Programmable discrimination with an error margin

G. Sentís,[1] E. Bagan,[1,2] J. Calsamiglia,[1] and R. Muñoz-Tapia[1]

[1]*Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*
[2]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*

The problem of optimally discriminating between two completely unknown qubit states is generalized by allowing an error margin. It is visualized as a device—the programmable discriminator—with one data and two program ports, each fed with a number of identically prepared qubits—the data and the programs. The device aims at correctly identifying the data state with one of the two program states. This scheme has the unambiguous and the minimum-error schemes as extremal cases, when the error margin is set to zero or it is sufficiently large, respectively. Analytical results are given in the two situations where the margin is imposed on the average error probability—weak condition—or it is imposed separately on the two probabilities of assigning the state of the data to the wrong program—strong condition. It is a general feature of our scheme that the success probability rises sharply as soon as a small error margin is allowed, thus providing a significant gain over the unambiguous scheme while still having high confidence results.

I. INTRODUCTION

Quantum state discrimination is one of the most basic yet fundamental tasks in quantum information [1]. In its simplest form, it consists in a protocol that tells in which out of two given states a quantum system was prepared. This is a primitive of great practical interest that has been investigated from many perspectives and for which many key results have been obtained. Theoretical results also abound in the literature, e.g., state discrimination provides an operational distance between any two states [2] based on the degree of difficulty of telling one from the other. It has also been shown that for multiple copies of pure states there exist individual adaptive measurements on each copy that provide exactly the same discrimination power as the optimal (global) measurement strategy [3]. This is however not so for mixed states, and there is numerical evidence that even the corresponding asymptotic exponential error rates are different in this case [4,5].

Generically, a discrimination protocol, to which we will refer throughout the paper as *device*, *machine*, or more explicitly as *discriminator*, is not universal but specifically designed for each given pair of possible states. A significant conceptual twist on discrimination was introduced in Refs. [6,7], where devices that work for arbitrary pairs of states were considered. These machines have two program ports through which multiple copies of the unknown quantum states are loaded ("the programs," for short). Multiple copies of a third state (guaranteed to coincide with one of the states loaded through the program ports) are fed into the data port of the machine. This so-called *programmable discriminator* is designed to report whether the state of the data is that of the first program, or whether it is that of the second program. The discrimination protocol exploits the difference between the permutation symmetry of the global state of the three ports in the two alternatives. These machines work for discrete [7,8] as well as for continuous variable systems [9]. Programmable discriminators can be regarded as machine-learning devices. It has recently been shown that, in some settings, optimal performance can be attained with a suitable measurement on the two programs followed by a measurement on the data, where only classical communication between the two separate measurements is required. Not only does this mean an important saving of resources, as conventional memory suffices to store the (classical) output of the first measurement, but also that programmable discriminators can be reused and still exhibit optimal performance without having to reload the program ports [10]. Interestingly, programmable discrimination is also formally equivalent to a change-point problem [11]. Let us assume that a source produces states of an unknown type and that either at time t_1 or at time t_2 the same source starts producing states of a different type. The change-point problem consists in identifying whether the time at which the change occurs is t_1 or t_2.

In most of the literature so far either the minimum error or the unambiguous discrimination scheme is considered. In the former, the discriminator always produces a conclusive answer about the identity of the input state, but sometimes this answer is wrong. In the unambiguous scheme no error is allowed, that is, the input state must be correctly identified with certainty. This can only happen at the expense of producing some inconclusive answers or, in other words, the machine sometimes must *abstain* [12] from giving an answer. In both cases optimality means that the machine attains maximum success probability. It is clear that, if we relax the unambiguous scheme by tolerating some error rate, we can increase the success probability. Likewise, by allowing some rate of inconclusive answers in the minimum-error scheme, we can also increase the reliability of the answers. Hence by introducing an error margin we can unify minimum error and unambiguous discrimination. Both become extremal points of the unified discrimination with error margin scheme [13–15]. Interpolating between these two extremal cases may have practical interest in some situations.

In this paper we combine the two concepts above and analyze the optimal performance of a qubit multiple-copy programmable machine when an error rate is allowed. We will show that by relaxing the zero error condition slightly the resulting scheme provides an important enhancement in

performance over the widely used unambiguous scheme. We will first review the standard problem, when the states between which we wish to discriminate are known. For the sake of self-containedness, we will rederive the success probability for a given error margin in both the so-called weak and strong senses. We will then present our results for programmable devices and obtain the analytical expression of the success probability as a function of the error margins. We will discuss our results in a separate section and will end the paper by stating our conclusions.

II. DISCRIMINATION WITH ERROR MARGINS

Consider two pure nonorthogonal states $\rho_1 = |\psi_1\rangle\langle\psi_1|$, $\rho_2 = |\psi_2\rangle\langle\psi_2|$ as hypotheses of a standard two-state discrimination problem, where for simplicity we assign equal prior probabilities to each state. The discrimination with an error margin protocol can be thought of as a generalized measurement on the system, described mathematically by a positive operator valued measure (POVM) with three elements $\mathcal{E} = \{E_1, E_2, E_0\}$, where the operator E_1 (E_2) is associated to the statement "the measured state is ρ_1 (ρ_2)," whereas E_0 is associated to the inconclusive answer or abstention. The overall success, error, and inconclusive probabilities are $P_s = \frac{1}{2}[\text{tr}(E_1\rho_1) + \text{tr}(E_2\rho_2)]$, $P_e = \frac{1}{2}[\text{tr}(E_2\rho_1) + \text{tr}(E_1\rho_2)]$, and $Q = \frac{1}{2}[\text{tr}(E_0\rho_1) + \text{tr}(E_0\rho_2)]$, respectively. The relation $P_s + P_e + Q = 1$ is guaranteed by the POVM condition $E_0 + E_1 + E_2 = \mathbb{1}$. The optimal discrimination with an error margin protocol is obtained by maximizing the success probability P_s over any possible POVM $\mathcal{E}$ that satisfies that certain errors occur with a probability not exceeding the given margin. Generically, these conditions imply a nonvanishing value of the inconclusive probability Q.

