rangle$) is said to be optimal if it maximizes F. For pure states, $r = 1$, the maximum fidelity is well known [3]:

$$F = \frac{N+1}{N+2} = 1 - \frac{1}{N} + \mathcal{O}(N^{-2}). \tag{5}$$

It is also known that the (continuous) covariant POVM [3]

$$\Pi(\vec{s}) = (2J+1)U(\vec{s})\,|J\,J\rangle\langle J\,J|\,U^\dagger(\vec{s}) \tag{6}$$

(with the obvious guessing rule $\Pi(\vec{s}) \to |\vec{s}\rangle$) is optimal. In (6), we use the standard notation, where $\{|jm\rangle\}_{m=-j}^j$ is the eigenbasis of the total angular momentum operators J^2 and J_z. We denote by $U(\vec{s}) = [u(\vec{s})]^{\otimes N}$, $u(\vec{s}) \in \mathrm{SU}(2)$ (the unitary representation of) the rotation that maps the unit (Bloch) vector $\hat{z}$ into $\vec{s}$ (thus $u(\vec{s})|\frac{1}{2}\frac{1}{2}\rangle = |\vec{s}\rangle$), and we have also introduced the definition $J \equiv N/2$. Note that the POVM $\{\Pi(\vec{s})\}$ acts on the symmetric subspace of largest total angular momentum J, of dimension $2J + 1 = N + 1$. In terms of J, (5) can also be written as

$$F = \frac{1}{2}\left(1 + \frac{J}{J+1}\right) \equiv \frac{1}{2}(1 + \Delta_J). \tag{7}$$

Mixed states span a much larger Hilbert space and the computation becomes more involved. Due to the permutational invariance of the input state, it greatly simplifies in the total angular momentum basis, where $\tau(\vec{n})$ is block-diagonal [9]. We have

$$\tau(\vec{n}) = \sum_{j=j_{\min}}^{J} p_j \tau_j(\vec{n}) \otimes \frac{\mathbb{1}_{\mathrm{rep}}^{(j)}}{n_j}, \tag{8}$$

where $\tau_j(\vec{n})$ is the normalized mixed state

$$\tau_j(\vec{n}) = \frac{1}{Z_j} \sum_{m=-j}^{j} R^m\, U(\vec{n})\,|jm\rangle\,\langle jm|\,U^\dagger(\vec{n}), \tag{9}$$

with the definitions

$$Z_j = \sum_{m=-j}^{j} R^m = \frac{R^{j+1} - R^{-j}}{R - 1}, \quad R = \frac{1+r}{1-r} > 1. \tag{10}$$

The projector $\mathbb{1}_{\text{rep}}^{(j)}$ stands for the various occurrences of the irreducible representation of total angular momentum j. It has dimension n_j given by

$$n_j = \begin{pmatrix} 2J \\ J - j \end{pmatrix} - \begin{pmatrix} 2J \\ J - j - 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2J \\ J - j \end{pmatrix} \frac{2j+1}{J+j+1}. \tag{11}$$

In the sum (9), j runs from $j_{\min} = 0$ ($j_{\min} = 1/2$) for N even (odd) to the maximum total angular momentum J, in contrast to the pure state case where only the maximum value J appears. The numbers $p_j > 0$ are the probabilities that the state $\tau(\vec{n})$ has quantum number j, i.e. $p_j = \text{tr}[\mathbb{1}_j \tau(\vec{n})]$, where $\mathbb{1}_j = \sum_{m=-j}^{j} |jm\rangle\langle jm| \otimes \mathbb{1}_{\text{rep}}^{(j)}$ is the projector onto the corresponding eigenspace, The probabilities read

$$p_j = \left(\frac{1-r^2}{4}\right)^J n_j Z_j$$

$$= \left(\frac{1-r^2}{4}\right)^J \begin{pmatrix} 2J \\ J - j \end{pmatrix} \frac{2j+1}{J+j+1} \frac{R^{j+1} - R^{-j}}{R - 1}. \tag{12}$$

One can easily check that $\sum_j p_j = 1$, as it should be.

Because of the block diagonal form of the input states, an obvious optimal measurement consists of a direct sum of covariant POVMs,

$$\Pi(\vec{s}) = \bigoplus_{j=j_{\min}}^{J} \Pi_j(\vec{s}) \otimes \mathbb{1}_{\text{rep}}^{(j)}, \tag{13}$$

where each of them is a straightforward generalization of equation(6):

$$\Pi_j(\vec{s}) = (2j+1) U(\vec{s}) |j\ j\rangle\langle j\ j| U^{\dagger}(\vec{s}). \tag{14}$$

One can easily check that the completeness condition $\int \mathrm{d}s\, \Pi(\vec{s}) = \mathbb{1}$ holds. The total fidelity then is

$$F = \frac{1}{2}\left(1 + \sum_{j=j_{\min}}^{J} p_j \Delta_j\right), \tag{15}$$

where [1]

$$\Delta_j = \frac{\langle J_z \rangle_j}{j+1} = \frac{\text{tr}[J_z \tau_j(\hat{z})]}{j+1}. \tag{16}$$

A straightforward calculation gives

$$\langle J_z \rangle_j = \frac{1}{Z_j} \sum_{m=-j}^{j} m R^m = j - \frac{1}{R - 1} + \frac{2j+1}{R^{2j+1} - 1}. \tag{17}$$

Note that for pure states, one has $R \to \infty$, and in turn $\langle J_z \rangle_J \to J$, in agreement with equation (7).

