Discrimination Power of a Quantum Detector

Christoph Hirche,[1] Masahito Hayashi,[2,3] Emilio Bagan,[1] and John Calsamiglia[1,*]
[1]Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain
[2]Graduate School of Mathematics, Nagoya University, Nagoya, Japan
[3]Centre for Quantum Technologies, National University of Singapore, Singapore
(Received 24 October 2016; published 17 April 2017)

We investigate the ability of a quantum measurement device to discriminate two states or, generically, two hypotheses. In full generality, the measurement can be performed a number n of times, and arbitrary preprocessing of the states and postprocessing of the obtained data are allowed. There is an intrinsic error associated with the measurement device, which we aim to quantify, that limits its discrimination power. We minimize various error probabilities (averaged or constrained) over all pairs of n-partite input states. These probabilities, or their exponential rates of decrease in the case of large n, give measures of the discrimination power of the device. For the asymptotic rate of the averaged error probability, we obtain a Chernoff-type bound, dual to the standard Chernoff bound for which the state pair is fixed and the optimization is over all measurements. The key point in the derivation is that identical copies of input states become optimal in asymptotic settings. Optimal asymptotic rates are also obtained for constrained error probabilities, dual to Stein's lemma and Hoeffding's bound. We further show that adaptive protocols where the state preparer gets feedback from the measurer do not improve the asymptotic rates. These rates thus quantify the ultimate discrimination power of a measurement device.

DOI: 10.1103/PhysRevLett.118.160502

Quantum-enabled technologies exploit the laws that govern the microscopic world to outperform their classical counterparts. Detectors, or measurement devices, are a key ingredient in quantum protocols. They are the interface that connects the microscopic world of quantum phenomena and the world of classical, macroscopically distinct, events that we observe. It is only through measurements that we can access the information residing in quantum systems and ultimately make use of any quantum advantage.

We often encounter experimental situations where measurement devices (e.g., Stern-Gerlach apparatus, heterodyne detectors, photon counters, fluorescence spectrometers) are a given. A natural question is then to ask about the ability or power of those devices to perform certain quantum information-processing tasks. The informational power of a measurement has been addressed in several ways [1], e.g., via the "intrinsic data" it provides [2] or the capacity of the quantum-classical channel it defines [1,3–6], or via some associated entropic quantities [7–10].

In this Letter we focus on what is arguably the most fundamental primitive in quantum information processing: state discrimination, or generically, quantum hypothesis testing. Our aim is to explore how well a quantum measurement device can discriminate two hypotheses. This problem is dual to that of exploring how well two given quantum states can be discriminated [11]. This is of practical interest since preparing states is often easier than tailoring optimal measurements for a given state pair.

Our main task is to discriminate two states ρ and σ using a given measurement device. We are interested in the most general scenario where the device can be used a number n of times. The given measurement device is the only means of extracting classical data from the quantum system; however, for better performance, one is free to apply any trace preserving quantum operation to the system prior to measuring. Likewise, we view data processing also as a free operation. It is then meaningful to ask what is the minimum error probability of discrimination between ρ and σ.

We wish to go a step further and minimize the error probability over all state pairs. This characterizes an intrinsic limitation on the discrimination performance of the measurement device since, in general, a device cannot perfectly discriminate two hypotheses, not even when they are given by orthogonal states. This characterization is of practical relevance since it sets the ultimate limit on the successful identification of two arbitrary states, for instance two critical phases of a quantum many-body system, when one is bound to a given type of measurements apparatus.

Special attention will be paid to asymptotically large n. We will prove that in this regime pairs of entangled states provide no advantage over those of i.i.d. states, of the form $\rho^{\otimes n}$. This is in sheer contrast with the dual problem where the measurement is optimized for fixed i.i.d. states; there is strong numerical evidence [12] that collective nonseparable measurements are required to attain the corresponding optimal exponential rate of the error probability, given by the quantum Chernoff bound [13,14]. In the proof, we approach state discrimination as a communication problem and allow for adaptive protocols. In this extended

hypothesis testing scenario, we show that adaptive protocols perform better than any fixed protocol, including those that use entangled inputs. A few examples for small n will be briefly discussed.

