

Received December 7, 2021, accepted January 4, 2022, date of publication January 14, 2022, date of current version February 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3143706

Quantum Multiple Hypothesis Testing Based on a Sequential Discarding Scheme

JORDI PÉREZ-GUIJARRO¹, ALBA PAGÈS-ZAMORA¹, (Senior Member, IEEE),
AND JAVIER RODRÍGUEZ FONOLLOSA¹, (Senior Member, IEEE)

Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya, ES-08034 Barcelona, Spain

Corresponding author: Jordi Pérez-Guijarro (jordi.perez.guijarro@upc.edu)

This work was supported in part by the Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación, of the Spanish Government, under Grant RED2018-102668-T and Grant PID2019-104958RB-C41; in part by the Catalan Government under Grant 2017 SGR 578 AGAUR; and in part by the QuantumCAT within the European Regional Development Fund (ERDF) Program of Catalunya under Grant 001-P-001644.

ABSTRACT We consider the quantum multiple hypothesis testing problem, focusing on the case of hypothesis represented by pure states. A sequential adaptive algorithm is derived and analyzed first. This strategy exhibits a decay rate in the error probability with respect to the expected value of measurements greater than the optimal decay rate of the fixed-length methods. A more elaborated scheme is developed next, by serially concatenating multiple implementations of the first scheme. In this case each stage considers as *a priori* hypothesis probability the *a posteriori* probability of the previous stage. We show that, by means of a fixed number of concatenations, the expected value of measurements to be performed decreases considerably. We also analyze one strategy based on an asymptotically large concatenation of the initial scheme, demonstrating that the expected number of measurements in this case is upper bounded by a constant, even in the case of zero average error probability. A lower bound for the expected number of measurements in the zero error probability setting is also derived.

INDEX TERMS Quantum sensing, quantum hypothesis testing, fixed-length algorithms, adaptive algorithms.

I. INTRODUCTION

The task of discriminating between hypothesis is essential in a wide variety of scientific fields. In particular, the quantum hypothesis testing problem has an important role in quantum information theory with applications to quantum communication [1] and cryptography [2]. Existing techniques that solve this testing problem can be classified into *fixed-length* methods, which measure a predetermined number of copies or samples of the state, and *sequential* schemes, where the number of copies is variable and depends on a stopping criterion. Moreover, a hypothesis test for quantum states is *adaptive* when the measurement procedure of a given state copy depends on the outcome of the measurements of the previous samples.

Fixed-length methods are the most extensively studied type of strategies. The optimal method in terms of error probability was presented in the seminal works of Helstrom [3] and Holevo [4] in the late 70s, and makes use of a collective

The associate editor coordinating the review of this manuscript and approving it for publication was Abdullah Ilyasu¹.

measure of all the state copies. In the binary case with states $\{\sigma_1, \sigma_2\}$ and L samples, the probability of error is given by $\frac{1}{2} \left(1 - \left\| \pi_1 \sigma_1^{\otimes L} - \pi_2 \sigma_2^{\otimes L} \right\|_1 \right)$, where π_i denotes the a priori probability of hypothesis i [3].

The results presented in [5], [6] show that the decay rate associated with this error probability, i.e., the exponent of the error probability as L goes to infinity, is $C_Q(\sigma_1, \sigma_2) = -\log \left(\min_{0 \leq s \leq 1} \text{Tr}(\sigma_1^s \sigma_2^{1-s}) \right)$, which defines the quantum Chernoff distance between two quantum states. This expression generalizes the classical result where the maximum decay rate is given by $C(p_1, p_2) = -\log \left(\min_{0 \leq s \leq 1} \sum_i (p_1(i)^s p_2(i)^{1-s}) \right)$ [7]. More recently, [8] shows that for a fixed-length scheme the maximum decay rate in the discrimination of N quantum states is $\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)$.

In general, the maximum decay rate is achieved by means of collective quantum measurements which are difficult to implement. This motivates the quest for techniques that limit the maximum number of states to be measured jointly. Specifically, authors in [9] propose a method for binary state dis-

crimination that measures each copy individually using an adaptive strategy, and attains the minimum error probability achievable with collective measurements. This result does not hold however for the case of two non-pure states, where the maximum decay rate is strictly lower than the quantum Chernoff distance [10].

An alternative strategy to fixed-length procedures are the variable-length or *sequential* techniques where the number of state copies to be measured depends on a stopping criterion and is typically modeled as a random variable. The error probability of sequential schemes is hence assessed in terms of the expected value of the number of copies required. In fact [11], [12] show that, for the binary case, the asymptotic rate of the error probability of sequential techniques might decay faster than for fixed-length methods.

In particular, [11] uses an optimal classical method called Sequential Probability Ratio Test where the algorithm stops and decides a hypothesis when the a posteriori probability of the state associated to this hypothesis is above a given threshold. The asymptotic rate of the error probability of first kind, $p(\hat{H}_1|H_1)$, and second kind, $p(\hat{H}_1|H_2)$, attain the optimal values given by $D(\sigma_1||\sigma_2)$ and $D(\sigma_2||\sigma_1)$, where $D(\cdot||\cdot)$ is the quantum relative entropy,¹ even though not simultaneously.

An adaptive sequential strategy is adopted in [12], and the asymptotic rate decays of $D(\sigma_1||\sigma_2)$ and $D(\sigma_2||\sigma_1)$ are achieved simultaneously for errors of the first and second kind, respectively. The latter is impossible using fixed-length schemes due to the Quantum Hoeffding Bound [13], [14], which implies that if the decay rate of the first kind error is $D(\sigma_2||\sigma_1)$ then the rate of the second kind will be 0 (or alternatively decay rate of $D(\sigma_1||\sigma_2)$ of the second kind error and 0 of the first).

In this paper we focus on sequential methods for the multiple hypothesis testing problem and present an adaptive sequential scheme, named Sequential Discarding Method (SDM), that attains an asymptotic decay rate with respect to the expected value of measurements higher than $\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)$. Unlike the techniques in [11], [12], the number of required state copies in the proposed method is random but bounded for any given error probability. It is further shown that by applying the SDM method c times in a serial fashion, the expected value of measurements to be performed decreases from $O(\ln 1/\epsilon)$ to $O(\ln^{(c)} 1/\epsilon)$, where ϵ is the mean error probability and $\ln^{(c)}(\cdot)$ indicates c -times composition of the $\ln(\cdot)$ function. Interestingly, we show that the expected number of samples is bounded even in the case of zero error probability.

The paper is organized as follows. In section II we introduce the problem of quantum hypothesis testing, and present the Sequential Discarding Method. Next, in section III the performance of the SDM algorithm is analyzed. Specifically, three aspects are studied, the number of measurements L , ana-

¹ The quantum relative entropy is defined as $D(\sigma||\rho) := \text{Tr}(\sigma(\log \sigma - \log \rho))$ if $\text{supp}(\sigma) \subseteq \text{supp}(\rho)$, and $D(\sigma||\rho) := \infty$ otherwise, where the support of state σ is the subspace spanned by the eigenvectors, $|\phi_k\rangle$, with associated eigenvalues $\lambda_k > 0$.

lyzing its probability mass function and expected value; the average error probability; and the decay rate of the average error probability for large values of L_U . In section IV the serial concatenation of SDM algorithms is presented, yielding new strategies with a significant reduction in the number of measurements. In section V, numerical experiments are performed supporting our theoretical findings. Conclusions are drawn in section VI.

II. QUANTUM HYPOTHESIS TESTING

The problem of quantum hypothesis testing for pure states consists in the discrimination of the true hypothesis among a set $\{H_s\}_{s=1}^N$. Hypothesis H_s corresponds to the observation of the pure quantum state² $\sigma_s = |\psi_s\rangle\langle\psi_s| \in \mathcal{D}(\mathcal{H})$, where $\mathcal{D}(\mathcal{H})$ represents the set of density operators acting on the Hilbert space \mathcal{H} . Each hypothesis occurs with prior probability $\{\pi_s\}_{s=1}^N$. Without loss of generality, the states are assumed sorted by decreasing probability, i.e., $\pi_1 \geq \pi_2 \geq \dots \geq \pi_N$. It is also assumed that an unlimited number of copies of the unknown state can be measured before taking a decision.

Quantum state measurements are modelled through Positive Operator-Valued Measure (POVM), each of which contains P operators. We then denote the POVM by the set $\mathcal{L} = \{\Lambda_x\}_{x=1}^P$, the outcome of which is $x \in [1 : P]$. Recall that the probability that a measurement of state σ using POVM \mathcal{L} gives outcome x is equal to $\text{Tr}(\Lambda_x \sigma)$, where $\text{Tr}(\cdot)$ is the trace operator. When measuring different copies of a given quantum state, x_i represents the measurement outcome associated to the i^{th} copy, and vector $x^m = [x_1 x_2 \dots x_m]^T$ gathers measurements from 1 to m .

The decision taken as the true hypothesis is denoted by random variable (r.v.) D , a particular realization of which is expressed as $d \in [1 : N]$. The total number of measured copies of the quantum state is denoted by integer L . In fixed-length schemes L is a constant. However, in sequential schemes, e.g. [12] and the one we present in this paper, L is a r.v. and realizations of this r.v. are denoted by l .