As will be shown in the next section, for asymptotically large N the probability p_j peaks at a value of $j \simeq rJ$, which gives the dominant and subdominant contributions to the sum in (15). Up to order $1/N$, and discarding exponentially vanishing contributions (e.g. $\sim R^{-rJ}$), the asymptotic fidelity turns out to be

$$F = 1 - \frac{1}{Nr}\frac{r+1}{2r} + \cdots. \tag{18}$$

This result is interesting on its own and, to the best of our knowledge, has not been presented before. Note that for pure states ($r = 1$), equation (18) agrees with the asymptotic expression of the fidelity in equation (5).

3. Abstention

In this section we focus on estimation protocols where the experimentalist is allowed to abstain, i.e. refrain from giving an answer, if the outcome of the measurement she has carried out cannot provide a good enough estimate of the unknown state. Obviously, F cannot decrease by excluding these abstentions from the average. In noisy scenarios, such as that considered in this paper, F actually increases, as will be shown below. Our aim is to quantify this gain and find the optimal protocol. In our approach, the probability of abstention, Q, is kept fixed, rather than unrestricted, since usually in practical situations one cannot afford to discard an unlimited amount of resources/state preparations.

3.1. The general framework

To enable the possibility of abstaining, the POVM representing the measurement must include the abstention operator, which we denote by Π_0, in addition to the operators $\{\Pi_\chi\}$, each of which are associated with a specific estimate $|\vec{n}_\chi\rangle$. Thus, the completeness relation reads

$$\sum_\chi \Pi_\chi + \Pi_0 = \mathbb{1}. \tag{19}$$

The probability of abstention (abstention rate) and that of producing an estimate (acceptance rate) are then given, respectively, by

$$Q = \int dn \, \text{tr}\left[\Pi_0 \tau(\vec{n})\right] \quad \text{and} \quad \bar{Q} = 1 - Q, \tag{20}$$

and the mean fidelity defined in (4) now becomes

$$F(Q) = \frac{1}{\bar{Q}} \sum_\chi \int dn \, f(\vec{n}, \vec{n}_\chi) \text{tr}\left[\Pi_\chi \tau(\vec{n})\right], \tag{21}$$

where note that the sum does not include the Π_0 operator and $\bar{Q}$ takes into account the abstentions excluded from the average.

We next note that for any unitary transformation U of the type defined after equation (6), the operators $\{U\Pi_\chi U^\dagger, U\Pi_0 U^\dagger\}$ give the same value of Q and $F(Q)$ as the original set $\{\Pi_\chi, \Pi_0\}$, provided we change the guessing rule as $\vec{n}_\chi \to \mathcal{R}_U \vec{n}_\chi$, where $\mathcal{R}_U$ is the SO(3) rotation whose unitary representation is U. Therefore, one can easily prove that Π_0 (the set

$\{\Pi_\chi\}$) can always be chosen to be SU(2) invariant (covariant) by simply averaging over U. In other words, with no loss of generality, the POVM elements that provide a guess $|\vec{s}\rangle$ can be chosen as

$$\tilde{\Pi}(\vec{s}) = U(\vec{s}) \, \Pi \, U^\dagger(\vec{s}), \tag{22}$$

where $\Pi \geqslant 0$ is the so-called seed of the POVM (in particular, note that $\tilde{\Pi}(\hat{z}) = \Pi$). The abstention operator then reads

$$\Pi_0 = \mathbb{1} - \int \mathrm{d}s \, \tilde{\Pi}(\vec{s}), \tag{23}$$

which is manifestly rotationally invariant (as claimed above). It is thus proportional to the identity on each invariant subspace

$$\Pi_0 = \bigoplus_{j=j_{\min}}^{J} a_j \mathbb{1}_j, \tag{24}$$

where a_j are coefficients that satisfy the condition $0 \leqslant a_j \leqslant 1$ and $\mathbb{1}_j$ is the projector onto the corresponding eigenspace j previously defined. We can also choose $\tilde{\Pi}(\vec{s})$ to have the block-diagonal form of the input state $\tau(\vec{n})$, namely,

$$\tilde{\Pi}(\vec{s}) = \bigoplus_{j=j_{\min}}^{J} \tilde{\Pi}_j(\vec{s}) \otimes \mathbb{1}_{\mathrm{rep}}^{(j)}. \tag{25}$$

For a given $\{a_j\}$, the optimality of $\Pi_j(\vec{s})$, defined in equation (14), clearly ensures that

$$\tilde{\Pi}_j(\vec{s}) = (1 - a_j)\Pi_j(\vec{s}) \tag{26}$$

are also optimal for estimation with abstention. We have from (20) that the abstention probability is simply

$$Q = \sum_{j=j_{\min}}^{J} p_j a_j, \tag{27}$$

where p_j is given in equation (12). The coefficients a_j can be understood as the probabilities of abstention conditional on the input state having total angular momentum j, i.e. $a_j = p(\text{abstention}|j)$. Similarly, for a given j, the probability of producing an estimate, or accepting, is $\bar{a}_j = 1 - a_j = p(\text{acceptance}|j)$.

