


---

This is the **accepted version** of the conference paper:

Salek, Farzin; Hayashi, Masahito; Winter, Andreas. «Asymptotic separation between adaptive and non-adaptive strategies in quantum channel discrimination». IEEE International Symposium on Information Theory, Melbourne, Australia, 12-20 July 2021, p. 1194-1199. DOI 10.1109/ISIT45174.2021.9517941

---

This version is available at <https://ddd.uab.cat/record/265370>

under the terms of the  **IN** COPYRIGHT license

# Asymptotic Separation Between Adaptive and Non-adaptive Strategies in Quantum Channel Discrimination

Farzin Salek

Department of Mathematics  
Technical University of Munich  
80333 Garching, Germany

&

Grup d'Informació Quàntica  
Departament de Física  
Universitat Autònoma de Barcelona  
08193 Barcelona, Spain

Masahito Hayashi

{Shenzhen Institute for  
Quantum Science and Engineering,  
Shenzhen Key Laboratory  
of Quantum Science and Engineering,  
Guangdong Provincial Key Laboratory  
of Quantum Science and Engineering}  
Southern University of Science and Technology  
Shenzhen 518055, China  
& Graduate School of Mathematics  
Nagoya University  
Nagoya, 464-8602, Japan

Andreas Winter

ICREA  
&  
Grup d'Informació Quàntica  
Departament de Física  
Universitat Autònoma de Barcelona  
08193 Barcelona, Spain

**Abstract**—We present a broad investigation of asymptotic binary hypothesis testing, when each hypothesis represents asymptotically many independent instances of a quantum channel, and the tests are based on using the unknown channel multiple times and observing its output at the end. Unlike the familiar setting of quantum states as hypotheses, there is a fundamental distinction between adaptive and non-adaptive strategies with respect to the channel uses, and we introduce a number of further variants of the discrimination tasks by imposing different restrictions on the test strategies. Our main result is the first separation between adaptive and non-adaptive symmetric hypothesis testing exponents for quantum channels, which we derive from a general lower bound on the error probability for non-adaptive strategies; the concrete example we analyze is a pair of entanglement-breaking channels. *Full details in [1].*

## I. INTRODUCTION

Hypothesis testing is one of the most fundamental primitives both in classical and quantum information processing. It is such a central task because a variety of other information processing problems can be cast in the framework of hypothesis testing; both direct coding theorems and converses can be reduced to it. In binary hypothesis testing, the two hypotheses are usually referred to as null and alternative hypotheses and accordingly, two error probabilities are defined: type-I error due to a wrong decision in favour of the alternative hypothesis (while the truth corresponds to the null hypothesis) and type-II error due to the alternative hypothesis being rejected despite being correct. The overall objective of the hypothesis testing is to minimize the error probability in identifying the hypotheses. Depending on the significance attributed to the two types of errors, several settings can be distinguished. An historical distinction is between the *symmetric* and the *asymmetric* hypothesis testing: in symmetric hypothesis testing, the goal

is to minimize both error probabilities simultaneously, while in asymmetric hypothesis testing, the goal is to minimize one type of error probability subject to a constraint on the other type of error probability.

This description of the problem presupposes that the two hypotheses correspond to objects in a probabilistic framework, in which also the possible tests (decision rules) are phrased, so as to give unambiguous meaning to the type-I and type-II error probabilities. The traditionally studied framework is that each hypothesis represents a probability distribution on a given set, and more generally a state on a given quantum system.

In the present paper, we consider the hypotheses to be described by two quantum channels, i.e. completely positive and trace preserving (cptp) maps, acting on a given quantum system, and more precisely  $n \gg 1$  independent realizations of the unknown channel. It is not hard to see that both the type-I and type-II error probabilities can be made to go to 0 exponentially fast, just as in the case of hypotheses described by quantum states, and hence the fundamental question is the characterization of the possible pairs of error exponents.

To spell out the precise questions, let us introduce a bit of notation. Throughout the paper,  $A, B, C$ , etc. denote quantum systems, but also their corresponding Hilbert space. We identify states  $\rho$  with their density operators and use superscripts to denote the systems on which the mathematical objects are defined. The set of density matrices (positive semi-definite matrices with unit trace) on  $A$  is written as  $\mathcal{S}^A$ , a subset of the trace class operators, denoted  $\mathcal{T}^A$ . When talking about tensor products of spaces, we may habitually omit the tensor sign, so  $A \otimes B = AB$ , etc. For the state  $\rho \in \mathcal{S}^{AB}$  in the composite system  $AB$ , the partial trace over system  $A$  (resp.  $B$ ) is denoted by  $\text{Tr}_A$  (resp.  $\text{Tr}_B$ ). We denote the identity

operator by  $I$ . We use  $\log$  and  $\ln$  to denote base 2 and natural logarithms, respectively. Moving on to quantum channels, these are linear, completely positive and trace preserving maps  $\mathcal{M} : \mathcal{S}^A \rightarrow \mathcal{S}^B$  for two quantum systems  $A$  and  $B$ ;  $\mathcal{M}$  extends uniquely to a linear map from trace class operators on  $A$  to those on  $B$ . We often denote quantum channels, by slight abuse of notation, as  $\mathcal{M} : A \rightarrow B$ . The ideal, or identity, channel on  $A$  is denoted  $\text{id}_A$ . Note furthermore that a state  $\rho^A$  on a system  $A$  can be viewed as a quantum channel  $\rho : 1 \rightarrow A$ , where  $1$  denotes the canonical one-dimensional Hilbert space, isomorphic to the complex numbers  $\mathbb{C}$ , which interprets a state operationally consistently as a state preparation procedure.

