Using postmeasurement information in state discrimination

Deepthi Gopal[1,*] and Stephanie Wehner[1,2,†]

[1]*Institute for Quantum Information, Caltech, Pasadena, California 91125, USA*
[2]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*

We consider a special form of state discrimination in which after the measurement we are given additional information that may help us identify the state. This task plays a central role in the analysis of quantum cryptographic protocols in the noisy-storage model, where the identity of the state corresponds to a certain bit string, and the additional information is typically a choice of encoding that is initially unknown to the cheating party. We first provide simple optimality conditions for measurements for any such problem and show upper and lower bounds on the success probability. For a certain class of problems, we furthermore provide tight bounds on how useful postmeasurement information can be. In particular, we show that for this class finding the optimal measurement for the task of state discrimination *with* postmeasurement information does in fact reduce to solving a different problem of state discrimination *without* such information. However, we show that for the corresponding *classical* state discrimination problems with postmeasurement information such a reduction is impossible, by relating the success probability to the violation of Bell inequalities. This suggests the usefulness of postmeasurement information as another feature that distinguishes the classical from a quantum world.

I. INTRODUCTION

One of the characteristic traits of quantum mechanics is that not all possible states of a physical system are perfectly distinguishable. This is in stark contrast to the classical world, but enables us to solve cryptographic problems such as key distribution [1,2] or two-party computation in the noisy-storage model [3,4]. Nevertheless, it is often possible to gain partial knowledge about the state. Imagine that a physical system is prepared in one out of several possible states chosen with a certain probability. The set of possible states as well as the distribution are thereby known to us. The goal of *state discrimination* is to identify which state was chosen by performing a measurement on the system, whereby our aim is to choose measurements that maximize the average probability of success. This fundamental problem has been studied extensively for the past 30 years, starting with the works of Helstrom [5], Holevo [6], and Belavkin [7] (see [8] for a survey of known results), and has found many applications in quantum information theory (see, e.g., [9]), cryptography [10], and algorithms [11,12].

Here, we consider a special twist to the standard state discrimination problem (Fig. 1) introduced in [13], in which we obtain additional information after the measurement that may help us to identify the state. This task is easily described in terms of the following game depicted in Fig. 2: Imagine that Alice chooses a state ρ_{xb} from a finite set $\mathcal{E}$ with probability p_{xb}, labeled by what we will call the *string* $x \in \mathcal{X}$ and the *encoding* $b \in \mathcal{B}$. Bob knows $\mathcal{E}$ as well as the distribution $P = \{p_{xb}\}_{xb}$. Alice then sends the state to Bob. Bob may now perform any measurement from which he obtains a classical measurement outcome k. Afterward, Alice informs him about the encoding b. The task of *state discrimination with postmeasurement information* (and no memory) for Bob is to identify the string x, using the encoding b and his classical measurement outcome k, where we are again interested in maximizing Bob's average probability of success over all measurements he may perform.[1] In [3] it was shown how bounds on this success probability can be used to prove security in the noisy-storage model.

Naturally, from a cryptographic standpoint it would be useful to know how much the additional information b can actually help Bob. Let $p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P)$ and $p_{\text{succ}}(\mathcal{E}, P)$ be the maximum average probabilities of success for the problem of state discrimination with and without postmeasurement information, respectively. Note that $p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \geqslant p_{\text{succ}}(\mathcal{E}, P)$, since we can always choose to ignore any additional information. We will measure how useful postmeasurement information is for Bob in terms of the difference in his success probability

$$\Delta(\mathcal{E}, P) := p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) - p_{\text{succ}}(\mathcal{E}, P). \quad (1)$$

Of course, even in a classical setting postmeasurement information can help Bob determine the string x. As a very simple example, suppose that $x \in \{0,1\}$ is a single classical bit, and we have only one encoding $b \in \{0,1\}$. Imagine that Alice chooses x and one of the two encodings uniformly at random and sends Bob the bit $x \oplus b = x + b$ mod 2. The states corresponding to this encoding are thus given by

$$\rho_{xb} = |x \oplus b\rangle\langle x \oplus b|, \quad (2)$$

where $p_{xb} = 1/4$. Note that Bob now has a randomly chosen bit in his possession and hence $p_{\text{succ}}(\mathcal{E}, P) = 1/2$. However, he can decode correctly once he receives the additional information b and thus $p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = 1$, giving us

---

*deepthi@caltech.edu
†wehner@nus.edu.sg

---

[1]Note that in [13] this problem was generalized to a setting where Bob may store not only classical information, but also a (limited) amount of quantum information. Here, however, we will only focus on the case of no storage,. which was enough to relate security of the noisy-storage model to a coding problem [3].
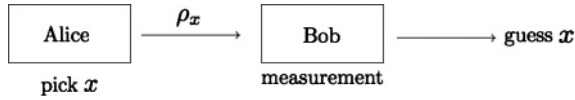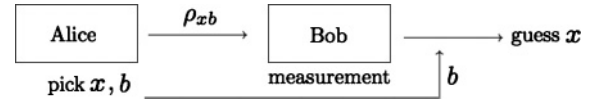
FIG. 1. Standard state discrimination.

FIG. 2. Using postmeasurement information.

$\Delta(\mathcal{E}, P) = 1/2$. As has been shown in [13] we always have $p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = 1$ in the classical world where all states ρ_{xb} are diagonal in the same basis and orthogonal for fixed b.

A. Results

We first provide a general condition for checking the optimality of measurements for our task (see Sec. II B). It was shown in [13] that the optimal measurements can be found numerically using semidefinite programming solvers; however, in higher dimensions this remains prohibitively expensive. We then focus on the case which is particularly interesting for cryptography, namely, when the string x is chosen uniformly and independently from the encoding b. First, we provide upper and lower bounds for the success probability $p_{\text{succ}}^{\text{PI}}$ (Secs. II D and II C).

In Sec. III, we then show that for a large class of encodings (so-called Clifford encodings) our lower bound is in fact tight. We thereby explicitly provide the optimal measurements for Clifford encodings. The class of encodings we consider includes any encodings into two orthogonal pure states in dimension $d = 2$ such as the well-known Bennett-Brassard 1984 (BB84) encodings [1], as well as the case where we have two possible strings and encodings which can be reduced to a problem in dimension $d = 2$ [13,14]. It was previously observed that for BB84 encodings postmeasurement information was useless [13]. Here, we see that this is no mere accident and give a general condition for when postmeasurement information is useless for Clifford encodings. We continue by showing that for Clifford encodings, we can always perform a relabeling of the strings x depending on the encoding b such that we obtain a new problem for which postmeasurement information is indeed useless. This is particularly appealing from a cryptographic perspective as it means the adversary cannot gain any additional knowledge from the postmeasurement information. This means that for Clifford encodings we no longer need to treat the problem *with* postmeasurement information any differently and can instead apply the well-studied machinery of state discrimination.

