New Journal of Physics
The open-access journal for physics

Optimal signal states for quantum detectors

Ognyan Oreshkov[1,2,5], **John Calsamiglia**[1], **Ramon Muñoz-Tapia**[1] **and Emili Bagan**[1,3,4]

[1] Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[2] QuIC, Ecole Polytechnique, CP 165, Université Libre de Bruxelles, 1050 Brussels, Belgium
[3] Department of Physics, Hunter College of the City University of New York, 695 Park Avenue, New York, NY 10021, USA
[4] HET Group, Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA
E-mail: oreshkov@ulb.ac.be

New Journal of Physics **13** (2011) 073032 (22pp)
Received 4 April 2011
Published 22 July 2011
Online at http://www.njp.org/
doi:10.1088/1367-2630/13/7/073032

Abstract. Quantum detectors provide information about the microscopic properties of quantum systems by establishing correlations between those properties and a set of macroscopically distinct events that we observe. The question of how much information a quantum detector can extract from a system is therefore of fundamental significance. In this paper, we address this question within a precise framework: given a measurement apparatus implementing a specific POVM measurement, what is the optimal performance achievable with it for a specific information readout task and what is the optimal way to encode information in the quantum system in order to achieve this performance? We consider some of the most common information transmission tasks—the Bayes cost problem, unambiguous message discrimination and the maximal mutual information. We provide general solutions to the Bayesian and unambiguous discrimination problems. We also show that the maximal mutual information is equal to the classical capacity of the quantum-to-classical channel describing the measurement, and study its properties in certain special cases. For a group covariant measurement, we show that the problem is equivalent to the problem of accessible information of a group covariant ensemble of states. We give analytical proofs of optimality in some relevant cases. The framework presented

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

here provides a natural way to characterize generalized quantum measurements in terms of their information readout capabilities.

Contents

1. Introduction

Quantum detectors provide the interface between the microscopic world of quantum phenomena and the world of macroscopically distinct events that we observe. A quantum detector is a device that interacts with the system under observation in a way that establishes correlations between certain properties of the system and a set of macroscopically distinct (orthogonal) states of the device. A general quantum detector can be described by a positive operator-valued measure (POVM), i.e. a set of positive operators $\{E_i\}$, $E_i \geqslant 0$, $i = 1, \ldots, M$, summing up to the identity, $\sum_i E_i = I$. For an input state ρ, the probability that the measurement yields outcome j is given by the Born rule, $p_j(\rho) = \mathrm{Tr}\{\rho E_j\}$.

A natural question is: To what extent is a given quantum detector able to provide information about the system it is used to observe? This question can be conveniently formulated in the context of a quantum communication scenario, where a sender (Alice) tries to send messages to a receiver (Bob) who is constrained to read those messages using the quantum detector in question. Concretely, let the source of classical information that Alice wants to communicate to Bob be characterized by a probability distribution $\pi_i > 0$, $i = 1, \ldots, N$, $\sum_i \pi_i = 1$, which specifies the probability of each classical message i. Alice encodes the different messages into quantum states via an encoding map $i \to \rho_i$, and Bob reads the information by carrying out the POVM measurement. If there are no constraints on the way Alice can prepare the signal states and these states can reach Bob undisturbed (i.e. Alice and Bob are connected through a noiseless channel), then the optimal performance they can achieve for a given task can be regarded as quantifying the readout capabilities of the measurement with respect to that task. In this respect, a problem of primary importance is to find the optimal encoding (or signal states ρ_i) for which the detector achieves its optimal performance.

The problem just outlined bears strong similarities to the problem of quantum state discrimination [1–8], where the encoding of Alice is fixed and Bob's task is to decide which message he has received by optimizing his measurement. In fact, we will see below that the two

problems can be regarded as dual to each other due to the symmetry that exists between the input ensembles and the POVM measurements. This allows one to adopt results from quantum state discrimination to the problem at hand. However, since in quantum state discrimination the space over which we optimize is more constrained due to the completeness relation $\sum_i E_i = I$, it turns out that in many cases the problem of optimal signal states for quantum detectors is easier to solve.

In addition to its application in characterizing detectors, the problem considered here is of natural practical interest for quantum communication, since generating different signal states [9] can be experimentally more accessible than carrying out different measurements. A quantum detector is usually fixed, while a preparation device, although possibly also based on a fixed (but nondestructive!) measurement, can be used together with post-selection, which provides additional flexibility to the preparation process. Furthermore, in the case of communication through a noiseless channel, any operation at the receiver's side prior to the detector can be equally done as part of the preparation strategy.

In this paper, we consider the above problem from the perspective of three different information transmission tasks—the task of optimal Bayes cost message discrimination (of which the well-known problem of minimum error discrimination is a special case), unambiguous message discrimination and the maximal mutual information. Due to the simplification mentioned above, we are able to provide solutions to the Bayesian and unambiguous discrimination problems in the general case. For the maximal mutual information, we show that this quantity is equal to the classical capacity of the quantum-to-classical channel corresponding to the measurement, which we term the '*capacity of the measurement*'. This quantity provides a general figure of merit for the information readout capabilities of a detector. Based on its relation to the accessible information [6], we prove a result similar to Davies's theorem [2] (proposition 2), which shows that the optimal ensemble can be chosen to consist of d^2 pure states, where d is the dimension of the system. For a group covariant measurement, we find that the problem is equivalent to that of accessible information of a group covariant ensemble of states. We apply our results to the case of a noisy two-level symmetric informationally complete measurement, for whose capacity we give analytical proofs of optimality.

2. The Bayes cost problem

In the Bayes cost problem, one is interested in minimizing an average cost function of the form

$$C(P) = \sum_{ij} C_{ij} P_{ij}, \tag{1}$$

where $P_{ij} = \mathrm{Tr}(\pi_i \rho_i E_j)$ are the joint probabilities for input i and measurement outcome j, and $C_{ij} \geqslant 0$ are the elements of the cost matrix (C_{ij} is the cost of choosing hypothesis j when hypothesis i is true). In what we will refer to as the *straight* version of this problem, one assumes that the encoding $i \to \rho_i$ is given, and the task is to find the measurement $\{E_j\}$ that minimizes the quantity in equation (1) [1]. An example of a Bayes cost problem is that of *minimum error discrimination*, i.e. minimizing the probability for incorrectly identifying the message. In this case, the probability for an error is given by $p_{\mathrm{err}} = \sum_{i \neq j} P_{ij}$, i.e. the elements of the cost matrix are $C_{ij} = 1 - \delta_{ij}$.

Here, we are concerned with the opposite scenario, which we will refer to as the *reverse* problem: we assume that the receiver has an apparatus that implements a particular POVM

measurement, and we ask what is the optimal way to encode the classical messages into quantum states so that, using only the given POVM measurement and possibly some side information processing, the receiver will identify the message at the lowest cost[6]. This side information processing involves finding the optimal way of choosing hypothesis i when the measurement outcome k takes place and includes the possibility of following a mixed strategy, i.e. assigning a hypothesis i randomly according to some prescribed probability distribution, which might of course depend on the outcome k. In other words, the receiver can use the given POVM $\{E_k\}$ to obtain new POVM measurements with elements of the form

$$\tilde{E}_j = \sum_k p(j|k)E_k, \quad \sum_j p(j|k) = 1 \quad \text{for all } k, \tag{2}$$

where $0 \leqslant p(j|k) \leqslant 1$ are conditional probabilities.

Up to a renormalization of the cost matrix, we can assume that $0 \leqslant C_{ij} \leqslant 1$. Hence, the problem is equivalent to that of maximizing the quantity

$$B(P) \equiv 1 - C(P) = \sum_{ij}(1 - C_{ij})P_{ij} \equiv \sum_{ij} B_{ij}P_{ij}, \tag{3}$$

where

$$0 \leqslant B_{ij} \leqslant 1, \quad \forall i, j. \tag{4}$$

For a given POVM measurement $\{E_i\}$, consider some encoding and decoding strategies given by the map $i \to \rho_i$ and the conditional probability distribution $p(j|k)$, respectively. For these strategies, the quantity $B(P)$ reads

$$B(P) = \sum_{ijk} B_{ij}\pi_i p(j|k) \operatorname{Tr}(\rho_i E_k). \tag{5}$$

Define $j^*(k)$ to be a value of j for which the quantity $\sum_i B_{ij}\pi_i \operatorname{Tr}(\rho_i E_k)$, for a fixed k, is maximal (if there are two or more such values, we can pick any one of them). Then,

$$B(P) \leqslant \sum_{ik} B_{ij^*(k)}\pi_i \operatorname{Tr}(\rho_i E_k), \tag{6}$$

which is achievable by choosing $p(j|k) = \delta_{jj^*(k)}$.

We see that for the purpose of achieving the maximum in equation (3), the receiver does not need a mixed strategy, i.e. the maximum can be achieved by choosing all conditional probabilities $p(j|k)$ to be either 0 or 1. This means that the receiver can associate more than one measurement outcome E_k with the same hypothesis j, but it does not help to associate two or more hypotheses with the same outcome. Note that this means, in particular, that in the case when the number of possible messages N is greater than the number M of different outcomes of the POVM, the best strategy is not to attempt to detect certain messages at all. In fact (see below), even when $M \leqslant N$, it may be advantageous to group different POVM elements for the detection of a single state.