A. BINARY HYPOTHESIS TEST: UNANIMITY VOTE

This section describes the unanimity vote method [9], which solves the binary hypothesis test problem for $N = 2$ pure states, as an introduction to the scheme presented later in Section II-B. The unanimity vote strategy belongs to the set of fixed-length techniques, i.e., it measures a fixed number L of copies of the unknown state and takes a decision about the true hypothesis, $d \in \{1, 2\}$, afterwards. The POVM used for the L measurements is $\mathcal{L} = \{\Lambda_1, \Lambda_2\} = \{|\psi_1\rangle\langle\psi_1|, I - |\psi_1\rangle\langle\psi_1|\}$ with $\pi_1 \geq \pi_2$ and where I is the identity operator. Hypothesis H_1 (or, equivalently, $d = 1$) is decided if all measurements are equal to 1, i.e., $\{x_i = 1; \forall i = 1, \dots, L\}$; otherwise, H_2 is assumed true.

Denoting by \mathcal{E} the error event, the probability of error of this scheme is

$$p(\mathcal{E}) = \pi_2 \langle \psi_1 | \psi_2 \rangle^{2L} \quad (1)$$

²We refer to a pure quantum state by $|\psi\rangle$ or $\sigma = |\psi\rangle\langle\psi|$, indistinctly.

which has an optimal asymptotic behaviour, since the decay rate associated to the error exponent as $L \rightarrow \infty$, coincides with the quantum Chernoff distance [9]. This follows from the fact that $C_Q(\sigma_1, \sigma_2) = -\log \text{Tr}(\sigma_1 \sigma_2)$ if at least one of the two states is pure [15].

Note that the number of measurements might be reduced without increasing the error probability by taking a decision as soon as one measurement outcome is equal to 2, so that state $|\psi_1\rangle$ can be discarded. This means that, in fact, the unanimity vote method can be reformulated as a sequential strategy with a stopping condition given by either measurement $x_i = 1$ is obtained for all $i \in [1, L]$, or any measurement gives $x_i = 2$ and thus the unanimous measurement outcome is not reached.

B. MULTIPLE HYPOTHESIS TEST: SEQUENTIAL DISCARDING

The previous observation motivates the proposed scheme, named sequential discarding method (SDM), which solves the multiple hypothesis testing problem, with $N \geq 2$ hypothesis, by discarding states in a sequential manner. The SDM method defines a total of $(N - 1)$ POVM sets, each of which includes only two operators, i.e., $P = 2$, as

$$\begin{aligned} \mathcal{L}^{(k)} &= \{\Lambda_1^{(k)}, \Lambda_2^{(k)}\} \\ &= \{|\psi_k\rangle\langle\psi_k|, I - |\psi_k\rangle\langle\psi_k|\} \end{aligned} \quad (2)$$

for $k \in [1 : N - 1]$. Note that each POVM $\mathcal{L}^{(k)}$ is matched to one quantum state $|\psi_k\rangle$ and the outcome of the measurements are $x_i \in \{1, 2\}$. The definition of the POVM in (2) is motivated by the fact that when $\mathcal{L}^{(k)}$ is used to measure state $|\psi_k\rangle$, the measurement outcome is deterministic and equal to 1. Clearly, state $|\psi_k\rangle$ can be discarded if any measurement outcome is different from 1 when $\mathcal{L}^{(k)}$ is used.

The SDM scheme proceeds as follows. It starts by measuring with POVM $\mathcal{L}^{(1)}$ associated with state $|\psi_1\rangle$. Additional measurements using the same POVM are performed unless either one measurement output is different from 1, or outcome 1 is unanimously obtained after L_U measurements, where $L_U \in \mathbb{N}$ is a predefined parameter of the method. If the second event occurs, decision $d = 1$ is taken and the method stops. If, on the contrary, one measurement $x_i = 2$ is obtained, state $|\psi_1\rangle$ is discarded as the true hypothesis, and the algorithm is repeated using POVM $\mathcal{L}^{(2)}$. Again, if measurement $x_i = 1$ is obtained L_U times, decision $d = 2$ is taken and the method stops; otherwise, i.e., if one measurement yields $x_i = 2$, state $|\psi_2\rangle$ is discarded. The procedure is iterated for increasing state indexes up to index $(N - 1)$ until, either the stopping condition is reached, i.e. $x_i = 1$ is obtained L_U times with the same POVM, or all states except $|\psi_N\rangle$ are discarded, so that the decision becomes $d = N$. A pseudo-code of the SDM scheme, denoted by $\mathcal{M}_{SDM}(L_U)$, is given in Algorithm 1.

Interestingly, the SDM scheme falls within the set of adaptive sequential methods but uses the predefined collection of POVMs shown in (2), unlike other adaptive methods that

Algorithm 1 SDM Scheme $\mathcal{M}_{SDM}(L_U)$

Input: The observed state $\sigma_s = |\psi_s\rangle\langle\psi_s|$, parameter L_U , and POVMs $\{\mathcal{L}^{(k)}\}_{k=1}^{N-1}$.

Output: Decision d .

$k = 1$

$N_{rep} = 0$

while $k < N$ and $N_{rep} < L_U$ **do**

 Measure state σ_s with $\mathcal{L}^{(k)}$ yielding outcome x

if $x=1$ **then**

$N_{rep} = N_{rep} + 1$

else

$N_{rep} = 0$

$k = k + 1$

end if

end while

$d = k$

solve an optimization problem to redesign the POVM at each iteration as, for instance, in [12].

III. PERFORMANCE ANALYSIS

In this section the performance of the SDM algorithm is analyzed, initially in terms of the required number of measurements and later looking into the average error probability. As in other sequential methods, the number of measurements L is a r.v. so that we study its probability mass function and expected value in the *large sample regime*, defined as $L_U \rightarrow \infty$. Afterwards, the average error probability of SDM is obtained and the decay rate for increasing values of L_U is evaluated.

A. NUMBER OF MEASUREMENTS

The number of measurements made by the SDM algorithm is a r.v. that can be expressed as

$$L = \begin{cases} L_U & \text{if } d = 1 \\ L_1 + \dots + L_{d-1} + L_U & \text{if } 2 \leq d \leq N - 1 \\ L_1 + \dots + L_{N-1} & \text{if } d = N \end{cases} \quad (3)$$

where $L_k \in [1 : L_U]$ is a r.v. that denotes the number of measurements using $\mathcal{L}^{(k)}$, if this POVM is used. Therefore, L can be lower and upper bounded as follows

$$\min \{L_U, N - 1\} \leq L \leq L_U (N - 1) \quad (4)$$

The lower bound is the minimum between L_U and $(N - 1)$, which are the number of measurements in two different situations: (a) $d = 1$ because the first L_U measurements are equal to 1, i.e. $\{x_i = 1; \forall i = 1, \dots, L_U\}$; and (b) states $\{|\psi_s\rangle\}_{s=1}^{N-1}$ are discarded, each after one single measurement, implying $\{x_i = 2; \forall i = 1, \dots, N - 1\}$ and, $\{L_i = 1; \forall i = 1, \dots, N - 1\}$ and therefore $d = N$. The upper bound in (4) corresponds also to the case when $d = N$, so that states $\{|\psi_s\rangle\}_{s=1}^{N-1}$ are discarded, but in this case after L_U measurements each, i.e., $\{L_i = L_U; \forall i = 1, \dots, N - 1\}$. We are interested in the performance analysis of the SDM scheme in the large sample

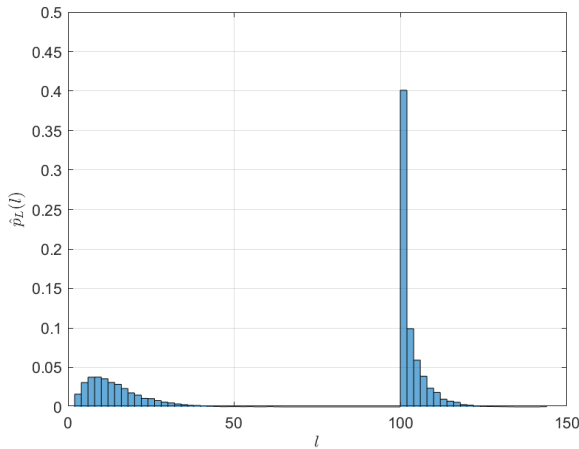


FIGURE 1. Histogram $\hat{p}_L(l)$, with 10.000 realizations of SDM algorithm with $N = 3$, $\pi_1 = \pi_2 = \pi_3 = 1/3$, and $L_U = 100$.

regime, defined in our case when $L_U \rightarrow \infty$. Assuming a limited number of states N , the upperbound in (4) shows that the SDM method might need an infinite number of copies only when $L_U \rightarrow \infty$. The following theorem assesses the behaviour of the expected value of L , denoted by $\mathbb{E}[L]$, in terms of L_U in the large sample regime.

Theorem 1: The expected value of number of measurements L of the SDM scheme in Algorithm 1 satisfies

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}[L]}{L_U} = 1 - \pi_N \quad (5)$$

Proof: See Appendix A. □

This result implies that $\mathbb{E}[L]$ in SDM increases proportionally with L_U with a factor equal to $(1 - \pi_N)$, where π_N is the minimum among all the a priori state probabilities. Note that if the states had been sorted by increasing probability (instead of decreasing), this factor would be smaller, but then, as we show later in Section III-C, the decay rate of the error probability would also be smaller.

As a complementary result, next corollary shows the moments of L in the large sample regime.

Corollary 1.1: The w^{th} -order moment of the number of measurements L of the SDM scheme in Algorithm 1 satisfies

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}[L^w]}{L_U^w} = 1 - \pi_N \quad (6)$$

Proof: See Appendix B. □

For illustrative purposes, Figure 1 shows the normalized histogram of L computed after 10.000 realizations of SDM algorithm with $N = 3$, $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, and $L_U = 100$. Clearly, the number of measurements tends to be concentrated in two intervals. This particular shape of the probability mass function of L , denoted by $p_L(l)$, is justified by the following theorem.