From equation (21) we obtain

$$F(Q) = \frac{1}{2}\left(1 + \sum_{j=j_{\min}}^{J} p_j \tilde{\Delta}_j\right), \tag{28}$$

where

$$\tilde{\Delta}_j = \frac{1 - a_j}{1 - Q} \Delta_j = \frac{\bar{a}_j}{\bar{Q}} \Delta_j, \tag{29}$$

and the quantity Δ_j is given in equations (16) and (17). Thus, we are only left with the free parameters a_j, which have to be optimized in order to maximize $F(Q)$, subject to the constraints $0 \leqslant a_j \leqslant 1$ and (27). Somehow expected, one can show that Δ_j is a monotonically increasing function of j, i.e. $\Delta_{j-1} < \Delta_j$; therefore the largest contribution to the fidelity is

given by Δ_J. This corresponds to $\bar{a}_J = 1$ and $\bar{a}_j = 0$, $j < J$. Hence, for an unrestricted probability of abstention, the optimal protocol discards any contribution with $j < J$. This protocol, however, would provide an estimate with a probability that decreases exponentially with N, for $r < 1$, as $p_J \simeq (1/r)[(1+r)/2]^{N+1}$. Note that in a noiseless scenario, $r = 1$, there is only the contribution $j = J$, which is already the optimal one and therefore abstention is of no use in such a case.

Clearly, for finite Q there can be contributions from other total angular momentum eigenspaces ($j < J$) compatible with equation (27). Recalling the monotonicity of Δ_j, and by convexity, it is obvious from equations (28) and (29) that there must exist an angular momentum threshold j^* such that $\bar{a}_j = 0$ ($\bar{a}_j = 1$), if $j < j^*$ ($j > j^*$). The value j^* is determined through equation (27) to be

$$j^* = \max \left\{ j \text{ such that } Q - \sum_{j'=j_{\min}}^{j-1} p_{j'} \geqslant 0 \right\}. \tag{30}$$

Thus, we have

$$a_j = \begin{cases} 1, & j < j^*, \\ p_j^{-1} \left(Q - \sum_{j'=j_{\min}}^{j^*-1} p_{j'} \right), & j = j^*, \\ 0, & j > j^*. \end{cases} \tag{31}$$

In a more physical language, the optimal strategy consists actually of two successive measurements. The experimentalist first makes a weak measurement to find the total angular momentum j of the input state $\tau(\vec{n})$ and decides to abstain (provide a guess) if $j < j^*$ ($j > j^*$). If $j = j^*$, she simply decides randomly, by tossing a Bernoulli coin with probability a_{j^*} of coming up heads, and if heads (tails) show up, abstain (provide a guess). In order to provide the actual guess, if she decides to do so, she makes the optimal POVM measurement $\{\Pi(\vec{s})\}$ (or just $\{\Pi_j(\vec{s})\}$) in equation (13) on the state $\tau_j(\vec{n})$ that resulted from the first measurement.

3.2. A small number of copies

In figure 1 we plot the fidelity gain due to abstention $[F(Q) - F(0)]/F(0) = \Delta F/F(0)$ versus Q for $N = 6, 8$, and purities of $r = 0.3$ and 0.7. The structure of equation (31) is apparent from these plots: at $Q = 0$ ($a_j = 0$ for all j) there is, naturally, no gain; kinks sequentially appear at the precise values of Q where a new coefficient a_j in (31) becomes positive (and j^* increases by one); the curves are convex between successive kinks, where the one a_j that has become positive, a_{j^*}, keeps increasing. This pattern repeats until the abstention rate Q reaches a critical value Q_{crit} at which $j^* = J$,

$$Q_{\text{crit}} = 1 - p_J = 1 - \frac{1}{r}\left(\frac{1+r}{2}\right)^{2J+1} \frac{1}{r}\left(\frac{1-r}{2}\right)^{2J+1} \tag{32}$$