Let K_j denote the set of those indices k for which $j^*(k) = j$, i.e. the indices k for which the outcomes E_k are associated with hypothesis j. Note that the sets K_j are non-intersecting as

[6] Obviously, in the Bayesian framework (straight or reverse problem) it makes no sense to optimize over priors $\{\pi_i\}$ since this renders the optimization problem trivial ($\pi_i = \delta_{ij}$ for some j). This type of figure of merit has an explicit dependence on the source. The same holds for the unambiguous discrimination problem studied in the next section. We consider a source-independent characterization in section 4.

shown above and that some sets may be empty. In other words, the set of possible assignments corresponds to that of all possible ways to distribute M elements into N groups $\{K_j^\alpha\}_{j=1}^N$, where the index α labels each of the N^M distributions. Then for any such choice we have

$$B_\alpha(P) = \max_{\{\rho_i\}} \sum_i \pi_i \operatorname{Tr}\left(\rho_i \sum_j B_{ij} \tilde{E}_j^\alpha\right), \tag{7}$$

where $\tilde{E}_j^\alpha = \sum_{k\in K_j^\alpha} E_k$. The maximum of this quantity is achieved when each of the signal states ρ_i is chosen to be an eigenstate corresponding to the maximal eigenvalue of the operator $\sum_j B_{ij}\tilde{E}_j^\alpha$, which we will denote by $\lambda^{\max}(\sum_j B_{ij}\tilde{E}_j^\alpha)$. Hence, we can write

$$B(P) = \max_\alpha \boldsymbol{\pi}\cdot\mathbf{s}_\alpha, \tag{8}$$

where we have defined the vectors $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_N\}$ and

$$\mathbf{s}_\alpha = \left\{\lambda^{\max}\left(\sum_j B_{1j}\tilde{E}_j^\alpha\right), \ldots, \lambda^{\max}\left(\sum_j B_{Nj}\tilde{E}_j^\alpha\right)\right\}.$$

We thus see that the problem reduces to that of finding the sets K_j for which the quantity in equation (8) is maximal. The corresponding partition specifies which outcomes k of the POVM measurement the receiver has to associate with a given classical message j. The optimal encoding strategy is to encode each classical message i into an eigenstate $\rho_i^{\max}$ corresponding to the maximal eigenvalue of $\sum_j B_{ij}\tilde{E}_j^\alpha$ (note that these states can always be chosen to be pure).

In general, the optimal grouping α^* of POVM elements, $\alpha^* = \arg\max_\alpha \boldsymbol{\pi}\cdot\mathbf{s}_\alpha$, will depend on the given priors $\boldsymbol{\pi}$. The region in the corresponding simplex where one particular grouping is optimal defines a polytope or, more precisely, a convex polytope when restricted to the region $\pi_1 \geqslant \pi_2 \geqslant \cdots \geqslant \pi_N$ (throughout the paper this ordering of prior probabilities will be always assumed), i.e. if $\boldsymbol{\pi}\cdot(\mathbf{s}_{\alpha^*} - \mathbf{s}_\alpha) \geqslant 0$ and $\boldsymbol{\pi}'\cdot(\mathbf{s}_{\alpha^*} - \mathbf{s}_\alpha) \geqslant 0$, then for $0 \leqslant p \leqslant 1$ one has $[p\boldsymbol{\pi} + (1-p)\boldsymbol{\pi}']\cdot(\mathbf{s}_{\alpha^*} - \mathbf{s}_\alpha) \geqslant 0$.

The described optimization procedure involves calculating and comparing a finite set of quantities. In contrast, the straight version of the problem in the general case is a linear program that requires maximization over a continuous set. Even though the task of finding the optimal encoding for a given decoding POVM exhibits an apparent similarity with the problem of finding the optimal POVM for a given encoding (see the symmetry of the cost function (1) with respect to interchanging the POVM elements and the input states), an important difference between the straight and reverse problems is that the quantities over which we maximize in the straight version have to satisfy the constraint $\sum_j E_j = I$, whereas in the reverse case there is no constraint on the signal states ρ_i.

Observe that in the case when $N < M$, the above optimal strategy requires at least one of the messages to be associated with multiple measurement outcomes. However, as mentioned earlier, even in the case when $N \geqslant M$, it may be advantageous to associate more than one outcome of the POVM with the same state. For example, in the problem of minimum error discrimination, two POVM elements may have very similar (or even identical) maximal eigenvalues and corresponding maximal eigenstates, but all prior probabilities of the different input messages may differ significantly. Then it is not difficult to see (see examples in the last section) that associating the two measurement outcomes in question with two different messages

would be worse than associating both of them with one of the messages—the one that has a higher prior probability.

Note that the special case of minimum error discrimination with a given POVM has been studied in [11] as part of the problem of optimal encoding of classical information in a quantum system for minimal error discrimination when both the encoding and the measurement can be optimized. However, the solution provided in [11] for a fixed POVM is not truly optimal since it assumes that different outcomes must be associated with different states.

We remark that in certain cases it may be possible to simplify the general procedure described above. For example, in the problem of minimum error discrimination, when the prior distribution is flat, $\pi_i = 1/N$, $i = 1, \ldots, N$ and $M \leqslant N$, all we need to do is encode M of the N different possible messages into the eigenstates corresponding to the maximal eigenvalues of the different POVM elements. In this case, associating multiple measurement outcomes with the same message does not help since $(1/N)\lambda^{\max}(E_j + E_k) \leqslant (1/N)\lambda^{\max}(E_j) + (1/N)\lambda^{\max}(E_k)$.

For a binary source, the minimum error probability can be written in a particularly simple form. In this case, the POVM grouping is $\{\tilde{E}^\alpha, I - \tilde{E}^\alpha\}$. We start discussing the unbiased case (i.e. $\pi_1 = \pi_2 = 1/2$) for which

$$p_s^\alpha = \max_{\{\rho_1,\rho_2\}} \tfrac{1}{2}[\mathrm{Tr}\,\tilde{E}^\alpha\rho_1 + \mathrm{Tr}(I - \tilde{E}^\alpha)\rho_2] = \tfrac{1}{2}[1 + \max_{\{\rho_1,\rho_2\}} \mathrm{Tr}\,\tilde{E}^\alpha(\rho_1 - \rho_2)]. \tag{9}$$

The maximum occurs when ρ_1 and ρ_2 are the states corresponding to the largest and lowest eigenvalues of $\tilde{E}^\alpha$, respectively. The difference between these two values is known as the spread of a matrix, defined for a generic matrix A as $\mathrm{Spr}(A) = \max_{ij}|\lambda_i - \lambda_j|$, where λ_i are the characteristic roots of A [10]. Hence,

$$p_s = \tfrac{1}{2}[1 + \max_\alpha \mathrm{Spr}(\tilde{E}^\alpha)]. \tag{10}$$

Note the resemblance to the well-known Helstrom's state discrimination formula [1], where the trace distance has been replaced by the (semi-norm) spread.

From equation (8), the success probability for arbitrary priors reads

$$p_s = \max_\alpha\{\pi_1\lambda^{\max}(\tilde{E}^\alpha) + \pi_2[1 - \lambda^{\min}(\tilde{E}^\alpha)]\}. \tag{11}$$

It is clear that when one signal is given with a prior probability larger than the success probability attained by a two-outcome POVM $\{E, I - E\}$, it pays to assign all outcomes to the most probable signal. In other words, the measurement does not add information to our prior knowledge and the optimal grouping results in the trivial POVM $\{I, 0\}$. The transition occurs at $\pi_1 = p_s$. More explicitly, the trivial POVM is optimal if

$$\pi_1 \geqslant \frac{1 - \lambda^{\min}(E)}{[1 - \lambda^{\min}(E)] + [1 - \lambda^{\max}(E)]}. \tag{12}$$

Note that if $\lambda^{\max}(E) = 1$, it is always advantageous to carry out the measurement, irrespective of the prior probabilities.

3. Unambiguous message discrimination

Unambiguous quantum state discrimination [3–5, 7, 8] concerns the task of identifying which of a set of possible states one has received so as to ensure zero error whenever a conclusive

answer is given. In general, such conclusive answers cannot always be given, and the problem consists in maximizing the probability with which they occur.

Let $\{E_i\}$ be the POVM the receiver has been provided with and let us allow, as in the previous section, some side information processing that will result in new POVMs, $\tilde{E}_j$ (see equation (2)). For the purpose of unambiguously identifying a given set of messages $i = 1, \ldots, N$, encoded in the quantum states ρ_i, $i = 1, \ldots, N$, these POVMs must consist of $N+1$ elements: $\tilde{E}_1, \ldots, \tilde{E}_N$ representing the conclusive answers and an additional element $\tilde{E}_?$ representing the inconclusive one. It must hold that

$$\mathrm{Tr}(\tilde{E}_i \rho_j) = \lambda_j \delta_{ij}, \quad i, j = 1, \ldots, N, \tag{13}$$

since errors are not allowed in conclusive answers. Any of the elements $\tilde{E}_1, \ldots, \tilde{E}_N, \tilde{E}_?$ can be the zero operator as a special case.