Theorem 2: There exist $\alpha, \beta \in \mathbb{R}^+$ with $\beta < 1$ such that

$$p_L(l) \leq (1 - p_D(N))\alpha\beta^{l-L_U}u(l - L_U) + p_D(N)\alpha\beta^l u(l) \quad (7)$$

where $u(a) = \mathbb{1}\{a \geq 0\}$ and $p_D(N)$ is the probability of deciding state N .

Proof: See Appendix C. □

Note that $\alpha\beta^{l-k}u(l - k)$ is an exponentially decaying window in the interval $l \in [k, k + \Delta]$, where Δ depends on β . Consequently, expression (7) shows that $p_L(l)$ is mostly concentrated in the intervals $[1, \Delta]$ and $[L_U, L_U + \Delta]$.

B. AVERAGE PROBABILITY OF ERROR

The average probability of error is given by

$$p(\mathcal{E}) = \sum_{d=1}^N p_{\mathcal{E}|D}(\mathcal{E}|d)p_D(d) \quad (8)$$

where \mathcal{E} denotes the error event and D the decision regarding the true hypothesis. Similarly, the conditional error probability $p_{\mathcal{E}|D}(\mathcal{E}|d)$ can be expressed as

$$\begin{aligned} p_{\mathcal{E}|D}(\mathcal{E}|d) &= \sum_{s=1}^N p_{\mathcal{E}|D,S}(\mathcal{E}|d, s)p_{S|D}(s|d) \\ &= \sum_{\substack{s=1 \\ s \neq d}}^N p_{S|D}(s|d) \end{aligned} \quad (9)$$

where S is the r.v. that represents the index of the observed state, and the second equality follows by $p_{\mathcal{E}|D,S}(\mathcal{E}|d, s) = \mathbb{1}\{d \neq s\}$. Using Bayes theorem and the fact that in the SDM algorithm $p_{D|S}(d|s) = 0 \forall d > s$, (9) becomes

$$p_{\mathcal{E}|D}(\mathcal{E}|d) = \sum_{s=d+1}^N \frac{p_{D|S}(d|s)}{p_D(d)} \pi_s \quad (10)$$

Substituting (10) in (8), we get

$$p(\mathcal{E}) = \sum_{d=1}^{N-1} \sum_{s=d+1}^N p_{D|S}(d|s)\pi_s \quad (11)$$

Then, noting that $\text{Tr}(|\psi_d\rangle\langle\psi_d| \sigma_s)^{L_U} = \text{Tr}(\sigma_d \sigma_s)^{L_U}$ is the probability of obtaining outcome $x_i = 1$ L_U times when POVM $\mathcal{L}^{(d)}$ is used, the conditional probabilities $p_{D|S}(d|s)$ for $d \in [1, N - 1]$ are equal to

$$\begin{cases} \text{Tr}(\sigma_1 \sigma_s)^{L_U} & \text{if } d = 1 \\ \text{Tr}(\sigma_d \sigma_s)^{L_U} \prod_{j=1}^{d-1} (1 - \text{Tr}(\sigma_j \sigma_s)^{L_U}) & \text{if } d > 1 \end{cases} \quad (12)$$

The expression of $p_{D|S}(d|s)$ for $d > 1$ comes from the observation that $\prod_{j=1}^{d-1} (1 - \text{Tr}(\sigma_j \sigma_s)^{L_U})$ is the probability of discarding the states $\{|\psi_i\rangle\}_{i=1}^{d-1}$, and $\text{Tr}(\sigma_d \sigma_s)^{L_U}$ is the probability of deciding d once lower indexed states are discarded.

Substituting (12) in (11), the probability of error becomes

$$\begin{aligned} p(\mathcal{E}) &= \sum_{s=2}^N \pi_s \text{Tr}(\sigma_1 \sigma_s)^{L_U} + \sum_{d=2}^{N-1} \sum_{s=d+1}^N \pi_s \text{Tr}(\sigma_d \sigma_s)^{L_U} \\ &\quad \times \prod_{j=1}^{d-1} (1 - \text{Tr}(\sigma_j \sigma_s)^{L_U}) \end{aligned} \quad (13)$$

Since $0 < \text{Tr}(\sigma_d \sigma_s) < 1$ for $d \neq s$, all terms of the sum decrease exponentially with L_U and, hence, the average error probability also decreases exponentially with L_U . This observation is further supported by the following proposition, to be used in next subsection where the decay rate of error probability is assessed.

Proposition 1: The average probability of error $p(\mathcal{E})$ of the SDM algorithm is upper and lower bounded as

$$B_L \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \leq p(\mathcal{E}) \leq B_U \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \quad (14)$$

where $B_L, B_U \in \mathbb{R}^+$ and $B_L < B_U$.

Proof: See Appendix D. \square

C. DECAY RATE OF THE AVERAGE ERROR PROBABILITY

The asymptotic decay rate of $p(\mathcal{E})$ as L_U tends to infinity is defined as

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{L_U} \quad (15)$$

To compute this limit we first observe that, using Proposition 1, the ratio can be lower and upper bounded as follows

$$\begin{aligned} &-\frac{\ln B_U}{L_U} + \min_{i,j:i \neq j} \{-\ln \text{Tr}(\sigma_i \sigma_j)\} \\ &\leq \frac{-\ln p(\mathcal{E})}{L_U} \leq -\frac{\ln B_L}{L_U} + \min_{i,j:i \neq j} \{-\ln \text{Tr}(\sigma_i \sigma_j)\} \end{aligned} \quad (16)$$

which implies that

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{L_U} = \min_{i,j:i \neq j} \{-\ln \text{Tr}(\sigma_i \sigma_j)\} \quad (17)$$

As the quantum Chernoff distance, in the case that at least one state is pure, can be written as $C_Q(\sigma_1, \sigma_2) = -\ln \text{Tr}(\sigma_1 \sigma_2)$ [15], the decay rate with respect to L_U becomes

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{L_U} = \min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j) \quad (18)$$

We are also interested in assessing the decay rate of the average error probability with respect to the expected number of samples, defined as

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{\mathbb{E}[L]} \quad (19)$$

Applying the product rule of the limit, (19) can be expressed as

$$\lim_{L_U \rightarrow \infty} \frac{L_U}{\mathbb{E}[L]} \lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{L_U} \quad (20)$$

since both limits are finite. Then, using (5) from Theorem 1 and (18), the final result is obtained

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{\mathbb{E}[L]} = \frac{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)}{1 - \pi_N} \quad (21)$$

This decay rate differs only in the factor $1/(1 - \pi_N)$ with respect to the optimal fixed-length method. Interestingly, since $(1 - \pi_N) < 1$, the decay rate with respect to the expected number of measurements of the sequential discarding method

is greater than the optimal decay rate of the fixed-length strategies.

It is worth noting that the decay rate obtained in (21) remains valid even if one of the states is not pure. In this scenario, the index of the mixed state must be N , and therefore no POVM associated with this state needs to be defined.

IV. SERIAL CONCATENATION OF SDM ALGORITHMS

In this section we study how the serial concatenation of sequential discarding methods can be employed to yield a new measuring strategy by properly transferring the posterior probabilities of the states. The concatenation operation of two SDM algorithms, namely $\mathcal{M}_{SDM}(L_U, 1)$ and $\mathcal{M}_{SDM}(L_U, 2)$, consists of:

- 1) Indexing states in increasing order of a priori probability.
- 2) Executing algorithm $\mathcal{M}_{SDM}(L_U, 1)$. The number of measurements performed in this stage is denoted l_1 .
- 3) Computing the posterior probabilities $p_{S|X^{l_1}}(s|x^{l_1})$ and reordering them in increasing order. In case any posterior probability is zero, the associated state is discarded and the total number of states is reduced.
- 4) Executing $\mathcal{M}_{SDM}(L_U, 2)$ using $p_{S|X^{l_1}}(s|x^{l_1})$ as a priori probability for each possible state.

This procedure can be iterated to concatenate an arbitrary number of SDM algorithms (see Appendix E details). Our analysis assesses the expected number of measurements in two cases: (i) when a predetermined number of SDM stages are concatenated; and (ii) when the number of concatenated SDM algorithms can be arbitrarily large such that the probability of error can be made equal to zero.

A. NUMBER OF MEASUREMENTS WITH SERIAL CONCATENATION

In this subsection, we first study the expected number of measurements when two SDM algorithms are concatenated. This value is then compared to the expected number of measurements obtained with a single SDM algorithm. Finally, the result is generalized to the case when multiple SDM algorithms are concatenated.