The most general operationally justified strategy to distinguish two channels  $\mathcal{M}, \overline{\mathcal{M}} : A \rightarrow B$  is to prepare a state  $\rho^{RA}$ , apply the unknown channel to  $A$  (and the identity channel  $\text{id}_R$  to  $R$ ), and then apply a binary measurement POVM  $(T, I - T)$  on  $BR$ , so that

$$\alpha = \text{Tr}((\text{id}_R \otimes \mathcal{M})\rho)(I - T) \quad \text{and} \quad \beta = \text{Tr}((\text{id}_R \otimes \overline{\mathcal{M}})\rho)T$$

are the error probabilities of type I and type II, respectively. It is easy to see that whatever state  $\rho^{AR}$  is considered as input, it can be purified to  $\psi^{ARR'}$ , with a suitable Hilbert space, and the latter state can be used to get the same error probabilities. Then, once there is a pure state, one only needs a subspace of  $R \otimes R'$  of dimension  $|A|$ , namely the support of  $\psi^{RR'}$ , which by the Schmidt decomposition is at most  $|A|$ -dimensional. Therefore, the state  $\rho$  is without loss of generality pure and that hence  $R$  has dimension at most that of  $A$ . The strategy is entirely described by the pair  $(\rho, (T, I - T))$  consisting of the initial state and the final measurement, and we denote it  $\mathcal{T}$ . Consequently, the above error probabilities are more precisely denoted  $\alpha(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T})$  and  $\beta(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T})$ , respectively.

These strategies use the unknown channel exactly once; to use it  $n > 1$  times, one could simply consider that  $\mathcal{M}^{\otimes n}$  and  $\overline{\mathcal{M}}^{\otimes n}$  are quantum channels themselves and apply the above recipe. While for states this indeed leads to the most general possible discrimination strategy, for general channels other, more elaborate procedures are possible. The most general strategy we shall consider in this paper is the *adaptive* strategy, applying the  $n$  channel instances sequentially, using quantum memory and quantum feed-forward, and a measurement at the end. This is called, variously, an adaptive strategy, a memory channel or a comb in the literature. It is defined as follows [2], [3], [4], [5], [6], [7].

**Definition 1:** A general adaptive strategy  $\mathcal{T}_n$  is given by an  $(n + 1)$ -tuple  $(\rho_1^{R_1 A_1}, \mathcal{F}_1, \dots, \mathcal{F}_{n-1}, (T, I - T))$ , consisting of an auxiliary system  $R_1$  and a state  $\rho_1$  on  $R_1 A_1$ , quantum channels  $\mathcal{F}_m : R_m B_m \rightarrow R_{m+1} A_{m+1}$  and a binary POVM  $(T, I - T)$  on  $R_n B_n$ . It encodes the following procedure (see Fig. 1): in the  $m$ -th round ( $1 \leq m \leq n$ ), apply the unknown channel  $\Xi \in \{\mathcal{M}, \overline{\mathcal{M}}\}$  to  $\rho_m = \rho_m^{R_m A_m}$ , obtaining

$$\omega_m^{R_m B_m} = \omega_m^{R_m B_m}(\Xi) = (\text{id}_{R_m} \otimes \Xi)\rho_m^{R_m A_m}.$$

Then, as long as  $m < n$ , use  $\mathcal{F}_m$  to prepare the state for the next channel use:

$$\rho_{m+1}^{R_{m+1} A_{m+1}} = \mathcal{F}_m(\omega_m^{R_m B_m}).$$

When  $m = n$ , measure the state  $\omega_n^{R_n B_n}$  with  $(T, I - T)$ , where the first outcome corresponds to declaring the unknown channel to be  $\mathcal{M}$ , the second  $\overline{\mathcal{M}}$ . Thus, the  $n$ -copy error probabilities of type I and type II are given by

$$\alpha_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n) := \text{Tr}(\omega_n^{R_n B_n}(\mathcal{M}))(I - T),$$

$$\beta_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n) := \text{Tr}(\omega_n^{R_n B_n}(\overline{\mathcal{M}}))T,$$

respectively.  $\square$

As in the case of a single use of the channel, one can without loss of generality (w.l.o.g.) simplify the strategy, by purifying the initial state  $\rho_1$ , hence  $|R_1| \leq |A|$ , and for each  $m > 1$  going to the Stinespring isometric extension of the cptp map  $\text{Tr}_{R_{m+1}} \circ \mathcal{F}_m : R_m B_m \rightarrow A_{m+1}$  that prepares the next channel input (and which by the uniqueness of the Stinespring extension is an extension of the given map  $\mathcal{F}_m$ ). This requires a system  $R_{m+1}$  with dimension no more than  $|R_{m+1}| \leq |R_m||A||B|$ , cf. [2]. This allows to efficiently parametrize all strategies in the case that  $A$  and  $B$  are finite dimensional. An equivalent description is in terms of so-called causal channels [2], which are ruled by a generalization of the Choi isomorphism. This turns many optimizations over adaptive strategies into semidefinite programs (SDP) [2], [6], [8], [9], which is relevant for practical calculations. See [10], [11] for recent comprehensive surveys of the concept of strategy and its history.