However, we will see that a relabeling that renders postmeasurement information useless is impossible when considering a *classical* ensemble.[2] In particular, we will see that as long as we are able to gain some information about the encoded string x without waiting for the postmeasurement information, then classically we cannot hope to find a nontrivial relabeling that makes postmeasurement information useless. We thereby focus on the case of encodings a single bit into two possible encodings in detail. Curiously, we will show this by relating the problem to Bell inequalities [15], such as for example the well-known Clauser-Horne-Shimony-Holt (CHSH) inequality [16]. This suggests that the usefulness of

postmeasurement information forms another intriguing property that distinguishes the quantum from the classical world.

II. GENERAL BOUNDS

Before investigating the use of postmeasurement information, we derive general conditions for the optimality of measurements for our task. We also provide a general bound on the success probability when the distribution over $\mathcal{X}$ is uniform (i.e., $p_x = 1/|\mathcal{X}|$) and independent of the choice of encoding.

A. Semidefinite program formalism

When considering state discrimination with postmeasurement information, we can without loss of generality assume that Bob performs a measurement whose outcomes correspond to vectors $\vec{x} = (x^{(1)}, \ldots, x^{(L)}) \in \mathcal{X}^{\times L}$ where each entry corresponds to the answer that Bob will give when he later learns which one of the $L = |\mathcal{B}|$ possible encodings was used. That is, when the encoding was b, Bob will output the guess $x^{(b)}$ of the vector $\vec{x}$ [13]. In [13] it was noted that the average probability that Bob outputs the correct guess $x^{(b)}$ when given the postmeasurement information b maximized over all possible positive operator valued measurements (POVMs) can be computed by solving the following semidefinite program (SDP). The primal of this SDP is given by

$$\text{maximize} \quad v_{\text{primal}} = \sum_{\vec{x}} \text{tr}(M_{\vec{x}} \tau_{\vec{x}})$$
$$\text{subject to} \quad \forall \vec{x} \in \mathcal{X}^{\times L}, \quad M_{\vec{x}} \geqslant 0,$$
$$\sum_{\vec{x}} M_{\vec{x}} = \mathbb{I},$$

where

$$\tau_{\vec{x}} = \sum_{b=1}^{L} p_{x^{(b)}b} \, \rho_{x^{(b)}b}. \tag{3}$$

By forming the Lagrangian, we can easily compute the dual of this SDP (see, e.g., [17], Appendix]) which is given by

$$\text{minimize} \quad v_{\text{dual}} = \text{tr}(Q)$$
$$\text{subject to} \quad \forall \vec{x} \in \mathcal{X}^{\times L}, \quad Q \geqslant \tau_{\vec{x}}.$$

SDPs can be solved in polynomial time (in the input size) using standard algorithms [18], which also provide us with the optimal measurement operators.

B. Optimality conditions

However, with the SDP formalism in mind, it is now also easy to provide necessary and sufficient conditions for when a set of measurement operators $\{M_{\vec{x}}\}_{\vec{x}}$ is in fact optimal. Similar conditions were derived for the case of state discrimination *without* postmeasurement information [7,19–23]. A proof can be found in the Appendix.

Lemma 1. A POVM with operators $\{M_{\vec{x}}\}_{\vec{x}}$ is optimal for state discrimination with postmeasurement information for

___

[2]An ensemble is classical if the states ρ_{xb} all commute.

the ensemble $\mathcal{E} = \{p_{xb}, \rho_{xb}\}$ if and only if the following two conditions hold:

1. $Q := \sum_{\vec{x}} \tau_{\vec{x}} M_{\vec{x}}$ is Hermitian.
2. $Q \geqslant \tau_{\vec{x}}$ for all $\vec{x} \in \mathcal{X}^{\times L}$.

C. Upper bound

We now derive a simple upper bound on the success probability of state discrimination with postmeasurement information when $p_{xb} = p_x p_b$ is a product distribution, and the string x is chosen uniformly at random (i.e., $p_x = 1/|\mathcal{X}|$). We will use a trick employed by Ogawa and Nagaoka [9] in the context of channel coding which was later rediscovered in the context of state discrimination [24]. A proof can be found in the Appendix.

Lemma 2. Let $N = |\mathcal{X}|$ be the number of possible strings, and suppose that the joint distribution over strings and encodings satisfies $p_{xb} = p_b/N$, where the distribution $\{p_b\}_b$ is arbitrary. Then

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \leqslant \frac{1}{N} \text{tr}\left[\left(\sum_{\vec{x}} \rho_{\vec{x}}^{\alpha} \right)^{1/\alpha} \right], \tag{4}$$

for all $\alpha > 1$, where $\mathcal{E} = \{\rho_{xb}\}_{xb}$, $P = \{p_{xb}\}_{xb}$, and $\rho_{\vec{x}} = \sum_{b=1}^{L} p_b \, \rho_{x^{(b)}b}$.

Note that the bound on the right-hand side (RHS) contains very many terms, and yet our normalization factor is only $1/N$. Nevertheless, for many interesting examples we can obtain a useful bound this way, by choosing α to be sufficiently large.

D. Lower bound

Similarly, if x is chosen uniformly at random and independent of the encoding, we can find a lower bound to $p_{\text{succ}}^{\text{PI}}$. The idea behind this lower bound is to subdivide the problem into a set of smaller problems which we can solve using standard techniques from state discrimination. Note that, without loss of generality, we can label the elements of $\mathcal{X}$ that we wish to encode from $0, \ldots, N-1$, where we let $N = |\mathcal{X}|$. The vector $\vec{x}$ can thus be written analogously as a vector $\vec{x} \in \{0, \ldots, N-1\}^{\times L}$. We now partition the set of all possible such vectors as follows. Consider a shorter vector of length $L - 1$, that is, $\vec{y} \in \{0, \ldots, N-1\}^{\times (L-1)}$. With every such vector, we associate the partition

$$T_{\vec{y}} = \{\vec{x} = (y^{(1)} + j \bmod N, \ldots, y^{(L-1)} + j \bmod N,$$
$$0 + j \bmod N) \mid j \in \{0, \ldots, N-1\}\}. \tag{5}$$