1) NUMBER OF MEASUREMENTS WITH TWO SDMS

Expression (21) implies that, for a sufficiently large value of $L_{U,1}$, the expected number of measurements required by $\mathcal{M}_{SDM}(L_U, 1)$ is

$$\mathbb{E}[L_1] \simeq \frac{1 - \max_{s \in [1:N]} \pi_s}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln \left(\frac{1}{\epsilon_1} \right), \quad (22)$$

where π_s for $s \in [1 : N]$ is the a priori state probability and ϵ_1 is the average probability of error of $\mathcal{M}_{SDM}(L_U, 1)$. Assuming that $\mathcal{M}_{SDM}(L_U, 1)$ performs l_1 measurements, then the a priori probabilities for $\mathcal{M}_{SDM}(L_U, 2)$ are given by the posteriors $p_{S|X^{l_1}}(s|x^{l_1})$. Consequently, the expected total number of measurements required by the concatenation of $\mathcal{M}_{SDM}(L_U, 1)$ and $\mathcal{M}_{SDM}(L_U, 2)$, conditioned by the fact that

l_1 measurements were performed in the first one, is given by

$$\begin{aligned} \mathbb{E}[L|x^{l_1}] &= \mathbb{E}[L_2|x^{l_1}] + l_1 \\ &\simeq \frac{1 - \max_s p_{S|x^{l_1}}(s|x^{l_1})}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln\left(\frac{1}{\epsilon_2}\right) + l_1 \end{aligned} \quad (23)$$

where ϵ_2 is the probability of error of $\mathcal{M}_{SDM}(L_{U,2})$ using a priori state probabilities equal to $p_{S|x^{l_1}}(s|x^{l_1})$. The expected total number of measurements, $\mathbb{E}[L]$, of this two stage concatenation scheme can be computed taking the expected value of (23) with respect to x^{l_1} as follows

$$\begin{aligned} \mathbb{E}[L] &= \mathbb{E}_{X^{L_1}} \left[\mathbb{E}[L|x^{l_1}] \right] \\ &\simeq \frac{\mathbb{E}_{X^{L_1}} \left[1 - \max_s p_{S|x^{l_1}}(s|x^{l_1}) \right]}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln\left(\frac{1}{\epsilon_2}\right) + \mathbb{E}_{X^{L_1}}[l_1] \end{aligned} \quad (24)$$

Note that $\mathbb{E}_{X^{L_1}} \left[1 - \max_s p_{S|x^{l_1}}(s|x^{l_1}) \right]$ is the probability of error of the first SDM algorithm when using a MAP criterion. Hence,

$$\mathbb{E}_{X^{L_1}} \left[1 - \max_s p_{S|x^{l_1}}(s|x^{l_1}) \right] \leq \epsilon_1 \quad (25)$$

Substituting (25) in (24), the expected total number of measurements becomes upper bounded as

$$\begin{aligned} \mathbb{E}[L] &\leq \frac{\epsilon_1}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln\left(\frac{1}{\epsilon_2}\right) + \mathbb{E}_{X^{L_1}}[l_1] \\ &\simeq \frac{1}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \left(\epsilon_1 \ln\left(\frac{1}{\epsilon_2}\right) + \epsilon_0 \ln\left(\frac{1}{\epsilon_1}\right) \right) \end{aligned} \quad (26)$$

where the approximation follows from (22) denoting $\epsilon_0 := 1 - \pi_N$.

2) COMPARISON WITH A SINGLE SDM

In order to assess the expected number of measurements required by the concatenation of two SDM algorithms, as compared to a single SDM, we fix the probability of error to a given value ϵ . Note that this value corresponds to ϵ_2 in the two-stage concatenation scheme since the decision is taken only after the second stage.

Again, using (21), for a sufficiently large value of L_U the expected number of measurements of one single SDM is

$$\mathbb{E}[L] \simeq \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln\left(\frac{1}{\epsilon}\right), \quad (27)$$

where $\epsilon_0 = 1 - \pi_N$. Therefore, comparing (27) with (26) and substituting $\epsilon_2 = \epsilon$, the two stage concatenation strategy requires a smaller expected number of measurements if ϵ_1 is designed such that

$$\epsilon_1 \ln\left(\frac{1}{\epsilon}\right) + \epsilon_0 \ln\left(\frac{1}{\epsilon_1}\right) < \epsilon_0 \ln\left(\frac{1}{\epsilon}\right) \quad (28)$$

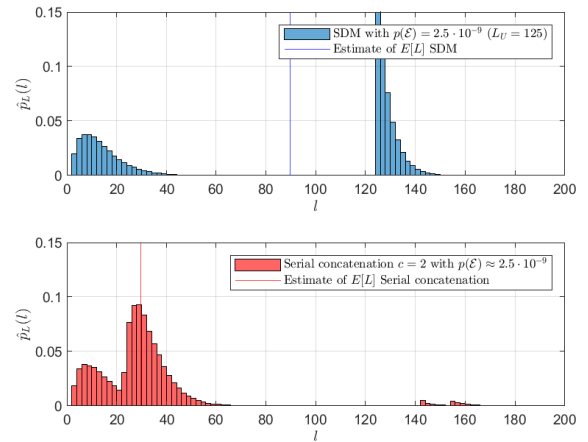


FIGURE 2. Normalized histogram $\hat{p}_L(l)$ of $\mathcal{M}_{SDM}(L_U)$ with $L_U = 125$ (upper) and of $[\mathcal{M}_{SDM}(L_{U,1}), \mathcal{M}_{SDM}(L_{U,2})]$ (lower), both with the same average error probability $\epsilon = 2.5 \cdot 10^{-9}$. Vertical lines indicate the average value of L .

In general there exist a range of values ϵ_1 that fulfill the previous inequality. Specifically, if ϵ_1 is set equal to

$$\epsilon_1^* = \operatorname{argmin}_{0 < \epsilon_1 < 1} \left(\epsilon_1 \ln\left(\frac{1}{\epsilon}\right) + \epsilon_0 \ln\left(\frac{1}{\epsilon_1}\right) \right) = \frac{\epsilon_0}{\ln\left(\frac{1}{\epsilon}\right)} \quad (29)$$

then the expected measurement number can be upper bounded as

$$\begin{aligned} \mathbb{E}[L] &\leq \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \left(1 + \ln \ln\left(\frac{1}{\epsilon}\right) + \ln\left(\frac{1}{\epsilon_0}\right) \right) \\ &\leq \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \left(\ln \ln\left(\frac{1}{\epsilon}\right) + \frac{1}{\epsilon_0} \right) \end{aligned} \quad (30)$$

where the second inequality follows from $\ln(1/\epsilon_0) \leq 1/\epsilon_0 - 1$. It is important to note that ϵ_1 cannot take any arbitrary value since it depends on parameter $L_{U,1}$, which must be integer. That is, a value $L_{U,1}$ such that ϵ_1 takes exactly the optimal value, $\epsilon_1^* = \epsilon_0 / \ln(1/\epsilon)$, does not exist in general. However, by taking $L_{U,1}$ as the smallest natural number that fulfills $\epsilon_1 \leq \epsilon_1^*$, the upper bound of the expected number of measurements grows as $O(\ln \ln 1/\epsilon)$ as well.

For illustrative purposes, Figure 2 plots the normalized histogram of L when one single SDM with $L_U = 125$ is used (upper plot) and when 2 SDM algorithms are concatenated (lower plot). A total of 10^5 realizations are evaluated and the average error probability is set to $\epsilon = 2.5 \cdot 10^{-9}$. The states and prior probabilities are the same as the ones used in Figure 1 i.e., $N = 3$ and $\pi_1 = \pi_2 = \pi_3 = 1/3$. Parameters $L_{U,1}$ and $L_{U,2}$ are computed as the smallest integers such that $\epsilon_1 < \frac{\epsilon_0}{\ln(1/\epsilon)}$ and $\epsilon_2 < \epsilon$.

The approximated probability mass function $\hat{p}_L(l)$ for the two stage concatenated strategy is mostly concentrated near the origin where two maxima appear. The first maximum is associated with the scenario where $\mathcal{M}_{SDM}(L_{U,1})$ decides $d = N$. As for the second maximum, it comes from the case where $\mathcal{M}_{SDM}(L_{U,2})$ discards $N - 1$ states and therefore the decision

has a zero probability of error. The difference between both maxima is approximately $L_{U,1} = 21$. Moreover, we can see that exists another region of values of l , around $l \approx L_{U,1} + L_{U,2}$, which is probable. This is due to the double band effect shown in Figure 1. Remarkably, the expected number of measurements of the two stage concatenation strategy is reduced by a factor of $1/3$ in comparison with the single SDM strategy.

3) NUMBER OF MEASUREMENTS WITH MULTIPLE SDMs

The expression of the expected number of measurements obtained by the concatenation of two sequential discarding methods given in (30) can be generalized by mathematical induction to an arbitrary number of stages. The resulting expected number of measurements is given by

$$\mathbb{E}[L] \leq \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln^{(c)}\left(\frac{1}{\epsilon}\right) + O(c) \quad (31)$$

where ϵ is the intended average error bound at the final stage, and $\ln^{(c)}(\cdot)$ indicates c -times composition of the $\ln(\cdot)$ function.

At the induction step, a concatenation of c SDM stages is built by adding one SDM block as the first stage to an existing concatenation of $c - 1$ SDM stages. We assume the expected number of measurements of the $c - 1$ concatenation is bounded by

$$\mathbb{E}[L] \leq \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln^{(c-1)}\left(\frac{1}{\epsilon}\right) + O(c-1) \quad (32)$$

Hence, following an analogous procedure to the one shown in (26), the expected number of measurements of the c stage concatenation becomes upper bounded by

$$\begin{aligned} \mathbb{E}[L] \leq & \frac{\epsilon_1}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln^{(c-1)}\left(\frac{1}{\epsilon}\right) \\ & + \frac{\epsilon_0}{\min_{i,j:i \neq j} C_Q(\sigma_i, \sigma_j)} \ln\left(\frac{1}{\epsilon_1}\right) + O(c-1) \end{aligned} \quad (33)$$

where the optimal probability of error of the additional SDM stage is $\epsilon_1^* = \frac{\epsilon_0}{\ln^{(c-1)} \frac{1}{\epsilon}}$. Substituting this value in (33), expression (31) is obtained.

In a nutshell, comparing (27) to (31) the number of measurements can be reduced from $O(\ln 1/\epsilon)$ to $O(\ln^{(c)} 1/\epsilon)$ using a concatenation of c SDM algorithms, where ϵ denotes the average error probability.