Before addressing the problem in detail, it will be helpful to recall a few definitions and results concerning hypothesis testing. Here, the so called null, H_0, and alternative, H_1, hypotheses refer respectively to two possible states, ρ, σ, of a quantum system S. In quantum hypothesis testing, one is confronted with the task of deciding which hypothesis holds by performing a measurement on S. With full generality, the measurement is defined by a two-outcome positive operator valued measure (POVM), $\mathcal{F} = \{F_0, F_1\}$ ($F_0,\ F_1 \geq 0$ and $F_0 + F_1 = \mathbb{1}$). Hypothesis H_0 (H_1) is accepted if and only if F_0 (F_1) clicks. Two error probabilities are defined: $\alpha = \mathrm{tr}(F_1\rho)$, false positive or type-I error; and $\beta = \mathrm{tr}(F_0\sigma)$, false negative or type-II error. Generically, a decrease in a type-I error results in an increase in a type-II error and vice versa. Depending on the problem at hand, one may need to know, e.g., β for a maximum allowed value of α, or one may instead be interested in the average error probability $p_{\mathrm{err}} = (\alpha + \beta)/2$, where for simplicity we assume equal priors for H_0 and H_1. This second possibility is known as symmetric hypothesis testing and leads to minimum error state discrimination [11], where p_{err} is minimized over all POVMs $\mathcal{F}$.

When H_0, H_1 refer to $\rho^{\otimes n}$, $\sigma^{\otimes n}$, i.e., to n i.i.d. copies of ρ, σ, the error probabilities generically fall off exponentially as n increases [15]. Then, one is usually interested in the corresponding exponential rates. For symmetric hypothesis testing, the optimal error rate is given by the quantum Chernoff bound [13,16]. Similarly, the exponential rate of β is given by Stein's lemma [17,18] if an upperbound is set on α, or by Hoeffding's bound [19–21] if instead a lowerbound is set on the exponential rate of α [22]. These asymptotic bounds have found many useful applications in quantum information theory, such as providing an alternative proof of the classical capacity of a quantum channel [23,24], giving operational meaning to abstract quantities [25], quantum reading [26], or in quantum illumination [27,28].

Coming back to our original problem, we wish to assess the discrimination power of a device given by a specific POVM, $\mathcal{E} = \{E_k\}_{k=1}^m$. Let us assume that the positive operators E_k (generically nonorthogonal) act on a finite d-dimensional Hilbert space, $\mathcal{H}_d$ [29], of the quantum system S. First, using free operations, we need to produce a valid POVM, $\mathcal{F} = \{F_0, F_1\}$, out of $\mathcal{E}$, to discriminate two states ρ and σ. This can be achieved [1] by grouping (postprocessing) the measurement outcomes, $\{1, 2, \ldots, m\}$, in two disjoint sets a, $\bar{a}$, and defining $E^a := \sum_{k\in a} E_k$, $E^{\bar{a}} := \sum_{k\in\bar{a}} E_k = \mathbb{1} - E^a$. Then, $\mathcal{F} = \{E^a_{\mathcal{M}}, E^{\bar{a}}_{\mathcal{M}}\}$, where $E^a_{\mathcal{M}} = \mathcal{M}^\dagger(E^a)$ (likewise for $E^{\bar{a}}_{\mathcal{M}}$), for a suitable trace preserving quantum operation $\mathcal{M}$ (preprocessing). The error probabilities thus read $\alpha = \mathrm{tr}(E^{\bar{a}}_{\mathcal{M}}\rho)$ and $\beta = \mathrm{tr}(E^a_{\mathcal{M}}\sigma)$.