One can readily see that all the conditional probabilities $p(j|k)$ that define $\{\tilde{E}_i\}$ in terms of the original POVM through equation (2) can be taken to be either 0 or 1, as for the Bayes cost problem, i.e. $\{\tilde{E}_i\}$ can be taken to be sums of certain subsets of the original POVM elements. This is so because there is no way one can unambiguously identify two or more messages that have been associated with a given E_i if the corresponding outcome takes place. (If some outcome i occurs with zero probability, we can add E_i to any of the elements $\tilde{E}_1, \ldots, \tilde{E}_N, \tilde{E}_?$, as this would not change the probabilities of the respective outcomes.) Similarly, if E_k is randomly associated with both a given message i (i.e. $0 < p(i|k)$) and the inconclusive answer (i.e. $0 < p(?|k)$), the probability of success would increase with the choice $p(i|k) = 1$.

Thus, for the unambiguous discrimination of N input states ρ_i, $i = 1, \ldots, N$, each occurring with prior probability π_i, consider some grouping of the original POVM elements into $N+1$ elements, $\tilde{E}_1^\alpha, \ldots, \tilde{E}_N^\alpha, \tilde{E}_?^\alpha$, where, as in the previous section, α labels the various grouping possibilities. Condition (13) requires that $\rho_i \in \mathscr{K}_i^\alpha \equiv \cap_{j\neq i}^N \ker \tilde{E}_j^\alpha$ for each i. Conversely, if each ρ_i is chosen to belong to this intersection (assuming that it is non-empty), then unambiguous discrimination would be possible with probability

$$p_s^\alpha = \sum_{i=1}^N \pi_i \,\mathrm{Tr}(\tilde{E}_i^\alpha \rho_i). \tag{14}$$

Let P_i^α denote the projector on $\mathscr{K}_i^\alpha$. Note that this projector can be easily computed because $\tilde{E}_j^\alpha \geqslant 0$ implies that $\mathscr{K}_i^\alpha = \ker(\sum_{j\neq i}^N \tilde{E}_j^\alpha)$. Since $\rho_i = P_i^\alpha \rho_i P_i^\alpha$, equation (14) can be written as

$$p_s^\alpha = \sum_{i=1}^N \pi_i \,\mathrm{Tr}[(P_i^\alpha \tilde{E}_i^\alpha P_i^\alpha)\rho_i], \tag{15}$$

and we can maximize each of the traces by choosing ρ_i to be an eigenstate of $P_i^\alpha \tilde{E}_i^\alpha P_i^\alpha$ with maximal eigenvalue. Let us denote this eigenvalue by $\lambda^{\max}(P_i^\alpha \tilde{E}_i^\alpha P_i^\alpha)$. Then, we have

$$p_s^\alpha = \sum_{i=1}^N \pi_i \lambda^{\max}(P_i^\alpha \tilde{E}_i^\alpha P_i^\alpha) = \boldsymbol{\pi} \cdot \mathbf{s}'_\alpha, \tag{16}$$

where, as before, $\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_N\}$, and $\mathbf{s}'_\alpha = \{\lambda^{\max}(P_1^\alpha \tilde{E}_1^\alpha P_1^\alpha), \ldots, \lambda^{\max}(P_N^\alpha \tilde{E}_N^\alpha P_N^\alpha)\}$ in decreasing order of value (this, actually, defines the labeling of the POVM elements $\tilde{E}_j^\alpha$). Note that this ordering ensures maximization of the overlap $\boldsymbol{\pi} \cdot \mathbf{s}'_\alpha$. The probability of success of the

optimal message discrimination protocol is

$$p_s = \max_{\alpha} p_s^{\alpha} = \max_{\alpha} \boldsymbol{\pi} \cdot \mathbf{s}'_{\alpha}. \tag{17}$$

Here, α takes $(N+1)^M/N!$ different values, namely the number of different ways of distributing M POVM elements in $N+1$ sets, where the sum of the elements in each of these sets is $\tilde{E}_1^{\alpha}$, $\ldots$, $\tilde{E}_N^{\alpha}$, $\tilde{E}_?^{\alpha}$, respectively ($N!$ takes into account the specific labeling defined above). Note that certain sets may be empty, i.e. we allow some of the new POVM elements to be the zero operator (the corresponding message will never be identified in these cases).

To compute p_s we may consider the following procedure. Pick a grouping α and construct each of the projectors P_i^{α} on the intersection $\mathscr{K}_i^{\alpha}$ for $i = 1, 2, \ldots, N$. If some $\mathscr{K}_i^{\alpha}$ is empty, terminate the calculation and consider a different grouping α'. If there is an empty intersection for all α, the problem does not have a solution (other than the trivial $\tilde{E}_? = I$), which means that the given POVM $\{E_i\}$ cannot be used to unambiguously discriminate N messages. For each grouping such that $\mathscr{K}_i^{\alpha} \neq \emptyset$, $i = 1, 2, \ldots, N$, compute $\mathbf{s}_{\alpha}$ and pick up the one, α^*, that maximizes (17). Optimal detection is attained with the POVM measurement $\{\tilde{E}_1^{\alpha^*}, \ldots, \tilde{E}_N^{\alpha^*}, \tilde{E}_?^{\alpha^*}\}$ and the optimal encoding of each classical message i is provided by an eigenstate ρ_i of $P_i^{\alpha^*} \tilde{E}_i^{\alpha^*} P_i^{\alpha^*}$ with maximal eigenvalue (note that the states ρ_i can always be chosen to be pure).

The above solution to the reverse unambiguous discrimination problem works for any POVM. In contrast, there is no known solution to the straight version of the same problem for an arbitrary ensemble of mixed input states (see e.g. [8]). As in the case of minimum error discrimination, there are certain similarities between the problem of finding the optimal encoding for a given POVM and that of finding the optimal POVM for a given encoding: for the latter, the POVM $\{\tilde{E}_i\}$ have to be chosen such that $\tilde{E}_i \in \cap_{j\neq i}^N \ker \rho_j$, which resembles the condition $\rho_i \in \cap_{j\neq i}^N \ker \tilde{E}_j$ in the reverse problem. Furthermore, in the two problems, one has to maximize the same quantity, equation (14), where states and POVM elements play essentially the same role (they are interchangeable). Recall, however, that in the straight case optimization has the additional constraint $\sum_i^N \tilde{E}_i \leqslant I$, which makes the problem more difficult.

4. Mutual information

The problems considered in the previous sections characterize the ability of a POVM measurement to perform certain information readout tasks (e.g. minimum error discrimination or unambiguous message discrimination) with respect to a given source of classical messages described by the prior probabilities $\{\pi_i\}$. These results are strongly dependent on the source. For example, if the source consists of only a single message, each of the tasks can be accomplished with unit probability using any measurement. Such a source, however, is trivial as it contains no information. In this section, we consider a source-independent characterization of the ability of a measurement to extract information which is provided by the maximum mutual information that can be established between the sender and the receiver over all possible sources and suitable encodings at the sender's side for the given POVM measurement at the receiver's side.

Consider an information source characterized by the probability distribution $\{\pi_i\}$, $i = 1, \ldots, N$, and an encoding $i \to \rho_i$. The joint probabilities of the input messages and the outcomes of the POVM measurement $\{E_j\}$, $j = 1, \ldots, M$, are

$$P_{ij} = \pi_i \operatorname{Tr}(\rho_i E_j). \tag{18}$$

The mutual information between the input and the output is given by

$$I(P) = \sum_i \eta\left(\sum_j P_{ij}\right) + \sum_j \eta\left(\sum_i P_{ij}\right) - \sum_{ij} \eta(P_{ij}), \tag{19}$$

where $\eta(x) = -x \log x$.

We will be interested in the maximum of $I(P)$ over all possible source distributions $\{\pi_i\}$ and encoding strategies $i \to \rho_i$, that is, over all input ensembles $\{\pi_i, \rho_i\}$,

$$C(\{E_i\}) = \max_{\{\pi_i, \rho_i\}} I(P). \tag{20}$$

Note that, according to the data processing inequality, post-processing of information at the receiver's side cannot increase the mutual information, so in this case it cannot help to group POVM elements (or randomize outcomes).

As shown by the following proposition, $C(\{E_i\})$ has a natural interpretation as the *capacity* of the measurement $\{E_i\}$ which for all practical purposes can be modeled by a quantum channel of the form $\mathcal{E}(\rho) = \sum_j \mathrm{Tr}(\rho E_j)|j\rangle\langle j|$, where $\{|j\rangle\}$ are orthogonal states that carry the classical information about the outcome of the measurement.

Proposition 1. *$C(\{E_i\})$ is equal to the classical capacity of the channel*

$$\mathcal{E}(\rho) = \sum_j \mathrm{Tr}(\rho E_j)|j\rangle\langle j|. \tag{21}$$

Proof. It is known [12, 13] that the classical capacity of a quantum channel $\mathcal{M}$ over independent uses of the channel (i.e. when no entanglement between multiple inputs to the channel is allowed) is given by the quantity

$$\chi(\mathcal{M}) = \max_{\{\pi_i, \rho_i\}} \left\{ S\left[\sum_i \pi_i \mathcal{M}(\rho_i)\right] - \sum_i \pi_i S\left[\mathcal{M}(\rho_i)\right] \right\}, \tag{22}$$

where $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$ denotes the von Neumann entropy of the state ρ. The general capacity of the channel, allowing possibly entangled inputs, is

$$C(\mathcal{M}) = \lim_{n\to\infty} \frac{\chi(\mathcal{M}^{\otimes n})}{n}, \tag{23}$$

where $\mathcal{M}^{\otimes n}$ denotes n uses of the channel. For entanglement-breaking channels [14], such as the quantum-to-classical channel $\mathcal{E}(\rho)$ above, it has been shown that the quantity $\chi(\mathcal{E})$ is additive [15–17], in particular $\chi(\mathcal{E} \otimes \mathcal{E}) = 2\chi(\mathcal{E})$, which implies that

$$C(\mathcal{E}) = \chi(\mathcal{E}). \tag{24}$$

Furthermore, for any input ensemble $\{\pi_i, \rho_i\}$, the channel $\mathcal{E}(\rho)$ outputs an ensemble of commuting quantum states, $\{\pi_i, \mathcal{E}(\rho_i)\}$, and for such an ensemble it is easy to verify that the quantity $S\left[\sum_i \pi_i \mathcal{E}(\rho_i)\right] - \sum_i \pi_i S[\mathcal{E}(\rho_i)]$ is equal to the mutual information in equation (19). The proposition then follows from definitions (20) and (22).

