Adaptive versus nonadaptive strategies for quantum channel discrimination

PHYSICAL REVIEW A **81**, 032339 (2010)

Adaptive versus nonadaptive strategies for quantum channel discrimination

Aram W. Harrow*
Department of Mathematics, University of Bristol, Bristol, United Kingdom

Avinatan Hassidim†
Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Debbie W. Leung‡ and John Watrous§
Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada
(Received 26 November 2009; published 31 March 2010)

We provide a simple example that illustrates the advantage of adaptive over nonadaptive strategies for quantum channel discrimination. In particular, we give a pair of entanglement-breaking channels that can be perfectly discriminated by means of an adaptive strategy that requires just two channel evaluations, but for which no nonadaptive strategy can give a perfect discrimination using any finite number of channel evaluations.

DOI: 10.1103/PhysRevA.81.032339 PACS number(s): 03.67.−a

I. INTRODUCTION

This article concerns the problem of *quantum channel discrimination*. In this problem, two quantum channels Φ_0 and Φ_1 are fixed and access to one of the two channels is made available. It is not known which of the two channels has been made available, however, and the goal is to correctly identify which of Φ_0 and Φ_1 it is. Several articles, including Refs. [1–13], discovered many interesting aspects of quantum channel discrimination. There exist related topics in the study of quantum information theory, including *quantum parameter estimation* (see, for instance Refs. [14–16] and the references therein), but this article will focus just on the specific problem of channel discrimination.

A *discrimination strategy* for a quantum channel discrimination problem is a step-by-step procedure consisting of channel evaluations, along with quantum state preparations, operations, and measurements that attempts to output the identity of the given channel. Generally speaking, one is typically interested in discrimination strategies that satisfy certain natural constraints; with one well-studied example being the discrimination strategies allowing a *single evaluation* of the unknown channel. An *optimal* discrimination strategy, among those satisfying a given collection of constraints, is simply one that maximizes the probability that the unknown channel is correctly identified, assuming it is selected according to a fixed distribution that is known ahead of time.

One interesting aspect of quantum channel discrimination is that the use of an *ancillary* system is generally necessary for the optimal discrimination of two quantum channels, assuming just a single evaluation of the unknown channel is made available [2,6,8,17]. In more precise terms, the optimal strategy to discriminate two channels may require that one first prepares the input system to the unknown channel in an entangled state with an ancillary system, followed by a joint measurement of that channel's output together with the ancillary system. Even *entanglement-breaking* channels are sometimes better discriminated through the use of an ancillary system, despite the fact that their output systems must necessarily be unentangled with the ancillary system after their evaluation [10]. There are two known special classes of channels that require no ancillary system for optimal discrimination: the unitary channels [2,4] and the classical channels.

There is a striking possibility for quantum channel discrimination problems that cannot occur in the classical setting. If a pair of classical channels cannot be perfectly distinguished with one evaluation, then they cannot be perfectly distinguished with any finite number of evaluations. (This fact is easily proved and a simple proof may be found later in the article.) In contrast, it is possible for a pair of quantum channels to be discriminated perfectly when multiple evaluations are available, but not in the single evaluation case. For example, this generally happens in the case of unitary channels [1].

Another interesting aspect of quantum channel discrimination is the distinction between *adaptive* and *nonadaptive* strategies when multiple uses of the unknown channel are made available. In an adaptive strategy, one may use the outputs of previous uses of the channel when preparing the input to subsequent uses, whereas a nonadaptive strategy requires that the inputs to all uses of the given channel are chosen before any of them is evaluated. It was found in Ref. [3] that unitary channels are insensitive to this distinction; adaptive strategies do not give any advantage over nonadaptive strategies for unitary channel discrimination. In the same article, a pair of *memory channels* was shown to require an adaptive scheme for optimal discrimination, but the question of whether or not there exist ordinary (memoryless) channels with a similar property was stated as an open question. Although an example of *three* channels that require adaptive strategies for an optimal identification was presented in Ref. [13], we were not able to find any example of a pair of (ordinary, memoryless) channels in the literature that require adaptive strategies for optimal discrimination; and so the question appears to have been unresolved prior to this work.