In this single-shot scenario, we can now quantify the discrimination power of $\mathcal{E}$ by the minimum average error probability. It can be written as

$$p^*_{\mathrm{err}} = \min_a \min_{(\rho,\sigma)} \frac{1}{2}\{1 + \mathrm{tr}[E^a(\sigma - \rho)]\}, \tag{1}$$

where the minimization is over all partitions $\{a, \bar{a}\}$ of the outcome set (over all postprocessing operations) and over all state pairs (ρ, σ), so $\mathcal{M}$ can be dropped in the minimization. One can readily check [1] that the minimum single-shot error probability is given by the spread of E^a: $p^*_{\mathrm{err}} = 1/2 - \min_a(\lambda^a_{\max} - \lambda^a_{\min})/2$. This value is attained when ρ (σ) is the eigenstate of E^a corresponding to its maximum (minimum) eigenvalue.

The single-shot scenario above is too restrictive since one can easily envision discrimination settings where the measurement $\mathcal{E}$ is performed a number n of times. In the most general setting, a system consisting of n copies of S is prepared (by, say, Alice) in one of the states of the pair $(\rho^n,\ \sigma^n)$, corresponding respectively to hypotheses H_0 and H_1. Here, $\rho^n,\ \sigma^n \in \mathcal{S}(\mathcal{H}_d^{\otimes n})$, where $\mathcal{S}(\mathcal{H})$ stands for the set of density matrices on a Hilbert space $\mathcal{H}$, can be fully general, not just of the form $\rho^{\otimes n}$, $\sigma^{\otimes n}$. The measurer's (say, Bob's) goal is to tell which hypothesis is true by performing n measurements, all of them given by the POVM $\mathcal{E}$. Free operations include again preprocessing of $(\rho^n,\ \sigma^n)$ and postprocessing of the classical data gathered after each measurement. As in Eq. (1), when minimizing over state pairs, it is enough to choose the discriminating POVM as $\mathcal{F} = \{E^a, E^{\bar{a}} = \mathbb{1} - E^a\}$, where E^a has now the form

$$E^a = \sum_{\mathbf{k}^n\in a} E_{\mathbf{k}^n} := \sum_{\mathbf{k}^n\in a} \bigotimes_{i=1}^n E_{k_i}. \tag{2}$$

Here $\mathbf{k}^r := \{k_1, k_2, \ldots, k_r\}$ denotes a sequence of outcomes of length r ($\mathbf{k}^0 := \emptyset$), so $\mathbf{k}^n$ is obtained after completing all measurements. The two disjoint sets a and $\bar{a}$ now contain all the sequences assigned to the hypotheses H_0 and H_1, respectively. Type-I and type-II error probabilities are $\alpha_n = \mathrm{tr}(E^{\bar{a}}\rho^n)$ and $\beta_n = \mathrm{tr}(E^a\sigma^n)$, and the error probability for symmetric hypothesis testing can be written as $p_{\mathrm{err}} = \min_a(\alpha_n + \beta_n)/2$ [30].

It is not hard to see that the errors fall off exponentially with n [15,31]. It is then natural to quantify the discrimination power of $\mathcal{E}$ by the optimal asymptotic exponential rate of p_{err}, defined as

$$\zeta_{\mathrm{CB}} = -\lim_{n\to\infty} \min_{(\rho^n,\sigma^n)} \frac{1}{n} \log p_{\mathrm{err}}. \tag{3}$$

Although p_{err} can still be written as the spread of the optimal grouping, the number of groupings grows superexponentially with n. Moreover, very little is known about the spectrum of operator sums such as those in Eq. (2)

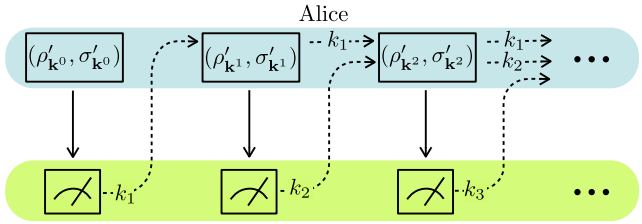

FIG. 1. Adaptive protocol. At each step (left to right), Alice sends to Bob (solid arrows) the state in Eq. (5), which she has prepared using Bob's feedback (dashed arrows).

and their eigenvectors (i.e., ρ^n and σ^n). A few examples that show the difficulties of dealing with finite n are given below. We will thus evaluate Eq. (3) following an alternative route, encapsulated in the following theorem, which is our main result.