In this paper, we consider two error margin conditions: *weak* and *strong*. The weak condition states that the *average* error probability cannot exceed a margin, i.e.,

$$P_e = \frac{1}{2}[\text{tr}(E_2\rho_1) + \text{tr}(E_1\rho_2)] \leqslant r. \quad (1)$$

The strong condition imposes a margin on the probabilities of misidentifying *each* possible state, i.e.,

$$p(\rho_2|E_1) = \frac{\text{tr}(E_1\rho_2)}{\text{tr}(E_1\rho_1) + \text{tr}(E_1\rho_2)} \leqslant r, \quad (2)$$

$$p(\rho_1|E_2) = \frac{\text{tr}(E_2\rho_1)}{\text{tr}(E_2\rho_1) + \text{tr}(E_2\rho_2)} \leqslant r, \quad (3)$$

where $p(\rho_2|E_1)$ and $p(\rho_1|E_2)$ are the probabilities that the state identified as ρ_1 is actually ρ_2 and the other way around, respectively. The strong condition is obviously more restrictive, as it sets a margin on both types of errors separately. However, as we will see, the two conditions are directly related: the strong one just corresponds to the weak one with a tighter error margin [14]. Note that both error margin schemes have the unambiguous (when $r = 0$) and the minimum-error schemes (when r is large enough) as extremal cases. We will denote by r_c the critical margin above which the success probability does not increase and thus coincides with that of (the unrestricted) minimum-error discrimination.

For the weak condition, it is straightforward to obtain the maximum success probability by taking into account that
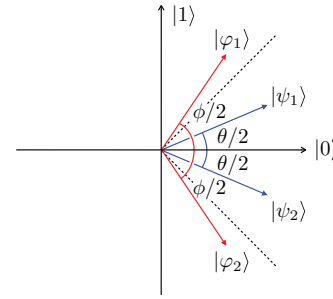

FIG. 1. (Color online) Parametrization of the states $|\psi_1\rangle$, $|\psi_2\rangle$, $|\varphi_1\rangle$, and $|\varphi_2\rangle$ as in Eqs. (4) and (5).

the corresponding error probability must saturate the margin condition (1) for $r \leqslant r_c$, namely, $P_e = r$. Furthermore, the symmetry of the problem dictates that $\text{tr}(E_1\rho_1) = \text{tr}(E_2\rho_2) = P_s$ and $\text{tr}(E_1\rho_2) = \text{tr}(E_2\rho_1) = P_e$. Without loss of generality (see Fig. 1), we can write the input states as

$$|\psi_i\rangle = \cos\frac{\theta}{2}|0\rangle - (-1)^i \sin\frac{\theta}{2}|1\rangle, \quad i = 1,2, \quad (4)$$

where $0 \leqslant \theta \leqslant \pi/2$, and the POVM elements as $E_i = \mu|\varphi_i\rangle\langle\varphi_i|$ for $i = 1,2$, with

$$|\varphi_i\rangle = \cos\frac{\phi}{2}|0\rangle - (-1)^i \sin\frac{\phi}{2}|1\rangle, \quad \frac{\pi}{2} \leqslant \phi \leqslant \pi. \quad (5)$$

The POVM condition implies $E_0 = \mathbb{1} - E_1 - E_2$, and the optimal value of μ is fixed by the extremal value of the inequality $E_0 \geqslant 0$. One obtains $\mu = 1/(1 - \cos\phi) \leqslant 1$ and finally the symmetry conditions fix ϕ to be

$$\tan\frac{\phi}{2} = \begin{cases} \frac{\sqrt{1+c}}{\sqrt{1-c+2\sqrt{r}}} & \text{if} \quad 0 \leqslant r \leqslant r_c, \\ 1 & \text{if} \quad r_c \leqslant r \leqslant 1, \end{cases} \quad (6)$$

where $c = |\langle\psi_1|\psi_2\rangle| = \cos\theta$ is the overlap of the states $|\psi_1\rangle$ and $|\psi_2\rangle$. Notice that in the unambiguous limit, $r = 0$, the POVM elements E_1 and E_2 are orthogonal to the states $|\psi_2\rangle$ and $|\psi_1\rangle$, respectively. In the other extreme case, when the error margin coincides with, or is larger than, the minimum error, $r \geqslant r_c$, one has $E_0 = 0$ (no abstention) and E_1 becomes orthogonal to E_2, i.e., $\phi = \pi/2$. In this range the measurement becomes of von Neumann type and the first case in Eq. (6) implies

$$r_c = \frac{1}{2}(1 - \sqrt{1 - c^2}). \quad (7)$$

Taking into account Eq. (6), the optimal success probability reads

$$P_s^W(r) = \begin{cases} (\sqrt{r} + \sqrt{1-c})^2 & \text{if} \quad 0 \leqslant r \leqslant r_c, \\ \frac{1}{2}(1 + \sqrt{1 - c^2}) & \text{if} \quad r_c \leqslant r \leqslant 1. \end{cases} \quad (8)$$

This result was derived in Ref. [13] and its generalization to arbitrary prior probabilities in Ref. [14] (also in Ref. [15], by fixing an inconclusive rate Q instead of an error margin). Note that the POVM $\mathcal{E}$ is fully determined by the angle ϕ, which in turn is fully determined by the margin r through Eq. (6).