The set of all adaptive strategies of  $n$  sequential channel uses is denoted  $\mathbb{A}_n$ . It quite evidently includes the  $n$  parallel uses described at the beginning, when a single-use strategy is applied to the channel  $\Xi^{\otimes n}$ ; the set of these non-adaptive or parallel strategies is denoted  $\mathbb{P}_n$  (See Fig. (2)).

For a given class  $\mathbb{S}_n \subset \mathbb{A}_n$  of adaptive strategies for any number  $n$  of channel uses, we define the Chernoff bound

$$\xi^{\mathbb{S}}(\mathcal{M}, \overline{\mathcal{M}}) := \inf_{\mathcal{T}_n \in \mathbb{S}_n} \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \left( \alpha_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n) + \beta_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n) \right).$$

Naturally, the first question in this search would be to investigate the existence of quantum channels for which some class  $\mathbb{S}_n \subset \mathbb{A}_n$  outperforms the parallel strategy when  $n \rightarrow \infty$ ; in other words, if there exists a separation between adaptive and non-adaptive strategies. We study this question in general, and in particular when the channels are entanglement-breaking of the following form:

$$\mathcal{M}(\xi) = \sum_x (\text{Tr } E_x \xi) \rho_x, \quad \overline{\mathcal{M}}(\xi) = \sum_x (\text{Tr } E'_x \xi) \sigma_x, \quad (1)$$

where  $\{E_x\}$  and  $\{E'_x\}$  are PVMs and  $\rho_x, \sigma_x$  are states on the output system. In [1], we show that when these two PVMs are the same,  $E_x = E'_x$ , then the largest class  $\mathbb{A}_n$  cannot outperform the parallel strategy as  $n \rightarrow \infty$ . This fact was shown by proving that any adaptive strategy cannot improve the parallel strategy for the discrimination of two cq-channels as  $n \rightarrow \infty$ . The aim of the present paper is to find pairs of entanglement-breaking channels of the following form (1) with different PVMs, such that the largest class  $\mathbb{A}_n$  outperforms the parallel strategies as  $n \rightarrow \infty$ .

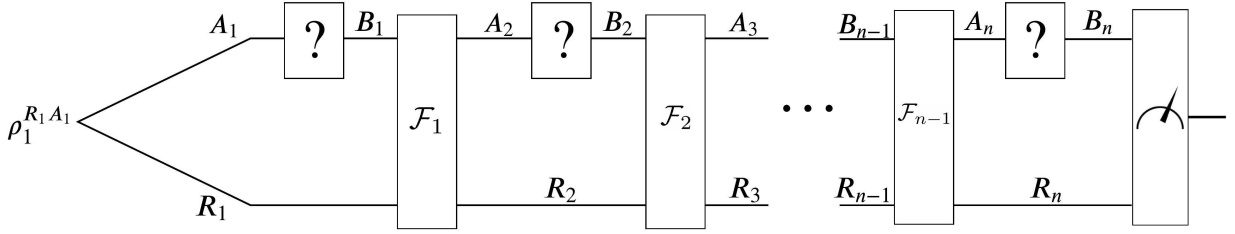


Fig. 1. The most general adaptive strategy for discrimination of qq-channels, from the class  $\mathbb{A}_n$ . After the  $m$ -th use of the unknown channel (denoted ‘?’), the output system  $B_m$  as well as the state on the memory, i.e. the reference system  $R_m$ , is processed by the cptp map  $\mathcal{F}_m$ , resulting in  $\rho_{m+1}^{R_{m+1} A_{m+1}}$ ; this continues as long as  $m < n$ . After the  $n$ -th use of the channel, the state  $\omega_n^{R_n B_n}$  is measured by a two-outcome POVM.

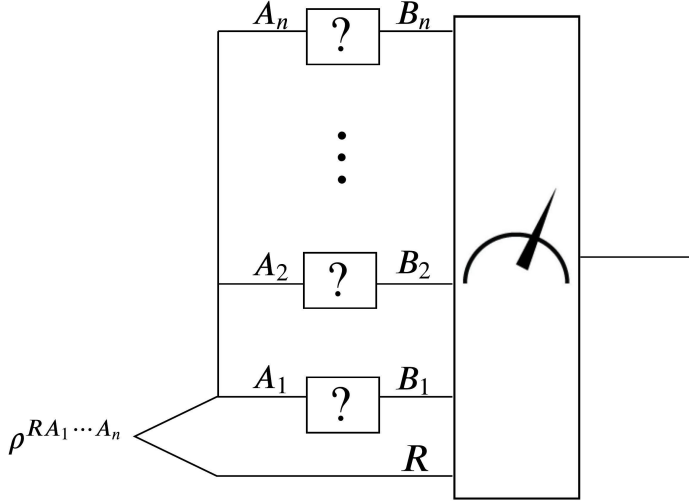


Fig. 2. The most general parallel strategy for discrimination of qq-channels, from the class  $\mathbb{P}_n$ . An  $(n+1)$ -partite state  $\rho$  on  $RA_1 \dots A_n$  is prepared and each system  $A_i$  is fed into a separate channel input; the final measurement is performed with a two-outcome POVM on  $RB_1 \dots B_n$ .