Note that $|T_{\vec{y}}| = N$ and if $\vec{y} \neq \vec{\hat{y}}$ we have $T_{\vec{y}} \cap T_{\vec{\hat{y}}} = \emptyset$. The union of all such partitions gives us the set of all possible vectors $\vec{x}$, that is,

$$\bigcup_{\vec{y}} T_{\vec{y}} = \{\vec{x} \mid \vec{x} \in \{0, \ldots, N-1\}^{\times L}\}. \tag{6}$$

With every partition $T_{\vec{y}}$ we can now associate a standard state discrimination problem *without* postmeasurement information in which we try to discriminate states

$$\rho_{\vec{x}} := \sum_{b=1}^{L} p_b \rho_{x^{(b)}b}, \tag{7}$$

such that $\vec{x} \in T_{\vec{y}}$. That is, the set of states is given by $\mathcal{E}_{T_{\vec{y}}} = \{\rho_{\vec{x}} \mid \vec{x} \in T_{\vec{y}}\}$ and $p_{\vec{x}} = 1/N$ is the uniform distribution. Note that the original problem of state discrimination where we do not receive any postmeasurement information corresponds to the partition given by $\vec{y} = (0, \ldots, 0)$, where we always give the same answer no matter what the postmeasurement information is going to be. We prove the following lemma in the Appendix.

Lemma 3. The success probability *with* postmeasurement information is at least as large as the success probability of a derived problem *without* postmeasurement information, i.e.,

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \geqslant \max_{\vec{y}} p_{\text{succ}}(\mathcal{E}_{T_{\vec{y}}}, P).$$

In particular, this allows us to apply any known lower bounds for the standard task of state discrimination [25] to this problem. Curiously, we will see that there exists a large class of problems for which this bound is tight, even though $\Delta(\mathcal{E}, P) > 0$, that is, even though postmeasurement information is useful.

III. TIGHT BOUNDS FOR SPECIAL ENCODINGS

We now consider a very special class of problems called *Clifford encodings*, for which we can determine the optimal measurement explicitly. In this problem, we will only ever encode a single bit $x \in \{0, 1\}$ chosen uniformly at random independent of the choice of encoding, and take $(d = 2^n)$-dimensional states of the form

$$\rho_{xb} = \frac{1}{d} \left(\mathbb{I} + \sum_{j=1}^{2n+1} \gamma_{xb}^{(j)} \Gamma_j \right), \tag{8}$$

where $\Gamma_1, \ldots, \Gamma_{2n+1}$ are generators of the Clifford algebra, that is, anticommuting operators[3] satisfying $(\Gamma_j)^2 = \mathbb{I}$ for all j. We also assume that the vector $\gamma_{xb} = (\gamma_{xb}^{(1)}, \ldots, \gamma_{xb}^{(2n+1)})$ satisfies $\gamma_{xb} = -\gamma_{(1-x)b}$ and $\|\gamma_{xb}\|_2 \leqslant 1$. The distribution over encodings can be arbitrary. Using the fact that the operators anticommute, it is not hard to see that $\text{tr}(\Gamma_j \Gamma_k) = 0$ for $j \neq k$ and the latter condition then ensures that ρ_{xb} is a valid quantum state [26], that is, ρ_{xb} is positive semidefinite satisfying $\text{tr}(\rho_{xb}) = 1$. The Clifford algebra has a unique representation by Hermitian matrices on n qubits (up to unitary equivalence) which we fix henceforth. This representation can be obtained via the famous Jordan-Wigner transformation [27]:

$$\Gamma_{2j-1} = Y^{\otimes(j-1)} \otimes Z \otimes \mathbb{I}^{\otimes(n-j)},$$
$$\Gamma_{2j} = Y^{\otimes(j-1)} \otimes X \otimes \mathbb{I}^{\otimes(n-j)},$$

for $j = 1, \ldots, n$, where we use X, Y, and Z to denote the Pauli matrices. We also use $\Gamma_{2n+1} = i\Gamma_1 \cdots \Gamma_{2n}$.

Note that in dimension $d = 2$, these operators are simply the Pauli matrices $\Gamma_1 = Z$, $\Gamma_2 = X$, and $\Gamma_{2n+1} = Y$ and *any* encoding of the bit x into two orthogonal pure states is of the above form. A simple example is the BB84 encoding [1] where we encode the bit x into the computational basis labeled by $b = 0$ and into the Hadamard basis labeled by $b = 1$.

---

[3]That is, $\{\Gamma_j, \Gamma_k\} = \Gamma_j \Gamma_k + \Gamma_k \Gamma_j = 0$ for $j \neq k$.

Furthermore, if we have only two possible strings and encodings, we can always reduce the problem to dimension $d = 2$ [13,14]. In higher dimensions, encodings of the above form were suggested for the use in cryptographic protocols [26].

A. Without postmeasurement information

We now first examine the setting of state discrimination *without* postmeasurement information, which will provide us with the necessary intuition. Again, we use $L = |\mathcal{B}|$ to denote the number of possible encodings. Recall the average state $\rho_{\vec{x}}$ from (7) for the vector $\vec{x} = (x^{(1)}, \ldots, x^{(L)})$, which tells us for every possible encoding which bit appears in the sum. We furthermore define the complementary vector $\underline{\vec{x}} = ((1 - x^{(1)}), \ldots, (1 - x^{(L)}))$, that is, $\vec{x} + \underline{\vec{x}} = \vec{0}$. As a warmup, suppose we are given $\rho_{\vec{x}}$ and $\rho_{\underline{\vec{x}}}$ chosen uniformly at random and wish to determine which one. Clearly, this is an example of state discrimination *without* postmeasurement information, which can also be written as an SDP [23,28]. The primal is of the form

$$\begin{aligned}
\text{maximize} \quad & \tfrac{1}{2}[\text{tr}(M_{\vec{x}}\rho_{\vec{x}}) + \text{tr}(M_{\underline{\vec{x}}}\rho_{\underline{\vec{x}}})] \\
\text{subject to} \quad & M_{\vec{x}} \geqslant 0 \\
& M_{\underline{\vec{x}}} \geqslant 0 \\
& M_{\vec{x}} + M_{\underline{\vec{x}}} = \mathbb{I}.
\end{aligned}$$

Its dual is easily found to be

$$\begin{aligned}
\text{minimize} \quad & \text{tr}(Q) \\
\text{subject to} \quad & Q \geqslant \tfrac{1}{2}\rho_{\vec{x}} \\
& Q \geqslant \tfrac{1}{2}\rho_{\underline{\vec{x}}}.
\end{aligned}$$