The optimal success probability under the strong condition can be obtained along the same lines of the weak case, but it will prove more convenient to use the connection between both conditions to derive it directly from (8). Let us denote by

r^S (r^W) the error margin of the strong (weak) condition. From the symmetry of the problem, Eqs. (2) and (3) can be written in the form of a weak condition with a margin r^W as

$$P_{\rm e} \leqslant r^S(P_{\rm e} + P_{\rm s}) \equiv r^W. \qquad (9)$$

Hence, if $\mathcal{E}$ is the optimal POVM for a strong margin r^S, it is also optimal for the weak margin r^W, where $P_{\rm e} = r^W$ and $P_{\rm s} = P_{\rm s}^W(r^W)$ is given by Eq. (8). In terms of the success probability, the relation between r^W and r^S reads

$$r^S = \frac{r^W}{P_{\rm s}^W(r^W) + r^W}. \qquad (10)$$

By solving for r^W and substituting into Eq. (8) one derives the success probability for a given r^S, which we denote by $P_{\rm s}^S(r^S)$. For the function $P_{\rm s}^S$ one readily obtains

$$P_{\rm s}^S(r) = \begin{cases} \left(\frac{\sqrt{1-r}}{\sqrt{r}-\sqrt{1-r}}\right)^2(1-c) & \text{if } 0 \leqslant r \leqslant r_c, \\ \frac{1}{2}(1+\sqrt{1-c^2}) & \text{if } r_c \leqslant r \leqslant 1, \end{cases} \qquad (11)$$

in agreement with [13]. Note that the critical margin is the same for both the weak and the strong conditions, i.e., $r_c^W = r_c^S = r_c$. Indeed, beyond the critical point inconclusive results are excluded by optimality ($Q = 0$ and $P_{\rm s} + P_{\rm e} = 1$) and thus there is no difference between the two types of conditions. As in the weak case, there is a correspondence between the angle ϕ and r^S; thus $\mathcal{E}$ can also be parametrized in terms of the strong margin:

$$\tan\frac{\phi}{2} = \begin{cases} \frac{\sqrt{1-r^S}-\sqrt{r^S}}{\sqrt{1-r^S}+\sqrt{r^S}}\frac{\sqrt{1+c}}{\sqrt{1-c}} & \text{if } 0 \leqslant r \leqslant r_c, \\ 1 & \text{if } r_c \leqslant r \leqslant 1. \end{cases} \qquad (12)$$

Note that an ambiguity arises for $c = 1$, as $\phi = \pi$ and then E_1 and E_2 become proportional to one another, independently of the value of r^S. Note also that for $r^S = 0$ and $r^S = r_c$ the values of ϕ for both, weak and strong conditions, coincide.

III. PROGRAMMABLE DISCRIMINATION

Let us elaborate on the definition of a programmable discriminator given in the Introduction. It is a device capable of identifying the state of a system (a qubit in our case) that is guaranteed to be prepared in one of two possible *unknown* pure states, say $\{|\psi_1\rangle, |\psi_2\rangle\}$. By unknown we mean that we lack all the information about their preparation. Instead, we assume that we are supplied with n copies of each of them, which can be fed into the device through two program ports labeled A and C for $|\psi_1\rangle$ and $|\psi_2\rangle$, respectively. In addition, a third port B is loaded with n' copies of the state to be identified. A programmable discriminator is assumed to be a universal device and it should thus work for any pair of states $\{|\psi_1\rangle, |\psi_2\rangle\}$. To make this paper self-contained, in this section we review the state of the art of this discrimination problem. A more general and detailed analysis can be found in Ref. [8].

A programmable discriminator is defined by a universal POVM with three elements $\mathcal{E} = \{E_1, E_2, E_0\}$. The operator E_1 (E_2) corresponds to the machine assigning the label 1 (2) to the copies in B, meaning that their state is identical to

that of the copies in A (C). Once again, the third operator, E_0, is associated to an inconclusive result. The optimal $\mathcal{E}$ is that which maximizes the averaged probability of success, $P_{\rm s} = \int d\psi_1 d\psi_2 P_{\rm s}(\psi_1, \psi_2)$, where $P_{\rm s}(\psi_1, \psi_2)$ is the success probability for a given pair of states $\{|\psi_1\rangle, |\psi_2\rangle\}$ and the average is taken over all possible pairs. Since $\mathcal{E}$ is state independent, $P_{\rm s}$ can be recast as the success probability of discrimination between the two effective global states (of the three-partite port system ABC) when the state in B is either $|\psi_1\rangle$ or $|\psi_2\rangle$. These effective states are given by the averages

$$\sigma_1 = \int d\psi_1 d\psi_2 [\psi_1^{\otimes n}]_A [\psi_1^{\otimes n'}]_B [\psi_2^{\otimes n}]_C,$$
$$\sigma_2 = \int d\psi_1 d\psi_2 [\psi_1^{\otimes n}]_A [\psi_2^{\otimes n'}]_B [\psi_2^{\otimes n}]_C, \qquad (13)$$

