Optimal quantum state discrimination via nested binary measurements

Matteo Rosati,[1] Giacomo De Palma,[2] Andrea Mari,[1] and Vittorio Giovannetti[1]

[1] *NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56127 Pisa, Italy.*
[2] *QMATH, Department of Mathematical Sciences, University of Copenhagen,*
Universitetsparken 5, 2100 Copenhagen, Denmark

A method to compute the optimal success probability of discrimination of N arbitrary quantum states is presented, based on the decomposition of any N-outcome measurement into sequences of nested two-outcome ones. In this way the optimization of the measurement operators can be carried out in successive steps, optimizing first the binary measurements at the deepest nesting level and then moving on to those at higher levels. We obtain an analytical expression for the maximum success probability after the first optimization step and examine its form for the specific case of $N = 3, 4$ states of a qubit. In this case, at variance with previous proposals, we are able to provide a compact expression for the success probability of any set of states, whose numerical optimization is straightforward; the results thus obtained highlight some lesser-known features of the discrimination problem.

I. INTRODUCTION

The discrimination of quantum states [1] is one of the fundamental problems in Quantum Information and a basic task for several applications in communication [2–6], cryptography [7–9], fundamental questions [10–13], measurement and control [14, 15] and algorithms [16]. Triggered by the observation that non-orthogonal quantum states cannot be perfectly discriminated, this subject has stimulated much work, both from a theoretical and practical point of view: the seminal works of Helstrom [17], Holevo [18] and Yuen *et al.* [19] formalized the problem, obtaining a set of conditions for the optimal measurement operators, which in turn provide the optimal success probability, then solved it for sets of states symmetric under a unitary transformation; more recently, acknowledging that a general analytical solution is hard to find, research focused on finding a solution for sets with more general symmetries [20–22], computing explicitly the optimal measurements for the most interesting sets of states [23–27] and studying the implementation of such measurements with available technology (see for example [28–40] for the case of two optical coherent states, the most relevant for optical communication). Also, the problem of discrimination has been identified as a convex optimization one, arguing that it can be solved efficiently with numerical optimization methods [41].

In this article we attempt to solve the optimal discrimination of N quantum states from a different perspective, by providing a structured expression for the N-outcome Positive Operator Valued Measure (POVM) used to discriminate the states. Indeed it can be shown [42, 43] that any POVM comprising N elements is equivalent to a collection of binary POVM's, i.e., comprising two elements, as the one employed in Ref. [6]: depending on the binary outcome of the first measurement, a second one is applied; its binary outcome in turn affects the choice of the third binary measurement and so on. In this way a sequence of nested binary POVM's can be constructed, where the POVM applied at a given level depends on the string of binary outcomes of the previous ones. This result was already obtained in Ref. [42]. When applied to state discrimination, it acquires a more operational meaning: each binary POVM can be seen as discriminating between two subsets of the initial set of states, identified by previous outcomes. Hence the sequence of measurements induces a sequence of discrimination probabilities, so that, if the optimization problem is solved independently for any set of a fixed number of states, the result can be employed in the optimization problem for larger sets of states.

In the second part of the article, employing this decomposition and the two-state optimal probability [17], we obtain an expression for the success probability of discrimination of any $N = 3, 4$ states, depending on a single measurement operator, and solve the problem analytically for specific sets of states. Then we restrict our attention to qubit states and obtain a compact expression which can be easily optimized numerically case by case, at variance with less compact results for $N = 3$ presented in previous works based on Bloch-space geometry [25–27]. We recover the results of those works and highlight in particular some interesting lesser-known implications of Ref. [26].

The article is structured as follows: in Sec. II we describe the decomposition in terms of nested binary POVM's and provide a proof of its validity, similar to that of Ref. [42]; in Sec. III we apply it to state discrimination and obtain an explicit expression for the case of $N = 3, 4$ arbitrary states, then discuss its optimization in some specific cases; in Sec. IV we treat the case of $N = 3, 4$ qubit states, computing a compact expression which can be optimized numerically and highlighting some results obtained in this way. Eventually in Sec. V we draw some conclusions. Detailed computations of the quantities appearing in the article are provided in the Appendices.

II. GENERAL DECOMPOSITION OF A N-OUTCOME MEASUREMENT INTO NESTED BINARY ONES

In this Section we prove that any quantum measurement with an arbitrary number of outcomes can be decomposed into a sequence of nested measurements with binary outcomes, where the previous results determine the choice of successive measurements. We stress that the same result was obtained in Ref. [42]. At variance with the latter, our proof does not make use of the spectral decomposition of the initial measurement operators; we present it here in a form adapted to the main purpose of the article. Let us suppose we want to perform a quantum measurement with N possible outcomes: it can be expressed in general as a POVM $\mathcal{M}^{(N)}$ of elements E_j, one for each outcome $j = 0, \cdots, N-1$, satisfying the positivity and completeness conditions, i.e., respectively $E_j \geq 0$ and $\sum_{j=0}^{N-1} E_j = \mathbf{1}$, where $\mathbf{1}$ is the identity operator on the Hilbert space of the system to be measured. This expression can be interpreted as a one-shot measurement with several possible results and its practical realization may often be very hard. On the other hand we could restrict to performing only measurements with two outcomes, as described by *binary* POVM's: $\mathcal{B} \equiv \mathcal{M}^{(2)} = \{B_0, B_1\}$. This may be useful when limited technological capabilities or specific theoretical requirements constrain the number of allowed outcomes and the complexity of our measurement. It is then natural to ask whether this smaller set of resources is sufficient to describe a general quantum measurement. We answer positively by showing that the more general N-outcome formalism can be broken up into several binary steps and interpreted as a sequence of nested POVM's with two outcomes, trading a *one-shot*, *multiple-outcome* measurement for a *multiple-step*, *yes-no* measurement.

The nested POVM can be expressed in terms of *conditional binary* POVM's $\mathcal{B}_{\vec{k}} = \{B_{\vec{k},0}, B_{\vec{k},1}\}$, each complete by itself, to be applied only if a specific string $\vec{k}$ of previous results is obtained. For example for $N = 4$ the nested POVM can be realized in two steps and written compactly as the collection of three binary POVM's: $\mathcal{N}^{(4)} = \left\{\mathcal{B}_0^{(2)}, \mathcal{B}_1^{(2)}\right\} \circ \left\{\mathcal{B}^{(1)}\right\}$, properly composed as follows and shown in Fig. 1. The measurement starts by applying the first-step binary POVM $\mathcal{B}^{(1)} = \left\{B_{k_1}^{(1)}\right\}_{k_1=0,1}$ then, depending on its outcome k_1, it selects $\mathcal{B}_{k_1}^{(2)} = \left\{B_{k_1,k_2}^{(2)}\right\}_{k_2=0,1}$ among the two POVM's available in the second-step collection $\left\{\mathcal{B}_0^{(2)}, \mathcal{B}_1^{(2)}\right\}$. Eventually the chosen second-step POVM is applied, receiving an outcome k_2. The total outcome is a string of two bits, i.e., k_1, k_2, whose value identifies one of four possible outcomes, as desired. Suppose now to apply this measurement on a state ρ of some physical system: if the first-step outcome is $k_1 = 0$, the resulting unnormalized evolved state is $\sqrt{B_0^{(1)}}\rho\sqrt{B_0^{(1)}}$; if then the second-step outcome is $k_2 = 0$, the final unnormalized state of the system is $\sqrt{B_{0,0}^{(2)}}\sqrt{B_0^{(1)}}\rho\sqrt{B_0^{(1)}}\sqrt{B_{0,0}^{(2)}}$. This means that the nested POVM has a more explicit representation as

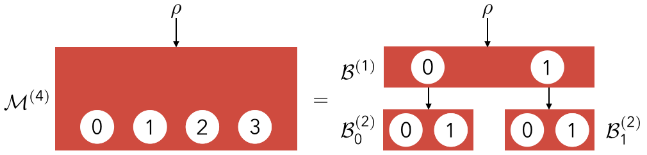

Figure 1. Schematic depiction of the nested decomposition for $N = 4$, explicitly discussed in the text. Any four-outcome measurement $\mathcal{M}^{(4)}$ acting on a state ρ is equivalent to the concatenation of two-outcome measurements: the first-step one $\mathcal{B}^{(1)}$, with result $k_1 = 0, 1$, and the second-step ones $\mathcal{B}_{k_1}^{(2)}$, which are mutually exclusive and applied only if the corresponding first outcome k_1 was obtained.