*a.harrow@bris.ac.uk
†avinatanh@gmail.com
‡wcleung@iqc.ca
§watrous@cs.uwaterloo.ca

The purpose of the present article is to demonstrate the necessity of adaptive schemes for optimal quantum channel discrimination. We do this by presenting an example of two quantum channels that can be perfectly discriminated given two adaptive channel evaluations, but for which *no finite number* of nonadaptive channel evaluations allows for a perfect discrimination. The channels in our example are *entanglement-breaking* channels, which provides further evidence suggesting that entanglement-breaking channels share similar properties to general quantum channels with respect to channel discrimination tasks. We note that a recent article of Duan, Feng, and Ying [5] provided a criterion for the perfect discrimination of pairs of quantum channels, as well as a general method to find adaptive strategies that allow for perfect discrimination. While no explicit examples were given in that article, the existence of pairs of channels with similar properties to those in our example is implied. Our example was, however, obtained independently from that article, and we hope that it offers some insight into the problem of quantum channel discrimination that is complementary to Ref. [5].

Finally, we note that a related (but weaker) phenomenon occurs in the context of classical channel discrimination. That is, there exist classical channels that can be better discriminated by adaptive strategies than by nonadaptive strategies and we provide three simple examples illustrating this phenomenon. While we suspect that similar examples illustrating the advantages of adaptive discrimination strategies may be known to some researchers, we did not find any in the literature. That such examples exist is also interesting when contrasted with the fact that adaptive strategies for classical channel discrimination cannot improve the asymptotic rate at which the error probability exponentially decays with the number of channel uses [7].

II. PRELIMINARIES

We will begin by summarizing some of the notation and terminology that is used in the subsequent sections of the article. We will let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{W}$ denote finite-dimensional complex Hilbert spaces, which will typically correspond to the input, output, and ancillary systems to be associated with channel discrimination tasks. The notation $\mathrm{L}(\mathcal{X},\mathcal{Y})$ refers to the space of all linear operators from $\mathcal{X}$ to $\mathcal{Y}$, $\mathrm{L}(\mathcal{X})$ is shorthand for $\mathrm{L}(\mathcal{X},\mathcal{X})$, and $\mathrm{D}(\mathcal{X})$ refers to the set of all density operators on $\mathcal{X}$. A similar notation is used for other spaces in place of $\mathcal{X}$ and $\mathcal{Y}$. The identity operator on $\mathcal{X}$ is denoted $\mathbb{1}_{\mathcal{X}}$.

For the example to be presented in the main part of the article, we will let $\mathcal{X}$ and $\mathcal{Y}$ be the spaces associated with two qubits and one qubit, respectively. The standard bases for these spaces are therefore $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ and $\{|0\rangle, |1\rangle\}$. As is common, we will also write

$$|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \quad \text{and} \quad |-\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle,$$

and we write tensor products of these states and standard basis states in a self-explanatory way (e.g., $|1+\rangle = |1\rangle|+\rangle$).

A *quantum channel* is a linear mapping of the form $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ that is both completely positive and trace preserving. Every such quantum channel Φ can be expressed in Kraus form as

$$\Phi(X) = \sum_{j=1}^{m} A_j X A_j^*,$$

for some choice of linear operators $A_1, \ldots, A_m \in \mathrm{L}(\mathcal{X},\mathcal{Y})$ satisfying the constraint

$$\sum_{j=1}^{m} A_j^* A_j = \mathbb{1}_{\mathcal{X}}.$$

The identity channel mapping $\mathrm{L}(\mathcal{W})$ to itself is denoted $\mathbb{1}_{\mathrm{L}(\mathcal{W})}$.