Theorem 1: The optimal exponential rate defined in Eq. (3) is given by the classical Chernoff bound (CB) [22]:

$$\zeta_{\mathrm{CB}} = -\min_{(\rho,\sigma)} \min_{0\le s\le 1} \phi(s|P\|\bar{P}), \tag{4}$$

where $P_k = \mathrm{tr}(E_k\rho)$ and $\bar{P}_k = \mathrm{tr}(E_k\sigma)$ are the outcome probability distributions of (a single use of) the POVM $\mathcal{E}$, and $\phi(s|P\|\bar{P}) := \log \sum_k P_k^s \bar{P}_k{}^{1-s}$. This rate can be attained using i.i.d. states, $\rho^{\otimes n}$ and $\sigma^{\otimes n}$.

The main ingredient of the proof of Theorem 1 is to show that our problem is a particular case of *classical* channel discrimination. We will then use Ref. [32] to complete the proof.

To this end, let us momentarily broaden the scope of our original problem. First, we view hypothesis testing as a communication protocol where Alice (the state preparer) sends one of two possible messages, H_0, H_1, to Bob (the measurer) using suitable states in $\mathcal{S}(\mathcal{H}_d^{\otimes n})$. Bob is allowed to perform n measurements with his detector to identify with minimum error which of the messages Alice sent. Second, in this communication context it is natural to allow classical feedback from Bob to Alice after each measurement. This enables an adaptive protocol (see Fig. 1) in which Alice sends one state at a time to Bob's detector and waits for him to provide feedback on the obtained outcome. Alice uses this information to prepare the succeeding state in a way that minimizes the identification error. Such protocols are widely used in quantum information theory [33–37], particularly in quantum channel discrimination [35,38,39]. It follows from the structure of $E_{\mathbf{k}^n}$ in Eq. (2), that

Lemma 1: for *any* $\rho^n \in \mathcal{S}(\mathcal{H}_d^{\otimes n})$ (analogously for σ^n) there is an adaptive protocol that gives the same outcome probability distribution.

Adaptive protocols are thus more general than those in which ρ^n is entangled, so the optimal protocol can be chosen to be adaptive with no loss of generality.

To prove Lemma 2, we define $\rho'_\emptyset := \mathrm{tr}_{[n]\setminus 1}(\rho^n)$, where we denote by $[n]\setminus s$ the set $\{1, 2, \ldots, s-1, s+1, \ldots, n\}$, $s = 1, 2, \ldots, n$. Then, ρ^n and $\rho'_\emptyset$ give the same probability distribution to the outcomes of Bob's first measurement: $P(k_1|\rho^n) := \mathrm{tr}[(E_{k_1} \otimes \mathbb{1})\rho^n] = \mathrm{tr}(E_{k_1}\rho'_\emptyset)$. With Bob's feedback (the value of k_1), Alice can next prepare the second (unnormalized) state as $\rho'_{k_1} := \mathrm{tr}_{[n]\setminus 2}[(E_{k_1} \otimes \mathbb{1})\rho^n]$. So, ρ^n and ρ'_{k_1} give the same outcome probabilities up to Bob's second measurements: $P(\mathbf{k}^2|\rho^n) := \mathrm{tr}[(E_{\mathbf{k}^2} \otimes \mathbb{1})\rho^n] = \mathrm{tr}(E_{k_2}\rho'_{k_1})$. Note that the probabilities of previous outcomes are implicit in the normalization of ρ'_{k_1}. We readily see that if Alice's preparation at an arbitrary step s is

$$\rho'_{\mathbf{k}^{s-1}} := \mathrm{tr}_{[n]\setminus s}[(E_{\mathbf{k}^{s-1}} \otimes \mathbb{1})\rho^n], \tag{5}$$

where we used the convention $E_{\mathbf{k}^0} = E_\emptyset := \mathbb{1}$, then $P(\mathbf{k}^s|\rho^n) = \mathrm{tr}(E_{k_s}\rho'_{\mathbf{k}^{s-1}})$, $s = 1, 2, \ldots, n$ (obviously, the analogous relation holds for σ^n, $\sigma'_{\mathbf{k}^{s-1}}$). This completes the proof of the lemma.