A comment is in order here. The classical capacity of a channel is the maximum rate at which information can be transmitted reliably through the channel in the limit of infinitely many uses. Since the optimal measurement for extracting information from the channel $\mathcal{E}(\rho)$ is a projective measurement in the basis $\{|j\rangle\}$, which preceded by $\mathcal{E}(\rho)$ is equivalent to the POVM measurement $\{E_j\}$, the quantity $C(\{E_j\})$ is equal to the maximum rate at which information can be read reliably using the POVM $\{E_j\}$. □

Corollary 1. *We have*

$$C(\{E_i\}) = \max_{\{\pi_i,\rho_i\}} \left\{ S\left[\sum_i \pi_i \mathcal{E}(\rho_i)\right] - \sum_i \pi_i S[\mathcal{E}(\rho_i)] \right\}. \tag{25}$$

Observe that we can write the joint probability (18) in the symmetric form

$$P_{ij} = \mathrm{Tr}(\breve{\rho}_i E_j), \tag{26}$$

where $\breve{\rho}_i \equiv \pi_i \rho_i$ are *unnormalized* positive operators satisfying $\mathrm{Tr}(\sum_i \breve{\rho}_i) = 1$. (Hereafter we will use the notation $\{\pi_i, \rho_i\}$ and $\{\breve{\rho}_i\}$ interchangeably to denote an ensemble of states.) In this notation, $C(\{E_i\}) = \max_{\{\breve{\rho}_i\}} I(P)$. Note further that the mutual information $I(P)$ is symmetric with respect to the indices i and j. Therefore, the problem we are considering can be regarded as dual to the one of accessible information of an ensemble of states $\{\breve{\rho}_i\}$ [6], which can be written as

$$A(\{\breve{\rho}_i\}) = \max_{\{E_i\}} I(P). \tag{27}$$

Note, however, that the two problems are not identical as the operators $\{E_i\}$ satisfy a stronger constraint than do the operators $\{\breve{\rho}_i\}$: $\sum_i E_i = I$. (A strict duality transformation between signal ensembles and POVM measurements has been established in [18, 19]. We will not be concerned with that correspondence here.)

The above suggests that certain results in the study of the accessible information of an ensemble of states may prove useful for the study of the capacity of a measurement. For example, the symmetry of the problems and the difference in constraints imply

$$C(\{E_i\}) \geqslant A(\{\breve{E}_i\}), \tag{28}$$

where $\breve{E}_i = E_i/d$. Therefore, any known lower bound of A can be used to obtain a lower bound of C. For example, the lower bound obtained in [20] yields

$$C(\{E_i\}) \geqslant Q\left(\sum_i m_i \bar{E}_i\right) - \sum_i m_i Q(\bar{E}_i), \tag{29}$$

where $m_i = \mathrm{Tr}(E_i)/d$, $\bar{E}_i = E_i/(m_i d)$, and $Q(\rho)$ is the *subentropy* of a density matrix ρ, which in terms of the eigenvalues λ_k of ρ reads [20]

$$Q(\rho) = -\sum_k \left(\prod_{l\neq k} \frac{\lambda_k}{\lambda_k - \lambda_l}\right) \lambda_k \log \lambda_k \tag{30}$$

(if two or more eigenvalues are equal, one takes the limit as they become equal).

Similarly, one may wonder if the Holevo quantity $S(\sum_i m_i \bar{E}_i) - \sum_i m_i S(\bar{E}_i)$ [21], which provides a simple upper bound to the accessible information $A(\{\breve{E}_i\})$, could also provide a useful bound for the capacity $C(\{E_i\})$. As we will see below, however, this quantity is neither an upper nor a lower bound to $C(\{E_i\})$.

Proposition 2. *The maximum in equation (20) can be achieved with an ensemble of pure input states $\rho_i = |\psi_i\rangle\langle\psi_i|$. Furthermore, the number N of input states can be made to satisfy $d \leqslant N \leqslant d^2$.*

This proposition is similar to theorem 3 in [2], where it is shown that for a given ensemble of input states, the optimal POVM measurement can be taken to have rank-one POVM elements whose number M satisfies $d \leqslant M \leqslant d^2$.

Proof. As noted in [2], $I(P)$ is a convex function over the convex set of $N \times M$ probability matrices P with fixed row sums. By a similar argument, $I(P)$ is a convex function over the convex set of $N \times M$ probability matrices P with fixed column sums. This implies that if P' is an $(N-1) \times M$ probability matrix obtained from P by replacing two rows by their row sum, then $I(P') \leqslant I(P)$, with equality when the two rows are proportional. Therefore, for any input ensemble $\{\pi_i, \rho_i\}$, where $\rho_i = \sum_k p_{ik}|\psi_{ik}\rangle\langle\psi_{ik}|$, we can consider the pure-state ensemble $\{\pi_i p_{ik}, |\psi_{ik}\rangle\langle\psi_{ik}|\}$ which has mutual information with the output not less than that of $\{\pi_i, \rho_i\}$. (Note that we can assume that no two states $|\psi_{ik}\rangle\langle\psi_{ik}|$ are identical, since if they are, we can combine them into a single state with prior probability equal to the sum of their prior probabilities, which does not change the mutual information.) Hence, the maximum in equation (20) is attained for an ensemble of different pure states.

Next, observe that equation (20) can be written as

$$C(\{E_i\}) = \max_{\rho} \max_{\{\pi_i, \psi_i\}_\rho} I(P), \tag{31}$$

where the left maximization is over all density matrices ρ, and the right maximization is over all ensembles $\{\pi_i, \psi_i\}_\rho$ of pure states $\psi_i \equiv |\psi_i\rangle\langle\psi_i|$, whose averages are equal to ρ, $\sum_i \pi_i|\psi_i\rangle\langle\psi_i| = \rho$. (We note that the quantity $\max_{\{\pi_i, \psi_i\}_\rho} I(P)$ for a fixed ρ has been previously considered in relation to methods for obtaining bounds on the mutual information [19].) Following closely the proof in [2], we will show that for any ρ, $\max_{\{\pi_i, \psi_i\}_\rho} I(P)$ can be achieved by an ensemble of at most d^2 states. Indeed, the latter maximization is equivalent to a maximization over the convex set Y of probability distributions with finite support on the set of pure states, whose average is equal to ρ. Note that the different ensembles $\{\pi_i, \psi_i\}_\rho$ give rise to joint probability matrices P with fixed row sums equal to $\mathrm{Tr}(\rho E_j)$, which according to the convexity property pointed out earlier implies that $I(P)$ is a convex function on Y. Hence, the maximum is achieved for an extreme point of Y, which by Caratheodory's theorem can be shown to be a probability distribution whose support has $\leqslant 1 + \dim \mathcal{A}$ points, where $\mathcal{A}$ is the convex set of density operators of which the pure states we are considering are extreme points. Since $\dim \mathcal{A} = d^2 - 1$, we obtain $N \leqslant d^2$.

To show that in general $d \leqslant N$, we will use the fact that for every d, there are certain types of POVMs for which the optimal ρ in equation (31) is full-rank (in particular, we will show below (theorem 1) that when the POVM is covariant under the irreducible representation of a finite group, the maximum in equation (31) is achieved for $\rho = I/d$). If we assume that $d > N$, there must exist a vector $|\psi\rangle$, $\langle\psi|\psi\rangle = 1$, such that $\langle\psi|\psi_i\rangle = 0$, $\forall i = 1, \ldots, N$. But then $\langle\psi|\rho|\psi\rangle = \sum_i \pi_i |\langle\psi|\psi_i\rangle|^2 = 0$, which is in contradiction to ρ being full-rank. □

We next consider the case of a group covariant POVM measurement, which is dual to the problem of accessible information for a group covariant input ensemble [2]. For this purpose, we need to introduce some terminology. Let S denote the set of all states on a Hilbert space $\mathcal{H}$ of dimension d. Following [2], we will regard a representation R of a group G as a homomorphism from G to the affine automorphisms of S, where every such automorphism is representable in the form $\alpha(\rho) = U\rho U^\dagger$ with U being a unitary or an antiunitary operator (we will consider the action of R automatically extended to all operators over $\mathcal{H}$ by linearity). A representation of G is irreducible if the only G-invariant point of S is I/d.