The distinguishability of two quantum states ρ, σ is quantified by the trace norm, $\|\rho - \sigma\|_1 = \mathrm{Tr}|\rho - \sigma|$, where $|M| = \sqrt{M^\dagger M}$. The distinguishability of two quantum channels $\Phi_0, \Phi_1 : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ may be quantified by the distance induced by the *diamond norm* (or *completely bounded trace norm*)

$$\|\Phi_0 - \Phi_1\|_\diamond = \max_{\rho \in \mathrm{D}(\mathcal{X}\otimes\mathcal{W})} \|(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho) - (\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\|_1, \quad (1)$$

where here $\mathcal{W}$ is assumed to have dimension at least that of $\mathcal{X}$. This quantity represents the greatest possible degree of distinguishability that can result by feeding an input state into the two channels, allowing for the possibility that the input system is entangled with an ancillary system. Assuming that a bit $a \in \{0, 1\}$ is uniformly chosen at random, the quantity

$$\tfrac{1}{2} + \tfrac{1}{4}\|\Phi_0 - \Phi_1\|_\diamond,$$

represents the optimal probability to correctly determine the value of a by means of a physical process involving just a single evaluation of the channel Φ_a. It therefore holds that Φ_0 and Φ_1 are perfectly distinguishable using a single evaluation if and only if $\|\Phi_0 - \Phi_1\|_\diamond = 2$.

III. SPECIFICATION OF THE EXAMPLE AND A PERFECT DISCRIMINATION PROTOCOL

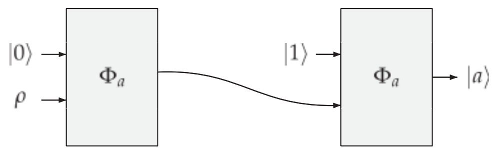
We now describe our example of two quantum channels that are better discriminated using an adaptive strategy than by any nonadaptive strategy. First, we will give an intuitive description of the channels. The two channels Φ_0 and Φ_1 both map two qubits to one and operate as follows.

(i) Channel Φ_0 measures the first input qubit with respect to the standard basis. If the result is 0, it outputs the state $|0\rangle$. If the result is 1, it measures the second qubit with respect to the standard basis. If the result is 0, then it outputs 0, and if the result is 1, then it outputs the completely mixed state $\mathbb{1}/2$.

(ii) Channel Φ_1 measures the first input qubit with respect to the standard basis. If the result is 0, it outputs the state $|+\rangle$. If the result is 1, it measures the second qubit with respect to the $\{|+\rangle, |-\rangle\}$ basis. If the result is $+$, then it outputs 1, and if the result is $-$, then it outputs the completely mixed state $\mathbb{1}/2$.

The intuition behind these channels is as follows. If the first input qubit is set to 0, then the output is a "key" state: $|0\rangle$ for channel Φ_0 and $|+\rangle$ for the channel Φ_1. If the first input is set to 1 and the second input qubit is the channel's "key" state, then the channel identifies itself (i.e., Φ_0 outputs 0 and Φ_1 outputs 1). If, however, the first input qubit is set to 1 and the second qubit's state is orthogonal to the channel's "key" state, then the channel outputs the completely mixed state. This effectively means that the channel provides no information about its identity in this case.

It is easy to discriminate these two channels with an adaptive strategy that requires two uses of the unknown channel. The following diagram describes such a strategy:



Here, the state ρ input as the second qubit of the first channel evaluation is arbitrary, as it is effectively discarded by both of the channels when the first input qubit is set to $|0\rangle$.