(see equation (12)). Increasing Q further will not provide any additional gain, as the flat plateaus of figure 1 illustrate. This is so, since one can view the optimal abstention protocol as a filtering process where the low angular momentum components of the input state are filtered out. Hence, keeping the maximum value of $j = J$ is the optimal filtering beyond which no further improvement is possible. Figure 1(a) shows that in noisy scenarios, e.g. $r = 0.3$, abstention can
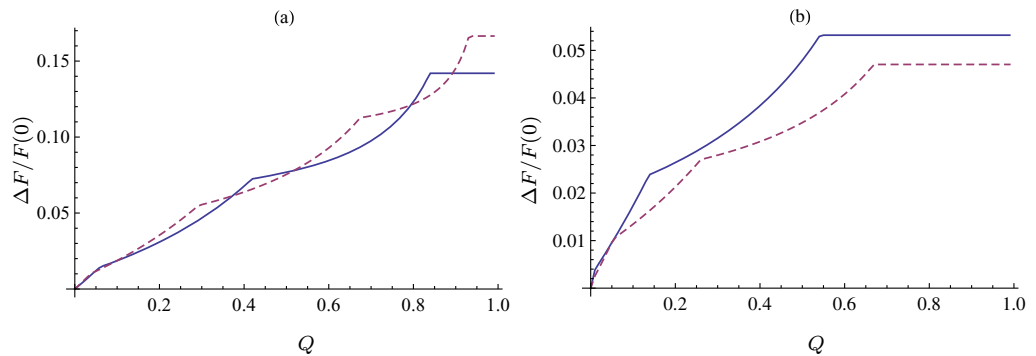
Figure 1. Fidelity gain $\Delta F/F(0) = [F(Q) - F(0)]/F(0)$ as a function of Q for $N = 6$ (solid line) and $N = 8$ (dashed line) and purities of $r = 0.3$ in (a) and $r = 0.7$ in (b).
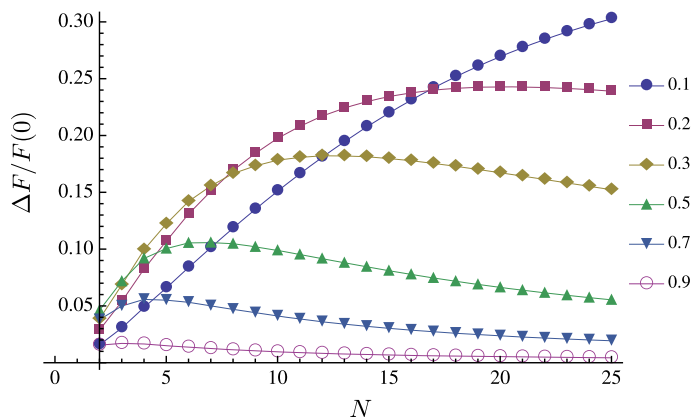

Figure 2. Fidelity gain as a function of N for various values of r, indicated in the legend, and for $Q \geqslant Q_{\text{crit}}$.

increase the fidelity quite notably, up to 15%. For higher purities the gain is more moderate, as shown in figure 1(b). The enhancement in this case is about 4–5% but with an abstention rate slightly above 50%. Further results are shown in figure 2, where we plot the fidelity gain as a function of the number of copies N for abstention rates larger than Q_{crit}, and for various values of the purity r. All the curves have a maximum at a value of N that varies with the purity. The lower the purity, the higher the value of N at which the maximum occurs (e.g. for $r = 0.3$ the maximum gain occurs at $N = 12$; for $r = 0.1$ the maximum is off scale at the right of the figure).

As we have seen, the possibility of abstaining enables us to reach values of the fidelity that otherwise we could only attain with lower levels of noise. To quantify this effective reduction of noise, let us define an effective purity r_{eff} by the implicit equation $F(r_{\text{eff}}, N, 0) = F(r, N, Q)$. That is, for an estimation setting, given by r, N and Q, r_{eff} is the purity of the input states that would provide the same fidelity if the standard strategy without abstention ($Q = 0$) were used instead. Since r is related to the probability of error η in our model of noisy measurements in (1), an increase of the effective purity corresponds to an effective reduction of the amount of noise in the measurement through the relation $\eta_{\text{eff}} = (3/4)(1 - r_{\text{eff}})$. Figure 3 shows a plot of the
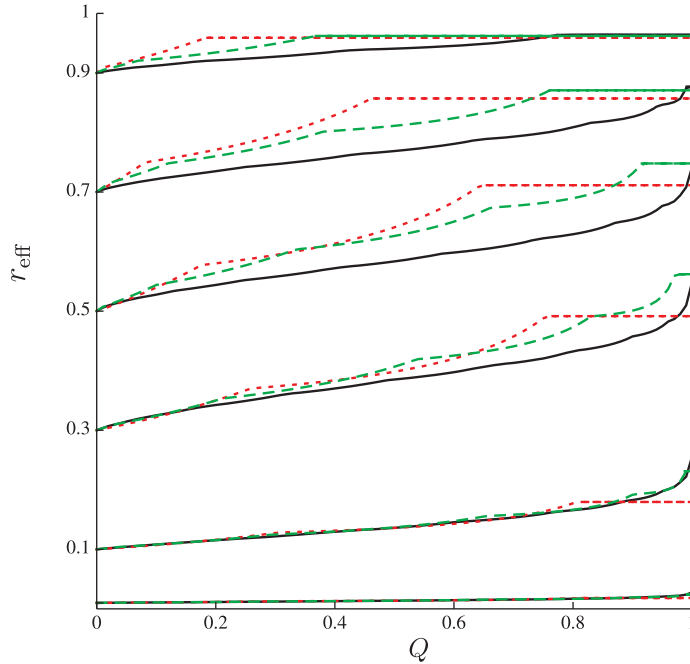
Figure 3. Effective purity r_{eff} as a function of the abstention rate Q for $N = 5$ (dotted), 10 (dashed) and 30 (solid), and for purities of $r = 0.01$, 0.1, 0.3, 0.5, 0.7 and 0.9, which can be read off from the values of r_{eff} at $Q = 0$.

effective purity r_{eff} as a function of Q for various values of r and N. As can be seen, r_{eff} increases faster at low values of N, but it saturates earlier (lower Q_{crit}), reaching a lower value. For low N and for a wide range of purities, $0.1 \lesssim r \lesssim 0.9$, we observe a constant effective increase of the purity, $r_{\text{eff}} \approx r + 0.2$, for reasonable values of the abstention rate Q. As N increases one has to go to higher values of the abstention rate, $Q \sim Q_{\text{crit}}$, to have a significant gain. Hence, a moderate abstention rate is most effective in noisy scenarios when a small, but fair, number of copies are available.