We will say that the POVM $\{E_j\}$, $j = 1, \ldots, M$, is G-covariant if there exists a surjection $f : G \to \{E_j\}$, where we denote $f(g) := E_g$, such that $R_g(E_h) = E_{gh}$, $\forall g, h \in G$. Note that every element E_j must be equal to E_g for at least one $g \in G$, but this correspondence may be degenerate, i.e. a given E_j may be associated with two or more elements of the group. The fact the G is a group implies that this degeneracy must be the same for every element E_j and hence M must be a factor of $|G|$.

Theorem 1 (the group covariant case). *If the POVM $\{E_j\}$ is covariant with respect to the finite group G that has an irreducible representation R on S, then there exists a pure state $|\psi\rangle\langle\psi|$, $\langle\psi|\psi\rangle = 1$, such that the maximum in equation (20) is achieved by the covariant ensemble of pure input states $\{|G|^{-1}, R_g^*(|\psi\rangle\langle\psi|)\}$, where $|G|$ is the number of elements of G and R^* denotes the representation of G dual to R. The capacity of $\{E_j\}$ is*

$$C(\{E_j\}) = \log d + M^{-1} d \sum_j \left\langle \psi \left| \frac{E_j}{\operatorname{Tr} E_j} \right| \psi \right\rangle \log \left\langle \psi \left| \frac{E_j}{\operatorname{Tr} E_j} \right| \psi \right\rangle. \tag{32}$$

Proof. Let $\{\pi_i, \psi_i\}$ be an ensemble of pure input states that maximizes the mutual information for the given covariant POVM measurement $\{E_j\}$. Construct a new input ensemble $\{\tilde{\pi}_{ig}, \tilde{\psi}_{ig}\}$, where

$$\tilde{\psi}_{ig} = R_g^*(\psi_i) \quad \text{and} \quad \tilde{\pi}_{ig} = \pi_i |G|^{-1}. \tag{33}$$

The new probability matrix $\tilde{P}$ obtained using this ensemble has the form

$$\tilde{P} = |G|^{-1} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_{|G|} \end{pmatrix}, \tag{34}$$

where each of the probability matrices $P_1, P_2, \ldots, P_{|G|}$ is obtained from P by a permutation of the rows and columns of P, and the column sums of $\tilde{P}$ are all equal to $|G|^{-1}$. A straightforward calculation shows that the new probability matrix yields a value for the mutual information that is no less than that obtained for P, i.e. $I(\tilde{P}) \geqslant I(P)$:

$$\begin{aligned} I(\tilde{P}) &\equiv \sum_i \eta \left(\sum_j \tilde{P}_{ij} \right) + \sum_j \eta \left(\sum_i \tilde{P}_{ij} \right) - \sum_{ij} \eta(\tilde{P}_{ij}) \\ &= |G| \sum_i \eta \left(\sum_j |G|^{-1} P_{ij} \right) + \sum_j \eta \left(|G|^{-1} \right) - |G| \sum_{ij} \eta(|G|^{-1} P_{ij}) \\ &= \left[\sum_i \eta \left(\sum_j P_{ij} \right) + \log |G|^{-1} \right] + \sum_j \eta \left(|G|^{-1} \right) - \left[\sum_{ij} \eta(P_{ij}) + \log |G|^{-1} \right] \\ &\geqslant \sum_i \eta \left(\sum_j P_{ij} \right) + \sum_j \eta \left(\sum_i P_{ij} \right) - \sum_{ij} \eta(P_{ij}) \equiv I(P). \end{aligned} \tag{35}$$

Now, consider the covariant input ensembles $\{|G|^{-1}, \bar{\psi}^i_g\}$, where

$$\bar{\psi}^i_g = R^*_g(\psi_i). \tag{36}$$

Let us denote by $\bar{P}_i$ the probability matrices that each of these ensembles yields. Since the ensemble $\{\tilde{\pi}_{ig}, \tilde{\psi}_{ig}\}$ is a convex combination of the ensembles $\{|G|^{-1}, \bar{\psi}^i_g\}$, and the mutual information is a convex function of the input ensemble, we obtain

$$I(P) \leqslant I(\tilde{P}) \leqslant \max_i I(\bar{P}_i), \tag{37}$$

i.e. the maximum in equation (20) is achieved for one of the covariant input ensembles $\{|G|^{-1}, \bar{\psi}^i_g\}$ that has the form stated in the theorem. The value of the capacity (equation (32)) is obtained by a straightforward calculation, taking into account the possible degeneracy in the correspondence between the group elements and the POVM elements. □

Note that since the average of a group covariant ensemble is G-invariant, from the irreducibility of R it follows that $\sum_g |G|^{-1}\bar{\psi}^i_g = I/d$. This shows that indeed for every d there are POVM measurements that require at least d optimal input states as argued in the proof of proposition 2.

Comment. The optimal 'seed' $|\psi\rangle\langle\psi|$ may be such that the input ensemble $\{|G|^{-1}, R^*_g(|\psi\rangle\langle\psi|)\}$ contains identical states, i.e. it may be that $R^*_g(|\psi\rangle\langle\psi|) = R^*_h(|\psi\rangle\langle\psi|)$ for certain $g \neq h$. The fact that G is a group implies that each maximal set of identical states in the ensemble must contain the same number of elements (and hence the number N of distinct states in the ensemble must be a factor of $|G|$). It is straightforward to see that the ensemble $\{N^{-1}, |\psi_i\rangle\langle\psi_i|\}$ obtained from $\{|G|^{-1}, R^*_g(|\psi\rangle\langle\psi|)\}$ by identifying the identical states and redefining their probabilities as the sum of the original probabilities is also optimal. This is because the joint probabilities resulting from the input ensemble $\{N^{-1}, |\psi_i\rangle\langle\psi_i|\}$ can be transformed into those resulting from $\{|G|^{-1}, R^*_g(|\psi\rangle\langle\psi|)\}$ by local postprocessing on the sender's side, which cannot increase the mutual information. Hence, the number of states in the optimal ensemble in general may be smaller than $|G|$ (just as the number of outcomes of a group covariant POVM may be smaller than $|G|$). This is the case, for example, with the optimal ensemble for the two-dimensional symmetric informationally complete (SIC)-POVM studied in section 5.3, which has four elements, whereas the symmetry group has 12.

Corollary 2. *In the group covariant case, we have*

$$C(\{E_i\}) = A(\{\breve{E}_i\}). \tag{38}$$

Moreover, if the POVM measurement $\{F_j\}$ optimizes the mutual information for the input ensemble $\{\breve{E}_i\}$, the input ensemble $\{\breve{F}_j\}$, where $\breve{F}_j \equiv F_j/d$, optimizes the mutual information for the measurement $\{E_i\}$.

Since under this symmetry the problem is equivalent to that of accessible information of a covariant input ensemble, any known results in the latter case can be applied here (see e.g. [2]). In particular, in section 5.3 we calculate the capacity of the two-dimensional SIC-POVM.

Another important case in which calculating the capacity of a measurement reduces to a well-known problem is that of a POVM $\{E_i\}$ with commuting elements, $[E_i, E_j] = 0$, $\forall i, j$.

In this case, we can assume that the optimal signal states ρ_i are diagonal in the eigenbasis of $\{E_j\}$, since for any ρ, the state $\rho' = \mathrm{diag}(\rho_{nn})$, where ρ_{nn} are the diagonal elements of ρ in the eigenbasis of $\{E_j\}$, yields the same values for the joint probabilities $\mathrm{Tr}(\rho E_j)$. Furthermore, as we saw in the proof of proposition 2, the optimal input ensemble can be taken to consist of the eigenstates of all ρ_i, which means that the maximum in equation (20) is achieved for an ensemble of input states which are the common eigenbasis of $\{E_i\}$. Hence, the joint probabilities are $P_{ij} = \pi_i \lambda_j^i$, where λ_j^i is the ith eigenvalue of E_j, and the problem reduces to finding $\max_{\{\pi_i\}} I(P)$ which is the capacity of the classical channel described by the conditional probability matrix $p(j|i) = \lambda_j^i$. Note that a measurement with two outcomes necessarily has commuting POVM elements, i.e. the capacity of a two-outcome measurement is always equal to the capacity of a classical channel with a binary output. Thus, for example, the capacity of a two-outcome qubit measurement that has elements $E_1 = \mathrm{diag}(\alpha, \beta)$, $E_2 = \mathrm{diag}(1-\alpha, 1-\beta)$ in some basis can be obtained from the formula for the capacity of a general binary channel [31]

$$C(\alpha, \beta) = \frac{\alpha H(\beta) - \beta H(\alpha)}{\beta - \alpha} + \log\left[1 + \exp\frac{H(\alpha) - H(\beta)}{\beta - \alpha}\right], \tag{39}$$

where $H(q) = -q\log q - (1-q)\log(1-q)$, $q \in [0, 1]$, is the entropy of a binary source. The optimal prior distribution in this case is $\{p, 1-p\}$, where [31]

$$p = \frac{\beta}{\beta - \alpha} - \frac{1}{(\beta - \alpha)\left[1 + \exp\frac{H(\beta) - H(\alpha)}{\beta - \alpha}\right]}. \tag{40}$$

We can now see that, as mentioned earlier, the naively constructed Holevo quantity $S(\sum_i m_i \bar{E}_i) - \sum_i m_i S(\bar{E}_i)$, where $m_i = \mathrm{Tr}(E_i)/d$, $\bar{E}_i = E_i/(m_i d)$, in general is neither an upper nor a lower bound to $C(\{E_i\})$. Indeed, it is known that the accessible information of an ensemble of density matrices is equal to the Holevo quantity of the ensemble if and only if all density matrices in the ensemble commute, and the maximal value of the mutual information is attained for a projective measurement in the common eigenbasis of the input ensemble. From the symmetry of the problem, we see that for a POVM with commuting elements, the quantity $S(\sum_i m_i \bar{E}_i) - \sum_i m_i S(\bar{E}_i)$ is equal to the mutual information between the equiprobable input ensemble of common eigenstates of $\{E_i\}$ and the outputs of the measurement $\{E_i\}$. However, from equation (40) it can be seen that an equiprobable prior distribution is generally suboptimal for this case, i.e. the quantity $S(\sum_i m_i \bar{E}_i) - \sum_i m_i S(\bar{E}_i)$ can be strictly smaller than $C(\{E_i\})$. On the other hand, in the group covariant case we have $C(\{E_i\}) = A(\{\breve{E}_i\})$, where in general $A(\{\breve{E}_i\})$ is strictly smaller than $S(\sum_i m_i \bar{E}_i) - \sum_i m_i S(\bar{E}_i)$.