## II. GENERAL BOUND FOR NON-ADAPTIVE STRATEGIES

In this section we exhibit an asymptotic separation between the Chernoff error exponents of discriminating between two channels by adaptive versus non-adaptive strategies. Concretely, we will show that two channels described in [12], and shown to be perfectly distinguishable by adaptive strategies of  $n \geq 2$  copies, hence having infinite Chernoff exponent, nevertheless have a finite error exponent under non-adaptive strategies.

The separation is based on a general lower bound on non-adaptive strategies for an arbitrary pair of channels. Consider two quantum channels, i.e. cptp maps,  $\mathcal{M}, \overline{\mathcal{M}} : A \rightarrow B$ . To fix notation, we can write their Kraus decompositions as

$$\mathcal{M}(\rho) = \sum_i E_i \rho E_i^\dagger, \quad \overline{\mathcal{M}}(\rho) = \sum_j F_j \rho F_j^\dagger.$$

The most general strategy to distinguish them consists in the preparation of a, w.l.o.g. pure, state  $\varphi$  on  $A \otimes R$ , where  $R \simeq$

$A$ , send it through the unknown channel, and make a binary measurement  $(T, I - T)$  on  $B \otimes R$ :

$$p = \text{Tr}((\text{id}_R \otimes \mathcal{M})\varphi)T, \quad q = \text{Tr}((\text{id}_R \otimes \overline{\mathcal{M}})\varphi)T,$$

and likewise  $1 - p$  and  $1 - q$  by replacing  $T$  in the above formulas with  $I - T$ . Note that for uniform prior probabilities on the two hypotheses, the error probability in inferring the true channel from the measurement output is  $\frac{1}{2}(1 - |p - q|)$ .

The maximum of  $|p - q|$  over state preparations and measurements gives rise to the (normalized) diamond norm distance of the channels [13], [14], [15], [8]:

$$\max_{\varphi, T} |p - q| = \frac{1}{2} \|\mathcal{M} - \overline{\mathcal{M}}\|_\diamond,$$

which in turn quantifies the minimum discrimination error under the most general quantum strategy:

$$P_e = \frac{1}{2} \left( 1 - \frac{1}{2} \|\mathcal{M} - \overline{\mathcal{M}}\|_\diamond \right).$$

We are interested in the asymptotics of this error probability when the discrimination strategy has access to  $n \gg 1$  many instances of the unknown channel in parallel, or in other words, in a non-adaptive way. This means effectively that the two hypotheses are the simple channels  $\mathcal{M}^{\otimes n}$  and  $\overline{\mathcal{M}}^{\otimes n}$ , so that the error probability is

$$P_{e, \mathbb{P}}^{(n)} = \frac{1}{2} \left( 1 - \frac{1}{2} \|\mathcal{M}^{\otimes n} - \overline{\mathcal{M}}^{\otimes n}\|_\diamond \right).$$

The (non-adaptive) Chernoff exponent is then given as

$$\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) = \lim_{n \rightarrow \infty} -\frac{1}{n} \log P_{e, \mathbb{P}}^{(n)},$$

the existence of the limit being guaranteed by general principles. Note that the limit can be  $+\infty$ , which happens in all cases where there is an  $n$  such that  $P_{e, \mathbb{P}}^{(n)} = 0$ . It is currently unknown whether this is the only case; cf. the case of the more flexible adaptive strategies, for which there is a simple criterion to determine whether there exists an  $n$  such that the adaptive error probability  $P_{e, \mathbb{A}}^{(n)} = 0$  [16], and then evidently  $C^{\mathbb{A}}(\mathcal{M}, \overline{\mathcal{M}}) = +\infty$ ; conversely, we know that in all other cases, the adaptive Chernoff exponent is  $C^{\mathbb{A}}(\mathcal{M}, \overline{\mathcal{M}}) < +\infty$  [17]. There exist also other lower bounds on the symmetric discrimination error by adaptive strategies, for instance [18, Thm. 3] geared towards finite  $n$ .

Duan *et al.* [19] have attempted a characterization of the channel pairs such that there exists an  $n$  with  $P_{e,\mathbb{P}}^{(n)} = 0$ , and have given a simple sufficient condition for the contrary. Namely, the existing result [19, Cor. 1] states that if  $\text{span}\{E_i^\dagger F_j\}$  contains a positive definite element, then for all  $n$  we have  $P_{e,\mathbb{P}}^{(n)} > 0$ . The following proposition, which makes the result of [19] quantitative, is the main result of this section.