Next, to prove Theorem 1, we show that the adaptive communication protocols introduced above can be cast as discrimination of two classical channels. To this end, we choose the classical (continuous) input alphabet as $\mathcal{X} = \mathcal{S}(\mathcal{H}_d) \times \mathcal{S}(\mathcal{H}_d)$, where each letter $x = (\rho, \sigma) \in \mathcal{X}$ is a classical description of the pair of states. The output alphabet $\mathcal{Y}$ is naturally given by the outcome labels of our fixed measurement (i.e., the POVM $\mathcal{E}$): $\mathcal{Y} = \{1, 2, \ldots, m\}$. We can then associate the null [alternate] hypothesis H_0 [H_1] with the classical channel $W_x(k) := \mathrm{tr}(E_k\rho)$ [$\bar{W}_x(k) := \mathrm{tr}(E_k\sigma)$], where $x \in \mathcal{X}$ and $k \in \mathcal{Y}$. These channels reproduce the same conditional probabilities, P_k and $\bar{P}_k$, that arise in our original problem. Hence the (single-shot) optimal state discrimination is formally equivalent to the optimal channel discrimination obtained by minimizing over the inputs $x \in \mathcal{X}$.

This analogy holds also for our general, multiple-shot problem. The adaptive protocol defined by the states $\rho'_{\mathbf{k}^{s-1}}, \sigma'_{\mathbf{k}^{s-1}} \in \mathcal{S}(\mathcal{H}_d)$, $s = 1, 2, \ldots, n$, translates into an adaptive channel discrimination strategy with n uses of either W or $\bar{W}$, where at each step s we feed the channel with an input letter $x_{\mathbf{k}^{s-1}} \in \mathcal{X}$, conditional on the previous outcomes $\mathbf{k}^{s-1} = \{k_1, k_2, \ldots, k_{s-1}\}$, $k_i \in \mathcal{Y}$.

We can now invoke the main result in Ref. [32]. It states that for the problem of classical channel discrimination no adaptive strategy can outperform the best nonadaptive or fix strategy. More precisely, it states that the optimal error rate can be attained by the simple sequence where all the letters are equal, $x_1 = x_2 = \cdots = x_n$. We hence conclude that the optimal error rates for our original problem can be achieved by i.i.d. state pairs, $(\rho^{\otimes n}, \sigma^{\otimes n})$. This holds for the Chernoff bound, Hoeffding's bound, and for Stein's lemma [see Eq. (6) below].

Computing the exponent rate in Eq. (3) is now identical to computing the analogous rate for the classical hypothesis testing problem of discriminating between the probability distribution $P_k = \mathrm{tr}(E_k\rho)$ and $\bar{P}_k = \mathrm{tr}(E_k\sigma)$ after n

samplings, which is given by the classical Chernoff bound (CB) [22]. This completes the proof of Theorem 1.

A simple, very explicit, alternative proof of this theorem for two-dimensional two-element POVMs can also be found in Ref. [40]. Also in Ref. [40], the reader will find upper- and lowerbounds to ζ_{CB} (and to ζ_{SL} and ζ_{HB}, defined below) that hold under mixing of POVMs.

It is plausible that Eq. (4) is minimized by an orthogonal pair, as such pairs arguably give rise to the most distinguishable P and $\bar{P}$ distributions. This would entail a simplification in the evaluation of ζ_{CB}. We have so far failed to prove the optimality of orthogonal pairs in full generality, though it is supported by extensive numerical analysis and it holds for the examples in the next paragraph [41].

To illustrate our results, in Ref. [40] we compute the discrimination power of the qubit covariant POVM, $\mathcal{E} = \{\mathbb{1} + \boldsymbol{n}\cdot\boldsymbol{\sigma}\}_{\boldsymbol{n}\in\mathbb{S}^2}$, where $\boldsymbol{\sigma}$ is the vector of Pauli matrices and $\mathbb{S}^2$ is the unit 2-sphere. The result is $\zeta_{\mathrm{CB}} = -\log(\pi/4)$ which can be compared to $\zeta_{\mathrm{CB}} = -(1/2)\log(1-r^2)$, corresponding to a noisy Stern-Gerlach apparatus of purity r. We see that $\mathcal{E}$ has the same discrimination power that a Stern-Gerlach apparatus with purity $r \approx 0.62$.