We remark that the maximal possible mutual information for an input ensemble of states on a Hilbert space of dimension d and any POVM measurement is $\log d$. This can be easily seen from Holevo's upper bound on the accessible information [21]. Moreover, this quantity is achievable only by an ensemble of pure commuting input states that sum up to the maximally mixed state, i.e. by an equiprobable ensemble of orthogonal basis states. The unique optimal measurement for such an ensemble is a projective measurement on the basis in question. Reversely, any rank-one projective measurement has capacity $\log d$ which is achievable by the equiprobable input ensemble of corresponding basis states. Hence, rank-one projective measurements have the highest capacity.

5. Example: the symmetric informationally complete positive operator-valued measure on a qubit

In this section, we apply the above results to the case of an SIC-POVM on a qubit, as well as to a noisy, or unsharp, version of this POVM. An SIC-POVM [22, 23] in dimension d consists of a set of d^2 rank-one positive operators, $E_i = (1/d)|\psi_i\rangle\langle\psi_i|$, where the pure states $|\psi_i\rangle$ are such that $|\langle\psi_i|\psi_j\rangle|^2 = 1/(d+1)$ for $i \neq j$. The measurement is called 'complete' in the sense that its statistics is sufficient for the full tomography of any quantum state [24, 25]. SIC-POVMs are of particular interest due to their various applications in quantum information, including quantum tomography [26], quantum cryptography [27] and the foundations of quantum mechanics [28].

Up to a change of basis, the POVM elements of such a measurement for $d = 2$ can be written as

$$E_i = \tfrac{1}{4}(I + \vec{n}_i \cdot \vec{\sigma}), \quad i = 1, 2, 3, 4, \tag{41}$$

where $\vec{\sigma}$ is the vector of Pauli matrices

$$\sigma_y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{42}$$

and

$$\begin{aligned} \vec{n}_1 &= \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{n}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \\ \vec{n}_3 &= \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}, \quad \vec{n}_4 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}. \end{aligned} \tag{43}$$

In order to illustrate the relation between the 'sharpness' of a measurement and its ability to read out information, we will consider a more general, noisy version of the above SIC-POVM, where each outcome is mixed with some amount of white noise,

$$E_i(\epsilon) = \epsilon E_i + (1-\epsilon)\frac{I}{4} = \frac{1}{4}(I + \epsilon\vec{n}_i \cdot \vec{\sigma}), \quad i = 1, 2, 3, 4, \quad 0 \leqslant \epsilon \leqslant 1. \tag{44}$$

When $\epsilon = 1$, the measurement reduces to the ideal SIC-POVM [$E_i(1) \equiv E_i$], whereas as $\epsilon \to 0$, the measurement becomes infinitesimally weak [29], approaching a trivial measurement, each of its outcomes occurring with probability $1/4$ independently of the input state. In this sense, ϵ can be regarded as parameterizing the 'sharpness' or 'strength' of the measurement [30].

5.1. Minimum error discrimination

For simplicity, let us start with the noiseless SIC-POVM (41). Given the symmetry of the problem, it is enough to consider four groupings, $\alpha \in \{A, B, C, D\}$:

$$\begin{aligned} A &: \{E_1, E_2, E_3, E_4\}, \\ B &: \{E_1 + E_2, E_3, E_4, 0\}, \\ C &: \{E_1 + E_2, E_3 + E_4, 0, 0\}, \\ D &: \{E_1 + E_2 + E_3, E_4, 0, 0\}. \end{aligned} \tag{45}$$

The corresponding vector of maximum eigenvalues (in decreasing order of value) are (see equation (8) with $B_{ij} = \delta_{ij}$)

$$\begin{aligned} \mathbf{s}_A &= \{1/2, 1/2, 1/2, 1/2\}, \\ \mathbf{s}_B &= \{(1+1/\sqrt{3})/2, 1/2, 1/2, 0, 0\}, \\ \mathbf{s}_C &= \{(1+1/\sqrt{3})/2, (1+1/\sqrt{3})/2, 0, 0\}, \\ \mathbf{s}_D &= \{1, 1/2, 0, 0\}, \end{aligned} \tag{46}$$

where it is understood that the vectors need to be padded with extra zeros if the number of signal states exceeds four ($N > 4$). For equiprobable signals, $\pi_i = 1/N$, the optimal success probability is given by $p_s = 1/N \max_\alpha \sum_{i=1}^N (\mathbf{s}_\alpha)_i$. In particular, $p_s = 1/2 + 1/(2\sqrt{3})$ for $N = 2$, $p_s = 1/2 + 1/(6\sqrt{3})$ for $N = 3$ and $p_s = 2/N$ for $N \geqslant 4$, which are attained by the groupings C, B and A, respectively. That is, for four signals ($N = 4$) no grouping is necessary and the signal states have to be chosen to point along the directions of the SIC-POVM (43). Any additional signals ($N > 4$) can be assigned to arbitrary states and will never contribute to the success probability. For $N = 3$ one has to group two POVM elements leading to an unsharp effective measurement, and leave the remaining two outcomes ungrouped (i.e. sharp). In that case the three signals lie on a plane: two signals point along, say, $\vec{n}_1$ and $\vec{n}_2$ (corresponding to the sharp POVM elements), and the third points along $-(\vec{n}_1 + \vec{n}_2)$. For $N = 2$ the optimal strategy is to encode the signals into orthogonal states pointing along the directions resulting from pairwise groupings, e.g. $\vec{n}_1 + \vec{n}_2$ and $\vec{n}_3 + \vec{n}_4 = -(\vec{n}_1 + \vec{n}_2)$.

In figure 1 we show the optimality regions for $N = 3$ and different priors. Within the region $\pi_1 \geqslant \pi_2 \geqslant \pi_3$, delimited by a dashed outline in this figure, we observe that the set of points where each particular grouping is dominant is a convex polytope. The corresponding maximum success probabilities are

$$\begin{aligned} p_s^B &= \frac{1}{2} + \frac{1}{2\sqrt{3}}\pi_1, \\ p_s^C &= \left(\frac{1}{2} + \frac{1}{2\sqrt{3}}\right)(\pi_1 + \pi_2), \\ p_s^D &= \pi_1 + \tfrac{1}{2}\pi_2. \end{aligned} \tag{47}$$

Note that regions C and D correspond to groupings where no outcome is assigned to the third signal state.

This illustrates the fact that there are cases (regions C and D) where it pays not to assign any measurement outcome to some of the messages ($i = 3$ in this example), even though the source emits them with nonzero prior probability. In particular, if the source is strongly biased towards one message ($\pi_1 \geqslant 1/\sqrt{3}$, in this example), all but one measurement outcome will be assigned to it (message $i = 1$).

In order to study the effect of noise, $\epsilon < 1$ in (44), one proceeds along the same lines as above. We first note that since the noise is isotropic, the optimal signal states, i.e. the eigenvectors with maximum eigenvalue of the sums of POVM elements in each grouping, A–D, are the same as those for the sharp case, thus independent of the sharpness parameter ϵ. Their corresponding maximum eigenvalues in (46) now have a noisy component that scales with the number $k_{\alpha i}$ of elements in those sums. More precisely, the vectors of eigenvalues have now components $\epsilon(\mathbf{s}_\alpha)_i + k_{\alpha i}(1-\epsilon)/4$.