B. NUMBER OF MEASUREMENTS FOR ZERO ERROR PROBABILITY

In this section, we let the number of SDM stages c to be arbitrarily large so that the algorithm only finishes once $N - 1$ states are discarded, which yields a zero probability of error. Note that in this case inequality (31) cannot be used to study the expected number of measurements since the upper bound tends to infinite.

The conducted analysis is summarized in Theorem 3 and 4 which provide an upper and lower bound for the expected number of measurements, respectively.

Theorem 3: The expected number of measurements, $\mathbb{E}[L]$, of an asymptotically large concatenation of SDM stages, all of which use $L_{U,\min} \leq L_{U,i} \leq L_{U,\max}$, is bounded as

$$\mathbb{E}[L] \leq L_{U,\max}(N - 1) \left(\frac{1}{p_+} + 1\right) - L_{U,\max} \quad (34)$$

where $p_+ = \frac{1}{2}(1 - \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j))^{L_{U,\min}}$.

Proof: Since zero error probability is required, all except one of the possible states must be discarded. This requires at most $L_{U,\max}$ measurements per state yielding a total of at most $(N - 1)L_{U,\max}$ measurements. In addition, all SDM stages $\mathcal{M}_{SDM}(L_{U,i})$, except the final one, end up with a non-discarding measurement and, therefore, use $L_{U,i}$ measurements each. The total number of required measurements is then upper bounded by

$$L \leq L_{U,\max}(N - 1) + \sum_{i=1}^{C_e-1} L_{U,i} \leq L_{U,\max}(N - 2 + C_e) \quad (35)$$

where C_e is a r.v. that denotes the number of stages used. As shown in appendix F, the expected number of required SDM stages $\mathbb{E}[C_e]$ satisfies

$$\mathbb{E}[C_e] \leq \frac{N - 1}{p_+} \quad (36)$$

Therefore,

$$\mathbb{E}[L] \leq L_{U,\max}(N - 1) \left(\frac{1}{p_+} + 1\right) - L_{U,\max} \quad (37)$$

□

Theorem 4: Any algorithm that decides with certainty the observed hypothesis, H_s , i.e., with $p(\mathcal{E}) = 0$, fulfills that

$$\mathbb{E}[L] \geq \sum_{s=1}^N \pi_s \sum_{\substack{i=1 \\ i \neq s}}^N \frac{1}{1 - |\langle \psi_s | \psi_i \rangle|^2} \quad (38)$$

if all the POVM used belong to the set $\{\mathcal{L}^{(k)}\}_{k=1}^N$ given in (2).

Proof: The expected value of measurements can be computed as

$$\mathbb{E}[L] = \sum_{s=1}^N \pi_s \mathbb{E}[L|s] \quad (39)$$

Since we require $p(\mathcal{E}) = 0$, all states except s must be discarded when the true state is s . Since POVM $\{\mathcal{L}^{(k)}\}_{k=1}^N$ can discard only one state, $\mathbb{E}[L|s]$ satisfies

$$\mathbb{E}[L|s] \geq \sum_{\substack{i=1 \\ i \neq s}}^N \mathbb{E}[M_i|s] \quad (40)$$

where M_i denotes the number of measurements used to discard state i when POVM $\mathcal{L}^{(i)} = \{|\psi_i\rangle\langle\psi_i|, I - |\psi_i\rangle\langle\psi_i|\}$ is used. The inequality follows since additional non-discarding

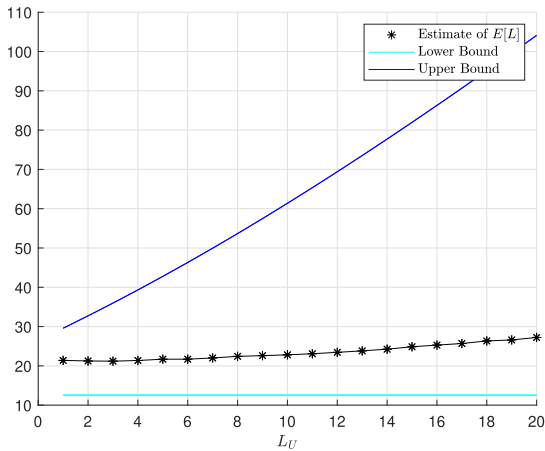


FIGURE 3. Empirical average of the number of measurements with the asymptotically large concatenation scheme as a function of L_U estimated using $2 \cdot 10^4$ realizations. Upper and lower bounds obtained in (34) and (38), respectively.

measures can also be performed. The value of $\mathbb{E}[M_i|s]$ is given by

$$\begin{aligned} \mathbb{E}[M_i|s] &= \sum_{m_i=1}^{\infty} m_i \left(1 - |\langle \psi_s | \psi_i \rangle|^2\right) |\langle \psi_s | \psi_i \rangle|^{2(m_i-1)} \\ &= \frac{1}{1 - |\langle \psi_s | \psi_i \rangle|^2} \end{aligned} \quad (41)$$

which implies that

$$\mathbb{E}[L] \geq \sum_{s=1}^N \pi_s \sum_{\substack{i=1 \\ i \neq s}}^N \frac{1}{1 - |\langle \psi_s | \psi_i \rangle|^2} \quad (42)$$

□

Setting $L_{U,i}$ equal in all stages, i.e. $L_{U,min} = L_{U,max} = L_U$, then the gap between upper and lower bounds in Theorem 3 and 4 is minimized for $L_U = 1$. This result follows since the lower bound does not depend on L_U , and the upper bound can be expressed as

$$(N - 1) \cdot g_{T^*}(L_U) + L_U \cdot (N - 2) \quad (43)$$

where $T^* = \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)$, and $g_T(L_U)$ is given by

$$g_T(L_U) = \frac{2L_U}{1 - T^{L_U}}, \quad (44)$$

which is an increasing function on $L_U, \forall T \in [0, 1)$. Figure 3 shows the empirical average of the number of measurements when the number of SDM stages is arbitrarily large until $p(\mathcal{E}) = 0$, and both the upper and lower bounds obtained in Theorem 3 and 4 for different values of L_U . Interestingly, the expected number of measurements of the asymptotically large concatenation strategy only moderately increases with L_U . Also, it is clear that the upper bound is not tight for large values of L_U .

C. NUMBER OF MEASUREMENTS WITH MIXED STATES

The reduction in the number of expected measurements obtained by the concatenation of multiple SDM stages can also be achieved when quantum states are mixed if certain requirements are met. In particular we can generalize the previous results to the multiple hypothesis testing problem for mixed states, denoted by $\{\rho_s\}_{s=1}^N, \rho_s \in \mathcal{D}(\mathcal{H})$, provided³ $\text{supp}(\rho_i) \not\subseteq \text{supp}(\rho_j) \forall i \neq j$, or equivalently, $D(\rho_i || \rho_j) = \infty \forall i \neq j$.

To show this, we consider the spectral decomposition of $\rho_k = \sum_q \lambda_q^k \left| \phi_q^k \right\rangle \left\langle \phi_q^k \right|$ and POVM $\{\Pi_k, I - \Pi_k\}$, where $\Pi_k = \sum_{q:\lambda_q^k > 0} \left| \phi_q^k \right\rangle \left\langle \phi_q^k \right|$ is the projection onto the support of ρ_k . Proceeding as in Section III, it is not difficult to show that the decay of (21) becomes

$$\lim_{L_U \rightarrow \infty} \frac{-\ln p(\mathcal{E})}{\mathbb{E}[L]} = \frac{-\ln \max_{i,j:i < j} \text{Tr}(\Pi_i \rho_j)}{1 - \pi_N} \quad (45)$$

Note that $\max_{i,j:i < j} \text{Tr}(\Pi_i \rho_j) < 1$ since $\text{supp}(\rho_i) \not\subseteq \text{supp}(\rho_j) \forall i \neq j$. Using (45), similar results to the ones presented in Section IV-A and IV-B can be obtained. In particular, the expected number of measurements when c SDM stages are concatenated is upper bounded by

$$\mathbb{E}[L] \leq \frac{\epsilon_0}{-\ln \max_{i,j:i < j} \text{Tr}(\Pi_i \rho_j)} \ln^{(c)}\left(\frac{1}{\epsilon}\right) + O(c), \quad (46)$$

Finally, the expected number of measurements of an asymptotically large concatenation of SDM stages, all of which use $L_{U,min} \leq L_{U,i} \leq L_{U,max}$, is bounded by

$$\mathbb{E}[L] \leq L_{U,max}(N - 1) \left(\frac{1}{p_+} + 1\right) - L_{U,max} \quad (47)$$

where $p_+ = \frac{1}{2}(1 - \max_{i,j:i < j} \text{Tr}(\Pi_i \rho_j)^{L_{U,min}})$.

V. NUMERICAL EXPERIMENTS

This section includes simulations that support our theoretical findings, both of the SDM algorithm and of the serial concatenation of SDM algorithms.