**Proposition 2:** Let  $\alpha_{ij} \in \mathbb{C}$  be such that  $\sum_{ij} |\alpha_{ij}|^2 = 1$  and  $P := \sum_{ij} \alpha_{ij} E_i^\dagger F_j > 0$ , i.e.  $P$  is assumed to be positive definite. Then for all  $n$ ,

$$P_{e,\mathbb{P}}^{(n)} \geq \frac{1}{4} \lambda_{\min}(P)^{4n},$$

where  $\lambda_{\min}(A)$  denotes the smallest eigenvalue of the Hermitian operator  $A$ . Consequently,

$$\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) \leq 4 \log \|P^{-1}\|_{\infty}.$$

*Proof:* We begin with a test state  $\varphi$  as in the above description of the most general non-adaptive strategy for the channels  $\mathcal{M}$  and  $\overline{\mathcal{M}}$ , so that the two output states are  $\rho = (\text{id}_R \otimes \mathcal{M})\varphi$ ,  $\sigma = (\text{id}_R \otimes \overline{\mathcal{M}})\varphi$ . By well-known inequalities [20], it holds

$$\frac{1}{2} \|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2} \leq 1 - \frac{1}{2} F(\rho, \sigma)^2,$$

where  $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$  is the fidelity. Thus, it will be enough to lower bound the fidelity between the output states of the two channels. With  $\tau = \text{Tr}_R |\varphi\rangle\langle\varphi|$ , we have:

$$\begin{aligned} F(\rho, \sigma) &= \|\sqrt{\rho}\sqrt{\sigma}\|_1 \geq \text{Tr} \sqrt{\rho}\sqrt{\sigma} \geq \text{Tr} \rho\sigma \\ &= \sum_{ij} |\text{Tr} E_i^\dagger F_j \tau|^2 \geq \left| \sum_{ij} \alpha_{ij} \text{Tr} E_i^\dagger F_j \tau \right|^2 = |\text{Tr} \tau P|^2. \end{aligned}$$

Here, in the first line, the first inequality is by standard inequalities for the trace norm, the second is because of  $\rho \leq \sqrt{\rho}$ , in the second line, the first equality is a formula from [19, Sec. II], the inequality follows Cauchy-Schwarz inequality and the last equality follows by the definition of  $P$ . Since  $\tau$ , like  $\varphi$ , ranges over all states, we get

$$F(\rho, \sigma)^2 \geq \lambda_{\min}(P)^4, \text{ and so } P_e \geq \frac{1}{4} \lambda_{\min}(P)^4.$$

We can apply the same reasoning to  $\mathcal{M}^{\otimes n}$  and  $\overline{\mathcal{M}}^{\otimes n}$ , for which the vector  $(\alpha_{ij})^{\otimes n}$  is eligible and leads to the positive definite operator  $P^{\otimes n}$ . Thus,

$$P_{e,\mathbb{P}}^{(n)} \geq \frac{1}{4} \lambda_{\min}(P^{\otimes n})^4 = \frac{1}{4} \lambda_{\min}(P)^{4n}.$$

Taking the limit and noting  $\lambda_{\min}(P)^{-1} = \|P^{-1}\|_{\infty}$  concludes the proof.  $\blacksquare$

### III. TWO EXAMPLES

**Example 3:** Next we show that two channels defined by Harrow *et al.* [12] yield an example of a pair with  $\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) < +\infty$ , yet  $\xi^{\mathbb{A}}(\mathcal{M}, \overline{\mathcal{M}}) = +\infty$  because indeed  $P_{e,\mathbb{A}}^{(2)} = 0$ . In [12], the following two entanglement-breaking

channels from  $A \otimes C = \mathbb{C}^2 \otimes \mathbb{C}^2$  (two qubits) to  $B = \mathbb{C}^2$  (one qubit) are considered:

$$\begin{aligned} \mathcal{M}(\rho^A \otimes \gamma^C) &= |0\rangle\langle 0| \langle 0|\gamma|0\rangle + |0\rangle\langle 0| \langle 1|\gamma|1\rangle \langle 0|\rho|0\rangle \\ &\quad + \frac{1}{2} I \langle 1|\gamma|1\rangle \langle 1|\rho|1\rangle, \\ \overline{\mathcal{M}}(\rho^A \otimes \gamma^C) &= |+\rangle\langle +| \langle 0|\gamma|0\rangle + |1\rangle\langle 1| \langle 1|\gamma|1\rangle \langle +|\rho|+\rangle \\ &\quad + \frac{1}{2} I \langle 1|\gamma|1\rangle \langle -|\rho|-\rangle, \end{aligned}$$

extended by linearity to all states. Here,  $|0\rangle, |1\rangle$  are the computational basis ( $Z$  eigenbasis) of the qubits, while  $|+\rangle, |-\rangle$  are the Hadamard basis ( $X$  eigenbasis).