$$\mathcal{N}^{(4)} = \left\{ F_{k_1,k_2} = \left| \sqrt{B_{k_1,k_2}^{(2)}}\sqrt{B_{k_1}^{(1)}} \right|^2 \right\}_{k_1,k_2=0,1}, \quad (1)$$

where $|X|^2 = X^\dagger X$ is the square of the absolute value of an operator X. In the general case, let us indicate a sequence of $b - a + 1$ bits as

$$k_{(a,b)} = \begin{cases} k_a, k_{a+1}, \cdots, k_b, & b \geq a \\ \emptyset, & b < a, \end{cases} \quad (2)$$

and define as $\mathcal{B}_{k_{(1,u-1)}}^{(u)} = \left\{B_{k_{(1,u-1)},0}^{(u)}, B_{k_{(1,u-1)},1}^{(u)}\right\}$ the binary POVM to be performed at the u-th step if the previous $u-1$ measurements had a sequence of results $k_{(1,u-1)}$. Then we can define a nested POVM $\mathcal{N}^{(N)}$ of order $N = 2^{u_F}$ as

$$\begin{aligned} \mathcal{N}^{(N)} &= \left\{\mathcal{B}_{k_{(1,u_F-1)}}^{(u_F)}\right\}_{k_1,\cdots,k_{u_F-1}=0,1} \circ \cdots \circ \left\{\mathcal{B}^{(1)}\right\} \\ &= \left\{ F_{k_{(1,u_F)}} = \left| \sqrt{B_{k_{(1,u_F)}}^{(u_F)}} \cdots \sqrt{B_{k_1}^{(1)}} \right|^2 \right\}, \quad (3) \end{aligned}$$

i.e., the collection of $2^{u_F}-1$ binary POVM's $\mathcal{B}_{k_{(1,u-1)}}^{(u)}$, for all previous outcomes $k_{(1,u-1)}$ at a given step u and all steps $u = 1, \cdots, u_F$. We can certify that $\mathcal{N}^{(N)}$ so constructed actually is a POVM by checking positivity and completeness of its elements $F_{k_{(1,u_F)}}$: the former requirement is trivial, while the latter follows from the fact that each binary POVM is complete, as shown in Appendix A.

In light of the previous discussion we can now state the main theorem:

Theorem 1. *Any N-outcome POVM $\mathcal{M}^{(N)} = \{E_j\}_{j=0,\cdots,N-1}$ is equivalent to a nested POVM $\mathcal{N}^{(\tilde{N})}$, $\tilde{N} = 2^{u_F}$, as in Eq.* (3)*, composed exclusively of binary POVM's $\mathcal{B}_{k_{(1,u-1)}}^{(u)}$, with a total number of steps u_F equal to:*

1. $\log_2 N$, *if* N *is a power of* 2;
2. $\lceil \log_2 N \rceil$ *otherwise, where* $\lceil \cdot \rceil$ *is the ceiling function, equal to the smallest integer following the argument.*

Proof. Consider the first case above, i.e., $N = 2^{u_F} \equiv \tilde{N}$. We start by providing a binary representation of the labels j of the initial POVM $\mathcal{M}^{(N)}$, i.e., we define $E_{k_{(1,u)}} \equiv E_{j^{(k)}}$, with $j^{(k)} = \sum_{u=1}^{u_F} 2^{u-1} k_u$. In order to prove the theorem we have to show that by combining the elements of the initial N-outcome POVM $\mathcal{M}^{(N)}$ one can always define a set of binary POVM's $\mathcal{B}^{(u)}_{k_{(1,u-1)}}$, for all $k_1, \cdots, k_{u-1} = 0, 1$ and $u = 1, \cdots, u_F$, such that: i) their nested composition is a POVM of the form $\mathcal{N}^{(N)}$, Eq. (3); ii) the elements of the latter are equal to the elements of $\mathcal{M}^{(N)}$.

First of all we construct the binary elements at each step u, by taking the sum of all the elements $E_{k_{(1,u)},k_{(u+1,u_F)}}$ with a fixed value of the first u bits, then renormalizing it by all previous binary elements, as in a Square Root Measurement [3, 44]. For example define the elements of the first-step POVM $\mathcal{B}^{(1)}$ as

$$B^{(1)}_{k_1} = \sum_{k_{(2,u_F)}} E_{k_1,k_{(2,u_F)}}, \tag{4}$$

for each value of the outcome $k_1 = 0, 1$. Being a sum of positive operators, the elements so defined are themselves positive; moreover their sum equals the sum of all the elements of $\mathcal{M}^{(N)}$, implying that they are complete. At the second step define the elements of the two possible POVM's $\mathcal{B}^{(2)}_{k_1}$ as

$$B^{(2)}_{k_{(1,2)}} = \sqrt{B^{(1)}_{k_1}}^{-1} \sum_{k_{(3,u_F)}} E_{k_{(1,2)},k_{(3,u_F)}} \sqrt{B^{(1)}_{k_1}}^{-1}, \tag{5}$$

where the inverse of an operator is to be computed only on its support, while it is equal to 0 on its kernel, i.e., its pseudo-inverse. Also in this case the defined elements are positive by construction, but they are not complete. Indeed it is easy to show, employing the definition (4), that $B^{(2)}_{k_1,0} + B^{(2)}_{k_1,1} = \mathbf{1}_{k_1}$. Here $\mathbf{1}_{k_1}$ is the projector on the support of the previous outcome operator, $B^{(1)}_{k_1}$, which may have a non-trivial kernel, so that in general it holds $\mathbf{1}_{k_1} \leq \mathbf{1}$. This problem may be overcome easily by redefining the POVM elements as $\tilde{B}^{(2)}_{k_{(1,2)}} = B^{(2)}_{k_{(1,2)}} \oplus (\mathbf{1} - \mathbf{1}_{k_1})/2$, i.e., trivially expanding the support of those already defined in (5), so that $\tilde{B}^{(2)}_{k_1,0} + \tilde{B}^{(2)}_{k_1,1} = \mathbf{1}_{k_1} \oplus (\mathbf{1} - \mathbf{1}_{k_1}) = \mathbf{1}$. This operation is trivial because, in the construction (3) of the nested POVM, the operators $B^{(2)}_{k_{(1,2)}}$ always act after the operator $B^{(1)}_{k_1}$, so that the value of the former outside the support of the latter is completely irrelevant. In other words, completeness of the binary POVM's is not necessary for the definition of $\mathcal{N}^{(N)}$ as a proper POVM; it is sufficient to ask for *weak completeness*, i.e., that $\mathcal{B}^{(u)}_{k_{(1,u-1)}}$ is complete on the support of the operator preceding it in the decomposition, $B^{(u-1)}_{k_{(1,u-1)}}$.

Generalizing the previous discussion, at the u-th step we can define the elements of the 2^{u-1} possible POVM's $\mathcal{B}^{(u)}_{k_{(1,u-1)}}$ as

$$\begin{aligned} B^{(u)}_{k_{(1,u)}} &= \sqrt{B^{(u-1)}_{k_{(1,u-1)}}}^{-1} \cdots \sqrt{B^{(1)}_{k_1}}^{-1} \\ &\cdot \sum_{k_{(u+1,u_F)}} E_{k_{(1,u)},k_{(u+1,u_F)}} \sqrt{B^{(1)}_{k_1}}^{-1} \cdots \sqrt{B^{(u-1)}_{k_{(1,u-1)}}}^{-1}. \end{aligned} \tag{6}$$

These elements are positive by construction and they satisfy the weak completeness relation $B^{(u)}_{k_{(1,u-1)},0} + B^{(u)}_{k_{(1,u-1)},1} = \mathbf{1}_{k_{(1,u-1)}}$, which is sufficient to define the POVM $\mathcal{N}^{(N)}$, as discussed in Appendix A. Hence we are left to show that, when combining the binary elements Eq. (6) as in Eq. (3), the elements $F_{k_{(1,u_F)}}$ so constructed are equal to the $E_{k_{(1,u_F)}}$. Indeed let us evaluate Eq. (6) for $u = u_F$, i.e., at the last step, noting that the sum contains only one term:

$$\begin{aligned} B^{(u_F)}_{k_{(1,u_F)}} &= \sqrt{B^{(u_F-1)}_{k_{(1,u_F-1)}}}^{-1} \cdots \sqrt{B^{(1)}_{k_1}}^{-1} \\ &\cdot\ E_{k_{(1,u_F)}} \sqrt{B^{(1)}_{k_1}}^{-1} \cdots \sqrt{B^{(u_F-1)}_{k_{(1,u_F-1)}}}^{-1}. \end{aligned} \tag{7}$$

Let us then successively invert the outer square roots on the left-hand side of the equation exactly $u_F - 1$ times, to obtain the relation

$$E_{k_{(1,u_F)}} = \left| \sqrt{B^{(u_F)}_{k_{(1,u_F)}}} \cdots \sqrt{B^{(1)}_{k_1}} \right|^2 \equiv F_{k_{(1,u_F)}}, \tag{8}$$

which demonstrates that we can recover the initial POVM with the procedure outlined above.