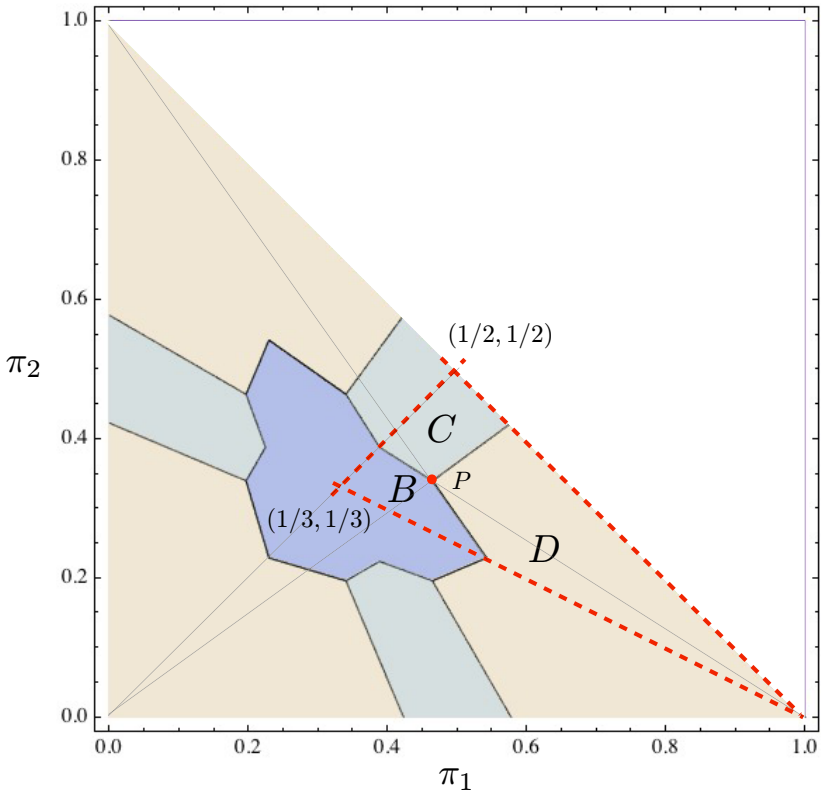

Figure 1. The colored regions in the prior-probability simplex for $N = 3$ indicate the various optimal groupings. The dashed outline delimits the region $\pi_1 \geqslant \pi_2 \geqslant \pi_3$. Within this region, the intersection point of A, B and C is $P = \{2\sqrt{3} - 3, 9 - 5\sqrt{3}\}$. Auxiliary thin lines are drawn to help understand the figure.

For equiprobable signals, $\pi_i = 1/N$, the optimal groupings are those that are optimal in the sharp case. Thus, they do not depend on ϵ, only on the number N of input states. The minimum errors are now: $p_s = 1/2 + \epsilon/(2\sqrt{3})$ for $N = 2$, $p_s = 1/3 + \epsilon(1 + 1/\sqrt{3})/6$ for $N = 3$ and $p_s = (1 + \epsilon)/N$ for $N \geqslant 4$.

In more generic cases, when the source emits symbols with arbitrary prior probabilities, the regions of optimality do depend on the noise or sharpness parameter ϵ. For the case of ternary sources, $N = 3$, it is straightforward to show that the overall structure of the optimality regions is that in figure 1, but the point $P(\epsilon)$ where B, C and D intersect moves monotonically away from $P(1) = P = \{2\sqrt{3} - 3, 9 - 5\sqrt{3}\}$ (when the POVM is sharp) to $P(0) = \{1/3, 1/3\}$ (when it is maximally unsharp).

5.2. *Unambiguous discrimination*

We now turn to unambiguous discrimination with the SIC-POVM on a qubit. Clearly, the slightest amount of noise ($\epsilon < 1$) will ruin any possibility of performing unambiguous discrimination since any signal state can trigger each of the outcomes with nonzero probability. We thus concentrate on the ideal sharp SIC-POVM. In a two-dimensional Hilbert space one can only hope to unambiguously discriminate two states ($N = 2$; $\pi_1 < 1$); hence, grouping A can

be excluded as it has too many outcomes. Moreover, we need only consider groupings B and D, since only they have at least one rank-one POVM element and have therefore a non empty kernel ($\mathscr{K}_j^\alpha \neq \emptyset$). If grouping B is used, two messages can be unambiguously identified by choosing the signals in the kernels of E_3 and E_4, respectively, that is, $\rho_1 = (1 - \vec{n}_3 \cdot \vec{\sigma})/2$ and $\rho_2 = (1 - \vec{n}_4 \cdot \vec{\sigma})/2$, so that outcome 4 can only be triggered by ρ_1 and outcome 3 by ρ_2 (i.e. $\tilde{E}_1 = E_4$, $\tilde{E}_2 = E_3$ and $\tilde{E}_? = E_1 + E_2$). This leads to a probability of successful identification given by

$$\begin{aligned} p_s^B &= \pi_1 \operatorname{Tr}(\tilde{E}_1 \rho_1) + \pi_2 \operatorname{Tr}(\tilde{E}_2 \rho_2) \\ &= \frac{1 - \vec{n}_4 \cdot \vec{n}_3}{4} = \frac{1}{3}, \end{aligned} \tag{48}$$

which is independent of the prior probabilities $\{\pi_1, \pi_2\}$.

Proceeding along the same lines, one finds that for grouping D one can only unambiguously identify the state $\rho_1 = (1 + \vec{n}_4 \cdot \vec{\sigma})/2$ with $\tilde{E}_1 = E_4$, by excluding $\rho_2 = (1 - \vec{n}_4 \cdot \vec{\sigma})/2$ (i.e. $\rho_2 \in \ker \tilde{E}_1$), while all other outcomes of the original POVM will be necessarily inconclusive ($\tilde{E}_? = I - E_4$). Obviously, no outcome will be associated with message $i = 2$ ($\tilde{E}_2 = 0$). The success probability is

$$p_s^D = \pi_1 \operatorname{Tr}(\tilde{E}_1 \rho_1) = \frac{\pi_1}{2}, \tag{49}$$

which beats that of grouping B for $\pi_1 > 2/3$.

5.3. Mutual information

The SIC-POVM on a qubit, including its noisy version, is covariant under the tetrahedral group (indeed, the tips of the Bloch vectors (43) corresponding to the POVM elements define the vertices of a tetrahedron). Therefore, according to theorem 1 in section 4, the mutual information for this POVM is maximized by an ensemble of pure input states possessing the same symmetry. Its maximal value, i.e. the capacity of the measurement, is given by equation (32) for a state ψ from the optimal ensemble (all other states in the ensemble are obtained from ψ by applying operators of the symmetry group, i.e. ψ plays the role of a 'seed' for the ensemble).

Theorem 2. (Capacity of the noisy two-level SIC-POVM). *For every value of $\epsilon \in [0, 1]$, the seed ψ that maximizes expression (32) can be chosen such that its Bloch vector is anti-parallel to the Bloch vector of any one of the four POVM elements (44), i.e. $\vec{v} = -\vec{n}_j$. The capacity of the (generally noisy) SIC-POVM is*

$$C_\epsilon = 1 + \frac{1-\epsilon}{4} \log \frac{1-\epsilon}{2} + 3 \frac{1+\epsilon/3}{4} \log \frac{1+\epsilon/3}{2}. \tag{50}$$

This result, which applies to both the straight and the reverse formulations of the problem, is interesting in its own right. As far as we are aware, previous results (for $\epsilon = 1$) relied on numerical optimization [2]. Here we provide an analytical proof for $0 \leqslant \epsilon \leqslant 1$.

Proof. Let us define

$$h(t) \equiv \eta\left(\frac{1+t}{2}\right).$$

We will first show that the following inequality holds for $-1 \leqslant t \leqslant 1$ and $0 \leqslant \epsilon \leqslant 1$:

$$h(\epsilon t) \geqslant a(\epsilon) + b(\epsilon)t + c(\epsilon)t^2 \equiv \wp_\epsilon(t), \tag{51}$$

where

$$a(\epsilon) = \tfrac{1}{16}[h(-\epsilon) + 15h(\epsilon/3) - 4\epsilon h'(\epsilon/3)],$$
$$b(\epsilon) = \tfrac{1}{8}[-3\,h(-\epsilon) + 3h(\epsilon/3) + 4\epsilon h'(\epsilon/3)],$$
$$c(\epsilon) = \tfrac{3}{16}[3h(-\epsilon) - 3h(\epsilon/3) + 4\epsilon h'(\epsilon/3)],$$

and h' is the derivative of h with respect to its argument.

We start by noting the following relations,

$$\wp_\epsilon(-1) = h(-\epsilon), \quad \wp_\epsilon(1/3) = h(\epsilon/3), \quad \wp'_\epsilon(1/3) = \epsilon h'(\epsilon/3) \tag{52}$$

and

$$\gamma(\epsilon) \equiv c(\epsilon) + \frac{\epsilon^2}{4\ln 2} \leqslant 0, \tag{53}$$

where the equality is attained only at $\epsilon = 0$. The first three of them are immediate. The last one is not so obvious and can be proved as follows. The function $\gamma(\epsilon)$ is concave in [0, 1] since

$$\gamma''(\epsilon) = -\frac{9}{2(1-\epsilon)(3+\epsilon)^2\ln 2} + \frac{1}{2\ln 2} = -\frac{\epsilon(3+5\epsilon+\epsilon^2)}{2(1-\epsilon)(3+\epsilon)^2\ln 2} \leqslant 0.$$

Differentiating the expression of $c(\epsilon)$ above, we readily obtain

$$c'(\epsilon) = \tfrac{3}{16}[-3h'(-\epsilon) + 3h'(\epsilon/3) + \tfrac{4}{3}\epsilon h''(\epsilon/3)],$$

which vanishes at $\epsilon = 0$. Thus $\gamma'(0) = \gamma''(0) = 0$ and $\gamma''(\epsilon) < 0$ if $\epsilon > 0$. Then, $\gamma(\epsilon)$ must necessarily decrease for $\epsilon > 0$, which in turn implies that $\gamma(\epsilon)$ has its unique maximum at $\epsilon = 0$. Since $\gamma(0) = 0$, equation (53) holds in the whole interval [0, 1].