Figure 4 shows results of the SDM algorithm with $N = 3$ and states $|\psi_i\rangle = \cos\left(\frac{\theta_i}{2}\right) |0\rangle + e^{j\theta_i} \sin\left(\frac{\theta_i}{2}\right) |1\rangle$, with

$$\begin{aligned} \phi_1 &= 5.40 \text{ rad}, & \phi_2 &= 0.45 \text{ rad}, & \phi_3 &= 5.91 \text{ rad} \\ \theta_1 &= 2.63 \text{ rad}, & \theta_2 &= 2.21 \text{ rad}, & \theta_3 &= 1.91 \text{ rad} \end{aligned}$$

The a priori state probabilities are $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$. Specifically, 10^4 simulations of the SDM algorithm are run for each value of L_U in the interval $[1 : 300]$. In each simulation, the average probability of error is computed using (13) and the number of measurements L is saved. Using these results, a histogram of L for each value of L_U is obtained, and represented using the gray scale on the right of the figure. For comparison, the optimal number of measurements for a fixed-length method, the estimated expected value of measurements of the SDM method, $\mathbb{E}[L]$, and its theoretical value on the large sample regime given in (5) are also indicated.

³ The support of a quantum state ρ_k is the subspace spanned by its eigenvectors $|\phi_q^k\rangle$ with positive associated eigenvalues, i.e. $\lambda_q^k > 0$.

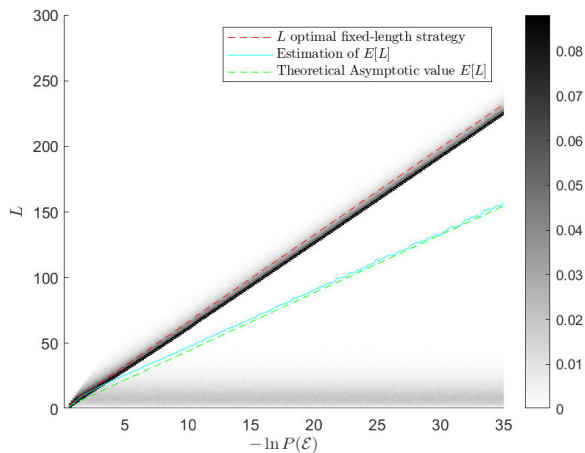


FIGURE 4. Histogram of the number of measurements L for $L_U \in [1 : 300]$, optimal values for fixed-length methods, and estimated and predicted values.

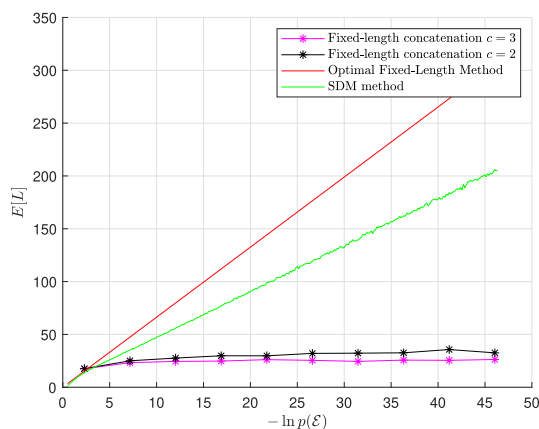


FIGURE 5. This figure shows a comparison between the expected value of measurements for SDM, serial concatenation of SDM, and optimal fixed-length schemes.

Figure 4 shows that the expected value of samples is very close to the asymptotic theoretical value. Moreover, the expected value of measurements is significantly smaller than the number of samples used by a fixed-length method for the same average probability of error.

Figure 5 compares the expected value of measurements of the SDM scheme, the serial concatenation of SDM stages with $c \in \{2, 3\}$, and the optimal fixed-length method. The states and their a priori probability are the same as the ones used in the previous example. Clearly, the expected number of measurements for the serial concatenation scheme is significantly smaller than the others.

Finally, Figure 6 illustrates the quotient between $\mathbb{E}[L]$ and the lower bound shown in Theorem 4, when an arbitrarily large concatenation of SDM algorithms is used. This ratio is calculated for different values of prior probabilities $\{\pi_1, \pi_2, \pi_3\}$, i.e, for coordinate values inside triangle $\text{co}(\{[1, 0, 0], [0, 1, 0], [0, 0, 1]\})$, where $\text{co}(\cdot)$ denotes the convex hull operator. For each value of the a priori

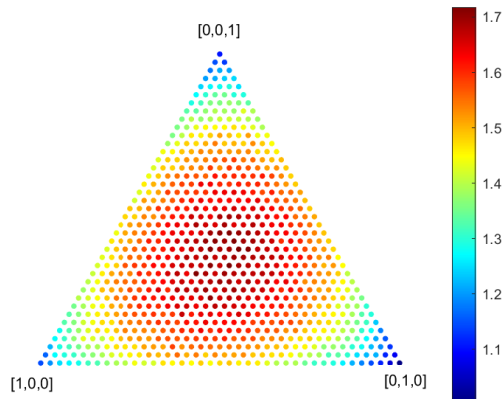


FIGURE 6. Quotient between $\mathbb{E}[L]$ and the lower bound shown in Theorem 4 for the method described in Section IV-B with $L_{U,i} = 1\forall i$.

probabilities, the method is run 5000 times to obtain an estimate of $\mathbb{E}[L]$.

As shown in Figure 6, for a priori probabilities close to the vertices, the expected number of measurements is approximately equal to the lower bound in Theorem 4, that is, the method is approximately optimal in terms of expected number of measurements. As for less informative prior probabilities, that is, those probability values represented close to the center of the triangle, a moderate increase in the number of measurements is obtained, specifically $\mathbb{E}[L]$ is 1.7 times higher than the lower bound in Theorem 4.

VI. CONCLUSION

This article demonstrates that sequential discarding methods for the discrimination of pure states, even when simple binary projective measurements are employed, can yield a sub-logarithmic increase of the expected number of measurements with respect to the error probability. The measuring strategy that achieves this decay rate consists of the concatenation of multiple stages of the SDM algorithm. This result also applies to the discrimination of mixed states under certain restrictions on their respective support subspaces. Moreover, this work also shows that, when a single SDM stage is used, the number of measurements required to reach a given probability of error is similar to the one required by the optimal fixed-length strategy, and smaller on average for sufficiently small error probability. Furthermore, in the case of employing an asymptotically large concatenation of SDM stages, the expected value of measurements remains bounded, even in the case of zero error probability.

APPENDIX A

EXPECTED VALUE OF L IN THE LARGE SAMPLE REGIME

The large sample regime is defined as $L_U \rightarrow \infty$ and thus we assume $L_U > (N - 1)$. The expected value of the number of measurements $\mathbb{E}[L]$ can be expressed as

$$\mathbb{E}[L] := \sum_{l=N-1}^{L_U(N-1)} l p_L(l)$$

$$\begin{aligned}
 &= \sum_{l=N-1}^{L_U(N-1)} \sum_{d=1}^N l p_{L|D}(l|d) p_D(d) \\
 &= \sum_{d=1}^N p_D(d) \mathbb{E}_{L|D}[L|d] \tag{48}
 \end{aligned}$$

where the first equality follows by the law of total probability, and the second by the definition of $\mathbb{E}_{L|D}[L|d] := \sum_{l=N-1}^{L_U(N-1)} l p_{L|D}(l|d)$, which denotes the average number of measurements when decision d is taken.

Let us define a r.v. A , which denotes the number of measurements made before POVM $\mathcal{L}^{(d)}$ is used and decision d is taken. This r.v. takes values

$$A = \begin{cases} 0 & \text{if } d = 1 \\ L_1 + L_2 + \dots + L_{d-1} & \text{if } 1 < d \leq N \end{cases} \tag{49}$$

Substituting (49) in (3), the total number of measurements becomes

$$L = \begin{cases} A + L_U & \text{if } 1 \leq d < N \\ A & \text{if } d = N \end{cases} \tag{50}$$

Consequently, for $1 < d < N$:

$$\begin{aligned}
 \mathbb{E}_{L|D}[L|d] &= \sum_{l=N-1}^{L_U(N-1)} l p_{L|D}(l|d) \\
 &= \sum_{a=d-1}^{(d-1)L_U} (a + L_U) p_{A|D}(a|d) \\
 &= L_U + \sum_{a=d-1}^{(d-1)L_U} a p_{A|D}(a|d) \tag{51}
 \end{aligned}$$

In the second equality we use the fact that, for a given decision d , A can take values between $d - 1$ and $L_U(d - 1)$. That is, $A = d - 1$ when a single measurement is used to discard each of the $d - 1$ previous states, and $L_U(d - 1)$, when all previous states are discarded after L_U measurements. Using (49) $p_{A|D}(a|d)$ becomes

$$p_{A|D}(a|d) = \sum_{l^{d-1} \in \mathcal{T}_a} p_{L_1|D}(l_1|d) \dots p_{L_{d-1}|D}(l_{d-1}|d) \tag{52}$$

where $l^{d-1} := (l_1, \dots, l_{d-1})$ and $\mathcal{T}_a = \{l^{d-1} \in \mathbb{N}^{d-1} : l_1 + \dots + l_{d-1} = a\}$. Since $d \leq s \leq N$, probabilities $p_{L_j|D}(l_j|d)$ are given by

$$\begin{aligned}
 p_{L_j|D}(l_j|d) &= \sum_{s=d}^N p_{L_j|S,D}(l_j|s, d) p_{S|D}(s|d) \\
 &= \sum_{s=d}^N \text{Tr}(\sigma_j \sigma_s)^{l_j-1} (1 - \text{Tr}(\sigma_j \sigma_s)) p_{S|D}(s|d)
 \end{aligned}$$

where $\text{Tr}(\sigma_j \sigma_s) = \text{Tr}(|\psi_j\rangle\langle\psi_j| \sigma_s)$ is the probability of obtaining the outcome $x = 1$ when POVM $\mathcal{L}^{(j)}$ is used, and