This completes the proof when N is an exact power of 2. If this is not the case, it means that $\log_2 N$ is not an integer and it suffices to consider the nested decomposition for the next higher integer, i.e., set $u_F = \lceil \log_2 N \rceil + 1$, $\tilde{N} = 2^{u_F}$. Let us then trivially expand the initial N-outcome POVM to a $\tilde{N}$-outcome one as

$$\mathcal{M}^{(\tilde{N})} = \mathcal{M}^{(N)} \cup \left\{ E_{k_{(1,u_F)}} = 0, \forall j^{(k)} > N - 1 \right\}, \tag{9}$$

by adding $\tilde{N} - N$ null elements. The nested decomposition $\mathcal{N}^{(\tilde{N})}$ equivalent to $\mathcal{M}^{(\tilde{N})}$ can be computed again by Eqs. (3,6) and it comprises $\tilde{N} - N$ null elements too. If we isolate these elements from the rest we obtain a decomposition

$$\mathcal{N}^{(\tilde{N})} = \mathcal{N}^{(N)} \cup \left\{ F_{k_{(1,u_F)}} = 0, \forall j^{(k)} > N - 1 \right\}, \tag{10}$$

where $\mathcal{N}^{(N)}$ can be interpreted as a nested representation of the initial POVM $\mathcal{M}^{(N)}$. □

III. AN APPLICATION: OPTIMAL QUANTUM STATE DISCRIMINATION

In this Section we apply the previous POVM decomposition to the problem of optimal state discrimination. Let us suppose we are given one copy of a quantum state, represented by a positive and trace-one operator ρ_j, chosen at random from a set $\mathcal{S}^{(N)} = \{\tilde{\rho}_j = p_j \rho_j\}_{j=0,\cdots,N-1}$ of N states weighted with probability p_j, so that $\sum_{j=0}^{N-1} p_j = 1$; we have to perform a measurement to decide which state was sent. If the states are not orthogonal, i.e., $\rho_j \rho_k \neq 0$ for some values of j, k, and we are constrained to give a conclusive answer, there exists no measurement that can succeed with unit probability. The average success probability of discriminating the set of states $S^{(N)}$ with a N-outcome POVM $\mathcal{M}^{(N)}$, as defined in Sec. II, can be computed as

$$P_{Succ}\left(S^{(N)}, \mathcal{M}^{(N)}\right) = \sum_{j=0}^{N-1} \mathrm{Tr}\left[E_j \tilde{\rho}_j\right], \qquad (11)$$

where each measurement outcome E_j is associated with the detection of the respective weighted state $\tilde{\rho}_j$. We are particularly interested in the optimal success probability, obtained by optimizing over all measurements:

$$\mathbb{P}_{Succ}(\mathcal{S}^{(N)}) = \max_{\mathcal{M}^{(N)}} P_{Succ}\left(\mathcal{S}^{(N)}, \mathcal{M}^{(N)}\right). \qquad (12)$$

Following Sec. II, we can always decompose the discrimination measurement into a sequence of nested binary ones, writing the success probability as

$$P_{Succ}\left(\mathcal{S}^{(N)}, \mathcal{N}^{(N)}\right) = \sum_{k_{(1,u_F)}} \mathrm{Tr}\left[F_{k_{(1,u_F)}} \tilde{\rho}_{k_{(1,u_F)}}\right]$$
$$= \sum_{k_{(1,u_F)}} \mathrm{Tr}\left[\left|\sqrt{B^{(u_F)}_{k_{(1,u_F)}}} \cdots \sqrt{B^{(1)}_{k_1}}\right|^2 \tilde{\rho}_{k_{(1,u_F)}}\right], \qquad (13)$$

where we have introduced the binary representation $k_{(1,u_F)}$ for the labels j of the states and measurement operators and employed the definition (3) for the elements of the nested POVM. This decomposition is interesting because it establishes a relation between the discrimination probability of a given set of states and that of its subsets of smaller size. Let us indeed suppose that the first measurement is successful, i.e., that an outcome k_1 occurs if one of the states $\rho_{k_1,k_{(2,u_F)}}$ with that value of the first bit was present. This happens with probability $p_{Succ}(k_1) = \sum_{k_{(2,u_F)}} \mathrm{Tr}\left[B^{(1)}_{k_1} \tilde{\rho}_{k_1,k_{(2,u_F)}}\right]$. In this case the possible weighted states after the measurement are

$$\tilde{\tau}_{k_1,k_{(2,u_F)}} = \sqrt{B^{(1)}_{k_1}} \tilde{\rho}_{k_1,k_{(2,u_F)}} \sqrt{B^{(1)}_{k_1}} / p_{Succ}(k_1), \qquad (14)$$

forming a set of size $N/2$: $\mathcal{S}^{(N/2)}_{k_1} = \left\{\tilde{\tau}_{k_1,k_{(2,u_F)}}\right\}_{k_2,\cdots,k_{u_F}=0,1}$. Moreover the collection of remaining measurements can be seen as a nested POVM of order $N/2$:

$$\mathcal{N}^{(N/2)}_{k_1} = \left\{\mathcal{B}^{(u_F)}_{k_1,k_{(2,u_F-1)}}\right\}_{k_2,\cdots,k_{u_F-1}=0,1} \circ \cdots \circ \left\{\mathcal{B}^{(2)}_{k_1}\right\}.$$

Hence we can easily rewrite the probability of discriminating the set of states $\mathcal{S}^{(N)}$ with the POVM $\mathcal{N}^{(N)}$, Eq. (13), as the probability of discriminating the set $\mathcal{S}^{(N/2)}_{k_1}$ with the POVM $\mathcal{N}^{(N/2)}_{k_1}$ if the first measurement had an outcome k_1, averaged over all values of $k_1 = 0, 1$:

$$P_{Succ}\left(\mathcal{S}^{(N)}, \mathcal{N}^{(N)}\right) = \sum_{k_1,k_{(2,u_F)}} p_{Succ}(k_1)$$
$$\cdot \mathrm{Tr}\left[\left|\sqrt{B^{(u_F)}_{k_1,k_{(2,u_F)}}} \cdots \sqrt{B^{(2)}_{k_1,k_2}}\right|^2 \tilde{\tau}_{k_1,k_{(2,u_F)}}\right]$$
$$= \sum_{k_1} p_{Succ}(k_1) P_{Succ}\left(\mathcal{S}^{(N/2)}_{k_1}, \mathcal{N}^{(N/2)}_{k_1}\right). \qquad (15)$$

This expression suggests a recursive optimization: if the optimal discrimination problem is solved for any set $\mathcal{S}^{(N/2)}$ of a fixed size $N/2$, possibly restricting to a specific Hilbert space, e.g., qubits or continuous-variable states, then the result can be plugged into (15) to obtain an expression for the discrimination probability of a set of double size, which depends on a single couple of measurement operators, i.e., the first-step binary POVM $\mathcal{B}^{(1)}$. However, if a general solution for the discrimination of a smaller set of states is not available, the problem remains hard, since when optimizing $P_{Succ}\left(\mathcal{S}^{(N/2)}_{k_1}, \mathcal{N}^{(N/2)}_{k_1}\right)$ one still has to take into account the dependence of the states of $\mathcal{S}^{(N/2)}_{k_1}$ on the first-step POVM, which is itself subject to optimization afterwards, thus making the states arbitrary.

Fortunately the first step of the recursion has a well-known solution [17]:

$$\mathbb{P}_{Succ}(\mathcal{S}^{(2)}) = \left(1 + ||\tilde{\rho}_0 - \tilde{\rho}_1||_1\right)/2, \qquad (16)$$

where $||\cdot||_1 = \mathrm{Tr}[|\cdot|]$ is the trace norm of the argument. Then by plugging this expression into the optimization of Eq. (15) for $N = 4$ states we can write:

$$\mathbb{P}_{Succ}\left(\mathcal{S}^{(4)}\right) = \max_{\mathcal{B}^{(1)}} \sum_{k_1} p_{Succ}(k_1) \frac{1 + ||\tilde{\tau}_{k_1,0} - \tilde{\tau}_{k_1,1}||_1}{2}$$
$$= \max_{\mathcal{B}^{(1)}} \sum_{k_1} \frac{1}{2}\Bigg(\mathrm{Tr}\left[B^{(1)}_{k_1}\left(\tilde{\rho}_{k_1,0} + \tilde{\rho}_{k_1,1}\right)\right]$$
$$+ \left\|\sqrt{B^{(1)}_{k_1}}\left(\tilde{\rho}_{k_1,0} - \tilde{\rho}_{k_1,1}\right)\sqrt{B^{(1)}_{k_1}}\right\|_1 \Bigg). \qquad (17)$$