Finally, let us point out that the protocol we have presented requires a projection on the total angular momentum eigenspaces. This is a non-local measurement that nonetheless can be implemented efficiently [31]. In a more extreme scenario where there are no restrictions on the abstention rate, one can attain the maximum fidelity with an even simpler strategy: make a local Stern–Gerlach measurement on every qubit (say, of the z-component of the spin) and abstain unless all outcomes agree. This strategy renders an abstention probability of $Q = 1 - [(1 + r)/2]^N$, which might be comparable to Q_{crit} in equation (32).

3.3. The asymptotic regime

We next compute the analytical expressions of the fidelity in the large N limit. Here it is useful to define the variable x as

$$x = \frac{j}{J}, \quad 0 \leqslant x \leqslant 1, \tag{33}$$

which becomes continuous in the limit $N \to \infty$ ($J \to \infty$). In this case, we can replace p_j by the continuous probability distribution in [0, 1] defined by

$$p(x) = J p_{j=xJ}, \tag{34}$$

so that $\int_0^1 \mathrm{d}x\, p(x) = 1$ as N goes to infinity. Equation (28) can then be approximated by its continuous version, which reads

$$F = \frac{1}{2}\left[1 + \int_0^1 \mathrm{d}x\, p(x)\tilde{\Delta}(x)\right], \tag{35}$$

where

$$\tilde{\Delta}(x) = \tilde{\Delta}_{j=xJ}, \tag{36}$$

where recall that $\tilde{\Delta}_j$ is given in equation (29). From equation (31) we see that asymptotically $\bar{a}_j$ becomes the step function $\theta(x - x^*)$, where $x^* = j^*/J$, and we have used the standard definition

$$\theta(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0. \end{cases} \tag{37}$$

With this, equation (29) becomes

$$\tilde{\Delta}(x) = \frac{\theta(x - x^*)}{\bar{Q}}\Delta(x), \tag{38}$$

and, in turn,

$$F = \frac{1}{2}\left[1 + \frac{1}{\bar{Q}}\int_{x^*}^1 \mathrm{d}x\, p(x)\Delta(x)\right]. \tag{39}$$

It also follows from (27) that

$$\bar{Q} = \int_0^1 \mathrm{d}x\, p(x)\theta(x - x^*) = \int_{x^*}^1 \mathrm{d}x\, p(x). \tag{40}$$

At this point, we need to find a good approximation to $p(x)$ that would enable us to obtain the explicit form for the asymptotic fidelity. From equation (12), and using the Stirling formula, we obtain

$$p(x) \simeq \sqrt{\frac{N}{2\pi}}\frac{1}{\sqrt{1-x^2}}\frac{x(1+r)}{r(1+x)}\,\mathrm{e}^{-NH\left(\frac{1+x}{2}\|\frac{1+r}{2}\right)}, \tag{41}$$

where $H(s \| t)$ is the (binary) relative entropy

$$H(s \| t) = s\log\frac{s}{t} + (1-s)\log\frac{1-s}{1-t}, \tag{42}$$

and the approximation is valid for both x and r in the open unit interval (0, 1). The appearance of a relative entropy[3] in equation (41) can be understood as follows. Our N-copy input state (diagonal in the canonical J_n basis) can be thought of as a classical coin tossing distribution of N identical coins with a bias of $(1+r)/2$. From the theory of types [30], it is well known that the probability to obtain k heads is given by the Kulback–Leibler distance (or relative entropy) between the empirical distribution $\{f = k/N, 1 - f\}$ and the distribution $\{(1+r)/2, (1-r)/2\}$. That is, $p(k) \sim \exp\{-NH[f \| (1+r)/2]\}$ to first order in the exponent.

---

[3] It has also been noticed in the more general setting [28] and known as 'the estimation theorem' [29].
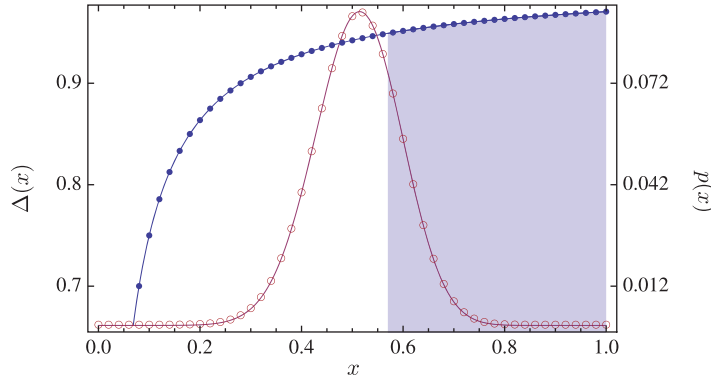
Figure 4. Plots of $\Delta(x)$ (blue line with solid circles) and $p(x)$ (red line with empty circles) for $N = 100$ and $r = 0.5$. The circles represent the quantities Δ_j and Jp_j as a function of $x = j/J$. The shaded area indicates the acceptance region for an abstention rate $Q \sim 93\%$.