In the interest of precision, and because it will be useful for the analysis of the next section, we note the following formal specifications of these channels. It holds that

$$\Phi_0(X) = \sum_{j=1}^{5} A_j X A_j^* \quad \text{and} \quad \Phi_1(X) = \sum_{j=1}^{5} B_j X B_j^*,$$

for

$$\begin{gathered} A_1 = |0\rangle\langle 00|, \quad A_2 = |0\rangle\langle 01|, \quad A_3 = |0\rangle\langle 10|, \\ A_4 = \frac{1}{\sqrt{2}}|0\rangle\langle 11|, \quad A_5 = \frac{1}{\sqrt{2}}|1\rangle\langle 11|, \\ B_1 = |+\rangle\langle 00|, \quad B_2 = |+\rangle\langle 01|, \quad B_3 = |1\rangle\langle 1+|, \\ B_4 = \frac{1}{\sqrt{2}}|0\rangle\langle 1-|, \quad B_5 = \frac{1}{\sqrt{2}}|1\rangle\langle 1-|. \end{gathered}$$

It is clear that Φ_0 and Φ_1 are both entanglement-breaking channels, as all of these Kraus operators have rank one [18].

IV. SUBOPTIMALITY OF NONADAPTIVE STRATEGIES

We now prove that nonadaptive strategies cannot allow for a perfect discrimination of the channels Φ_0 and Φ_1 defined in the previous section, for any finite number n of channel uses. In more precise terms, we have

$$\left\| \Phi_0^{\otimes n} - \Phi_1^{\otimes n} \right\|_{\diamond} < 2$$

for all choices of $n \in \mathbb{N}$.

We first prove a simpler mathematical fact, which is that there does not exist a two-qubit density operator ρ for which $\Phi_0(\rho)$ and $\Phi_1(\rho)$ are perfectly distinguishable. As we will see, the proof is similar when taking the tensor product of the channel with itself or with an identity channel that acts on an auxiliary system. This handles the multiple-copy case with the possible use of an ancillary space, thus establishing the more general statement earlier.

Assume toward contradiction that there exists a density operator ρ such that $\Phi_0(\rho)$ and $\Phi_1(\rho)$ are perfectly distinguishable. By a simple convexity argument, we may assume that the same is true for a pure state $|\psi\rangle\langle\psi|$ in place of ρ. In other words, there exists a unit vector $|\psi\rangle$ satisfying

$$\mathrm{Tr}[\Phi_1(|\psi\rangle\langle\psi|)\Phi_0(|\psi\rangle\langle\psi|)] = 0. \tag{2}$$

Expanding this equation in terms of the Kraus operators of Φ_0 and Φ_1 yields

$$\sum_{j=1}^{5}\sum_{k=1}^{5} |\langle\psi| B_j^* A_k |\psi\rangle|^2 = 0.$$

Each of the terms in this sum is nonnegative and must therefore be zero (i.e., $\langle\psi|B_j^* A_k|\psi\rangle = 0$ for all choices of $j,k \in \{1,\ldots,5\}$). It follows that

$$\langle\psi| \sum_{j=1}^{5}\sum_{k=1}^{5} \alpha_{j,k} B_j^* A_k |\psi\rangle = 0, \tag{3}$$

for every choice of complex numbers $\{\alpha_{j,k} : 1 \leqslant j,k \leqslant 5\}$.

We will now obtain a contradiction by choosing the coefficients $\{\alpha_{j,k} : 1 \leqslant j,k \leqslant 5\}$ in such a way that Eq (3) cannot hold. In particular, by letting

$$\alpha_{1,1} = \alpha_{2,2} = \sqrt{2}, \quad \alpha_{3,5} = \alpha_{4,3} = 1, \quad \text{and} \quad \alpha_{4,4} = -2\sqrt{2},$$

and letting $\alpha_{j,k} = 0$ for all of the remaining values of j and k, we find that

$$\sum_{j=1}^{5}\sum_{k=1}^{5} \alpha_{j,k} B_j^* A_k = P,$$

for

$$\begin{aligned} P &= |00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 11| + |1-\rangle\langle 1-| \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & -1/2 & 3/2 \end{pmatrix}. \end{aligned}$$

The operator P is positive-definite and therefore $\langle\psi|P|\psi\rangle > 0$ for every nonzero vector $|\psi\rangle$, which is in contradiction with (3). Having established a contradiction, we conclude that there cannot exist a density operator ρ such that $\Phi_0(\rho)$ and $\Phi_1(\rho)$ are perfectly distinguishable as claimed.