We can write the latter equation more compactly by introducing the function

$$\mathcal{F}_Q(A, B, C) = \mathrm{Tr}\Big[QA + \left|\sqrt{Q} B \sqrt{Q}\right|$$
$$+ \left|\sqrt{\mathbf{1} - Q} C \sqrt{\mathbf{1} - Q}\right|\Big], \qquad (18)$$

where Q is a positive and less-than-one operator, while the arguments A, B, C are hermitian operators, and its maximum over Q, i.e.,

$$\mathcal{F}(A,B,C) = \max_{\mathbf{1}\geq Q\geq 0} \mathcal{F}_Q(A,B,C). \tag{19}$$

Setting $B_0^{(1)} = Q$ and $B_1^{(1)} = \mathbf{1} - Q$, we obtain:

$$\mathbb{P}_{Succ}\Big(\mathcal{S}^{(4)}\Big) = \frac{p_{1,0}+p_{1,1}}{2} + \mathcal{F}\left(A^{(4)}, B^{(4)}, C^{(4)}\right), \tag{20}$$

with $A^{(4)} = (\tilde{\rho}_{0,0} + \tilde{\rho}_{0,1} - \tilde{\rho}_{1,0} - \tilde{\rho}_{1,1})/2$, $B^{(4)} = (\tilde{\rho}_{0,0} - \tilde{\rho}_{0,1})/2$ and $C^{(4)} = (\tilde{\rho}_{1,0} - \tilde{\rho}_{1,1})/2$. Similarly for $N = 3$ states we have:

$$\mathbb{P}_{Succ}\Big(\mathcal{S}^{(3)}\Big) = p_{1,0} + \mathcal{F}\left(A^{(3)}, B^{(3)}, C^{(3)}\right), \tag{21}$$

with $A^{(3)} = (\tilde{\rho}_{0,0} + \tilde{\rho}_{0,1})/2 - \tilde{\rho}_{1,0}$, $B^{(3)} = B^{(4)}$ as before and $C^{(3)} = 0$. Thus the optimal discrimination problem of $N = 3, 4$ states has been reduced to the evaluation of the function $\mathcal{F}$, which requires an optimization over a single operator Q.

As already discussed, if the problem of Eq. (20) were to be solved exactly for any set of states $\mathcal{S}^{(4)}$, then the result could be plugged into Eq. (15), obtaining an expression for the optimal discrimination probability of $N = 8$ states dependent only on the first binary POVM. Unfortunately a solution of Eqs. (20,21) can be found only in some specific cases, listed below and discussed in detail in Appendix B. In the following we employ the positive part of an operator X, defined as $X_+ = (X + |X|)/2$.

Proposition 1. *The value of the function $\mathcal{F}(A, B, C)$ of Eq. (19) is*

$$\mathcal{F}(A,B,C) = \mathrm{Tr}[(A + |B| - |C|)_+] + ||C||_1, \tag{22}$$

when at least one of the following conditions holds: i) the operators B and C have support respectively on the positive and negative support of A; ii) B and C have a definite sign; iii) A, B and C all commute with each other.

Remark 1. *In the first case of Proposition 1, i.e., that the operators B and C have support respectively on the positive and negative support of A, the expression* (22) *can be simplified as*

$$\mathcal{F}(A,B,C) = \mathrm{Tr}[A_+] + ||B||_1 + ||C||_1. \tag{23}$$

Remark 2. *The optimal success probability is invariant under exchange of the states, i.e., under relabelling of the indices k_1, k_2 in our case $N = 3, 4$. Hence it can happen that the conditions listed in Proposition 1 are valid only for A, B, C given by a specific ordering of the states.*

The previous remark implies that, when checking whether a set of states satisfies the conditions of Proposition 1 or not, one has to consider all possible sets of A, B, C obtainable by different orderings of the states, not only the conventional one employed in Eqs. (20,21). Alternatively, one can apply this symmetry under exchange of the states to obtain recursive relations for $\mathcal{F}(A, B, C)$, e.g., for $N = 3$ and by exchanging $(0, 0) \leftrightarrow (1, 0)$, it holds

$$\mathcal{F}(A,B,0) = \frac{\mathcal{F}(-3B - A, B - A, 0)}{2} + \mathrm{Tr}[A + B]; \tag{24}$$

then Proposition 1 holds on the right-hand side of (24) when $B' = B - A$ has a definite sign, but the latter is simply $B' = (\tilde{\rho}_{1,0} - \tilde{\rho}_{0,1})/2$, an expression of the operator B for the new ordering of the states.

Remark 3. *In all the cases listed in Proposition 1, with the conventional ordering of the states of Eqs. (20,21), the optimal success probabilities for the discrimination of $N = 3, 4$ states become*

$$\mathbb{P}_{Succ}\Big(\mathcal{S}^{(3)}\Big) = p_{1,0} + \mathrm{Tr}\left[\frac{(\tilde{\rho}_{0,0} + \tilde{\rho}_{0,1} - 2\tilde{\rho}_{1,0} + |\tilde{\rho}_{0,0} - \tilde{\rho}_{0,1}|)_+}{2}\right], \tag{25}$$

$$\mathbb{P}_{Succ}\Big(\mathcal{S}^{(4)}\Big) = \frac{p_{1,0}+p_{1,1}}{2} + \mathrm{Tr}\left[\frac{(\tilde{\rho}_{0,0} + \tilde{\rho}_{0,1} - \tilde{\rho}_{1,0} - \tilde{\rho}_{1,1} + |\tilde{\rho}_{0,0} - \tilde{\rho}_{0,1}| - |\tilde{\rho}_{1,0} - \tilde{\rho}_{1,1}|)_+}{2}\right] + \left|\left|\frac{\tilde{\rho}_{1,0} - \tilde{\rho}_{1,1}}{2}\right|\right|_1. \tag{26}$$

IV. A NUMERICAL EXAMPLE: THE $N = 3, 4$ QUBIT CASE

In this Section we analyze the discrimination probability obtained with the nested POVM decomposition in the case of $N = 3, 4$ *qubit* states. Indeed, since Eqs. (20,21) seem not to be solvable analytically for generic sets of states, it is interesting to tackle the problem by choosing the simplest possible Hilbert space for the measured system, i.e., the qubit space $\mathcal{H}_2$ of dimension two. It is well known that the density matrices ρ of this system can be represented as a real vector $\vec{v}_\rho$ inside a three-dimensional unit sphere (the Bloch sphere), i.e. $\rho = (\mathbf{1}_2 + \vec{v}_\rho \cdot \vec{\sigma})/2$ where $\mathbf{1}_2$ is the identity operator and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In particular, pure states are situated on the sphere's surface, i.e., $v_\rho = |\vec{v}_\rho| = 1$ for $\rho = |\psi\rangle\langle\psi|$, while the completely mixed state $\mathbf{1}_2/2$ is at the origin. More generally, any hermitian operator X on the qubit space can be expressed in terms of four real coefficients: a scalar c_X, which represents the normalization coefficient of the operator, and a vector $\vec{r}_X$, which represents the operator in the Bloch space, i.e.

$$X = c_X \mathbf{1}_2 + \vec{r}_X \cdot \vec{\sigma}, \tag{27}$$

the trace of the operator being determined by $\mathrm{Tr}[X] = 2c_X$, while its eigenvalues by $\lambda_X^{(\pm)} = c_X \pm r_X$ with $r_X = |\vec{r}_X|$.

Employing the representation described above we can hence rewrite the function $\mathcal{F}_Q(A,B,C)$ as (see Appendix C for details)

$$\begin{aligned}\mathcal{F}_Q(A,B,C)\Big|_{\mathcal{H}_2} = 2\Bigg(& c_Q c_A + \vec{r}_Q \cdot \vec{r}_A + \sqrt{(c_Q c_B + \vec{r}_Q \cdot \vec{r}_B)^2 + \Big((r_B)^2 - (c_B)^2\Big)\Big((c_Q)^2 - (r_Q)^2\Big)} \\ & + \sqrt{((1-c_Q)c_C - \vec{r}_Q \cdot \vec{r}_C)^2 + \Big((r_B)^2 - (c_B)^2\Big)\Big((1-c_Q)^2 - (r_Q)^2\Big)}\Bigg),\end{aligned} \tag{28}$$

when B and C do not have a definite sign, or

$$\begin{aligned}\mathcal{F}_Q(A,B,C)\Big|_{\mathcal{H}_2} &= 2\Bigg(c_Q c_A + \vec{r}_Q \cdot \vec{r}_A + c_Q c_{|B|} + \vec{r}_Q \cdot \vec{r}_{|B|} + (1-c_Q)\, c_{|C|} - \vec{r}_Q \cdot \vec{r}_{|C|}\Bigg) \\ &= 2\left(c_Q c_{A+|B|-|C|} + \vec{r}_Q \cdot \vec{r}_{A+|B|-|C|} + c_{|C|}\right),\end{aligned} \tag{29}$$

when both B and C have a definite sign, with $c_{|B|} = \pm c_B$, $\vec{r}_{|B|} = \pm \vec{r}_B$ respectively for $B \geq 0$ and $B \leq 0$ and similar definitions for $|C|$.