The number of heads k is in one-to-one correspondence with the magnetic quantum number, $m = k - J$, and the conditioned probability $p(j|m)$ is strongly peaked at $m = j$, as one can easily check. It follows that the probability that the input state has total angular momentum j, given by $p(j) = \sum_m p(j|m)p(m)$, will be asymptotically determined by the probability distribution $p(m)$, which has a convenient expression in terms of the typical and the empirical distribution of up/down outcomes.

From equation (41) it follows that $p(x)$ is peaked at the value $x = r$, i.e. at $j = rJ$, as shown in figure 4 and stated without a proof in section 2. Actually, around the peak, $x \sim r$, the exponent becomes quadratic and $p(x)$ approaches the Gaussian distribution

$$p(x) \simeq \sqrt{\frac{N}{2\pi(1 - r^2)}}\, \mathrm{e}^{-N \frac{(x-r)^2}{2(1-r^2)}}, \tag{43}$$

as also follows from the central limit theorem, whereas it falls off exponentially elsewhere.

It is now apparent that, asymptotically, abstention has negligible impact if components with j below rJ are filtered out ($x^* < r$), since the main contribution to the fidelity, which comes from the peak around $x \simeq r$, is not excluded from the integral in equation (39) (only the left exponentially decaying tail is). For the same reason (see equation (40)), $\bar{Q} \simeq 1$ (the abstention rate Q is exponentially small), and equation (39) yields

$$F = 1 - \frac{1}{2N}\frac{r+1}{r^2} + \cdots \quad \text{for} \quad x^* < r, \tag{44}$$

which is the same expression as the asymptotic fidelity of the protocol without abstention, equation (18).

It is then clear that, in order to have a discernible improvement in the fidelity, the abstention threshold x^* must lie on the right of the peak of the probability distribution. The fidelity in (39) then can be written as

$$F \simeq \frac{1}{2}\left[1 + \frac{p(x^*)}{\bar{Q}}\Delta(x^*)\right] \simeq \frac{1}{2}\left[1 + \Delta(x^*)\right], \quad x^* > r, \tag{45}$$

where we have used that for $x \geqslant x^* > r$ and for large enough N, $p(x)$ falls off exponentially and the integral can be approximated by the value of the integrand at its lower limit. By the very same argument equation (40) gives

$$\bar{Q} \simeq p(x^*),\tag{46}$$

which has also been used in (45). Using now (41) we obtain that in the asymptotic limit of many copies, the rate at which our protocol provides a guess is

$$\bar{Q} \sim \exp\left[-NH\left(\frac{1+x^*}{2}\middle\|\frac{1+r}{2}\right)\right].\tag{47}$$

Recalling equations (16) and (17) we obtain the optimal fidelity

$$F = 1 - \frac{1}{2Nx^*}\frac{r+1}{r} + \cdots, \quad \text{for } r \leqslant x^* \leqslant 1,\tag{48}$$

for a value of Q given by (47). For $x^* = r$ the results (18) and (44) are recovered, whereas for $x^* \to 1$ ($Q \geqslant Q_{\text{crit}}$) the maximum average fidelity is attained:

$$F_{\max} = 1 - \frac{1}{2N}\frac{r+1}{r} + \cdots.\tag{49}$$

The advantage provided by our estimation with the abstention protocol can be quantified by the effective number of copies that the standard protocol without abstention would require to achieve the same fidelity: $N_{\text{eff}} = (x^*/r)N$, where $x^* \in [r, 1)$ is determined by the abstention rate Q through (47). For high noise levels (low purity, $r \ll 1$) our protocol provides an important saving of resources/copies, as $N_{\text{eff}}/N = 1/r \gg 1$, whereas for nearly ideal detectors the saving in this asymptotic regime is more modest.

Alternatively, the advantage discussed above can also be quantified by the effective measurement-noise reduction, or equivalently, the effective purity r_{eff} (see section 3.2). Using (48) one can easily find a simple expression for the effective purity in the asymptotic limit and for large abstention rate: $r_{\text{eff}} = (r + \sqrt{4r + 5r^2})/[2(1+r)]$. In the limit of very low noise levels the errors probability η (recall equation (1)) is effectively reduced by a factor of three, i.e. $\eta_{\text{eff}} = \eta/3$, while in the opposite limit of very noisy measurements, one finds that $r_{\text{eff}} = \sqrt{r}$.

3.4. Other regimes

In the previous section we have seen how a gain in fidelity can be obtained provided the 'acceptance' rate $\bar{Q}$ falls off exponentially as N becomes very large. Here we give an example where this gain takes place even at finite $\bar{Q}$.