respectively, where the notation $[\cdot]$ stands for $|\cdot\rangle\langle\cdot|$. The integrals can be easily computed using the Schur lemma (see [8]) and one obtains

$$\sigma_1 = \frac{1}{d_{AB}d_C} \mathbb{1}_{AB} \otimes \mathbb{1}_C \qquad (14)$$

and the analogous expression for σ_2, where the labels A and C are exchanged. Here $\mathbb{1}_X$ ($\mathbb{1}_{XY}$) is the projector onto the completely symmetric subspace of $\mathcal{H}_X$ ($\mathcal{H}_X \otimes \mathcal{H}_Y$) and $d_X = \text{tr}\,\mathbb{1}_X$ ($d_{XY} = \text{tr}\,\mathbb{1}_{XY}$) is its dimension. In our case we have $d_A = d_C = n + 1$ and $d_{AB} = d_{BC} = n + n' + 1$. The states σ_1 and σ_2 are diagonal in the angular momentum basis $\{|j\,m\rangle\}$, but extra labels are needed to specify how the various subsystems A, B, and C are coupled to each other. In particular, we use the basis $|(j_A j_B)j_{AB} j_C; jm\rangle$ to diagonalize σ_1 and $|j_A(j_B j_C)j_{BC}; jm\rangle$ to diagonalize σ_2, where $j_A = j_C = n/2$, $j_B = n'/2$, and $j_{AB} = j_{BC} = (n + n')/2$. The diagonal form of σ_1 is

$$\sigma_1 = \frac{1}{d_{AB}d_C} \sum_{j=n'/2}^{n'/2+n} \sum_{m=-j}^{j} [(j_A j_B)j_{AB} j_C; jm], \qquad (15)$$

and the analogous form of σ_2 is obtained by coupling j_B and j_C instead of j_A and j_B. The key property of the angular momentum basis is that it satisfies the orthogonality relation

$$\langle (j_A j_B)j_{AB} j_C; jm | j_A(j_B j_C)j_{BC}; j'm'\rangle = c_j \delta_{jj'}\delta_{mm'}, \qquad (16)$$

where the overlaps c_j can be obtained from the Wigner $6j$ symbols [16] [see Eq. (19) below]. Bases obeying an orthogonality relation of the form (16) exist for any two subspaces and are known as Jordan bases [17]. Since a state of the first basis has nonzero overlap with only one element of the second basis, the problem of discriminating σ_1 from σ_2 can be cast as pure state discrimination in each Jordan subspace, which we label by j (note that the overlaps c_j do not depend on the magnetic number m). Hence the optimal POVM can be chosen to be of the form $\mathcal{E} = \bigoplus_j \mathcal{E}_j$, where each $\mathcal{E}_j$ is itself a POVM acting on the subspace $\mathcal{H}_j$ of total angular momentum j, and the total success probability is simply the sum of all the contributions. The success probability for both, the unambiguous ($P_{\rm e} = 0$) and the minimum-error ($Q = 0$)

schemes, are given respectively by [8]

$$P_s^{UA} = \frac{nn'}{(n+1)(n'+2)}, \tag{17}$$

$$P_s^{ME} = \frac{1}{2} + \frac{1}{2}\sum_{k=0}^{n} \frac{n'+2k+1}{(n+1)(n+n'+1)}\sqrt{1 - \left[\frac{(n'+k)!n!}{(n'+n)!k!}\right]^2}, \tag{18}$$

where equal prior probabilities are assumed.

IV. ERROR MARGINS IN PROGRAMMABLE DISCRIMINATION

In this section, we generalize programmable discrimination by allowing an error margin. To ease the notation, rather than labeling the various subspaces $\mathcal{H}_j$ by their total angular momentum j, we will simply enumerate them by natural numbers, $\alpha = 1, 2, \ldots, n+1$, and sort them by increasing value of j. Hence $j = \alpha + n'/2 - 1$. With a slight abuse of notation, we will accordingly write $\mathcal{H}_\alpha$ and enumerate the corresponding POVMs and overlaps as $\mathcal{E}_\alpha$ and c_α, respectively, where one has [8]

$$c_\alpha = \frac{\binom{n'+\alpha-1}{n'}}{\binom{n+n'}{n'}}. \tag{19}$$

A direct consequence of the block structure of the averaged states and $\mathcal{E}$ is that the overall success probability of a programmable discriminator can be expressed as

$$P_s = \sum_{\alpha=1}^{n+1} p_\alpha P_{s,\alpha}, \tag{20}$$

$$p_\alpha = \mathrm{tr}\,(\sigma_i \mathbb{1}_\alpha) = \frac{2\alpha + n' - 1}{(n+1)(n+n'+1)}, \quad i = 1, 2, \tag{21}$$

where $P_{s,\alpha}$ is the success probability of discrimination in the subspace $\mathcal{H}_\alpha$ and p_α is the probability of σ_1 and σ_2 projecting onto that subspace upon performing the measurement $\{\mathbb{1}_\alpha\}$. Likewise, P_e and Q can be expressed as a convex combination of the form (20).