So far, special emphasis has been placed on the asymptotics of the problem at hand. It is illustrative to examine with a few examples the difficulties arising for finite n, where some of the asymptotic results do not hold. Let us focus on two-element POVMs, $\mathcal{E} = \{E_1, E_2 = \mathbb{1} - E_1\}$. In this case, E_1 and E_2 commute and can be diagonalized simultaneously. In the multiple-shot scenario, the groupings $\{E^a, E^{\bar{a}}\}$ will also be diagonal in the very same local basis that diagonalizes E_1 and E_2 and thus each state of the optimal pair, (ρ^n, σ^n), in necessarily a product state of elements of that basis. In this case, however, one can show that i.i.d. states are not necessarily optimal. In Ref. [40] we give a concrete example for $n = 3$ where the optimal states are $\rho^3 = |001\rangle\langle 001|$ and $\sigma^3 = |110\rangle\langle 110|$, rather than $|000\rangle\langle 000|$ and $|111\rangle\langle 111|$. Furthermore, we also show that there exists an adaptive protocol with yet a smaller error rate, thus outperforming the optimal nonadaptive protocol for $n = 3$.

Though in this Letter we have focused on the problem dual to symmetric hypothesis testing, which led us to Eq. (4), the very same arguments concerning the optimality of i.i.d. state pairs apply to the dual Stein's lemma and Hoeffding's bound, whose asymptotic rates are defined as $\zeta_{\mathrm{SL/HB}} = -\lim_{n\to\infty} \min_{(\rho^n,\sigma^n)} (1/n)\log\beta_n$, where the minimization is subject to $\alpha_n \leq \epsilon$ and $\alpha_n \leq \mathrm{e}^{-nr}$, respectively. It follows from our analysis that they can be computed simply as

$$\zeta_{\mathrm{SL}} = \max_{(\rho,\sigma)} D(P\|\bar{P}),$$
$$\zeta_{\mathrm{HB}} = \max_{(\rho,\sigma)} \sup_{0\leq s\leq 1} \frac{-sr - \phi(s|P\|\bar{P})}{1-s}, \tag{6}$$

where $D(P\|\bar{P}) = \sum_k P_k \log(P_k/\bar{P}_k)$ is the relative entropy.

In summary, we have introduced a class of problems dual to quantum hypothesis testing where the measurement device is a given. We have derived simple expressions for the asymptotic (exponential) error rates, which quantify the discrimination power of the measurement device when it can be used multiple times. We have briefly discussed two paradigmatic examples for qubits, covariant POVMs, and noisy Stern-Gerlach apparatus, and addressed the nonasymptotic regime with some examples. As final remarks, we point out that these dual problems complement our understanding of quantum hypothesis testing and state discrimination and we believe they will find many applications in quantum information theory. Open problems include a deeper understanding of the structure of the optimal pairs and extending the analysis to infinite dimensional systems such as light fields, where constraints on the mean energy are necessary.

C. H. acknowledges support by Spanish MINECO Grant No. BES-2014-068888 and C. H., E. B., and J. C. by the Spanish MINECO through Projects No. FIS2013-40627-P and No. FIS2016-80681-P (AEI/FEDER, UE), as well as by the Generalitat de Catalunya, CIRIT 2014-SGR966. M. H. is partially supported by a MEXT Grant-in-Aid for Scientific Research (A) No. 23246071, a MEXT Grant-in-Aid for Scientific Research (B) No. 16KT0017, and the Okawa Reserach Grant. Centre for Quantum Technologies is a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore.