At a fixed noise level (purity r), the fidelity is an increasing function of N. However, one could imagine an experimental setup where the noise (purity) also increases (decreases) with N. If this is so, the asymptotic fidelity could be strictly less than one, or in other words, perfect estimation could be unattainable even with unbounded resources. This is the case in our example, were we assume that $r = a/\sqrt{N}$, a being a positive constant. Note that the threshold x^* must also scale as $1/\sqrt{N}$ in order to have a reasonably low abstention rate. Therefore, it is convenient to use a new variable $\xi = \sqrt{N}x = \sqrt{N}\,j/J = 2j/\sqrt{N}$ instead. Then, the probability distribution in this new variable is

$$p(\xi) = \frac{\sqrt{N}}{2}p_{j=\xi\sqrt{N}/2}, \quad \text{with } r = \frac{a}{\sqrt{N}}.\tag{50}$$

Recalling equation (12) and using the Stirling formula this equation gives

$$p(\xi) = \frac{e^{-\left(\frac{\xi-a}{\sqrt{2}}\right)^2} - e^{-\left(\frac{\xi+a}{\sqrt{2}}\right)^2}}{\sqrt{2\pi}\,a}\xi \tag{51}$$

to leading order in inverse powers of N. The subleading terms are of order $N^{-1/2}$ and will be neglected here. For a given threshold value $\xi^* = 2j^*/\sqrt{N}$ the abstention rate is

$$Q = \int_0^{\xi^*} p(\xi)\,d\xi = \frac{1}{2}\left(\operatorname{erf}\xi_+^* + \operatorname{erf}\xi_-^*\right) - \frac{e^{-\xi_-^{*2}} - e^{-\xi_+^{*2}}}{\sqrt{2\pi}\,a}, \tag{52}$$

where $\xi_\pm^* = (\xi^* \pm a)/\sqrt{2}$ and $\operatorname{erf} x$ is the error function.

From equations (16) and (17) we have in this same regime and at leading order

$$\Delta(\xi) = \Delta_{j=\xi\sqrt{N}/2} = 1 - \frac{2}{1-e^{2a\xi}} - \frac{1}{a\xi}. \tag{53}$$

With the above, the fidelity (28) (or rather, the counterpart of (35)) is

$$F = \frac{1}{2}\left[1 + \int_0^\infty d\xi\, p(\xi)\tilde{\Delta}(\xi)\right] = \frac{1}{2}\left[1 + \frac{1}{Q}\int_{\xi^*}^\infty d\xi\, p(\xi)\Delta(\xi)\right], \tag{54}$$

where the last integral can be computed to be

$$\Delta^* \equiv \int_{\xi^*}^\infty \Delta(\xi)p(\xi)\,d\xi \tag{55}$$

$$= \frac{1-a^2}{2a^2}\left(\operatorname{erf}\xi_-^* - \operatorname{erf}\xi_+^*\right) + \frac{e^{-\xi_-^{*2}} + e^{-\xi_+^{*2}}}{\sqrt{2\pi}\,a}.$$

We can finally write the fidelity as

$$F = \frac{1}{2}\left(1 + \frac{\Delta^*}{1-Q}\right) + \mathcal{O}(N^{-1/2}). \tag{56}$$

As shown in (52) and (55), both Q and Δ^* are functions of the filtering threshold ξ^*, which is just a properly scaled version of the original threshold j^*. Finding the maximum fidelity for a given rate of abstention Q requires inverting equation (52) to obtain $\xi^*(Q)$, but this cannot be done analytically and one has to resort to numerical methods.

In figure 5 we plot F as a function of Q for $a = 1$. The increase of the fidelity in the asymptotic regime of large N is clearly seen: e.g., an abstention rate of 50% yields a rise of about 10%, and it goes up to about 30% for higher (but still reasonable) values of Q. The figure also shows the agreement between the approximate form of the fidelity given by equations (52)–(56) and the numerical evaluation of its exact expression in (28).

It should be noted that in the regime described here a rise of the input size N fails to replicate the fidelity improvement that results from increasing the rate of abstention (no N_{eff} can be defined in this regime); thus abstention appears to be the only means by which one can improve estimation.
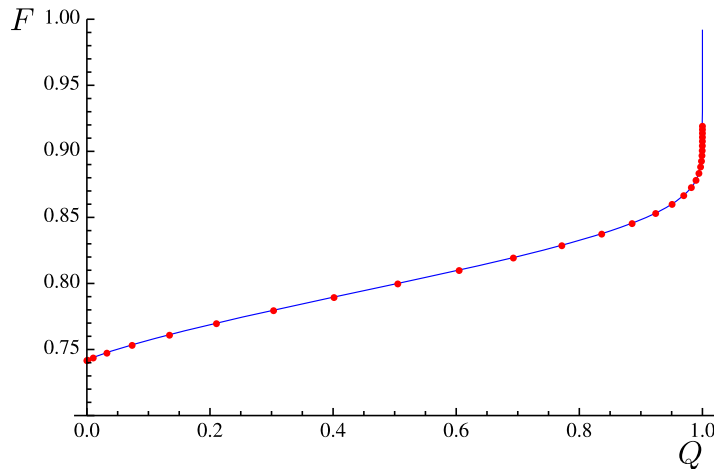
Figure 5. Plot of the fidelity as a function of Q for $r = a/\sqrt{N}$, with the choice $a = 1.0$, $N = 10^6$ (red circles). The solid line (in blue) is the leading term in equation (56) plotted as a function of Q (a parametric plot of the pairs (Q, F), as given by equations (52) and (56)).