Analogous to Lemma 1 with $\tau_{\vec{x}} = \tfrac{1}{2}\rho_{\vec{x}}$ one can derive optimality conditions which for the case of state discrimination were previously obtained in [7,19–23]. In our case they tell us that $Q = \tfrac{1}{2}(\rho_{\vec{x}}M_{\vec{x}} + \rho_{\underline{\vec{x}}}M_{\underline{\vec{x}}})$ must be Hermitian, and Q is a feasible dual solution. All we have to do is thus to guess an optimal measurement, and use these conditions to prove its optimality. Consider the operators

$$\begin{aligned}
M_{\vec{x}} &= \frac{1}{2}\left(\mathbb{I} + \sum_j a_{\vec{x}}^{(j)}\Gamma_j\right), \\
M_{\underline{\vec{x}}} &= \frac{1}{2}\left(\mathbb{I} - \sum_j a_{\vec{x}}^{(j)}\Gamma_j\right),
\end{aligned} \quad (9)$$

where $\vec{a}_{\vec{x}} = \vec{v}_{\vec{x}}/\|\vec{v}_{\vec{x}}\|_2$ is the normalized average vector

$$\vec{v}_{\vec{x}} = \sum_{b=1}^{L} p_b \gamma_{x^{(b)}b}. \quad (10)$$

Note that, since the generators of the Clifford algebra anticommute, we have that $M_{\vec{x}}, M_{\underline{\vec{x}}} \geqslant 0$ and $M_{\vec{x}} + M_{\underline{\vec{x}}} = \mathbb{I}$. Hence, these operators do form a valid measurement. In the Appendix, we derive two lemmas which show that $Q = \tfrac{1}{2}(\rho_{\vec{x}}M_{\vec{x}} + \rho_{\underline{\vec{x}}}M_{\underline{\vec{x}}})$ is Hermitian (Lemma 9) and satisfies $Q \geqslant \tfrac{1}{2}\rho_{\vec{x}}$ for all $\vec{x}$ (Lemmas 9 and 10),[4] which are the conditions we needed for optimality. All proofs can be found in the Appendix.

––––––––––

[4]Recall that for any Hermitian operator we have $\lambda_{\max}(A)\mathbb{I} \geqslant A$, where $\lambda_{\max}(A)$ is the largest eigenvalue of A.

Theorem 1. The measurements given in (9) are optimal to discriminate $\rho_{\vec{x}}$ from $\rho_{\underline{\vec{x}}}$ chosen with equal probability.

B. With postmeasurement information

We are now ready to determine the optimal measurements for the case *with* postmeasurement information. First of all, recall from Lemma 3 that we can subdivide our problem into smaller parts by partitioning the set of strings $\vec{x}$. Applied to the present case, these partitions are simply given by

$$\tilde{T}_{\vec{x}} = \{\vec{x}, \underline{\vec{x}}\}, \quad (11)$$

where for simplicity we here use the vector $\vec{x}$ itself to label the partition. Note that by Lemma 3 we thus have that

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \geqslant \max_{\vec{x}} p_{\text{succ}}(\mathcal{E}_{\tilde{T}_{\vec{x}}}). \quad (12)$$

We show in the Appendix that this bound is in fact tight.

Lemma 4. For Clifford encodings

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = \max_{\vec{x}} p_{\text{succ}}(\mathcal{E}_{\tilde{T}_{\vec{x}}}), \quad (13)$$

and postmeasurement information is useless if and only if the maximum on the RHS is attained by $\vec{x} = (0, \ldots, 0)$.

Note that the optimal measurement is thus given by (9) for the vector $\vec{x}$ maximizing the RHS of (12), and letting all other $M_{\underline{\vec{x}}} = 0$. This shows that for our class of problems the problem of finding the optimal measurement can be simplified considerably and is easily evaluated.

It is a very useful consequence of our analysis that for any cryptographic application that makes use of such encodings, we can always perform a relabeling of states ρ_{xb} such that postmeasurement information becomes useless. More precisely, we will associate $\vec{x}$ with the new all $(0, \ldots, 0)$ vector and $\underline{\vec{x}}$ with the new $(1, \ldots, 1)$ vector. That is, for the optimal vector $\vec{x}$ we let

$$\rho_{0b}^{\text{new}} := \rho_{x^{(b)}b}, \quad (14)$$
$$\rho_{1b}^{\text{new}} := \rho_{(1-x^{(b)})b}. \quad (15)$$

Clearly, by Lemma 4 we then have for $\mathcal{E}^{\text{new}} = \{\rho_{xb}^{\text{new}}\}_{xb}$ that

$$\Delta(\mathcal{E}^{\text{new}}, P) = 0, \quad (16)$$

as desired.

C. Example

We now consider a small example that illustrates how our statement applies to the case where we have only two possible encodings $\mathcal{B} = \{0,1\}$ into two orthogonal pure states in dimension $d = 2$, and we choose the encoding uniformly at random ($p_b = 1/2$). A simple example is encoding into the BB84 bases [1], where we pick the computational basis for $b = 0$ and the Hadamard basis for $b = 1$. We now show that in two dimensions, postmeasurement information is useless if and only if the angle between the Bloch vectors for the states ρ_{00} and ρ_{01} obeys $\theta \leqslant \frac{\pi}{2}$ as illustrated in Figs. 3 and 4.

Note that in this example the average states are given by

$$\rho_{(0,0)} = \tfrac{1}{2}(\rho_{00} + \rho_{01}), \quad (17)$$
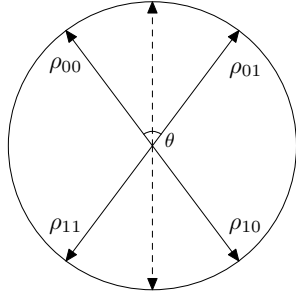$$\rho_{(1,1)} = \tfrac{1}{2}(\rho_{10} + \rho_{11}), \quad (18)$$

FIG. 3. Postmeasurement information is useless if and only if $\theta \leqslant \pi/2$. The dashed line corresponds to the Bloch vector of the optimal measurement using postmeasurement information consisting of two rank-1 projectors M_{00} and M_{11}, which is the same measurement one would make for standard state discrimination. We output the same bit, no matter what encoding information b we receive.