Now let us consider the general setting where an arbitrary finite number n of (nonadaptive) channel uses, as well as an ancillary system of arbitrary size, are permitted. We may follow a similar proof to the one previous to show that there cannot exist a unit vector $|\psi\rangle$ such that

$$\mathrm{Tr}\left[\left(\Phi_1^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\right)(|\psi\rangle\langle\psi|)\left(\Phi_0^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\right)(|\psi\rangle\langle\psi|)\right] = 0, \tag{4}$$

where $\mathcal{W}$ is the space (of arbitrary finite dimension) that is to be associated with the ancillary system. We may express the relevant mappings in this expression in terms of the Kraus operators of Φ_0 and Φ_1 as follows

$$\begin{aligned} \left(\Phi_0^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\right)(X) = &\sum_{1\leqslant j_1,\ldots,j_n\leqslant 5} \left(A_{j_1} \otimes \cdots \otimes A_{j_n} \otimes \mathbb{1}_{\mathcal{W}}\right) \\ &\times X\left(A_{j_1} \otimes \cdots \otimes A_{j_n} \otimes \mathbb{1}_{\mathcal{W}}\right)^*, \\ \left(\Phi_1^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\right)(X) = &\sum_{1\leqslant j_1,\ldots,j_n\leqslant 5} \left(B_{j_1} \otimes \cdots \otimes B_{j_n} \otimes \mathbb{1}_{\mathcal{W}}\right) \\ &\times X\left(B_{j_1} \otimes \cdots \otimes B_{j_n} \otimes \mathbb{1}_{\mathcal{W}}\right)^*. \end{aligned}$$

Now, for the same coefficients $\{\alpha_{j,k} : 1 \leqslant j,k \leqslant 5\}$ that were defined previously, we find that

$$\sum_{\substack{1\leqslant j_1,\ldots,j_n\leqslant 5\\ 1\leqslant k_1,\ldots,k_n\leqslant 5}} \alpha_{j_1,k_1}\cdots\alpha_{j_n,k_n}\, B_{j_1}^* A_{k_1} \otimes \cdots \otimes B_{k_n}^* A_{j_n} \otimes \mathbb{1}_{\mathcal{W}} = P^{\otimes n} \otimes \mathbb{1}_{\mathcal{W}},$$

which is again positive-definite. Therefore, there cannot exist a nonzero vector $|\psi\rangle$ for which

$$\langle\psi|\, B_{j_1}^* A_{k_1} \otimes \cdots \otimes B_{k_n}^* A_{j_n} \otimes \mathbb{1}_{\mathcal{W}}\, |\psi\rangle = 0,$$

for all $j_1,\ldots,j_n,k_1,\ldots,k_n$. Consequently, Eq. (4) does not hold for any nonzero vector $|\psi\rangle$, which implies that Φ_0 and Φ_1 cannot be perfectly discriminated by means of a nonadaptive strategy.

When the number of evaluations n of the unknown channel is small, one can efficiently compute the value $\|\Phi_0^{\otimes n} - \Phi_1^{\otimes n}\|_\diamond$ because it is the optimal value of a semidefinite programming problem [19]. For instance, it holds that

$$\|\Phi_0 - \Phi_1\|_\diamond = 1 + \frac{1}{\sqrt{2}},$$

and therefore the channels can be discriminated with a probability

$$\tfrac{1}{2} + \tfrac{1}{4}\,\|\Phi_0 - \Phi_1\|_\diamond \approx 0.9268,$$

of correctness with just a single channel evaluation. For two nonadaptive queries, we used CVX [20,21], a package for specifying and solving convex programs in Matlab, to approximate the value

$$\tfrac{1}{2} + \tfrac{1}{4}\,\|\Phi_0\otimes\Phi_0 - \Phi_1\otimes\Phi_1\|_\diamond \approx 0.9771.$$

One can also obtain an upper bound on the probability of success using any feasible solution to the dual problem. In fact, even obvious choices give fairly tight upper bounds. Thus, we establish a small, but finite, advantage of an adaptive strategy over a nonadaptive one for discriminating these channels.