We can now turn to proving (51). We assume that $\epsilon > 0$, since $\epsilon = 0$ is a trivial case. If $f(t) = h(\epsilon t) - \wp_\epsilon(t)$, then

$$f''(t) = -2c(\epsilon) - \frac{\epsilon^2}{2(1+\epsilon\, t)\ln 2}.$$

It follows from this equation that there is only one value of t for which $f''(t)$ vanishes. But using (53), we see that $f''(t) > 0$ for $t \geqslant 0$. Therefore, $f''(t)$ can only change sign at some $t_0 < 0$. Hence, $f(t)$ is convex in $(t_0, 1]$ and concave in $[-1, t_0)$. It can have only one minimum in $(t_0, 1]$, and according to the third relation (52), it must be at $t = 1/3$. Using the second relation (52), we see that this minimum value is 0. Thus $f(t) \geqslant 0$ if $t \in [t_0, 1]$. Because of the concavity of f in the other interval, we just need to check the value of f at the end point $t = -1$ (by continuity we must have $f(t_0) \geqslant 0$). The first relation (52) ensures that $f(t) \geqslant 0$ also in $[-1, t_0]$.

Now, using the inequality (51), one can show that the mutual information for the POVM (44),

$$I = 1 - \frac{1}{2}\sum_{j=1}^{4} \eta\left(\frac{\langle\phi|E_j(\epsilon)|\psi\rangle}{\operatorname{tr} E_j(\epsilon)}\right) = 1 - \frac{1}{2}\sum_{j=1}^{4} h\left(\epsilon\, \vec{v}\cdot\vec{n}_j\right),$$

is bounded as

$$I \leqslant 1 - \frac{1}{2}\sum_{j=1}^{4} \wp_\epsilon(\vec{v}\cdot\vec{n}_j) = 1 - \frac{1}{2}\left[4a(\epsilon) + \frac{4}{3}c(\epsilon)\right] = 1 - \frac{h(-\epsilon) + 3h(\epsilon/3)}{2}.$$

This bound is attained with any one of the four choices $\vec{v} = -\vec{n}_j$. The value of the capacity (50) is obtained by a straightforward substitution. □

Note that in the minimum error scenario, the optimal signal ensemble is such that each state and its corresponding POVM element have maximum overlap (i.e. they are aligned with each other). In contrast, here we find that it pays to have a signal ensemble where each state would be excluded by one of the POVM outcomes in the absence of noise (i.e. states and POVM elements are anti-aligned with each other). This configuration minimizes the (average) conditional entropy of the output (the POVM outcomes) given the input signal ensemble (recall that the mutual information (32) can be obtained by subtracting this conditional entropy from the entropy of the output, which is constant here).

As expected, the capacity attains its maximal value $C_1 = \log 4/3$ for $\epsilon = 1$ (the ideal SIC-POVM) and monotonically decreases towards 0 as ϵ approaches 0. Note that, as pointed out in corollary 2, the capacity of such a group covariant POVM is equal to the accessible information of an equiprobable ensemble of states proportional to the original POVM elements,

$$\rho_i = \frac{I + \epsilon\vec{n}_i\cdot\vec{\sigma}}{2}, \quad i = 1, 2, 3, 4. \tag{54}$$

The latter problem, in the case $\epsilon = 1$, was studied in [2], where it was shown that the accessible information of the corresponding ensemble is $A = \log 4/3$, which is equal to C_1. The capacity of the ideal SIC-POVM has also been obtained by a different approach in [19].

6. Conclusion

In summary, we have studied the problem of optimal signal states for information readout with a given quantum detector. We considered some of the most common information transmission problems—the Bayes cost problem, unambiguous message discrimination and the maximal mutual information. We provided solutions to the Bayesian and unambiguous discrimination strategies. We also showed that the maximal mutual information is equal to the classical capacity of the measurement and studied its properties in certain special cases. For a group covariant measurement, we obtained that the problem is equivalent to the problem of accessible information of a group covariant ensemble of states. As an example, we applied our results on the different discrimination strategies to the case of a SIC-POVM on a qubit, including a noisy version of that POVM.

An interesting question for a future investigation is whether and under what conditions the optimal solutions provided here are unique. Another question of significant interest would be to obtain an upper bound on the capacity of a measurement. We provided a lower bound which is obtained from a lower bound on the accessible information, but that lower bound could also be improved. It would also be interesting to investigate the continuity properties of the optimal quantities considered in this paper. For example, if two measurements are close in terms of the distance functions introduced in [32], are their capacities also close?

Finally, we note that the capacity of a POVM provides a very natural and source-independent means of giving a quantitative characterization of a generalized quantum measurement. However, it cannot be used as the unique figure of merit against which measurement devices should be benchmarked. Ultimately, the performance of a given measurement apparatus depends strongly on the task that it is meant to accomplish. For instance, a noisy Stern–Gerlach measurement might have a higher capacity than that of an ideal SIC-POVM; however, it would be misleading to claim that such a Stern–Gerlach measurement outperforms the SIC-POVM since the latter can carry out tasks (e.g. full single-qubit tomography or unambiguous state discrimination) that are impossible to achieve with the former.

Note added. Almost simultaneously with the posting of this paper, two concurrent works appeared—by M Dall'Arno, G M D'Ariano and M F Sacchi (see [33]) and by A S Holevo (see [34])—which also introduce and study the capacity of a POVM measurement.

Acknowledgments

We thank Alex Monras for valuable discussions. This work was supported by the Spanish MICINN through the Ramón y Cajal program (to JC), contract number FIS2008-01236, and project QOIT (CONSOLIDER2006-00019), and by the Generalitat de Catalunya through CIRIT 2009SGR-0985. OO was partially supported by the Interuniversity Attraction Poles program of the Belgian Science Policy Office, under the grant IAP P6-10 'photonics@be'. EB thanks the HET group at BNL and Hunter College of the CUNY for their hospitality during the final stages of this work. EB acknowledges financial support from the Spanish MICINN, reference number PR2010-0367.

References

[1] Helstrom C W and Kennedy R S 1974 *IEEE Trans. Inf. Theory* **20** 16
[2] Davies E B 1978 *IEEE Trans. Inf. Theory* **24** 596
[3] Ivanovic I D 1987 *Phys. Lett.* A **123** 257
[4] Dieks D 1988 *Phys. Lett.* A **126** 303
[5] Peres A 1988 *Phys. Lett.* A **128** 19
[6] Fuchs C A 1996 Distinguishability and accessible information in quantum theory *PhD Thesis* University of New Mexico, Albuquerque, NM
[7] Chefles A and Barnett S M 1998 *Phys. Lett.* A **250** 223
[8] Raynal P and Lütkenhaus N 2005 *Phys. Rev.* A **72** 022342
[9] Lundeen J S *et al* 2009 *Nat. Phys.* **5** 27
[10] Marcus M and Minc H 1992 *A Survey of Matrix Theory and Matrix Inequalities* (New York: Dover)
[11] Elron N and Eldar Y C 2007 *IEEE Trans. Inf. Theory* **53** 1900
[12] Holevo A S 1998 *IEEE Trans. Inf. Thy.* **44** 269–73
[13] Schumacher B and Westmoreland M D 1997 *Phys. Rev.* A **56** 131–8
[14] Horodecki M, Shor P W and Ruskai M B 2003 *Rev. Math. Phys.* **15** 629
[15] Shor P W 2002 *J. Math. Phys.* **43** 4334
[16] Holevo A S 1998 *Russ. Math. Surv.* **53** 1295 (arXiv:quant-ph/9809023)
[17] King C 2002 *J. Math. Phys.* **43** 1247
[18] Hughston L P, Josza R and Wootters W K 1993 *Phys. Lett.* A **183** 14

[19] Hall M J W 1997 *Phys. Rev.* A **55** 100
[20] Jozsa R, Robb D and Wootters W K 1994 *Phys. Rev.* A **49** 668
[21] Kholevo A S 1979 *Probl. Peredachi Inf.* **15** 3 (in Russian)
[22] Zauner G 1999 Quantum designs—foundations of a non-commutative theory of designs *PhD Thesis* University of Vienna (in German)
[23] Renes J M, Blume-Kohout R, Scott A J and Caves C M 2004 *J. Math. Phys.* **45** 2171
[24] Prugoveki E 1977 *Int. J. Theor. Phys.* **16** 321
[25] Busch P 1991 *Int. J. Theor. Phys.* **30** 1217
[26] Caves C M, Fuchs C A and Schack R 2002 *J. Math. Phys.* **43** 4537
[27] Fuchs C A and Sasaki M 2003 *Quantum Inf. Comput.* **3** 377
[28] Fuchs C A 2002 arXiv:quant-ph/0205039
[29] Oreshkov O and Brun T A 2005 *Phys. Rev. Lett.* **95** 110409
[30] Rapcan P, Calsamiglia J, Munoz-Tapia R, Bagan E and Buzek V 2011 arXiv:1105.5326
[31] Reza F M 1961 *An Introduction to Information Theory* (New York: McGraw-Hill)
[32] Oreshkov O and Calsamiglia J 2009 *Phys. Rev.* A **79** 032336
[33] Dall'Arno M, D'Ariano G M and Sacchi M F 2011 *Phys. Rev.* A **83** 062304
[34] Holevo A S 2011 arXiv:1103.2615.