$p_{S|D}(s|d)$ is the probability of observing state σ_s when d is decided. Substituting this result in (52) gives

$$p_{A|D}(a|d) = \sum_{l^{d-1} \in \mathcal{T}_a} \prod_{j=1}^{d-1} \sum_{s=d}^N \text{Tr}(\sigma_j \sigma_s)^{l_j-1} \times (1 - \text{Tr}(\sigma_j \sigma_s)) p_{S|D}(s|d) \tag{53}$$

Now the term

$$\prod_{j=1}^{d-1} \sum_{s=d}^N \text{Tr}(\sigma_j \sigma_s)^{l_j-1} (1 - \text{Tr}(\sigma_j \sigma_s)) p_{S|D}(s|d) \tag{54}$$

which incorporates $d - 1$ factors, can be expressed by the following sum of $N_d = (N - d + 1)^{d-1}$ terms

$$\sum_{i=1}^{N_d} c_i \cdot \beta_{1,i}^{l_{1,i}-1} \beta_{2,i}^{l_{2,i}-1} \dots \beta_{d-1,i}^{l_{d-1,i}-1} \tag{55}$$

where $0 < c_i < 1$ gathers the product of $d - 1$ factors of the form $(1 - \text{Tr}(\sigma_j \sigma_s)) p_{S|D}(s|d)$, and $0 < \beta_{j,i} < 1$ is equal to $\text{Tr}(\sigma_j \sigma_s)$. Defining $\hat{\beta}_i := \max_j \beta_{j,i}$ then expression (55) can be upper bounded by

$$\sum_{i=1}^{N_d} \hat{\beta}_i^{a-d+1} \tag{56}$$

since $l_1 + \dots + l_{d-1} = a$. Noting that this upper bound does not depend on the element $l^{(d-1)}$, $p_{A|D}(a|d)$ in (53) can be then upper bounded by

$$p_{A|D}(a|d) \leq \sum_{i=1}^{N_d} |\mathcal{T}_a| \hat{\beta}_i^{a-d+1} \tag{57}$$

where $|\mathcal{T}_a|$ is the cardinality of the set \mathcal{T}_a . Substituting (57) in (51), we have

$$\mathbb{E}_{L|D}[L|d] \leq L_U + \sum_{a=d-1}^{(d-1)L_U} \sum_{i=1}^{N_d} a |\mathcal{T}_a| \hat{\beta}_i^{a-d+1} \tag{58}$$

Now, note that $|\mathcal{T}_a|$ is the number of solutions of equation $a = l_1 + l_2 + \dots + l_{d-1}$ for $l_j \in \mathbb{N}$. Defining $\bar{l}_j = (l_j - 1)$, we have $\sum_{j=1}^{d-1} \bar{l}_j = \sum_{j=1}^{d-1} l_j - d + 1$ and, therefore, $|\mathcal{T}_a|$ also amounts the number of solutions of $\sum_{j=1}^{d-1} \bar{l}_j = a - d + 1$ for $\bar{l}_j \geq 0$. Then denoting $v = a - d + 1$, the cardinality of the set \mathcal{T}_a is equal to

$$|\mathcal{T}_a| = \binom{v + d - 2}{v} = \binom{a - 1}{a - d + 1}, \tag{59}$$

where $\binom{x}{y} = \frac{x!}{y!(x-y)!}$. Substituting (59) in (58), we have

$$\begin{aligned}
 \mathbb{E}_{L|D}[L|d] &\leq L_U + \sum_{i=1}^{N_d} \sum_{a=d-1}^{(d-1)L_U} a \binom{a - 1}{a - d + 1} \hat{\beta}_i^{a-d+1} \\
 &= L_U + \sum_{i=1}^{N_d} \sum_{a=d-1}^{(d-1)L_U} \frac{a!}{(a - d + 1)!(d - 2)!} \hat{\beta}_i^{a-d+1}
 \end{aligned}$$

$$\leq L_U + \sum_{i=1}^{N_d} \frac{1}{(d-2)!} \sum_{a=0}^{\infty} \left(\prod_{j=0}^{d-2} (a-j) \right) \hat{\beta}_i^{a-d+1}$$

where the second inequality follows increasing the range of the summation from $[d-1, (d-1)L_U]$ to $[0, \infty)$. Next, using that

$$\frac{d^{d-1}}{d \hat{\beta}_i^{d-1}} \sum_{a=0}^{\infty} \hat{\beta}_i^a = \sum_{a=0}^{\infty} \left(\prod_{j=0}^{d-2} (a-j) \right) \hat{\beta}_i^{a-d+1}, \quad (60)$$

and $\sum_{a=0}^{\infty} u^a = 1/(1-u)$ if $|u| < 1$, gives

$$\mathbb{E}_{L|D}[L|d] \leq L_U + \sum_{i=1}^{N_d} \frac{d-1}{(1-\hat{\beta}_i)^d} \quad (61)$$

Since $L = L_U$ for $d = 1$, the conditional expected value $\mathbb{E}_{L|D}[L|d]$ for $1 \leq d < N$ is lower and upper bounded by

$$L_U \leq \mathbb{E}_{L|D}[L|d] \leq L_U + \sum_{i=1}^{N_d} \frac{d-1}{(1-\hat{\beta}_i)^d}$$

meaning that

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}_{L|D}[L|d]}{L_U} = 1 \quad (62)$$

for $d < N$.

The case $d = N$ can be analyzed similarly, except for the fact that L_U does not appear in $\mathbb{E}_{L|D}[L|d = N]$ since the algorithm stops if the other $N - 1$ states have been discarded. Thus,

$$N - 1 \leq \mathbb{E}_{L|D}[L|N] \leq \sum_{i=1}^{N_N} \frac{N-1}{(1-\hat{\beta}_i)^N}$$

which gives

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}_{L|D}[L|N]}{L_U} = 0 \quad (63)$$

Results (62) and (63) can be used to compute the limit

$$\begin{aligned} \lim_{L_U \rightarrow \infty} \frac{\mathbb{E}[L]}{L_U} &= \lim_{L_U \rightarrow \infty} \frac{\sum_{d=1}^N \mathbb{E}_{L|D}[L|d] p_D(d)}{L_U} \\ &= \sum_{d=1}^N \lim_{L_U \rightarrow \infty} p_D(d) \lim_{L_U \rightarrow \infty} \frac{\mathbb{E}_{L|D}[L|d]}{L_U} \\ &= \sum_{d=1}^{N-1} \lim_{L_U \rightarrow \infty} p_D(d) \end{aligned} \quad (64)$$

Using (12), it is not difficult to check that $\lim_{L_U \rightarrow \infty} p_D(d) = \pi_d$, giving the final result and completing the proof.

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}[L]}{L_U} = \sum_{d=1}^{N-1} \pi_d = 1 - \pi_N \quad (65)$$

APPENDIX B MOMENTS OF L IN THE LARGE SAMPLE REGIME

This annex shows the proof of Collorary III-A. First, let's analyze, $\mathbb{E}_{L|D}[L^w|d]$, when $w > 1$ and $d < N$, which can be expressed as

$$\begin{aligned} \mathbb{E}_{L|D}[L^w|d] &= \sum_{l=N-1}^{L_U(N-1)} l^w p_{L|D}(l|d) \\ &= \sum_{a=d-1}^{(d-1)L_U} (a + L_U)^w p_{A|D}(a|d) \\ &= L_U^w + \sum_{i=1}^w L_U^{w-i} \left(\binom{w}{i} \sum_{a=d-1}^{(d-1)L_U} a^i p_{A|D}(a|d) \right) \end{aligned} \quad (66)$$

where the second equality follows from (50), i.e., $L = A + L_U$ if $d < N$, and the third equality is obtained by expanding $(a + L_U)^w$. Following an analogous procedure to the one shown in Appendix A, the following inequality is obtained

$$\sum_{a=d-1}^{(d-1)L_U} a^i p_{A|D}(a|d) \leq q_{i,d} \quad (67)$$

where $q_{i,d}$ is a positive constant. Therefore,

$$\begin{aligned} L_U^w \leq \mathbb{E}_{L|D}[L^w|d] &\leq L_U^w + \sum_{i=1}^w L_U^{w-i} \binom{w}{i} q_{i,d} \\ &= L_U^w + Q_d(L_U) \end{aligned} \quad (68)$$

where Q_d is a polynomial of degree $w - 1$. The case $d = N$ can be computed using $L = A$ instead of $L = A + L_U$, to yield

$$\mathbb{E}_{L|D}[L^w|N] = \sum_{a=d-1}^{(d-1)L_U} a^w p_{A|D}(a|d) \leq q_{w,N} \quad (69)$$

Hence, combining (68) and (69),

$$\lim_{L_U \rightarrow \infty} \frac{\mathbb{E}_{L|D}[L^w|d]}{L_U^w} = \mathbb{1}\{d \neq N\} \quad (70)$$

Using (70), the limit of the ratio $E[L^w]/L_U^w$ is computed as follows

$$\begin{aligned} \lim_{L_U \rightarrow \infty} \frac{\mathbb{E}[L^w]}{L_U^w} &= \lim_{L_U \rightarrow \infty} \frac{\sum_{d=1}^N \mathbb{E}_{L|D}[L^w|d] p_D(d)}{L_U^w} \\ &= \sum_{d=1}^N \lim_{L_U \rightarrow \infty} p_D(d) \lim_{L_U \rightarrow \infty} \frac{\mathbb{E}_{L|D}[L^w|d]}{L_U^w} \\ &= \sum_{d=1}^{N-1} \lim_{L_U \rightarrow \infty} p_D(d) = 1 - \pi_N \end{aligned} \quad (71)$$

where the last equality is obtained since $\lim_{L_U \rightarrow \infty} p_D(d) = \pi_d$.