$$\rho_{(0,1)} = \tfrac{1}{2}(\rho_{00} + \rho_{11}), \tag{19}$$

$$\rho_{(1,0)} = \tfrac{1}{2}(\rho_{10} + \rho_{01}). \tag{20}$$

The two partitions we are considering are $\tilde{T}_{(0,0)} = \{(0,0),(1,1)\}$ and $\tilde{T}_{(0,1)} = \{(0,1),(1,0)\}$. Let $\vec{v}_0$ and $\vec{v}_1$ be the Bloch vectors corresponding to the states ρ_{00} and ρ_{01} respectively. We have from Lemma 10 that

$$\lambda_{\max}(\rho_{\vec{x}}) = \lambda_{\max}(\rho_{\underline{\vec{x}}}) \tag{21}$$

$$= \begin{cases} \tfrac{1}{2}(\mathbb{I} + \|v_0 + v_1\|_2) & \text{for } \vec{x} = (0,0), \\ \tfrac{1}{2}(\mathbb{I} + \|v_0 - v_1\|_2) & \text{for } \vec{x} = (0,1). \end{cases} \tag{22}$$

Hence, by Lemma 4 postmeasurement information is useless if and only if

$$\|v_0 + v_1\|_2 \geqslant \|v_0 - v_1\|_2. \tag{23}$$

Since $\|v_0\|_2 = \|v_1\|_2 = 1$ for pure states, we have $\|v_0 + v_1\|_2 = 2 + 2\cos\theta$ and $\|v_0 - v_1\|_2 = 2 - 2\cos\theta$ and thus (23) holds if and only if $\theta \leqslant \frac{\pi}{2}$. The optimal measurement is again given by (9). Note that this is rather intuitive, since for partition $\tilde{T}_{(0,0)}$ we always give the same answer, no matter what postmeasurement information we receive.
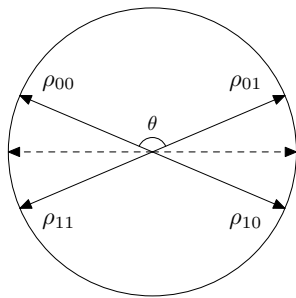

FIG. 4. Postmeasurement information is useful for $\theta > \pi/2$. The dashed line corresponds to the Bloch vector of the optimal measurement using postmeasurement information consisting of two rank-1 projectors M_{01} and M_{10}, which is the measurement one would make in standard state discrimination, if we were to distinguish $(\rho_{00} + \rho_{11})/2$ from $(\rho_{01} + \rho_{10})/2$. Which bit we output depends on the postmeasurement information we receive.

IV. CLASSICAL ENSEMBLES

We saw above that for the case of Clifford encodings even if postmeasurement information was useful for the original problem, that is, $p_{\text{succ}}(\mathcal{E}, P) < p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P)$, we could always perform a relabeling to obtain a new problem for which postmeasurement information is useless. We now show that this is a unique quantum feature and is not present in analogous classical problems as long as we are able to gain some information even without postmeasurement information, i.e., $p_{\text{succ}}(\mathcal{E}, P) > 1/|\mathcal{X}|$. We thereby call a problem *classical* if and only if all states ρ_{xb} commute.

We again focus on the case where we wish to encode a single bit $x \in \{0,1\}$. Let Π_{xb} be a projector onto the support of ρ_{xb}. For simplicity, we will assume in the following that $\Pi_{0b} + \Pi_{1b} = \mathbb{I}$ for all encodings b, and that the projectors are of equal rank $r = \text{rank}(\Pi_{0b}) = \text{rank}(\Pi_{1b})$. We also assume that $\rho_{xb} = \Pi_{xb}/r$. It is straightforward to extend our argument to a more general case, but makes it more difficult to follow our idea.

In [[13], Lemma 5.1] it was shown that if $[P_{xb}, P_{x'b'}] = 0$ for all bits x, x' and encodings b, b' of this form

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = 1. \tag{24}$$

Recall that we are interested in the case where $p_{\text{succ}}(\mathcal{E}, P) < p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P)$. Hence, our goal will be to show that there exists no relabeling as in the previous section that allows us to create a new problem $\mathcal{E}^{\text{new}}$ for which $p_{\text{succ}}(\mathcal{E}^{\text{new}}, P) = p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = 1$.

A. Nonlocal games

To show our result, we will need the notion of nonlocal games which are a different way of looking at Bell inequalities [15]. For example, the well-known CHSH inequality [16] takes the following form when converted to a game. Imagine two spacelike separated parties, Alice and Bob. We choose two questions $s, t \in \{0,1\}$ uniformly at random and send them to Alice and Bob, respectively. The rules are that they win the game if and only if they manage to return answers $a, b \in \{0,1\}$ such that $st = a + b \bmod 2$. Without loss of generality, we may thereby assume that Alice and Bob perform a measurement depending on the question they receive, and simply return the outcome of that measurement. To help them win the game, Alice and Bob may thereby agree on any shared state and measurements ahead of time, but are no longer able to communicate once the game starts. The average probability that they win the game is thus

$$p_{\text{win}} = \max \frac{1}{4} \sum_{s,t} \sum_{\substack{a,b \\ a+b \bmod 2 = st}} \Pr[a,b \mid s,t], \tag{25}$$

where $\Pr[a,b \mid s,t]$ is the probability that they return answers a and b given questions s and t, and the maximization is over all states and measurements allowed in a particular theory. Classically, we have

$$p_{\text{win}}^{\text{classical}} = \tfrac{3}{4}. \tag{26}$$

In a quantum world, however, Alice and Bob can achieve

$$p_{\text{win}}^{\text{quantum}} = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.853. \tag{27}$$

More general nonlocal games are of course possible, where we may have a larger number of questions and answers, and the rules of the game may be more complicated.

Of central importance to us will be the fact that if Alice's (or Bob's) measurements commute, then there exists a classical strategy that achieves the same winning probability (see, e.g., [17]). We now use this fact to prove our result.

B. A classical-quantum gap

To explain the main idea behind our construction, we focus on the case where we only have two possible encoding $L = 2$. That is, $\mathcal{B} = \{0,1\}$ and $\mathcal{X} = \{0,1\}$. We also assume that the bit x, as well as the encoding b is chosen uniformly and independently at random. The states defining our problem are thus $\rho_{00}, \rho_{01}, \rho_{10}$, and ρ_{11}. We again consider the two partitions labeled by $\vec{x} \in \{0,1\}^2$ given by

$$\tilde{T}_{(0,0)} = \{(0,0),(1,1)\}, \tag{28}$$
$$\tilde{T}_{(0,1)} = \{(0,1),(1,0)\}. \tag{29}$$

As before, we can associate a standard state discrimination problem with each of these partitions. For the first partition $\tilde{T}_{(0,0)}$ as wish to discriminate between the states $\rho_{(0,0)}$ and $\rho_{(1,1)}$ specified by (17) and (18) where we are given one of the two states with equal probability. Let p_1 denote the success probability of solving this problem, maximized over all possible measurements. Note that our condition of being able to gain some information in the state discrimination problem corresponds to having

$$\tfrac{1}{2} < p_1. \tag{30}$$

For the second partition $\tilde{T}_{(0,1)}$, we wish to discriminate between $\rho_{(0,1)}$ and $\rho_{(1,0)}$ from (19) and (20), again given with equal probability. Let p_2 denote the corresponding success probability for the second partition. Note that since we have only two possible partitions here constructed in the way outlined in Sec. III, our goal of showing that there exists no relabeling that makes postmeasurement information useless can be rephrased as showing that $p_2 < 1$.