For each set of $N = 3, 4$ qubit states to discriminate, the operators A, B, C, i.e., their coefficients c and $\vec{r}$, are fixed and the optimization of Eqs. (28, 29) is to be carried out only over Q, i.e., on its coefficients c_Q and $\vec{r}_Q$ under the constraints

$$1 \geq c_Q \geq 0, \qquad r_Q \leq \min[c_Q, 1 - c_Q]\ , \tag{30}$$

that ensure the positivity of Q and the fact that it must be smaller than one. In particular for $N = 3$ states $C = 0$ and one can show that $c_Q + r_Q = \lambda_Q^{(+)} = 1$ is optimal. Moreover the optimal $\vec{r}_Q$ lies on the plane of $\vec{r}_A$ and $\vec{r}_B$, so that it can be defined in terms of its norm r_Q and a single angle ϕ_Q as $\vec{r}_Q \cdot \vec{r}_A = r_Q r_A \cos\phi_Q$. Then in this case it is only required to optimize two parameters, namely c_Q and ϕ_Q. For $N = 4$ instead there are no further simplifications and one has to optimize four parameters, with constraints.

Let us now consider the case in which Eq. (29) is valid, i.e., B and C have a definite sign. It is one of the situations considered in Proposition 1, thus the optimization of (29) must match the expression (22). This fact is already quite clear if we express Eq. (29) directly in terms of the initial operators; furthermore it can also be shown by direct analytical optimization that in this case

$$\mathcal{F}(A,B,C)\Big|_{\mathcal{H}_2} = |c_{A+|B|-|C|}| + c_{A+|B|-|C|} + 2c_C \tag{31}$$

and the optimal value of Q is $r_Q = 0$ and $c_Q = \theta(c_{A+|B|-|C|})$, with $\theta(\cdot)$ the step function, valued 1 when its argument is positive and zero otherwise.

As for the other case, in which Eq. (28) is valid, unfortunately the function cannot be completely optimized analytically. Nevertheless its numerical optimization is straightforward and thus, together with Eqs. (20, 21), it provides a convenient method to obtain the optimal success probability of discrimination and the optimal measurement operators for $N = 3, 4$ qubit states. As an example let us consider $N = 3$ *equiprobable pure* qubit states situated on the (x, y) plane of the Bloch sphere, i.e., a combination of $\mathbf{1}_2$, σ_1 and σ_2; this is a simple choice for the sake of clarity, but we stress that no additional optimization difficulties are met when considering non-equiprobable and mixed states. Let us fix the first state to be on the x axis, i.e., $\vec{r}_{\rho_1} = (1, 0, 0)$, without loss of generality. Then we can study the optimal success probability by varying the angles of the other two vectors with respect to the first one: $\vec{r}_{\rho_2} = (\cos\phi_2, \sin\phi_2, 0)$ and $\vec{r}_{\rho_3} = (\cos\phi_3, \sin\phi_3, 0)$. If the states are also symmetrically distributed at constant angles along the circumference, i.e., $\phi_2 = \phi_3 = 2\pi/3$, the result is well-known [17]: $\mathbb{P}_{Succ}\left(\mathcal{S}_{sym}^{(3)}\right) = 2/3$, which is the maximum success probability of discrimination for any three equiprobable qubit states (because no other configuration can achieve a lower average state-overlap than this one). For more general states the results are shown in Fig. 2 where we plot the optimal success probability (21), computed by numerical optimization of (28) over c_Q, $\vec{r}_Q$, as a function of the third angle ϕ_3 and for several choices of $\phi_2 = 0, \pi/6, \pi/2, 2\pi/3, \pi$. It can be seen that, for all values of ϕ_2, there is a range of values of ϕ_3 for which

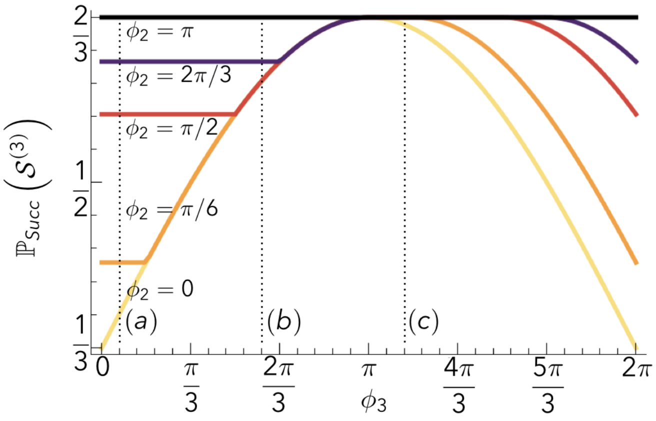

Figure 2. Plot of $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}\right)$, the optimal success probability (21) for a set of three equiprobable pure qubit states, identified by the Bloch vectors $\vec{r}_{\rho_1} = (1,0,0)$, $\vec{r}_{\rho_2} = (\cos\phi_2, \sin\phi_2, 0)$ and $\vec{r}_{\rho_3} = (\cos\phi_3, \sin\phi_3, 0)$, as a function of ϕ_3 for several values of $\phi_2 = 0, \pi/6, \pi/2, 2\pi/3, \pi$ (respectively from yellow/light-gray to black). The results are obtained by numerical optimization of Eq. (28) over c_Q, $\vec{r}_Q$. Observe that, for all values of ϕ_2, there is a range of values of ϕ_3 where $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}\right)$ attains the maximum allowed for non-orthogonal states, i.e., the same as for symmetric states. Outside of this range the quantity decreases, reaching a constant minimum for $\phi_3 \leq \phi_2$ (see text and Fig. 3 for an explanation). The cases $\phi_2 = \pi/6, 2\pi/3$ are explicitly depicted in Fig. 3 for three values of ϕ_3 identified by the labelled dotted lines.

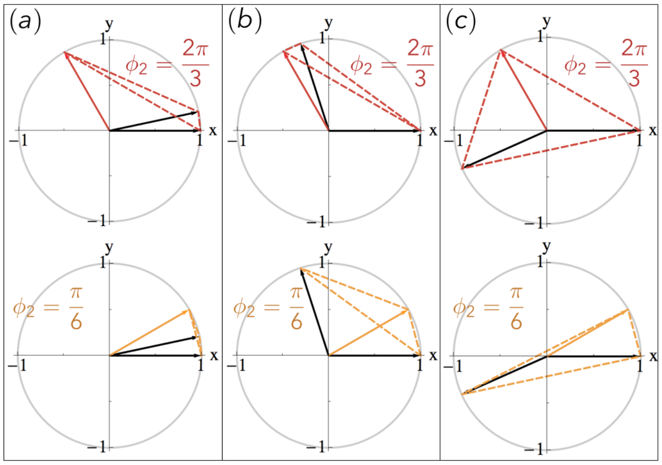

Figure 3. Plot of the vectors of the three states $\vec{r}_{\rho_1} = (1,0,0)$ (black fixed on x axis), $\vec{r}_{\rho_3} = (\cos\phi_3, \sin\phi_3, 0)$ (black) and $\vec{r}_{\rho_2} = (\cos\phi_2, \sin\phi_2, 0)$ (red/dark gray at the top and orange/light gray at the bottom) on the (x, y) Bloch plane, as well as the triangles formed by them (same color codes as $\vec{r}_{\rho_2}$). The labels a, b, c refer respectively to values of $\phi_3 = \pi/15, 3\pi/5, 7\pi/5$, also highlighted in Fig. 2, while two values of $\phi_2 = 2\pi/3, \pi/6$ (respectively red/dark gray and orange/light gray figures) are considered. By comparison with Fig. 2 it is evident that $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}\right)$ is maximum when the triangle formed by the states contains the origin (c), while it is lower otherwise. In particular, when $\phi_3 \leq \phi_2$ (a, b top, a bottom) the largest side of the triangle formed by the states is always $\vec{r}_{\rho_2} - \vec{r}_{\rho_1}$ and this determines completely the optimal success probability.