A. Weak error margin

Let us start by considering the weak condition. If we denote the error margin by R, the weak condition reads $P_e \leqslant R$. According to the previous paragraph, the optimal strategy and the corresponding success probability P_s are defined through the maximization problem

$$P_s = \max_{\mathcal{E}} \sum_{\alpha=1}^{n+1} p_\alpha P_{s,\alpha} \quad \text{subject to} \quad \sum_{\alpha=1}^{n+1} p_\alpha P_{e,\alpha} \leqslant R. \tag{22}$$

Recall now that the POVMs $\mathcal{E}_\alpha$ are independent and each of them is parametrized through Eq. (6) by a margin $r = r_\alpha$

which, moreover, satisfies the constraint $P_{e,\alpha} \leqslant r_\alpha$. Therefore, Eq. (22) can be cast as

$$P_s = \max_{\{r_\alpha\}} \sum_{\alpha=1}^{n+1} p_\alpha P_{s,\alpha}^W(r_\alpha) \quad \text{subject to} \quad \sum_{\alpha=1}^{n+1} p_\alpha r_\alpha = R, \tag{23}$$

where the functions $P_{s,\alpha}^W$ are defined as in Eq. (8) with $c = c_\alpha$. In other words, these functions give the success probability of discrimination in the subspaces $\mathcal{H}_\alpha$ with *weak* error margins r_α. The maximization of the success probability translates into finding the optimal set of weak margins $\{r_\alpha\}_{\alpha=1}^{n+1}$ whose average, $\sum_{\alpha=1}^{n+1} p_\alpha r_\alpha$, equals a (global) margin R.

Let us start by discussing the extreme cases of this scheme. On the unambiguous side, $R = 0$, the only possible choice is $r_\alpha = 0$ for all values of α, and the success probability is given by (17). At the other end point, if $R \geqslant R_c = \sum_{\alpha=1}^{n+1} p_\alpha r_{c,\alpha}$, where $r_{c,\alpha}$ is the critical margin in the subspace $\mathcal{H}_\alpha$, given by (7) with $c = c_\alpha$, we immediately recover the minimum-error result (18). We will refer to R_c as the global critical margin.

An explicit expression for P_s if $0 < R < R_c$ is most easily derived by starting at the unambiguous end and progressively increasing the margin R. For a very small error margin, the Lagrange multiplier method provides the maximum. It occurs at $r_\alpha = r_\alpha^{(1)}$, where

$$r_\alpha^{(1)} = \frac{1 - c_\alpha}{\sum_{\alpha=1}^{n+1} p_\alpha(1 - c_\alpha)} R. \tag{24}$$

This solution is valid only when all (partial) error margins are below their critical values, $r_\alpha^{(1)} \leqslant r_{c,\alpha}$. If this inequality holds, the maximum success probability is $P_s = \sum_\alpha p_\alpha P_{s,\alpha}^W(r_\alpha^{(1)})$. The use of the superscript "(1)" will become clear shortly.

If we keep on increasing the global margin R, it will eventually reach a value $R = R_1$ at which the error margin of the first subspace $\mathcal{H}_1$ is saturated, namely, where $r_1^{(1)} = r_{c,1}$. This is so because the overlaps, given in Eq. (19), satisfy $c_1 < c_2 < \cdots < c_{n+1} = 1$. Hence we have $r_1^{(1)} > r_2^{(1)} > \cdots > r_{n+1}^{(1)}$ and $r_{c,1} < r_{c,2} < \cdots < r_{c,n+1}$, according to (24) and (7), respectively. The expression for R_1 can be read off from Eq. (24):

$$R_1 = \frac{r_{c,1}}{1 - c_1} \sum_{\alpha=1}^{n+1} p_\alpha(1 - c_\alpha). \tag{25}$$

For $R > R_1$, the optimal value of the margin of subspace $\mathcal{H}_1$ is then frozen at the value $r_1 = r_{c,1}$, and the remaining margins are obtained by excluding the fixed contribution of the subspace $\mathcal{H}_1$, i.e., by computing the maximum on the right-hand side of

$$P_s - p_1 P_{s,1}^W(r_{c,1}) = \max_{\{r_\alpha\}} \sum_{\alpha=2}^{n+1} p_\alpha P_{s,\alpha}^W(r_\alpha) \quad \text{subject to} \quad \sum_{\alpha=2}^{n+1} p_\alpha r_\alpha = R - p_1 r_{c,1}. \tag{26}$$

The location of this maximum, which we denote by $\{r_\alpha^{(2)}\}_{\alpha=2}^{n+1}$, is formally given by (24) with R replaced by $R - p_1 r_{c,1}$ and the sum in the denominator running from $\alpha = 2$ to $n + 1$. In this case, we have

$$P_s = p_1 P_{s,1}^W(r_{c,1}) + \sum_{\alpha=2}^{n+1} p_\alpha P_{s,\alpha}^W\left(r_\alpha^{(2)}\right). \tag{27}$$

Again, this is valid only until R reaches a second saturation point R_2, i.e., provided $R_1 < R < R_2$, and so on. Clearly, the margins r_α saturate in an orderly fashion as we increase R.