In words, both channels measure the qubit  $C$  in the computational basis. If the outcome is ‘0’, they each prepare a pure state on  $B$  (ignoring the input in  $A$ ):  $|0\rangle\langle 0|$  for  $\mathcal{M}$ ,  $|+\rangle\langle +|$  for  $\overline{\mathcal{M}}$ . If the outcome is ‘1’, they each make a measurement on  $A$  and prepare an output state on  $B$  depending on its outcome: standard basis measurement for  $\mathcal{M}$  with  $|0\rangle\langle 0|$  on outcome ‘0’ and the maximally mixed state  $\frac{1}{2}I$  on outcome ‘1’; Hadamard basis measurement for  $\overline{\mathcal{M}}$  with  $|1\rangle\langle 1|$  on outcome ‘+’ and the maximally mixed state  $\frac{1}{2}I$  on outcome ‘-’. In [12], a simple adaptive strategy for  $n = 2$  uses of the channel is given that discriminates  $\mathcal{M}$  and  $\overline{\mathcal{M}}$  perfectly: The first instance of the channel is fed with  $|0\rangle\langle 0| \otimes |0\rangle\langle 0|$ , resulting in an output state  $\rho_1$ ; the second instance of the channel is fed with  $|1\rangle\langle 1| \otimes \rho_1$ ; the output state  $\rho_2$  of the second instance is  $|0\rangle\langle 0|$  if the unknown channel is  $\mathcal{M}$ , and  $|1\rangle\langle 1|$  if the unknown channel is  $\overline{\mathcal{M}}$ , so a computational basis measurement reveals it. Note that no auxiliary system  $R$  is needed, but the feed-forward nevertheless requires a qubit of quantum memory for the strategy to be implemented. In any case, this proves that  $P_{e,\mathbb{A}}^{(2)} = 0$ . In [12], it is furthermore proved that for all  $n \geq 1$ ,  $P_{e,\mathbb{P}}^{(n)} > 0$ .

We now show that Proposition 2 is applicable to yield an exponential lower bound on the non-adaptive error probability. The Kraus operators of the two channels can be chosen as follows:

$$\begin{aligned} \mathcal{M} : E_i &\in \left\{ |0\rangle^B \langle 00|^A, |0\rangle^B \langle 10|^A, |0\rangle^B \langle 01|^A, \right. \\ &\quad \left. |0\rangle^B \langle 11|^A / \sqrt{2}, |1\rangle^B \langle 11|^A / \sqrt{2} \right\}, \\ \overline{\mathcal{M}} : F_j &\in \left\{ |+\rangle^B \langle 00|^A, |+\rangle^B \langle 10|^A, |1\rangle^B \langle +1|^A, \right. \\ &\quad \left. |0\rangle^B \langle -1|^A / \sqrt{2}, |1\rangle^B \langle -1|^A / \sqrt{2} \right\}. \end{aligned}$$

Thus, the products  $E_i^\dagger F_j$  include the matrices

$$\begin{aligned} E_1^\dagger F_1 &= \sqrt{\frac{1}{2}} |00\rangle\langle 00|, & E_2^\dagger F_2 &= \sqrt{\frac{1}{2}} |10\rangle\langle 10|, \\ E_5^\dagger F_3 &= \sqrt{\frac{1}{2}} |11\rangle\langle +1|, & E_5^\dagger F_5 &= \frac{1}{2} |11\rangle\langle -1|, \\ E_3^\dagger F_4 &= \sqrt{\frac{1}{2}} |01\rangle\langle -1|, \end{aligned}$$

from which we can form, by linear combination, the operators

$$\begin{aligned} E_1^\dagger F_1 &= \sqrt{\frac{1}{2}} |0\rangle\langle 0| \otimes |0\rangle\langle 0|, & E_2^\dagger F_2 &= \sqrt{\frac{1}{2}} |1\rangle\langle 1| \otimes |0\rangle\langle 0|, \\ \sqrt{\frac{1}{2}} E_5^\dagger F_3 - E_5^\dagger F_5 &= \sqrt{\frac{1}{2}} |1\rangle\langle 1| \otimes |1\rangle\langle 1|, \\ \sqrt{\frac{1}{2}} E_3^\dagger F_4 - E_5^\dagger F_5 &= \sqrt{\frac{1}{2}} |-\rangle\langle -| \otimes |1\rangle\langle 1|, \end{aligned}$$

whose sum is indeed positive definite, so we get an exponential lower bound on  $P_{e,\mathbb{P}}^{(n)}$  and hence a finite value of  $\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}})$ .

To get a concrete upper bound on  $\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}})$  from the above method, we make the ansatz

$$\begin{aligned} P &= \alpha E_1^\dagger F_1 + \alpha E_2^\dagger F_2 + \beta \sqrt{\frac{1}{2}} E_5^\dagger F_3 + \beta \sqrt{\frac{1}{2}} E_3^\dagger F_4 - 2\beta E_5^\dagger F_5 \\ &= \alpha \sqrt{\frac{1}{2}} I \otimes |0\rangle\langle 0| + \beta \sqrt{\frac{1}{2}} (|1\rangle\langle 1| + |-\rangle\langle -|) \otimes |1\rangle\langle 1|, \end{aligned}$$

where  $\alpha, \beta > 0$  and  $2\alpha^2 + 5\beta^2 = 1$ . The minimum eigenvalue of  $P$  is easily calculated: it turns out the smaller of  $\alpha\sqrt{\frac{1}{2}}$  and  $\beta\sqrt{2}\sin^2 \frac{\pi}{8}$ . Letting  $\beta^2 = (8\sin^4 \frac{\pi}{8} + 5)^{-1}$  makes the latter two values equal, hence

$$\lambda_{\min}(P) = \sqrt{\frac{2}{8\sin^4 \frac{\pi}{8} + 5}} \sin^2 \frac{\pi}{8} = \frac{2 - \sqrt{2}}{4\sqrt{4 - \sqrt{2}}} \approx 0.091,$$

where we have used the identity  $\sin^2 \frac{\pi}{8} = \frac{1}{2}(1 - \sqrt{\frac{1}{2}})$ . Hence we conclude

$$\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) \leq 4 \log \frac{4\sqrt{4 - \sqrt{2}}}{2 - \sqrt{2}} \approx 13.83.$$