$\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}\right)$ is equal to the maximum value of 2/3, even though the states are not symmetrically distributed on the circumference. In other words there is a wide class of states that can be discriminated with performance as good as if they were symmetric. Outside of this range, whose width depends on ϕ_2, the value of $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}\right)$ decreases and it reaches a constant minimum when $\phi_3 \leq \phi_2$. These peculiarities can be explained by referring to Refs. [25–27]. In particular Ref. [26] states that the optimal success probability of discrimination of a set $\mathcal{S}^{(N)}_{eq} = \{\rho_j/N\}_{j=0,\cdots,N-1}$ of N *equiprobable* qubit states can be found by: i) considering the geometric figure determined by the weighted states in the Bloch space, i.e., their polytope of vertices $\{\vec{r}_{\rho_j}/N\}$; ii) finding the polytope similar to the latter and that is also maximal in the Bloch sphere; iii) computing the ratio R between the original and the maximal polytope. Then the optimal success probability is $\mathbb{P}_{Succ}\left(\mathcal{S}^{(N)}_{eq}\right) = \frac{1}{N} + R$. In light of this observation, we can explain the results of Fig. 2 by plotting the states in the Bloch sphere and studying the polygon formed by their vertices, as done in Fig. 3 for two values of $\phi_2 = \pi/6, 2\pi/3$ used in Fig. 2 and $\phi_3 = \pi/15, 3\pi/5, 7\pi/5$, corresponding to the dotted lines labelled a, b, c in Fig. 2. If the polytope determined by the qubits, usually a triangle, contains the origin, then the optimal success probability is still maximum, i.e., $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}_{\geq 0}\right) \equiv \mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}_{sym}\right)$, as in Fig. 3$c$. Indeed in this case the polytope formed by the states is already maximal in the Bloch sphere and $R = 1/3$, as for the symmetric set. On the other hand, if the polytope does not contain the origin, the optimal success probability is strictly lower than the maximum one, i.e., $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}_{\ngeq 0}\right) < \mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}_{sym}\right)$, as in Fig. 3$a, b$. Indeed in this second case the polytope formed by the states is not maximal and can be expanded until its largest side matches a diameter of the circumference, so that $R < 1/3$. As for the region $\phi_3 \leq \phi_2$ where $\mathbb{P}_{Succ}\left(\mathcal{S}^{(3)}_{\ngeq 0}\right)$ is minimum and constant (as in Fig. 3a, b top and a bottom), it can be explained by observing that the largest side of the triangle determined by the states, which in turn determines R, is always the one that connects $\vec{r}_{\rho_1}$ and $\vec{r}_{\rho_2}$, independently of $\vec{r}_{\rho_3}$. Since $\vec{r}_{\rho_2} - \vec{r}_{\rho_1}$ is constant for constant ϕ_2, R is constant too in this case.

V. CONCLUSIONS

In this article we proposed a method to compute the optimal discrimination probability and optimal measurement operators of an arbitrary set of N states. We showed how to decompose any multiple-outcome measurement into several binary-outcome steps, which could be of interest also in other contexts. For the discrimination problem this decomposition introduces a connection between the success probabilities of sets of different sizes, possibly simplifying the optimization procedure, but does not allow to reach a general analytical solution. Nevertheless it proves to be a useful tool for quickly determining the optimal discrimination probability of $N = 3, 4$ qubit states, requiring just a simple numerical optimization. Indeed with this method we were able to highlight some interesting properties, explicitly verifying the validity and geometric insight of some previous results [25–27]. Future lines of work could focus on simplifying the optimization for higher-dimensional systems and larger sets of states or investigating different kinds of measurement decompositions.

VI. ACKNOWLEDGMENTS

GdP acknowledges financial support from the European Research Council (ERC Grant Agreement no 337603), the Danish Council for Independent Research (Sapere Aude) and VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059).

[1] S. M. Barnett and S. Croke, Adv. Opt. Phot. **1**, 238-278 (2009).

[2] A. S. Holevo, Probl. Peredachi Inf. **9**, 3 (1973); Probl. Inf. Transm. (Engl. Transl.) **9**, 110 (1973).

[3] P. Hausladen and W. K. Wooters, J. Mod. Opt. **41**, 2385 (1994).

[4] S. Lloyd, V. Giovannetti, L. Maccone, Phys. Rev. Lett. **106**, 250501 (2011).

[5] M. M. Wilde, Saikat Guha, IEEE Trans. Inf. Theory **59**, 1175 (2013).

[6] M. Rosati and V. Giovannetti, J. Math. Phys. **57**, 062204 (2016).

[7] C. H. Bennett and G. Brassard, Proceedings of International Conference on Com- puter Systems and Signal Processing **175**, 8 (1984).

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

[9] R. Konig, R. Renner and C. Schaffner, IEEE Trans. Info. Theory **55**, 4337 (2009).

[10] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

[11] J. Bae and A. Acín, Phys. Rev. Lett. **97**, 030402 (2006).

[12] J. Bae, W.-Y. Hwang and Y.-D. Han, Phys. Rev. Lett. **107**, 170403 (2011).

[13] N. Brunner, M. Navascués and T. Vértesi, Phys. Rev. Lett. **110**, 150501 (2013).

[14] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, UK, 2010).

[15] S.-H. Tan, B. I. Erkmen, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, S. Pirandola and J. H. Shapiro, Phys. Rev. Lett. **101**, 253601 (2008).

[16] D. Bacon, A. M. Childs and W. van Dam, 46th Annual IEEE Symposium on Foundations of Computer Science, 469 (2005).

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

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

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

[20] Y. C. Eldar and G. D. Forney, IEEE Trans. Inform. Theory **47**, 858 (2001).

[21] G. Chiribella and G. M. D'Ariano, J. Math. Phys. **47**, 092107 (2006).

[22] H. Krovi, S. Guha, Z. Dutton and M. P. da Silva, Phys. Rev. A **92**, 062333 (2015).

[23] M. Ban, K. Kurokawa, R. Momose and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).

[24] K. Kato, M. Osaki, M. Sasaki and O. Hirota, IEEE Trans. Comm. **47**, 248 (1999).

[25] K. Hunter, arXiv:quant-ph/0410228v1 (2004).

[26] J. Bae, New J. Phys. **15**, 073037 (2013).

[27] D. Ha and Y. Kwon, Phys. Rev. A **87**, 062302 (2013).

[28] R. Kennedy, MIT Res. Lab. Electron. Quart. Progr. Rep. **108**, 219 (1973).

[29] M. Takeoka and M. Sasaki, Phys. Rev. A **78**, 022320 (2008).

[30] C. Wittmann, M. Takeoka, K. N. Cassemiro, M. Sasaki, G. Leuchs and U. L. Andersen, Phys. Rev. Lett. **101**, 210501 (2008).

[31] C. R. Müller and Ch. Marquardt, New J. Phys. **17**, 032003 (2015).

[32] M. Takeoka, M. Sasaki, P. van Loock and N. Lütkenhaus, Phys. Rev. A **71**, 022318 (2005).

[33] A. Acín, E. Bagan, M. Baig, L. Masanes and R. Muñoz-Tapia, Phys. Rev. A **71**, 032338 (2005).

[34] M. Osaki, M. Ban and O. Hirota, Phys. Rev. A **54**, 1691 (1996).

[35] S. Dolinar, MIT Res. Lab. Electron. Quart. Progr. Rep. **111**, 115 (1973).

[36] K. Banaszek, Phys. Lett. A **253**, 12 (1999).

[37] M. Rosati, A. Mari and V. Giovannetti, Phys. Rev. A **93**, 062315 (2016).

[38] R. Nair, S. Guha and S.-H. Tan, Phys. Rev. A **89**, 032318 (2014).

[39] R. L. Cook, P. J. Martin and J. M. Geremia, Nature **446**, 774 (2007).

[40] F. E. Becerra, J. Fan, G. Baumgartner, S. V. Polyakov, J. Goldhar, J. T. Kosloski and A. Migdall, Phys. Rev. A **84**, 062324 (2011).

[41] Y. C. Eldar, A. Megretski and G. C. Verghese, IEEE Trans. Info. Theory **49**, 1007 (2003).

[42] E. Andersson and D. K. L. Oi, Phys. Rev. A **77**, 052104 (2008).

[43] C. Shen, K. Noh, V. V. Albert, S. Krastanov, M. H. Devoret, R. J. Schoelkopf, S. M. Girvin, L. Jiang, arXiv:1611.03463 [quant-ph] (2016).

[44] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997); P. Hausladen, R. Jozsa, B. W. Schu-

macher, M. Westmoreland, and W. K. Wootters, ibid. **54**, 1869 (1996).