*Corresponding author.
John.Calsamiglia@uab.cat

[1] O. Oreshkov, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, New J. Phys. **13**, 073032 (2011).
[2] A. Winter, Commun. Math. Phys. **244**, 157 (2004).
[3] M. Dall'Arno, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **83**, 062304 (2011).
[4] M. Dall'Arno, Phys. Rev. A **92**, 012328 (2015).
[5] A. S. Holevo, Probl. Inf. Transm. **48**, 1 (2012).
[6] M. Berta, J. M. Renes, and M. M. Wilde, IEEE Trans. Inf. Theory **60**, 7987 (2014).
[7] A. Szymusiak, J. Phys. A **47**, 445301 (2014).
[8] W. Slomczynski and A. Szymusiak, Quantum Inf. Process. **15**, 565 (2016).
[9] A. Szymusiak and W. Slomczynski, Phys. Rev. A **94**, 012122 (2016).
[10] M. Dall'Arno, F. Buscemi, and M. Ozawa, J. Phys. A **47**, 235302 (2014).
[11] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, New York, 1976).
[12] J. Calsamiglia, J. I. de Vicente, R. Muñoz-Tapia, and E. Bagan, Phys. Rev. Lett. **105**, 080504 (2010).
[13] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).

[14] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan, Phys. Rev. A **77**, 032311 (2008).
[15] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. (Springer-Verlag, Berlin Heidelberg, 2017).
[16] M. Nussbaum and A. Szkoła, Ann. Stat. **37**, 1040 (2006).
[17] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).
[18] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).
[19] M. Hayashi, Phys. Rev. A **76**, 062301 (2007).
[20] T. Ogawa and M. Hayashi, IEEE Trans. Inf. Theory **50**, 1368 (2004).
[21] H. Nagaoka, arXiv:quant-ph/0611289.
[22] The errors and rates for classical hypothesis testing can be recovered from those above by simply taking the matrices ρ and σ to be diagonal with entries given by two probability distributions $P = \{P_k\}$, $\bar{P} = \{\bar{P}_k\}$, associated with H_0 and H_1, respectively.
[23] M. Hayashi and H. Nagaoka, IEEE Trans. Inf. Theory **49**, 1753 (2003).
[24] L. Wang and R. Renner, Phys. Rev. Lett. **108**, 200501 (2012).
[25] T. Cooney, C. Hirche, C. Morgan, J. P. Olson, K. P. Seshadreesan, J. Watrous, and M. M. Wilde, Phys. Rev. A **94**, 022310 (2016).
[26] S. Pirandola, Phys. Rev. Lett. **106**, 090504 (2011).
[27] S. Lloyd, Science **321**, 1463 (2008).
[28] S.-H. Tan, B. I. Erkmen, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, S. Pirandola, and J. H. Shapiro, Phys. Rev. Lett. **101**, 253601 (2008).
[29] For simplicity and to ease up the notation we will assume POVM with a finite number of outcomes, however, the results hold for any POVM, including those with continuous outcome.
[30] The corresponding asymptotic error rate, Eq. (3), is independent of the prior probabilities η_1 and η_2, provided they do not vanish. Here we take $\eta_1 = \eta_2 = 1/2$ for simplicity.
[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Hoboken, N.J, 2006).
[32] M. Hayashi, IEEE Trans. Inf. Theory **55**, 3807 (2009).
[33] F. Dupuis, S. Fehr, P. Lamontagne, and L. Salvail, in *Adaptive Versus Non-Adaptive Strategies in the Quantum Setting with Applications* (Springer, Berlin Heidelberg, 2016), pp. 33–59.
[34] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Phys. Rev. Lett. **83**, 3081 (1999).
[35] S. Pirandola and C. Lupo, Phys. Rev. Lett. **118**, 100502 (2017).
[36] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
[37] M. Hayashi, Commun. Math. Phys. **304**, 689 (2011).
[38] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Phys. Rev. A **81**, 032339 (2010).
[39] T. Cooney, M. Mosonyi, and M. M. Wilde, Commun. Math. Phys. **344**, 797 (2016).
[40] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.118.160502 for details.
[41] However, one can easily find counterexamples of the optimality of orthogonal pairs for the statistical overlap, given by $-\min_{(\rho,\sigma)}\phi(\frac{1}{2}, P||\bar{P})$.