We now show that these two state discrimination problems arise naturally in the CHSH game. In particular, we prove the following lemma in the Appendix.

Lemma 5. There exists a strategy for Alice and Bob to succeed at the CHSH game with probability $(p_1 + p_2)/2$, where Alice's measurements are given by the projectors $\{P_{00}, P_{10}\}$ and $\{P_{01}, P_{11}\}$.

However, recall that if the ensemble of states is classical the projectors P_{xb} all commute, and hence there exists a classical strategy for Alice and Bob that also achieves a winning probability of $(p_1 + p_2)/2$. Hence, by (26) we must have

$$\frac{p_1 + p_2}{2} \leqslant \frac{3}{4}. \tag{31}$$

Using (30) this implies $p_2 \leqslant 3/2 - p_1 < 1 = p_{\text{succ}}^{\text{PI}}$, and hence the relabeling corresponding to the second partition cannot make postmeasurement information useless. To summarize we obtain the following theorem.[5]

---

[5]Any relabeling that relabels at least one ρ_{xb} is called nontrivial.

Theorem 2. For the case of two encodings of a single bit chosen uniformly at random (i.e., $p_{xb} = 1/4$), which do allow us to gain some information even without postmeasurement information ($p_{\text{succ}} > 1/2$), there exists no nontrivial relabeling that renders postmeasurement information useless.

Note that if we are able to gain some information in both state discrimination problems, i.e., $p_1, p_2 > 1/2$, the preceding discussion also implies that $p_1, p_2 < 1$, that is, postmeasurement information is never useless. Bounds on Bell inequalities corresponding to bounds on the maximum winning probability that can be achieved in a classical world can thus allow us to place bounds on how well we can solve state discrimination problems *without* postmeasurement information.

This is in stark contrast to the quantum setting. For example, for the BB84 encodings it is not hard to see that $p_1 = p_2 = p_{\text{succ}}^{\text{PI}} \approx 0.853$ [13], and hence postmeasurement information is always useless. Yet there exist classical encodings [13] for which $p_1 = p_2 = 3/4$ but $p_{\text{succ}}^{\text{PI}} = 1$.

To analyze the case of multiple encodings, we have to consider more complicated games than the one obtained from the CHSH inequality. A natural choice is to consider games in which Bob has to solve different state discrimination problems corresponding to different partitions of the vectors $\vec{x}$ depending on his question t in the game. To make a fully general statement we would like to include all possible partitions. Clearly, however the above approach can also be used to place bounds on the average of success probabilities for a subset of partitions by defining a game with less questions, and evaluating it's maximum classical winning probability.

V. CONCLUSIONS

Our work raises several immediate open questions. First of all, can we obtain sharper bounds? Since solving an SDP numerically is still very expensive in higher dimensions, it would also be interesting to prove bounds on how well generic measurements such as the square-root measurement (also known as the pretty good measurement [29]) perform. The pretty good measurement is a special case of Belavkin's weighted measurements [7,30,31], which was already used in its cube weighted form in [13] to provide bounds on the state discrimination with postmeasurement information. Such bounds have most recently been shown by Tyson [32] for standard state discrimination. Yet no good bounds are known on how well such measurements perform for our task. More generally, it would be very interesting to see whether one can adapt the iterative procedures investigated in [33–36] to find optimal measurements for the case of standard state discrimination without postmeasurement information to this setting. Concerning such iterative procedures, we would like to draw special attention to the recent work by Tyson [25] generalizing monotonicity results for such iterates [37], which could be applied here.

Naturally, it would be very interesting to know if our results for Clifford encodings can be extended to a more general setting. Our discussion of classical ensembles shows that there exist problems for which $p_{\text{succ}} < p_{\text{succ}}^{\text{PI}}$ no matter what relabeling we perform [13], and hence we cannot hope that

a similar statement holds in general. Nevertheless, it would be interesting to obtain necessary and sufficient conditions for when postmeasurement is already useless, or otherwise can be made useless by performing a relabeling.

APPENDIX

In this appendix, we provide the technical details of our claims. For ease of reading, we thereby provide the proofs together with the statement of the lemmas.

1 Proofs of Sec. II

A. Optimality conditions

Lemma 6. A POVM with operators $\{M_{\vec{x}}\}_{\vec{x}}$ is optimal for state discrimination with postmeasurement information for the ensemble $\mathcal{E} = \{p_{xb}, \rho_{xb}\}$ if and only if the following two conditions hold:

1. $Q := \sum_{\vec{x}} \tau_{\vec{x}} M_{\vec{x}}$ is Hermitian.
2. $Q \geqslant \tau_{\vec{x}}$ for all $\vec{x} \in \mathcal{X}^{\times L}$.

Proof. Suppose first that the two conditions hold. Note that condition (2) tells us that Q is a feasible solution, that is, it satisfies all constraints for the dual SDP. By weak duality of SDPs we thus have $v_{\text{primal}} \leqslant v_{\text{dual}} \leqslant \text{tr}(Q)$, and from condition (1) we also have that $\text{tr}(Q) = \sum_{\vec{x}} \text{tr}(M_{\vec{x}} \tau_{\vec{x}}) \leqslant v_{\text{primal}}$. Hence the POVM forms an optimal solution for the SDP.