Iterating the procedure described above, the optimal error margins in the interval $R_{\beta-1} \leqslant R \leqslant R_\beta$ (throughout the paper, Greek indexes run from 1 to $n + 1$), where $R_0 \equiv 0$ and $R_{n+1} \equiv R_c$, are found to be

$$r_\alpha^{(\beta)} = \frac{1 - c_\alpha}{\chi_\beta}\left(R - \xi_\beta\right), \tag{28}$$

where

$$R_\beta = \frac{r_{c,\beta}}{1 - c_\beta}\chi_\beta + \xi_\beta, \tag{29}$$

and

$$\xi_\beta = \sum_{\alpha=1}^{\beta-1} p_\alpha r_{c,\alpha}, \quad \chi_\beta = \sum_{\alpha=\beta}^{n+1} p_\alpha(1 - c_\alpha). \tag{30}$$

The success probability in this interval [analogous to Eq. (27)] is

$$P_s = P_{s,\beta}^{sat} + \sum_{\alpha=\beta}^{n+1} p_\alpha P_{s,\alpha}^W\left(r_\alpha^{(\beta)}\right), \tag{31}$$

where

$$P_{s,\beta}^{sat} = \sum_{\alpha=1}^{\beta-1} p_\alpha P_{s,\alpha}(r_{c,\alpha}) = \frac{1}{2}\sum_{\alpha=1}^{\beta-1} p_\alpha\left(1 + \sqrt{1 - c_\alpha^2}\right) \tag{32}$$

is the contribution to the success probability of the subspaces where the error margins are frozen at their critical values. After some algebra, we find that the success probability can be written in a quite compact form as

$$P_s = P_{s,\beta}^{sat} + (\sqrt{R - \xi_\beta} + \sqrt{\chi_\beta})^2, \quad R_{\beta-1} \leqslant R \leqslant R_\beta. \tag{33}$$

Equations (28) to (33) comprise our main result.

B. Strong error margin

The concept of a strong margin for programmable machines requires a more careful formulation than that of a weak margin since, in principle, there are different conditions one can impose on the various probabilities involved. For instance, one could require the strong conditions (2) and (3) for *every* possible pair of states fed into the machine, that is, for every given $\{\rho_1 = [\psi_1], \rho_2 = [\psi_2]\}$. This approach is quickly seen to be trivial since the machine, whose performance is independent of the states, is required to satisfy the condition in a worst case scenario, in which $|\psi_1\rangle$ and $|\psi_2\rangle$ are arbitrarily close to each other. For any value of the error margin less than $1/2$ the inconclusive probability must then approach unity, i.e., $Q \to 1$. This implies that both P_s and P_e vanish. A similar

argument leads to the trivial solution $P_s = P_e = 1/2$ if the margin is larger than or equal to $1/2$.

The task performed by a programmable discriminator can be most naturally viewed as state labeling: the machine attaches the label 1 (2) to the data if its state is identified, by a "clicking" of the operator E_1 (E_2), to be that of the qubits loaded through program port A (C); i.e., the state of the ports has the pattern $[\psi_1^{\otimes n}][\psi_1^{\otimes n'}][\psi_2^{\otimes n}]$ ($[\psi_1^{\otimes n}][\psi_2^{\otimes n'}][\psi_2^{\otimes n}]$). For this task, the relevant error probabilities are $p(2|E_1)$ and $p(1|E_2)$, namely, the probability of wrongly assigning the labels 1 and 2, respectively. It seems, therefore, more suitable for programmable discrimination to impose the strong margin conditions $p(2|E_1) \leqslant R$ and $p(1|E_2) \leqslant R$. In terms of the average states σ_1 and σ_2 in Eq. (13) these conditions are

$$p(2|E_1) = \frac{\operatorname{tr} E_1\sigma_2}{\operatorname{tr} E_1\sigma_1 + \operatorname{tr} E_1\sigma_2} \leqslant R, \tag{34}$$

and likewise for $p(1|E_2)$.

Note that in contrast to the weak case, here the conditional probabilities are nonlinear functions of the POVM elements, and thus the maximization of the success probability under these conditions is *a priori* more involved. To circumvent this problem, we can use the relation (10), which for programmable discrimination also holds, and reads

$$R^S = \frac{R^W}{P_s(R^W) + R^W} \tag{35}$$

to express the (global) weak error margin R^W in terms of the strong one R^S. Then, one simply uses Eqs. (28) to (33) to obtain the maximum success probability. The inversion of Eq. (35) is somewhat lengthy but straightforward. The difficulty arises from the fact that the success probability, Eq. (33), is a piecewise function whose expression depends specifically on how many margins r_α have reached their critical value $r_{c,\alpha}$ for a given R^S. Thus we need to compute the strong saturation points R_β^S, analogous to (29), through the relation (35).

V. DISCUSSION OF THE RESULTS

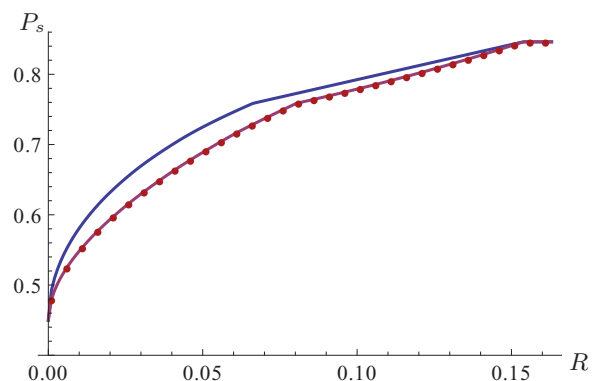In Fig. 2 we plot the maximum success probabilities for both the weak and the strong conditions as a function of a

FIG. 2. (Color online) P_s vs R for a weak (upper line) and a strong (lower line) condition, for $n = 9$ and $n' = 2$. The global critical margin is $R_c \simeq 0.154$. A numerical maximization of the success probability under the strong condition (34) (points) is seen to agree with our analytical solution.

common (global) margin R, for nine program and two data copies. We also show in Fig. 2 the results of a numerical optimization with the strong condition (dots), which exhibit perfect agreement with our analytical solution. We observe that by allowing just a 5% error margin, the success probability increases by more than 50%. This is just an example of a general feature of programmable discrimination with an error margin: the success probability increases sharply for small values of the error margin.