Appendix A: Completeness of the nested POVM

In this appendix we show that the nested POVM defined in (3) is complete, i.e., the sum of its elements is the identity on the whole Hilbert space of the system. This can be shown by employing the completeness of each binary POVM's $\mathcal{B}^{(u)}_{k_{(1,u-1)}}$ at each step u. Indeed we can start by summing over the last bit $k_{u_F} = 0, 1$, coupling elements that differ only for its value, i.e., $F_{k_{(1,u_F-1)},0}$ and $F_{k_{(1,u_F-1)},1}$. These are made of the same sequence of operators apart from the most interior ones, $B^{(u_F)}_{k_{(1,u_F-1)},0}$ and $B^{(u_F)}_{k_{(1,u_F-1)},1}$, which are instead two different elements of the same binary POVM $\mathcal{B}^{(u_F)}_{k_{(1,u_F-1)}}$, thus satisfy a completeness relation and their sum can be simplified. The same procedure is then applied recursively on previous bits as follows:

$$\begin{aligned}
\sum_{k_1,\cdots,k_{u_F}} F_{k_{(1,u_F)}} &= \sum_{k_1,\cdots,k_{u_F-1}} \left(F_{k_{(1,u_F-1)},0} + F_{k_{(1,u_F-1)},1}\right) \\
&= \sum_{k_1,\cdots,k_{u_F-1}} \sqrt{B^{(1)}_{k_1}} \cdots \sqrt{B^{(u_F-1)}_{k_{(1,u_F-1)}}} \left(B^{(u_F)}_{k_{(1,u_F-1)},0} + B^{(u_F)}_{k_{(1,u_F-1)},1}\right) \sqrt{B^{(u_F-1)}_{k_{(1,u_F-1)}}} \cdots \sqrt{B^{(1)}_{k_1}} \\
&= \sum_{k_1,\cdots,k_{u_F-1}} \sqrt{B^{(1)}_{k_1}} \cdots \sqrt{B^{(u_F-2)}_{k_{(1,u_F-2)}}} B^{(u_F-1)}_{k_{(1,u_F-1)}} \sqrt{B^{(u_F-2)}_{k_{(1,u_F-2)}}} \cdots \sqrt{B^{(1)}_{k_1}} \\
&= \sum_{k_1,\cdots,k_{u_F-2}} \sqrt{B^{(1)}_{k_1}} \cdots \sqrt{B^{(u_F-2)}_{k_{(1,u_F-2)}}} \left(B^{(u_F-1)}_{k_{(1,u_F-2)},0} + B^{(u_F-1)}_{k_{(1,u_F-2)},1}\right) \sqrt{B^{(u_F-2)}_{k_{(1,u_F-2)}}} \cdots \sqrt{B^{(1)}_{k_1}} \\
&= \cdots = \sum_{k_1} \sqrt{B^{(1)}_{k_1}} \left(B^{(2)}_{k_1,0} + B^{(2)}_{k_1,1}\right) \sqrt{B^{(1)}_{k_1}} = B^{(1)}_0 + B^{(1)}_1 = \mathbf{1}. \qquad \text{(A1)}
\end{aligned}$$

We note that the previous result does not change if instead of employing complete binary POVM's, we relax to weak completeness, as defined in Sec. II, i.e., that each measurement $\mathcal{B}^{(u)}_{k_{(1,u-1)}}$ is complete on the support of the operator that preceeds it in the nested decomposition, $B^{(u-1)}_{k_{(1,u-1)}}$. In this case it still holds $\sqrt{B^{(u-1)}_{k_{(1,u-1)}}} \left(B^{(u)}_{k_{(1,u-1)},0} + B^{(u)}_{k_{(1,u-1)},1}\right) \sqrt{B^{(u-1)}_{k_{(1,u-1)}}} = B^{(u-1)}_{k_{(1,u-1)}}$ and the equalities in (A1) are unchanged.

Appendix B: Detailed study of $\mathcal{F}$

In this appendix we study the function $\mathcal{F}_Q(A, B, C)$ appearing in Eqs. (18), and discuss its optimization in the cases mentioned in Sec. III. The optimization of $\mathcal{F}_Q(A, B, C)$ is difficult because of the competing interests of the three terms composing it. Indeed each single term of Eq. (18) is maximized by a different operator Q: the first one is maximum when Q is the projector on the positive support of A; the second one is maximum when Q is the identity on the whole Hilbert space of the system; the third one is maximum when Q is zero. Hence we can solve the problem exactly only if the three operators exhibit specific properties.

We start by observing that the function is subadditive in all its arguments, i.e., for any set of operators $\{A_j, B_j, C_j\}_{j=1,\cdots,n}$ it holds:

$$\mathcal{F}(\sum_j A_j, \sum_j B_j, \sum_j C_j) \leq \sum_j \mathcal{F}(A_j, B_j, C_j). \qquad \text{(B1)}$$

This follows from the subadditivity of the trace norm. We can now state some lemmas that help demonstrate Proposition 1. Throughout the Appendix, the notation $\mathbf{1}_X$ represents the projector on the support of the operator X.

Lemma 1. *Let us suppose that A is positive semidefinite, B has support inside the support of A and that C and A have orthogonal supports. Then $\mathcal{F}(A, B, C) = \mathrm{Tr}[A] + ||B||_1 + ||C||_1$.*

Proof. Consider the set of operators $\{A_j = A\delta_{j,1}, B_j = B\delta_{j,2}, C_j = C\delta_{j,3}\}_{j=1,2,3}$ and apply the subadditivity property (B1):

$$\begin{aligned}
\mathcal{F}(A, B, C) &\leq \mathcal{F}(A, 0, 0) + \mathcal{F}(0, B, 0) + \mathcal{F}(0, 0, C) \\
&= \mathrm{Tr}[A] + ||B||_1 + ||C||_1 . \qquad \text{(B2)}
\end{aligned}$$

The latter inequality can be saturated under the hypotheses of this lemma, by taking $Q = \mathbf{1}_A$. □

Lemma 2. *Let us suppose that A is negative semidefinite, C has support inside the support of A and that*

B and A have orthogonal supports. Then $\mathcal{F}(A,B,C) = ||B||_1 + ||C||_1$.

Proof. Consider the same set of operators of Lemma 1 and apply again the subadditivity property (B2), then use the fact that $A \leq 0$:

$$\begin{aligned}\mathcal{F}(A,B,C) &\leq \mathcal{F}(A,0,0) + \mathcal{F}(0,B,0) + \mathcal{F}(0,0,C) \quad \text{(B3)}\\ &= ||B||_1 + ||C||_1 \,.\end{aligned}$$

The latter inequality can be saturated under the hypotheses of this lemma, by taking $Q = \mathbf{1}_B$. □

Hence we can prove the first case of Proposition 1: let $A = A_+ \oplus (-A_-)$ be the decomposition of A in terms of its positive and negative parts, with $A_\pm \geq 0$, and suppose that B, C have support respectively inside the support of A_+, A_-. Then consider the set of operators $\left\{A_j = (-1)^{j+1} A_{(-1)^{j+1}}, B_j = B\delta_{j,1}, C_j = C\delta_{j,2}\right\}_{j=1,2}$ and apply the subadditivity property (B1), together with Lemmas 1, 2:

$$\begin{aligned}\mathcal{F}(A,B,C) &\leq \mathcal{F}\left(A_+, B, 0\right) + \mathcal{F}\left(-A_-, 0, C\right)\\ &= \mathrm{Tr}\left[A_+\right] + ||B||_1 + ||C||_1 \,, \qquad \text{(B4)}\end{aligned}$$

which is saturated by a measurement operator $Q = \mathbf{1}_{A_+}$. This expression is equivalent to that given in (22) under the current hypotheses, indeed in this case it holds

$$\begin{aligned}(A + |B| - |C|)_+ &= ((A_+ + |B|) \oplus (-A_- - |C|))_+\\ &= A_+ + |B|, \qquad \text{(B5)}\end{aligned}$$

so that (B4) becomes

$$\begin{aligned}\mathcal{F}\left(A,B,C\right) &= \mathrm{Tr}\left[A_+ + |B|\right] + ||C||_1\\ &= \mathrm{Tr}\left[(A_+ + |B| - |C|)_+\right] + ||C||_1 \,. \text{(B6)}\end{aligned}$$

As for the second and third cases of Proposition 1, let us first note that

$$\begin{aligned}\left|\left|\sqrt{Q}B\sqrt{Q}\right|\right|_1 &\leq \left|\left|\sqrt{Q}B_+\sqrt{Q}\right|\right|_1 + \left|\left|\sqrt{Q}B_-\sqrt{Q}\right|\right|_1\\ &= \mathrm{Tr}[Q(B_+ + B_-)] = \mathrm{Tr}[Q|B|] \,, \qquad \text{(B7)}\end{aligned}$$

where $B_\pm$ are the positive and negative parts of B as defined above for A, and analogously

$$\left|\left|\sqrt{\mathbf{1}-Q}C\sqrt{\mathbf{1}-Q}\right|\right|_1 \leq \mathrm{Tr}[(\mathbf{1}-Q)|C|] \,. \qquad \text{(B8)}$$

We then have

$$\begin{aligned}\mathcal{F}_Q(A,B,C) &\leq \mathrm{Tr}[Q(A + |B| - |C|)] + ||C||_1 \quad \text{(B9)}\\ &\leq \mathrm{Tr}[(A + |B| - |C|)_+] + ||C||_1 \,.\text{(B10)}\end{aligned}$$

The inequality (B10), is saturated by taking Q equal to the projector onto the support of $(A + |B| - |C|)_+$. The inequalities (B7,B8) and hence (B9) are saturated in both the second and third cases of Proposition 1, though for different reasons:

- If B and C have a definite sign, then it holds $B = B_+$ or $B = B_-$, so that Eq. (B7) is saturated and analogously (B8);
- If A, B, C all commute with each other, then Eqs. (B7,B8) are saturated by any operator Q which commutes with both B and C. Eventually the choice $Q = \mathbf{1}_{(A+|B|-|C|)_+}$ necessary to saturate Eq. (B10) satisfies this latter condition in the case considered.