4. Conclusions

In this work we have addressed optimal estimation of pure qubit states when abstention from providing an outcome is allowed. We have considered a reasonably realistic multiple copy scenario, where a sample of N identically prepared systems go through a non-ideal (noisy) process of measurement. We have shown that in the limit of zero noise, abstention does not help to improve estimation (it does not hamper it either). However, abstention turns out to counterbalance the adverse effect of errors in a noisy process of measurement. We have shown that, in general, abstention is most useful for inputs of a few copies and for error rates of the order of a few per cent. For example, for $N = 6$ and a value of the error probability of $\eta = 0.5$ (per qubit), one can easily attain fidelity gains of the order of 15% with an abstention rate of $Q = 4/5$. As N increases, one needs to allow for higher abstention rates to obtain a significant improvement. We have given analytical asymptotic expressions of the fidelity valid in the limit of a large number of copies. In this limit, abstention can have the effect of increasing the number of copies by a constant fraction: $N_{\mathrm{eff}}/N = x^*/r$ ($x^* > r$), with an acceptance rate $\bar{Q}$ given by the relative entropy: $-(1/N) \log \bar{Q} = H[(1+x^*)/2 \,\|\, (1+r)/2]$. For low levels of noise, this amounts to reducing the error probability η by a factor of up to three.

We have also considered a scenario where the noise (per qubit) increases with the number of copies in such a way that perfect estimation is unattainable ($\lim_{N\to\infty} F < 1$). In this case one can obtain a significant enhancement of the asymptotic fidelity (a few per cent) even for finite abstention probabilities $Q < 1$. Moreover, in such a scenario abstention appears to be the only way to improve estimation.

In broader parameter estimation contexts, where, e.g., phase or direction information is encoded in more general many-particle states, abstention may have a much more dramatic effect. These issues will be analyzed in a separate publication [24].

IOP Institute of Physics ⏀ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Acknowledgments

References

[1] Holevo A S 1982 Probabilistic and statistical aspects of quantum theory *North-Holland Series in Statistics and Probability* (Amsterdam: North-Holland)
[2] Paris M and Řeháček J (ed) 2004 *Quantum State Estimation* (*Lecture Notes in Physics* vol 649) (Berlin: Springer)
[3] Massar S and Popescu S 1995 *Phys. Rev. Lett.* **74** 1259
[4] Hradil Z 1997 *Phys. Rev.* A **55** 1561
[5] Derka R, Bužek V and Ekert A K 1998 *Phys. Rev. Lett.* **80** 1571
[6] Bruss D and Macchiavello C 1999 *Phys. Lett.* A **253** 249
[7] Fischer D G, Kienle S-H and Freyberger M 2000 *Phys. Rev.* A **61** 032306
[8] Bagan E, Baig M and Munoz-Tapia R 2002 *Phys. Rev. Lett.* **89** 277904
[9] Bagan E, Ballester M A, Gill R D, Monras A and Munoz-Tapia R 2006 *Phys. Rev.* A **73** 032301
[10] Blume-Kohout R 2010 *New. J. Phys.* **12** 043034
[11] Adamson R B A and Steinberg A M 2010 *Phys. Rev. Lett.* **105** 030406
[12] Gross D, Liu Y K, Flammia S T, Becker S and Eisert J 2010 *Phys. Rev. Lett.* **105** 150401
[13] Bergou J A 2010 *J. Mod. Opt.* **57** 160
[14] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[15] Ivanovic I D 1987 *Phys. Lett.* A **123** 257
Dieks D 1988 *Phys. Lett.* A **126** 303
Peres A 1988 *Phys. Lett.* A **128** 19
Jaeger G and Shimony A 1995 *Phys. Lett.* A **197** 83
[16] Chefles A and Barnett S M 1998 *J. Mod. Opt.* **45** 1295
[17] Zhang C-W, Li C-F and Guo G-C 1999 *Phys. Lett.* A **261** 25
[18] Fiurášek J and Ježek M 2003 *Phys. Rev.* A **67** 012321
[19] Eldar Y C 2003 *Phys. Rev.* A **67** 042309
[20] Hayashi A, Hashimoto T and Horibe M 2008 *Phys. Rev.* A **78** 012333
[21] Sugimoto H, Hashimoto T, Horibe M and Hayashi A 2009 *Phys. Rev.* A **80** 052322
[22] Bagan E, Munoz-Tapia R, Olivares-Renteria G A and Bergou J A arXiv:1206.4145
[23] Fiurášek J 2006 *New. J. Phys.* **8** 192
[24] Gendra B, Ronco-Bonvehi E, Calsamiglia J, Munoz-Tapia R and Bagan E arXiv:1209.5736
[25] D'Ariano G M, Macchiavello C and Perinotti P 2005 *Phys. Rev. Lett.* **95** 060503
[26] Fiurášek J 2004 *Phys. Rev.* A **70** 032308
[27] Cirac J I, Ekert A K and Macchiavello C 1999 *Phys. Rev. Lett.* **82** 4344
[28] Keyl M and Werner R F 2001 *Phys. Rev.* A **64** 052311
Heckman G J 1982 *Invent. Math.* **67** 333–56
[29] Christandl M, Harrow A and Mitchison G 2007 *Commun. Math. Phys.* **270** 575–85
[30] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd edn (*Wiley Series in Telecommunications and Signal Processing*) (Hoboken, NJ: Wiley)
[31] Bacon D, Chuang I and Harrow A 2006 *Phys. Rev. Lett.* **97** 170502