Conversely, suppose that $\{M_{\vec{x}}\}_{\vec{x}}$ is an optimal solution for the primal SDP. Let Q be the optimal solution for the dual SDP. Note that this means that Q already satisfies condition (2), and all that remains is to show that Q has the desired form given by condition (1). Since $M_{\vec{x}} = \mathbb{I}/|\mathcal{X}^{\times L}|$ is a feasible solution for the primal SDP, we have by Slater's condition [18] that the optimal values v_{primal}^* and v_{dual}^* are equal, i.e., $v_{\text{primal}}^* = v_{\text{dual}}^*$. Using the fact that $\sum_{\vec{x}} M_{\vec{x}} = \mathbb{I}$ and that the trace is cyclic we thus have

$$\text{tr}(Q) - \sum_{\vec{x}} \text{tr}(M_{\vec{x}} \tau_{\vec{x}}) = \sum_{\vec{x}} \text{tr}[(Q - \tau_{\vec{x}}) M_{\vec{x}}] = 0. \quad \text{(A1)}$$

Since $Q \geqslant \tau_{\vec{x}}$ (equivalently $Q - \tau_{\vec{x}} \geqslant 0$), and $M_{\vec{x}} \geqslant 0$ for all $\vec{x}$ we have that all the terms $\text{tr}[(Q - \tau_{\vec{x}}) M_{\vec{x}}]$ in the sum are positive and hence we must have for all $\vec{x}$ that $\text{tr}[(Q - \tau_{\vec{x}}) M_{\vec{x}}] = 0$. Again using the fact that the two operators are positive semidefinite, and the cyclicity of the trace, we thus have for the optimal solution that

$$(Q - \tau_{\vec{x}}) M_{\vec{x}} = M_{\vec{x}} (Q - \tau_{\vec{x}}) = 0. \quad \text{(A2)}$$

Summing the LHS over all $\vec{x}$ and noting that $\sum_{\vec{x}} M_{\vec{x}} = \mathbb{I}$ then gives us condition (1). ∎

B. Upper bound

Lemma 7. Let $N = |\mathcal{X}|$ be the number of possible strings, and suppose that the joint distribution over strings and encodings satisfies $p_{xb} = p_b/N$, where the distribution $\{p_b\}_b$ is arbitrary. Then

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \leqslant \frac{1}{N} \text{tr}\left[\left(\sum_{\vec{x}} \rho_{\vec{x}}^{\alpha} \right)^{1/\alpha} \right], \quad \text{(A3)}$$

for all $\alpha > 1$, where $\mathcal{E} = \{\rho_{xb}\}_{xb}$, $P = \{p_{xb}\}_{xb}$ and $\rho_{\vec{x}} = \sum_{b=1}^{L} p_b \rho_{x_b b}$.

Proof. Note that since $y^{1/\alpha}$ is operator monotone for $\alpha > 1$ [[38], Theorem V.1.9] we have

$$\rho_{\vec{x}} = \left(\rho_{\vec{x}}^{\alpha} \right)^{1/\alpha} \leqslant \left(\sum_{\vec{x}} \rho_{\vec{x}}^{\alpha} \right)^{1/\alpha}. \quad \text{(A4)}$$

Using the fact that $\sum_{\vec{x}} M_{\vec{x}} = \mathbb{I}$ we hence obtain

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = \frac{1}{N} \sum_{\vec{x}} \text{tr}(M_{\vec{x}} \rho_{\vec{x}}) \quad \text{(A5)}$$

$$\leqslant \frac{1}{N} \sum_{\vec{x}} \text{tr}\left[M_{\vec{x}} \left(\sum_{\vec{x}} \rho_{\vec{x}}^{\alpha} \right)^{1/\alpha} \right] \quad \text{(A6)}$$

$$= \frac{1}{N} \text{tr}\left[\left(\sum_{\vec{x}} \rho_{\vec{x}}^{\alpha} \right)^{1/\alpha} \right], \quad \text{(A7)}$$

as promised. ∎

C. Lower bound

Lemma 8. The success probability *with* postmeasurement information is at least as large as the success probability of a derived problem *without* postmeasurement information, i.e.,

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) \geqslant \max_{\vec{y}} p_{\text{succ}}(\mathcal{E}_{T_{\vec{y}}}, P).$$

Proof. This follows immediately from the discussion by noting that

$$\sum_{\vec{x}} \text{tr}(M_{\vec{x}} \rho_{\vec{x}}) = \sum_{\vec{y}} \sum_{\vec{x} \in P_{\vec{y}}} \text{tr}(M_{\vec{x}} \rho_{\vec{x}}). \quad \text{(A8)}$$

∎

2 Proofs of Sec. III

A. Without postmeasurement information

Lemma 9. For the measurement defined by (9) we have

$$Q = \frac{1}{2}(\rho_{\vec{x}} M_{\vec{x}} + \rho_{\underline{x}} M_{\underline{x}}) = \frac{1}{2d}(1 + \|\vec{v}_{\vec{x}}\|_2)\mathbb{I}, \quad \text{(A9)}$$

and hence Q is Hermitian.

Proof. We use the shorthand $\vec{a}_{\vec{x}} \cdot \vec{\Gamma} = \sum_j a_{\vec{x}}^{(j)} \Gamma_j$. We have

$$\rho_{\vec{x}} M_{\vec{x}} = \frac{1}{2d}[\mathbb{I} + (\vec{v}_{\vec{x}} + \vec{a}_{\vec{x}}) \cdot \vec{\Gamma} + (\vec{v}_{\vec{x}} \cdot \vec{a}_{\vec{x}})\mathbb{I}], \quad \text{(A10)}$$

$$\rho_{\underline{x}} M_{\underline{x}} = \frac{1}{2d}[\mathbb{I} - (\vec{v}_{\vec{x}} + \vec{a}_{\vec{x}}) \cdot \vec{\Gamma} + (\vec{v}_{\vec{x}} \cdot \vec{a}_{\vec{x}})\mathbb{I}], \quad \text{(A11)}$$

where the equality follows from the fact that

$$(\vec{v}_{\vec{x}} \cdot \vec{\Gamma})(\vec{a}_{\vec{x}} \cdot \vec{\Gamma}) = \frac{1}{2} \sum_{jk} v_{\vec{x}}^{(j)} a_{\vec{x}}^{(k)} \{\Gamma_j, \Gamma_k\} \tag{A12}$$

$$= (\vec{v}_{\vec{x}} \cdot \vec{a}_{\vec{x}}) \mathbb{I}. \tag{A13}$$

Using that $\vec{v}_{\vec{x}} \cdot \vec{v}_{\vec{x}} = \|\vec{v}_{\vec{x}}\|_2^2$ gives our claim. ∎

Lemma 10. The largest eigenvalue of $\rho_{\vec{x}}$ and $\rho_{\underline{\vec{x}}}$ is given by

$$\lambda_{\max}(\rho_{\vec{x}}) = \lambda_{\max}(\rho_{\underline{\vec{x}}}) = \frac{1}{d}(1 + \|\vec{v}_{\vec{x}}\|_2). \tag{A14}$$