V. REMARKS ON CLASSICAL CHANNEL DISCRIMINATION

The channels in our previous example are entanglement-breaking channels, yet the optimal adaptive discriminating strategy operates in a distinctively quantum way: One out of two nonorthogonal key states is extracted from the first channel evaluation and coherently input to the second. A natural question arises, which is whether adaptive strategies also help when discriminating *classical* channels. It turns out that adaptive strategies indeed are better in the classical setting, although in a more limited respect. This section discusses a few basic facts and examples that illustrate this claim.

A classical channel can, of course, be succinctly represented by a stochastic matrix M, where the vector $M|k\rangle$ represents the output distribution when the input is k. Throughout this section, we will let M_0 and M_1 denote the two possible channels in a classical channel discrimination problem.

Advantages of adaptive classical strategies

We will present three examples illustrating that adaptive strategies may give advantages over nonadaptive strategies for classical channel discrimination, restricting our attention to the special case where just two channel evaluations are permitted and where one of two channels is given with equal probability. We have the following expression for the optimal success probability using an adaptive strategy in this setting

$$\frac{1}{2} + \frac{1}{4}\max_{k,f}\sum_j \|M_0(j,k)M_0\,|f(j)\rangle - M_1(j,k)M_1\,|f(j)\rangle\|_1\,, \tag{5}$$

where $M_i(j,k)$ is the probability that channel i outputs j given input k. In this expression, j and k range over all outputs and inputs, respectively, of the channels M_0 and M_1 (i.e., they are row and column indices). The function f ranges over all maps from outputs to inputs (or row indices to column indices).

An alternate expression for the optimal success probability (5) is

$$\frac{1}{2}+\frac{1}{4}\max_k\sum_j q(j,k)\max_l \|p_0(j,k)M_0\,|l\rangle - p_1(j,k)M_1\,|l\rangle\|_1,$$

where

$$q(j,k) = \frac{M_0(j,k)+M_1(j,k)}{2},$$

and where

$$p_a(j,k) = \frac{M_a(j,k)}{M_0(j,k)+M_1(j,k)},$$

is the probability that the unknown channel is M_a, conditioned on k being chosen as the input and j being obtained as the output. This illustrates that, at least for strategies allowing just two channel evaluations, the optimal adaptive strategy for two uses of a classical channel can be readily found by first finding the optimal input for each prior distribution over the chosen channel (this may be the input in the second use). We then compute the success probability given every prior distribution and one use of the channel. Finally, to choose an input to the first use of the channel, we choose an input which maximizes the probability of getting each prior times the success probability given that prior.

Example 1. This "minimal" example shows that adaptive strategies are better than nonadaptive ones in some cases. The two channels are given by

$$M_0 = \begin{pmatrix} 1/3 & 8/9 \\ 2/3 & 1/9 \end{pmatrix}, \quad M_1 = \begin{pmatrix} 0 & 1/3 \\ 1 & 2/3 \end{pmatrix}.$$

One can verify that the best two-evaluation nonadaptive strategy is to input 1 to both of the channel uses, which leads to a correct identification with probability 7/9. The best adaptive strategy is to take $k=2$ and $f(1)=2,\ f(2)=1$ in the formula (5), which gives a correct identification with probability 65/81. Similar examples are abundant.

Example 2. Here, the optimal one-shot input is never used in the optimal nonadaptive scheme. The idea is to start with two optimal one-shot inputs k,k' such that using k' becomes more informative with two parallel uses. Then we perturb the kth column slightly so that k becomes the unique optimal one-shot input. In this example, the optimal one-shot input k still serves as the first input to the optimal adaptive scheme.