Note that a lower bound is the Chernoff bound of the two pure output states  $|0\rangle\langle 0| = \mathcal{M}(|00\rangle\langle 00|)$  and  $|+\rangle\langle +| = \overline{\mathcal{M}}(|00\rangle\langle 00|)$ , which is  $\log 2 = 1$ , so  $\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) \geq 1$ . It seems reasonable to conjecture that this is optimal, but we do not have at present a proof of it.  $\square$

*Example 4:* We briefly discuss another example due to Krawiec *et al.* [21], which consists of two quantum-classical channels (qc-channels) implementing two rank-one POVMs on a qutrit  $A$ , and the output  $Y$  is a nine-dimensional Hilbert space. They are given by vectors  $|x_i\rangle \in A$  and  $|y_i\rangle \in A$  ( $i = 1, \dots, 9$ ) such that  $\sum_{i=1}^9 |x_i\rangle\langle x_i| = \sum_{j=1}^9 |y_j\rangle\langle y_j| = I$ :

$$\mathcal{P}(\rho) = \sum_{i=1}^9 \langle x_i | \rho | x_i \rangle |i\rangle\langle i|, \quad \overline{\mathcal{P}}(\rho) = \sum_{j=1}^9 \langle y_j | \rho | y_j \rangle |j\rangle\langle j|.$$

The Kraus operators are  $E_i = |i\rangle\langle x_i|$  and  $F_j = |j\rangle\langle y_j|$ , which makes it easy to calculate  $\text{span}\{E_i^\dagger F_j\} = \text{span}\{|x_i\rangle\langle y_j|\}$ .

In [21] it is shown how to choose the two POVMs in such a way that this subspace does not contain the identity  $I$  and indeed satisfies the “disjointness” condition of Duan *et al.* [16] for perfect finite-copy distinguishability of the two channels using adaptive strategies. Thus,  $\xi^{\mathbb{A}}(\mathcal{P}, \overline{\mathcal{P}}) = +\infty$ . On the other hand, it is proven in [21] that the subspace contains a positive definite matrix  $P > 0$ , showing by Proposition 2 that  $\xi^{\mathbb{P}}(\mathcal{P}, \overline{\mathcal{P}}) < +\infty$ .  $\square$

## IV. CONCLUSION

So indeed there are channels, entanglement-breaking channels at that, for which the adaptive and the non-adaptive Chernoff exponents are different; in fact, the separation is maximal, in that the former is  $+\infty$  while the latter is finite: They lend themselves easily to experiments, as the channels of Example 3 are composed of simple qubit measurement and state preparations. It should be noted that this separation is a robust phenomenon, and not for example related to the perfect finite-copy distinguishability. Namely, by simply mixing our example channels with the same small fraction  $\epsilon > 0$  of the completely depolarizing channel  $\tau$ , we get two new channels  $\mathcal{M}' = (1 - \epsilon)\mathcal{M} + \epsilon\tau$  and  $\overline{\mathcal{M}}' = (1 - \epsilon)\overline{\mathcal{M}} + \epsilon\tau$  with only smaller non-adaptive Chernoff bound,  $\xi^{\mathbb{P}}(\mathcal{M}', \overline{\mathcal{M}}') \leq \xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) < +\infty$ , but the fully general adaptive strategies yield arbitrarily large  $\xi^{\mathbb{A}}(\mathcal{M}', \overline{\mathcal{M}}')$ , as it is based on a two-copy strategy. On the other hand,  $\xi^{\mathbb{A}}(\mathcal{M}', \overline{\mathcal{M}}') < +\infty$ , because the Kraus operators of the channels satisfy  $I \in \text{span}\{E_i^\dagger F_j\}$ , which according to Duan *et al.* [16] implies that  $\mathcal{M}'$  and  $\overline{\mathcal{M}}'$  are not perfectly distinguishable under any  $\mathbb{A}_n$  for any finite  $n$ , and the result Yu and Zhou [17] gives a finite upper bound on the Chernoff exponent  $\xi^{\mathbb{A}}(\mathcal{M}', \overline{\mathcal{M}}')$ .

Since the error rate tradeoff function

$$\begin{aligned} &B_e^{\mathbb{S}}(r|\mathcal{M}||\overline{\mathcal{M}}) \\ &:= \sup \left\{ s \left| \begin{array}{l} \exists \mathcal{T}_n \in \mathbb{S}_n, \\ r \leq \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n), \\ s \leq \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \alpha_n(\mathcal{M}||\overline{\mathcal{M}}|\mathcal{T}_n) \end{array} \right. \right\} \end{aligned}$$

is continuous near  $r = \xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}})$ , whereas the adaptive variant  $B_e^{\mathbb{A}}(r|\mathcal{M}||\overline{\mathcal{M}})$  is infinite everywhere, we automatically get separations in the Hoeffding setting, as well. Note that there is no contradiction with the results of [22], [23], which showed equality of the adaptive and the non-adaptive Stein’s exponents, which are indeed both  $+\infty$ : for the non-adaptive one this follows from the fact that the channels on the same input prepare different pure states,  $|0\rangle\langle 0|$  for  $\mathcal{M}$ ,  $|+\rangle\langle +|$  for  $\overline{\mathcal{M}}$ .