Proof. We now prove our claim for $\rho_{\vec{x}}$. Our goal is to evaluate

$$\lambda_{\max}(\rho_{\vec{x}}) = \max_{\sigma} \operatorname{tr}(\sigma \rho_{\vec{x}}), \tag{A15}$$

where the maximization is taken over all states σ. Using the fact that the set of operators $\{\mathbb{I}, \Gamma_j, i\Gamma_j\Gamma_k, \ldots\}_{jk\ldots}$ forms an orthonormal (with respect to the Hilbert-Schmidt inner product) basis for the $d \times d$ Hermitian matrices we can write

$$\sigma = \frac{1}{d}\left(\mathbb{I} + \sum_j s^{(j)}\Gamma_j + \cdots\right). \tag{A16}$$

Since $\operatorname{tr}(\Gamma_j\Gamma_k) = 0$ for $j \neq k$, and we can rewrite $\rho_{\vec{x}} = \frac{1}{d}(\mathbb{I} + \sum_j a_{\vec{x}}^{(j)}\Gamma_j)$, this gives us

$$\operatorname{tr}(\sigma \rho_{\vec{x}}) = \frac{1}{d}(1 + \vec{s} \cdot \vec{v}_{\vec{x}}), \tag{A17}$$

where $\vec{s} = (s^{(1)}, \ldots, s^{(2n+1)})$ and $\cdot$ denotes the Euclidean inner product. Since $\sigma \geqslant 0$ if and only if $\|\vec{s}\|_2 \leqslant 1$ [13], we have that the maximum in (A17) is attained for $\sigma = (\mathbb{I} + \sum_j s^{(j)}\Gamma_j)/d$ with

$$\vec{s} = \frac{\vec{v}_{\vec{x}}}{\|\vec{v}_{\vec{x}}\|_2}, \tag{A18}$$

which gives our claim. The argument for $\rho_{\underline{\vec{x}}} = \frac{1}{d}(\mathbb{I} - \sum_j a_{\vec{x}}^{(j)}\Gamma_j)$ is analogous. ∎

B. With postmeasurement information

Lemma 11. For our class of problems

$$p_{\text{succ}}^{\text{PI}}(\mathcal{E}, P) = \max_{\vec{x}} p_{\text{succ}}(\mathcal{E}_{\bar{T}_{\vec{x}}}), \tag{A19}$$

and postmeasurement information is useless if and only if the maximum on the RHS is attained by $\vec{x} = (0, \ldots, 0)$.

Proof. Let $\vec{x}$ be the string that achieves the optimum on the RHS of (12). We now claim that $Q = \frac{1}{2}(\rho_{\vec{x}} M_{\vec{x}} + \rho_{\underline{\vec{x}}} M_{\underline{\vec{x}}})$ is an optimal solution to the SDP for the problem of state discrimination *with* postmeasurement information. First of all, note that Lemma 9 gives us that Q is Hermitian. We then have by Lemma 10 that $Q \geqslant \frac{1}{2}\rho_{\tilde{\vec{x}}}$ *for all* possible $\tilde{\vec{x}}$. Our claim now follows from Lemma 1, and by noting that for the partition $\vec{x} = (0, \ldots, 0)$ we will always give the same answer, no matter what postmeasurement information we receive later on. ∎

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984), pp. 175–179.

[2] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[3] R. König, S. Wehner, and J. Wullschleger, e-print arXiv:0906.1030.

[4] S. Wehner, C. Schaffner, and B. M. Terhal, Phys. Rev. Lett. **100**, 220502 (2008).

[5] C. W. Helstrom, Inf. Control **10**, 254 (1967).

[6] A. S. Holevo, Prob. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Trans. **9**, 177 (1973)].

[7] V. P. Belavkin, Stochastics **1**, 315 (1975).

[8] S. M. Barnett and S. Croke, Adv. Opt. Photon. **1**, 238 (2009).

[9] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **45**, 2486 (1999).

[10] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).

[11] D. Bacon and T. Decker, Phys. Rev. A **77**, 032335 (2008).

[12] C. Moore and A. Russell, Quantum Inf. Comput. **7**, 752 (2007).

[13] M. Ballester, S. Wehner, and A. Winter, IEEE Trans. Inf. Theory **54**, 4183 (2008).

[14] P. Halmos, Trans. Am. Math. Soc. **144**, 381 (1969).

[15] J. S. Bell, Physics (Long Island City, NY) **1**, 195 (1965).

[16] J. Clauser, M. Horne, A. Shimony, and R. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[17] S. Wehner, Ph.D. thesis, University of Amsterdam, 2008, e-print arXiv:0806.3483.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[19] S. M. Barnett and S. Croke, J. Phys. A **42**, 062001 (2009).

[20] V. P. Belavkin and A. G. Vancjan, Radio Eng. Electron. Phys. **19**, 1397 (1974).

[21] A. S. Holevo, J. Multivariate Anal. **3**, 337 (1973).

[22] A. S. Holevo, Prob. Peredachi Inf. **10**, 51 (1974) [Probl. Inf. Trans. **10**, 317 (1974)].

[23] H. P. Yuen, R. S. Kennedy, and M. Lax, IEEE Trans. Inf. Theory **21**, 125 (1975).

[24] J. Tyson, J. Math. Phys. **50**, 032106 (2009).

[25] J. Tyson, e-print arXiv:0907.3386.

[26] S. Wehner and A. Winter, J. Math. Phys. **49**, 062105 (2008).

[27] P. Jordan and E. Wigner, Z. Phys. **47**, 631 (1928).

[28] Y. Eldar, IEEE Trans. Inf. Theory **49**, 446 (2003).

[29] P. Hausladen and W. Wootters, J. Mod. Opt. **41**, 2385 (1994).

[30] V. P. Belavkin, Radio Eng. Electron. Phys. **20**, 39 (1975).

[31] C. Mochon, Phys. Rev. A **75**, 042313 (2007).

[32] J. Tyson, Phys. Rev. A **79**, 032343 (2009).

[33] M. Jezek, J. Fiurasek, and Z. Hradil, Phys. Rev. A **68**, 012305 (2003).

[34] M. Jezek, J. Rehacek, and J. Fiurasek, Phys. Rev. A **65**, 060301 (2002).

[35] M. Reimpell and R. F. Werner, Phys. Rev. Lett. **94**, 080501 (2005).

[36] J. Tyson, e-print arXiv:0902.0395.

[37] M. Reimpell, Ph.D. thesis, Technische Universität Braunschweig, 2007.

[38] R. Bhatia, *Matrix Analysis* (Springer, Berlin, 1996).