Let the two channels be given by

$$M_0 = \begin{pmatrix} 0.86 & 0.45 & 1 & 0.5 \\ 0.14 & 0.1 & 0 & 0.5 \\ 0 & 0.45 & 0 & 0 \end{pmatrix},$$

$$M_1 = \begin{pmatrix} 0.15 & 0.1 & 0.5 & 0 \\ 0.85 & 0.8 & 0.5 & 1 \\ 0 & 0.1 & 0 & 0 \end{pmatrix}.$$

The best one-shot input is $k = 1$ (probability of success is 0.855) (whereas $k' = 2$). The best parallel input pairs are (2, 3) and (3, 2) (probability of success is 0.9). Allowing adaptation and using $k = 1$ as the first input, $f(1) = 3$, $f(2) = 4$, $f(3) = 1$, the probability of success is 0.9275.

Example 3. In this final example, the optimal one-shot input is *not* the first input to the optimal adaptive scheme. The idea is to have two optimal one-shot inputs in which one is more informative than the other if given a second use. Then, we perturb the column corresponding to the less informative input to be slightly better for the one-shot case.

Let the two channels be given by

$$M_0 = \begin{pmatrix} 1 & 0.5 & 0.828 & 0.76 \\ 0 & 0.5 & 0.092 & 0.04 \\ 0 & 0 & 0.08 & 0.2 \end{pmatrix},$$

$$M_1 = \begin{pmatrix} 0.5 & 0 & 0.092 & 0.04 \\ 0.5 & 1 & 0.828 & 0.76 \\ 0 & 0 & 0.08 & 0.2 \end{pmatrix}.$$

The best one-shot input is 3 (probability of success is 0.868), but the best parallel input pairs to two uses are (3, 4) and (4, 3) (probability of success is 0.9336). The optimal adaptive scheme uses $k = 4$ as the first input and $f(j) = j$ for $j = 1, 2, 3$, resulting in a probability of success of 0.9536.

Perfect classical strategies

Finally, we give a simple proof of a fact claimed in the Introduction of this article, which is that if two classical channels are not perfectly distinguishable with a single evaluation, then they cannot be perfectly distinguished by any finite number of evaluations, even using an adaptive strategy. We will prove the contrapositive of this statement.

Suppose that two classical channels M_0 and M_1 are perfectly discriminated by a discrimination strategy that uses n channel evaluations. Without loss of generality we may assume the strategy takes the general form suggested in Fig. 1.

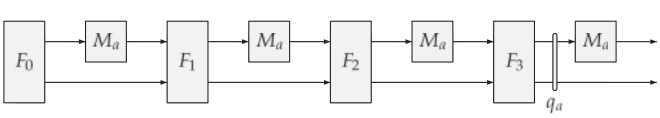


FIG. 1. The structure of a general discrimination strategy for classical channels. This example makes four channel evaluations, each illustrated by a box labeled M_a, but in general any finite number n of evaluations may be considered. Each arrow represents a register that may be in a random mixture over some finite set of classical states and the boxes labeled F_0, F_1, F_2, and F_3 represent arbitrary functions (or random processes) that must be independent of the value $a \in \{0, 1\}$ that indicates which of the two channels is given.

The assumption that the strategy perfectly discriminates M_0 and M_1 means that the final output distributions for the cases $a = 0$ and $a = 1$ have disjoint support. Our goal is to prove that M_0 and M_1 are perfectly discriminated with a single evaluation.

The proof of this statement proceeds by induction on n. In case $n = 1$ there is nothing to prove, so assume that $n \geqslant 2$. Consider the two distributions q_0 and q_1 that are illustrated in the figure. Each distribution q_a represents the state of the discrimination strategy immediately before the final channel evaluation takes place, assuming the unknown channel is given by M_a. There are two cases: q_0 and q_1 have disjoint support, or they do not. If q_0 and q_1 do have disjoint support, then terminating the discrimination strategy after $n - 1$ channel evaluations allows for a perfect discrimination, so by the induction hypothesis it is possible to discriminate the channels with a single evaluation. In the other case, where q_0 and q_1 do not have disjoint supports, there must exist a classical state x of the strategy at the time under consideration for which $q_0(x)$ and $q_1(x)$ are both positive. Given that the discrimination strategy is perfect, and therefore has final distributions with disjoint supports, it must hold that evaluating M_0 and M_1 on x results in distributions with disjoint supports. Therefore, M_0 and M_1 can be discriminated with a single channel evaluation as required.