A comment about the effect of the subspace $\mathcal{H}_{n+1}$ on the shape of the plots is in order. This subspace contains the completely symmetric states of the whole system ABC and, hence, it is impossible to tell if the state of the data (B) coincides with that of one program (A) or that of the other (C); more succinctly, $c_{n+1} = 1$. Therefore, half the number of conclusive answers will be correct and half of them will be wrong, and $P_{s,n+1}^W = r_{n+1}$, provided $r_{n+1} \leqslant r_{c,n+1} = 1/2$. Increasing the error margin simply allows for an equal increase in the success probability. This is reflected in the linear stretch in the upper curve in Fig. 2, right before the (rightmost) flat plateau. For the strong condition, the same situation arises in the interval $R_n^S \leqslant R \leqslant R_c$, but the plot of the success probability is *not* a straight line due to the nonlinear relation (35) between the weak and the strong margin.

An alternative (though completely equivalent) way to compute the maximum success probability with a strong margin is based on the observation that the POVMs $\mathcal{E}_\alpha$ are also fully determined by strong margins, r_α^S, through Eq. (12), with the exception of $\mathcal{E}_{n+1}$, for which $c = c_{n+1} = 1$ [giving rise to an ambiguity, as discussed after Eq. (12)]. In this approach, the success probability becomes a convex combination of $P_{s,\alpha}^S(r_\alpha^S)$, as in Eq. (20), where these functions are given in Eq. (11) with $c = c_\alpha$. The optimal set $\{r_\alpha^{S\,(\beta)}\}$ can be readily obtained from the weak margins in Eq. (28) using the relation (10). The strategy in the last subspace $\mathcal{H}_{n+1}$ can be easily seen to consist in abstention with a certain probability, and a random choice of the labels 1 and 2 otherwise.

The bar chart in Fig. 3 represents an optimal strategy in terms of the corresponding weak and strong error margins. For this example, we have chosen 11 program and two data copies. For illustration purposes, the (global) margin is set to a low value of 0.0055. The wide vertical bars in the background depict the critical margins $r_{c,\alpha}$. There are 12 of them, displayed in increasing order of α (the first one is not visible because of the small value of $r_{c,1}$). On their left (right) halves, a narrow green (orange) bar depicts the optimal weak (strong) margin r_α^W (r_α^S) (we attach the subscripts W and S through the rest of the paper to avoid confusion). We note that the first five margins ($\alpha \leqslant 5$) have reached their critical value. For $\alpha > 5$, the weak margins decrease monotonically according to Eq. (28). For the last one, we have $r_{n+1}^W = r_{12}^W = 0$, which holds for any value of R, provided $R \leqslant R_n$. This must be so, since we recall that the projections of σ_1 and σ_2 onto the subspace with maximum angular momentum are indistinguishable. Clearly, allowing for $r_{n+1}^W > 0$ while there is still room for the other margins to increase cannot be optimal.

Also noticeable in Fig. 3 is that the set of strong margins that have not reached their critical value $r_{c,\alpha}$ has a flat profile (this does not apply to r_{n+1}^S that is always frozen to its critical value of 1/2). To provide an explanation for this, we write the
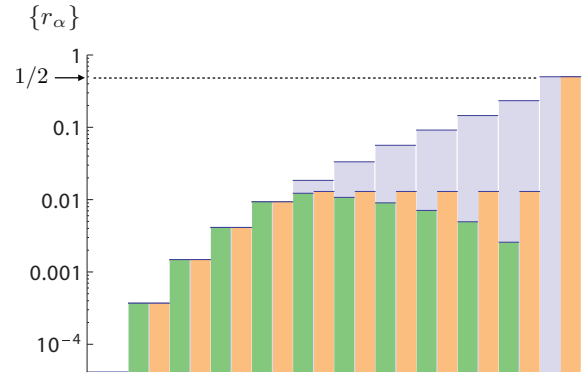

FIG. 3. (Color online) The various error margins for $n = 11, n' = 2$, and a (global) margin $R = 0.0055$. The full heights of the wide bars in the background (blue) represent the values of the critical margins $r_{c,\alpha}$, starting from $\alpha = 1$ (leftmost) up to $\alpha = 12$ (rightmost). For the same values of α, each pair of narrow bars represents the weak margin r_α^W [left (green)] and the strong margin r_α^S [right (orange)]. We note that the first five error margins have reached their critical value. The values for $\alpha = 1$ are very small, which explains why the corresponding bars do not show up in the chart.

equality in Eq. (34), which is attained if $R \leqslant R_c$, as $R P_s - (1 - R)P_e = 0$, using once again the symmetry of the problem. We next write the success and error probabilities as a convex sum over α and use the equality in the strong conditions (2) and (3) for each subspace $\mathcal{H}_\alpha$ to express $P_{e,\alpha}^S$ in terms of $P_{s,\alpha}^S$. We obtain the strong condition

$$\sum_\alpha p_\alpha P_{s,\alpha}^S(r_\alpha^S)\left[R - (1 - R)\frac{r_\alpha^S}{1 - r_\alpha^S}\right] = 0. \quad (36)$$