Finally we note that the previous case of commuting operators, as well as further results, can also be derived by applying the symmetry property of the optimal success probabilities (20, 21) to obtain recursive formulas, as discussed after Remark 2, but still a full solution cannot be found in this way.

Appendix C: Computation of $\mathcal{F}_Q$ in the qubit case

In this Section we derive the results (28, 29) explicitly. As a preliminary recall that, for any three vectors $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$ and the Pauli matrices $\vec{\sigma}$ it holds:

$$\left(\vec{a}\cdot\vec{\sigma}\right)\left(\vec{b}\cdot\vec{\sigma}\right) = \left(\vec{a}\cdot\vec{b}\right) + i\left(\vec{a}\times\vec{b}\right)\cdot\vec{\sigma}, \qquad \text{(C1)}$$

$$\left(\vec{a}\times\left(\vec{b}\times\vec{c}\right)\right) = \left(\vec{a}\cdot\vec{c}\right)\vec{b} - \left(\vec{a}\cdot\vec{b}\right)\vec{c}. \qquad \text{(C2)}$$

Moreover, given a positive operator Q on $\mathcal{H}_2$, the coefficients $c_{\sqrt{Q}}$, $\vec{r}_{\sqrt{Q}}$ of its square root $\sqrt{Q}$ can be expressed in terms of its coefficients c_Q, $\vec{r}_Q$ as:

$$\begin{cases} c_Q = \left(c_{\sqrt{Q}}\right)^2 + \left(r_{\sqrt{Q}}\right)^2 \\ r_Q = 2c_{\sqrt{Q}}r_{\sqrt{Q}} \end{cases} \leftrightarrow \begin{cases} c_{\sqrt{Q}} = \frac{\sqrt{c_Q+r_Q}+\sqrt{c_Q-r_Q}}{2} \\ r_{\sqrt{Q}} = \frac{\sqrt{c_Q+r_Q}-\sqrt{c_Q-r_Q}}{2} \end{cases}. \quad \text{(C3)}$$

with $\vec{r}_Q \parallel \vec{r}_{\sqrt{Q}}$. In order to evaluate $\mathcal{F}_Q(A,B,C)$ we can compute its first two terms, while the third one is similar to the second one. Let us start with the product QA: it is a generic operator with coefficients

$$c_{QA} = c_Q c_A + \vec{r}_Q \cdot \vec{r}_A \qquad \text{(C4)}$$

$$\vec{r}_{QA} = c_Q\vec{r}_A + c_A\vec{r}_Q + i\left(\vec{r}_Q \times \vec{r}_A\right), \qquad \text{(C5)}$$

computed by applying Eq. (C2). Thus the first term of $\mathcal{F}_Q$ is simply $\mathrm{Tr}[QA] = 2c_{QA}$.
As for the product $\sqrt{Q}B\sqrt{Q}$, its first coefficient is simple: $c_{\sqrt{Q}B\sqrt{Q}} = \mathrm{Tr}[\sqrt{Q}B\sqrt{Q}]/2 = c_{QB}$, easily obtained by relabelling Eq. (C4). The vector of coefficients instead is

$$\begin{aligned}\vec{r}_{\sqrt{Q}B\sqrt{Q}} &= c_{\sqrt{Q}B}\vec{r}_{\sqrt{Q}} + c_{\sqrt{Q}}\vec{r}_{\sqrt{Q}B} + i\left(\vec{r}_{\sqrt{Q}B} \times \vec{r}_{\sqrt{Q}}\right) \text{(C6)}\\ &= c_B\vec{r}_Q + 2\left(\vec{r}_{\sqrt{Q}} \cdot \vec{r}_B\right)\vec{r}_{\sqrt{Q}} + \left(\left(c_{\sqrt{Q}}\right)^2 - \left(r_{\sqrt{Q}}\right)^2\right)\vec{r}_B,\end{aligned}$$

where we have first computed the product between $\sqrt{Q}B$ and $\sqrt{Q}$, then substituted the expression for the former by relabelling once again the product QA and employed (C3).

We are interested in the absolute value of $\sqrt{Q}B\sqrt{Q}$, i.e., the sum of the absolute value of its eigenvalues $\lambda^{(\pm)}_{\sqrt{Q}B\sqrt{Q}} = c_{\sqrt{Q}B\sqrt{Q}} \pm r_{\sqrt{Q}B\sqrt{Q}}$. Hence the only dependence of the final expression on (C6) is through its norm:

$$\begin{aligned}\left(r_{\sqrt{Q}B\sqrt{Q}}\right)^2 &= \left(c_B r_Q\right)^2 + 4(\vec{r}_{\sqrt{Q}} \cdot \vec{r}_B)^2 \left(c_{\sqrt{Q}}\right)^2 + \left(\left(c_{\sqrt{Q}}\right)^2 - \left(r_{\sqrt{Q}}\right)^2\right)^2 \left(r_B\right)^2 + 2c_B \left(\left(c_{\sqrt{Q}}\right)^2 + \left(r_{\sqrt{Q}}\right)^2\right) \left(\vec{r}_Q \cdot \vec{r}_B\right) \\ &= \left(c_Q c_B + \vec{r}_Q \cdot \vec{r}_B\right)^2 + \left(\left(r_B\right)^2 - \left(c_B\right)^2\right)\left(\left(c_Q\right)^2 - \left(r_Q\right)^2\right), \end{aligned} \quad \text{(C7)}$$

which we have simplified by employing the relations (C3). Eventually we have to distinguish between two cases:

- If both B and C have definite sign then they can always be taken to be positive semidefinite, up to a relabeling 0 ↔ 1 of the second bits k_2 of the original states. Then we have

$$\begin{aligned}\left|\left|\sqrt{Q}B\sqrt{Q}\right|\right|_1 &= \lambda^{(+)}_{\sqrt{Q}B\sqrt{Q}} + \lambda^{(-)}_{\sqrt{Q}B\sqrt{Q}} = 2c_{\sqrt{Q}B\sqrt{Q}}, \\ \left|\left|\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}\right|\right|_1 &= 2c_{\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}}; \end{aligned} \quad \text{(C8)}$$

- If instead B and C do not have a definite sign, then it must hold $\lambda^{(+)}_{\sqrt{Q}B\sqrt{Q}} \geq 0$ and $\lambda^{(-)}_{\sqrt{Q}B\sqrt{Q}} \leq 0$ and similar relations for C, so that

$$\begin{aligned}\left|\left|\sqrt{Q}B\sqrt{Q}\right|\right|_1 &= \lambda^{(+)}_{\sqrt{Q}B\sqrt{Q}} - \lambda^{(-)}_{\sqrt{Q}B\sqrt{Q}} = 2r_{\sqrt{Q}B\sqrt{Q}}, \\ \left|\left|\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}\right|\right|_1 &= 2r_{\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}}. \end{aligned} \quad \text{(C9)}$$

Note that the third term $\left|\left|\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}\right|\right|_1$ can be expressed in terms of the coefficients of Q by observing that $c_{\boldsymbol{1}-Q} = (1 - c_Q)$ and $\vec{r}_{\boldsymbol{1}-Q} = -\vec{r}_Q$.
We can conclude that for B and C of non-definite sign

$$\mathcal{F}_Q(A, B, C) = 2(c_{QA} + r_{\sqrt{Q}B\sqrt{Q}} + r_{\sqrt{\boldsymbol{1}-Q}C\sqrt{\boldsymbol{1}-Q}}), \quad \text{(C10)}$$

while for B and C of definite sign

$$\mathcal{F}_Q(A, B, C) = 2\left(c_{QA} + c_{QB} + c_{QC}\right), \quad \text{(C11)}$$

which give respectively Eqs. (28, 29) after inserting the values of the coefficients computed above, i.e., Eqs. (C4, C7).