VI. CONCLUSION

In this article, we presented a pair of quantum channels that can be discriminated perfectly by a strategy making two adaptive channel evaluations, but which cannot be perfectly discriminated nonadaptively with any finite number of channel evaluations.

One natural question that arises is whether our example can be generalized to show a similar advantage of general adaptive strategies making n channel evaluations versus strategies that make channel evaluations with depth at most $n - 1$. Although our example can be generalized in a natural way, we did not prove that it has the required properties with respect to depth $n - 1$ strategies.

Finally, for the example we present, we find that, although strategiesr making two nonadaptive channel evaluationsrr cannot be perfect, they can be correct with high rprobability (about 97.7%). What is the largest possible gap between optimal adaptive versus nonadaptive strategies making two (or any other number of) channel evaluations? Ther only upper bound we have on this gap is that channels Φ_0 and Φ_1 that are perfectly discriminated by two (adaptive or nonadaptive) evaluations must satisfy $\|\Phi_0 - \Phi_1\|_\diamond \geqslant 1$ and can therefore be discriminated (with a single evaluation) with probability at least 3/4 of correctness.

ACKNOWLEDGMENTS

D.L. thanks Chris Fuchs for helpfulr comments. A.W.H. was funded by the EPSRC grant "QIP IRC" and is grateful for the hospitality of the Perimeter Institute where some of this work was carried out. A.H. received support from the xQIT Keck fellowship. D.L. was funded by CRC, CFI, ORF, CIFAR, NSERC, and QuantumWorks. J.W. was funded by NSERC, CIFAR, and QuantumWorks.

[1] A. Acín, Phys. Rev. Lett. **87**, 177901 (2001).
[2] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing* (ACM, New York, 1997), pp. 20–30.
[3] G. Chiribella, G. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180501 (2008).
[4] A. M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).
[5] R. Duan, Y. Feng, and M. Ying, e-print arXiv:0908.0119.
[6] G. D'Ariano, P. Presti, and M. Paris, Phys. Rev. Lett. **87**, 270404 (2001).
[7] M. Hayashi, IEEE Trans. Inf. Theory **55**, 8, 3807 (2009).
[8] A. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[9] M. Piani and J. Watrous, Phys. Rev. Lett. **102**, 250501 (2009).
[10] M. Sacchi, Phys. Rev. A **72**, 014305 (2005).
[11] M. Sacchi, Phys. Rev. A **71**, 062340 (2005).
[12] J. Watrous, Quantum Inf. Comput. **8**, 819 (2008).
[13] G. Wang and M. Ying, Phys. Rev. A **73**, 042301 (2006).
[14] A. Fujiwara and H. Imai, J. Phys. A **36**, 8093 (2003).
[15] H. Imai and M. Hayashi, New J. Phys. **11**, 043034 (2009).
[16] Z. Ji, G. Wang, R. Duan, Y. Feng, and M. Ying, IEEE Trans. Inf. Theory **54**, 5172 (2008).
[17] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation*, Vol. 47 of *Graduate Studies in Mathematics* (American Mathematical Society, Providence, RI, 2002).
[18] M. Horodecki, P. Shor, and M. Ruskai, Rev. Math. Phys. **15**, 629 (2003).
[19] J. Watrous, e-print arXiv:0901.4709.
[20] M. Grant and S. Boyd, in *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, edited by V. Blondel, S. Boyd, and H. Kimura, Lecture Notes in Control and Information Sciences (Springer, New York, 2008), pp. 95–110.
[21] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming. Available from http://stanford.edu/boyd/cvx, 2009.