The terms in square brackets can be positive or negative depending on r_α^S being smaller or larger than R, both of which are possible. So, at face value, this equation cannot explain the flat profile of r_α^S and more work is needed. Next, we use the Lagrange multiplier method to maximize $P_s = \sum_\alpha p_\alpha P_{s,\alpha}^S(r_\alpha^S)$ and note that the dependence of $P_{s,\alpha}^S$ on α (i.e., the term $1 - c_\alpha$) factorizes, as can be checked from Eq. (11). Without further calculation, we can anticipate that the optimal margins will be determined by $n + 1$ equations of the form $p_\alpha(1 - c_\alpha)f(r_\alpha^S) = 0$, where f can be a function only of R, the Lagrange multiplier, and the number of margins below their critical value. Hence all the (unfrozen) margins will have the same optimal value. For $\beta = 1$ (no frozen margins) we have the simple solution $r_\alpha^{S,(1)} = R$ for all α, and the corresponding success probability is

$$P_s = \left(\frac{\sqrt{1 - R}}{\sqrt{R} - \sqrt{1 - R}}\right)^2 \frac{nn'}{(n + 1)(n' + 2)} \quad (37)$$

for a sufficiently small strong margin R.

VI. CONCLUSIONS

In this paper, we have provided two generalizations of programmable state discrimination that enable control on the rate with which errors inevitably arise because of the very principles of quantum mechanics. In the first, a margin is set on the average error probability of mislabeling the input

data states (weak condition). In the second, a more stringent condition is required that, for each label, the probability of it being wrongly assigned is within a given margin (strong condition). Generically, in both cases, the discrimination protocol may result sometimes in an inconclusive outcome (i.e., in being unable to assign a label to the data). We have shown that there is a one-to-one correspondence between these two margins, so that weak and strong conditions turn out to be the same if their margins are related by a simple equation. These generalizations extend the range of applicability of programmable discriminators to scenarios where some rate of errors and some rate of inconclusive outcomes are both affordable; or more specifically, to situations where a trade-off between these two rates is acceptable, which departs from the standard unambiguous (zero error) and minimum-error (zero abstention) discrimination scenarios.

Our results include the analytical expression of the success probability for the optimal programmable device as a function of both weak and strong error margins, as well as the characterization of the POVM that specifies such optimal device. From the analysis of these results, we conclude that small error margins can significantly boost the success probability; i.e., a small departure from the unambiguous scheme can translate into an important increase of the success rate while still having very reliable results (very low error rate). We provide an example of this, where a mere error margin value of 5% adds about 50% to the success probability.

A future extension of this work is, e.g., the asymptotic analysis of programmable discrimination with an error margin, when the data and/or program ports are fed with an asymptotically large number of copies. Also relevant is the analysis of programmable discriminators when the measurement is restricted to those compatible with a machine learning scenario. These devices require only classical memory to store the information about the state of the programs, and use it in a later test stage to fix the measurement on the unknown data. They can be reused an arbitrary number of times without reloading the program ports [10].

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[2] K. M. R. Audenaert *et al.*, Phys. Rev. Lett. **98**, 160501 (2007); J. Calsamigilia *et al.*, Phys. Rev. A **77**, 032311 (2008).

[3] A. Acin, E. Bagan, M. Baig, Ll. Masanes, and R. Munoz-Tapia, Phys. Rev. A **71**, 032338 (2005).

[4] J. Calsamiglia, J. I. de Vicente, R. Munoz-Tapia, and E. Bagan, Phys. Rev. Lett. **105**, 080504 (2010).

[5] B. L. Higgins, A. C. Doherty, S. D. Bartlett, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A **83**, 052314 (2011).

[6] M. Dusek and V. Buzek, Phys. Rev. A **66**, 022112 (2002).

[7] J. A. Bergou and M. Hillery, Phys. Rev. Lett. **94**, 160501 (2005).

[8] G. Sentis, E. Bagan, J. Calsamiglia, and R. Munoz-Tapia, Phys. Rev. A **82**, 042312 (2010); **83**, 039909(E) (2011).

[9] M. Sedlak, M. Ziman, V. Buzek, and M. Hillery, Phys. Rev. A **79**, 062305 (2009).

[10] G. Sentis, J. Calsamiglia, R. Munoz-Tapia, and E. Bagan, Sci. Rep. **2**, 708 (2012).

[11] D. Akimoto and M. Hayashi, Phys. Rev. A **83**, 052328 (2011).

[12] The benefits of abstention in quantum metrology have been recently analyzed in B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Munoz-Tapia, and E. Bagan, New J. Phys. **14**, 105015 (2012); Phys. Rev. Lett. **110**, 100501 (2013); Phys. Rev. A **88**, 012128 (2013).

[13] A. Hayashi, T. Hashimoto, and M. Horibe, Phys. Rev. A **78**, 012333 (2008).

[14] H. Sugimoto, T. Hashimoto, M. Horibe, and A. Hayashi, Phys. Rev. A **80**, 052322 (2009).

[15] E. Bagan, R. Munoz-Tapia, G. A. Olivares-Renteria, and J. A. Bergou, Phys. Rev. A **86**, 040303 (2012).

[16] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1960).

[17] J. A. Bergou, E. Feldman, and M. Hillery, Phys. Rev. A **73**, 032107 (2006).