## ACKNOWLEDGMENTS

FS and AW acknowledge partial financial support by the Baidu-UAB collaborative project ‘Learning of Quantum Hidden Markov Models’, the Spanish MINECO (projects FIS2016-86681-P and PID2019-107609GB-I00/AEI/10.13039/501100011033) with the support of FEDER funds, and the Generalitat de Catalunya (project 2017-SGR-1127). FS was also supported by the Catalan Government 001-P-001644 QuantumCAT within the ERDF Program of Catalunya. MH is supported in part by Guangdong Provincial Key Laboratory (grant no. 2019B121203002), a JSPS Grant-in-Aids for Scientific Research (A) no. 17H01280 and for Scientific Research (B) no. 16KT0017, and Kayamori Foundation of Information Science Advancement.

## REFERENCES

- [1] Farzin Salek, Masahito Hayashi, and Andreas Winter. When are Adaptive Strategies in Asymptotic Quantum Channel Discrimination Useful? [arXiv\[quant-ph\]:2011.06569](#), 2020.
- [2] Dennis Kretschmann and Reinhard F. Werner. Quantum channels with memory. *Physical Review A*, 72(6):062323, Dec 2005.
- [3] Gus Gutoski and John Watrous. Toward a general theory of quantum games. In *Proc. 39th annual ACM Symposium on Theory of Computing (STOC)*, pages 565–574, 2007.
- [4] Giulio Chiribella, G. Mauro D’Ariano, and Paolo Perinotti. Memory effects in quantum channel discrimination. *Physical Review Letters*, 101:180501, Oct 2008.
- [5] Giulio Chiribella, G. Mauro D’Ariano, and Paolo Perinotti. Transforming quantum operations: Quantum supermaps. *Europhysics Letters*, 83:30004, Aug 2008.
- [6] Giulio Chiribella, G. Mauro D’Ariano, and Paolo Perinotti. Theoretical framework for quantum networks. *Physical Review A*, 80:022339, Aug 2009.
- [7] Masahito Hayashi. Discrimination of Two Channels by Adaptive Methods and Its Application to Quantum System. *IEEE Transactions on Information Theory*, 55(8):3807–3820, Aug 2009.
- [8] John Watrous. Semidefinite Programs for Completely Bounded Norms. *Theory of Computing*, 5(11):217–238, Nov 2009.
- [9] Gus Gutoski. On a measure of distance for quantum strategies. *Journal of Mathematical Physics*, 53(3):032202, Mar 2012.
- [10] Stefano Pirandola, Bhaskar Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12:724–733, Nov 2018. [arXiv\[quant-ph\]:1811.01969](#).
- [11] Vishal Kataryia and Mark M. Wilde. Geometric distinguishability measures limit quantum channel estimation and discrimination. [arXiv\[quant-ph\]:2004.10708](#), 2020.
- [12] Aram W. Harrow, Avinatan Hassidim, Debbie W. Leung, and John Watrous. Adaptive versus nonadaptive strategies for quantum channel discrimination. *Physical Review A*, 81:032339, Mar 2010.
- [13] Alexei Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys* 52:6 1191-1249, 52(6):1191–1249, 1997.
- [14] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. In *Proc. 13th Annual ACM Symposium on Theory of Computation (STOC)*, pages 20–30, 1997.
- [15] Vern I. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.
- [16] Runyao Duan, Yuan Feng, and Mingsheng Ying. Perfect Distinguishability of Quantum Operations. *Physical Review Letters*, 103:210501, Nov 2009.
- [17] Nengkun Yu and Li Zhou. Chernoff Bound for Quantum Operations is Faithful. [arXiv\[quant-ph\]:1705.01642](#), May 2017.
- [18] Stefano Pirandola, Riccardo Laurenza, Cosmo Lupo, and Jason L. Pereira. Fundamental limits to quantum channel discrimination. *npj Quantum Information*, 5:50, June 2019. [arXiv\[quant-ph\]:1803.02834](#).
- [19] Runyao Duan, Chen Guo, Chi-Kwong Li, and Yinan Li. Parallel distinguishability of quantum operations. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2259–2263, July 2016.
- [20] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.
- [21] Aleksandra Krawiec, Łukasz Paweł, and Zbigniew Puchała. Discrimination of POVMs with rank-one effects. [arXiv\[quant-ph\]:2002.05452](#), Feb 2020.
- [22] Xin Wang and Mark M. Wilde. Resource theory of asymmetric distinguishability for quantum channels. *Physical Review Research*, 1:033169, Dec 2019.
- [23] Mario Berta, Christoph Hirche, Eneet Kaur, and Mark M. Wilde. Amortized channel divergence for asymptotic quantum channel discrimination. *Letters in Mathematical Physics*, 110(8):2277–2336, 2020.