**APPENDIX C
BOUND OF THE PROBABILITY MASS FUNCTION OF L**

In order to prove Theorem 2, we hereafter derive the probability mass function of L , denoted by $p_L(l)$. According to (50), the total number of required measurements is given by

$$L = \begin{cases} A + L_U & \text{if } 1 \leq d < N \\ A & \text{if } d = N \end{cases} \quad (72)$$

Then, the probability mass function of L can be expressed as

$$p_L(l) = \sum_{d=1}^N p_{L|D}(l|d)p_D(d) \\ = \sum_{d=1}^{N-1} p_{A|D}(l-L_U|d)p_D(d) + p_{A|D}(l|N)p_D(N) \quad (73)$$

As shown in equation (57) of Appendix A, there exist $\alpha, \beta \in \mathbb{R}^+$ with $\beta < 1$ such that $p_{A|D}(a|d) \leq \alpha\beta^a u(a)$, where $u(a) = \mathbb{1}\{a \geq 0\}$. Substituting this inequality in (73), the probability mass function of L can be upper bounded as follows

$$p_L(l) \leq \sum_{d=1}^{N-1} \alpha\beta^{l-L_U} u(l-L_U)p_D(d) + \alpha\beta^l u(l)p_D(N) \\ = (1-p_D(N))\alpha\beta^{l-L_U} u(l-L_U) + p_D(N)\alpha\beta^l u(l). \quad (74)$$

**APPENDIX D
BOUNDS FOR THE PROBABILITY OF ERROR**

This appendix derives upper and lower bounds for the average error probability. The upper bound is obtained using equation (13) from section III – B:

$$p(\mathcal{E}) = \sum_{s=2}^N \pi_s \text{Tr}(\sigma_1 \sigma_s)^{L_U} + \sum_{d=2}^{N-1} \sum_{s=d+1}^N \pi_s \text{Tr}(\sigma_d \sigma_s)^{L_U} \\ \times \prod_{j=1}^{d-1} (1 - \text{Tr}(\sigma_j \sigma_s)^{L_U}) \quad (75)$$

Since $\prod_{j=1}^{d-1} (1 - \text{Tr}(\sigma_j \sigma_s)^{L_U}) \leq 1$, the two sums can be combined to yield an upper bound as

$$p(\mathcal{E}) \leq \sum_{d=1}^{N-1} \sum_{s=d+1}^N \pi_s \text{Tr}(\sigma_d \sigma_s)^{L_U} \quad (76)$$

Now, using $\text{Tr}(\sigma_d \sigma_s) \leq \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)$,

$$p(\mathcal{E}) \leq \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \sum_{d=1}^{N-1} \sum_{s=d+1}^N \pi_s \\ = B_U \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \quad (77)$$

where $B_U = \sum_{d=1}^{N-1} \sum_{s=d+1}^N \pi_s$.

Similarly, for the lower bound, starting with the exact expression for the average probability of error

$$p(\mathcal{E}) = \sum_{s=2}^N \pi_s \text{Tr}(\sigma_1 \sigma_s)^{L_U} + \sum_{d=2}^{N-1} \sum_{s=d+1}^N \pi_s \text{Tr}(\sigma_d \sigma_s)^{L_U}$$

$$\times \prod_{i=1}^{d-1} (1 - \text{Tr}(\sigma_i \sigma_s)^{L_U}) \quad (78)$$

Since all the elements of the sum are positive, the sum is greater or equal than any of its terms, in particular than the term corresponding to $\max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U}$. Also, as $\text{Tr}(\sigma_i \sigma_j) < 1$, or equivalently $1 - \text{Tr}(\sigma_i \sigma_j)^{L_U} > 1 - \text{Tr}(\sigma_i \sigma_j)$, the following bound is obtained

$$p(\mathcal{E}) \geq \left(\pi_j \prod_{l=1}^{i-1} (1 - \text{Tr}(\sigma_l \sigma_j)) \right) \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \quad (79)$$

Therefore,

$$p(\mathcal{E}) \geq B_L \max_{i,j:i \neq j} \text{Tr}(\sigma_i \sigma_j)^{L_U} \quad (80)$$

where $B_L = \pi_j \prod_{l=1}^{i-1} (1 - \text{Tr}(\sigma_l \sigma_j))$.

**APPENDIX E
COMPUTATION OF POSTERIOR PROBABILITIES**

This annex shows the procedure followed to compute the posterior probabilities, $p_{S|X^l}(s|x^l)$. To do this, we rewrite $p_{S|X^l}(s|x^l)$ in terms of $p_{S|X^{l-1}}(s|x^{l-1})$, thus obtaining a recursive expression.

$$p_{S|X^l}(s|x^l) = \frac{p_{S,X^{l-1},X_l}(s,x^{l-1},x_l) p_{X^{l-1}}(x^{l-1})}{p_{X^{l-1}}(x^{l-1}) p_{X^l}(x^l)} \\ = \frac{p_{S,X_l|X^{l-1}}(s,x_l|x^{l-1})}{p_{X_l|X^{l-1}}(x_l|x^{l-1})} \\ = \frac{p_{X_l|X^{l-1},S}(x_l|x^{l-1},s) p_{S|X^{l-1}}(s|x^{l-1})}{\sum_{s=1}^N p_{X_l|X^{l-1},S}(x_l|x^{l-1},s) p_{S|X^{l-1}}(s|x^{l-1})} \\ = \frac{\text{Tr}(\Lambda_{x_l}(x^{l-1}) \sigma_s) p_{S|X^{l-1}}(s|x^{l-1})}{\sum_{s=1}^N \text{Tr}(\Lambda_{x_l}(x^{l-1}) \sigma_s) p_{S|X^{l-1}}(s|x^{l-1})} \quad (81)$$

where the first and second equalities follow from the definition of conditional probability, the third equality from the law of total probability, and the last step follows from substituting probabilities $p_{X_l|X^{l-1},S}(x_l|x^{l-1},s)$ by their corresponding values, i.e, $\text{Tr}(\Lambda_{x_l}(x^{l-1}) \sigma_s)$. Matrix $\Lambda_{x_l}(x^{l-1})$ denotes the x_l element of POVM $\mathcal{L}(x^{l-1})$, which depends on the previous outcomes x^{l-1} .

Using this recursive expression together with the rules defined by the implemented algorithm to decide the POVMs $\mathcal{L}(x^l) \in \{\mathcal{L}^{(k)}\}_{k=1}^N$, the posterior probabilities $p_{S|X^l}(s|x^l)$ can be computed.

**APPENDIX F
UPPER BOUND ON THE EXPECTED NUMBER OF STAGES FOR ZERO ERROR PROBABILITY**

This section shows that the average number of sequential discarding stages $\mathbb{E}[C_e]$ required to attain zero error probability is bounded. Note that we assume that $L_{U,min} \leq L_{U,i} \leq L_{U,max} \forall i$.

Denoting by \bar{N}_c the number of discarded states after the execution of the stages from 1 to c , we have:

$$\bar{N}_c = \sum_{i=1}^c \bar{n}_i = \bar{N}_{c-1} + \bar{n}_c \tag{82}$$

where \bar{n}_c denotes the number of discarded states at stage c and $\bar{N}_0 := 0$. Note that in general $0 \leq \bar{N}_{c-1} < N - 1$ for $c \leq C_e$, since the method stops when $N - 1$ states are discarded. Let us assume without loss of generality that $\bar{N}_{c-1} = 0$, that is, no states were discarded in the previous stages, and that the state with the lowest probability after execution of state $c - 1$ is $s = 1$, i.e., $1 = \operatorname{argmin}_s p_{S|X^{L_{c-1}}}(s|x^{L_{c-1}})$, where $x^{L_{c-1}}$ is the outcome from all measurements performed at stages up to $c - 1$. Note that this implies

$$p_{S|X^{L_{c-1}}}(1|x^{L_{c-1}}) \leq \frac{1}{2} \tag{83}$$

since the rest of the states must accumulate more probability. Then, the probability that at least one state is discarded at stage c can be expressed as

$$p(\bar{n}_c > 0) = p(d_c \neq 1) \tag{84}$$

where d_c indicates the decision taken at stage c . The equality follows since if a decision $d_c \neq 1$ is taken, then at least state 1 is discarded. The probability $p(\bar{n}_c > 0)$, taking into account all measurements completed before stage c , can then be written as

$$p(\bar{n}_c > 0) = 1 - \sum_{s=1}^N p_{D_c|S}(1|s) p_{S|X^{L_{c-1}}}(s|x^{L_{c-1}}) \tag{85}$$

Now, using $p_{D_c|S}(1|s) = \operatorname{Tr}(\sigma_1 \sigma_s)^{L_{U,c}}$ (85) becomes

$$1 - p_{S|X^{L_{c-1}}}(1|x^{L_{c-1}}) - \sum_{s=2}^N \operatorname{Tr}(\sigma_1 \sigma_s)^{L_{U,c}} p_{S|X^{L_{c-1}}}(s|x^{L_{c-1}}) \tag{86}$$

This expression is lower bounded by

$$1 - p_{S|X^{L_{c-1}}}(1|x^{L_{c-1}}) - \max_{j \neq 1} \operatorname{Tr}(\sigma_1 \sigma_j)^{L_{U,c}} (1 - p_{S|X^{L_{c-1}}}(1|x^{L_{c-1}})) = (1 - p_{S|X^{L_{c-1}}}(1|x^{L_{c-1}})) (1 - \max_{j \neq 1} \operatorname{Tr}(\sigma_1 \sigma_j)^{L_{U,c}}) \tag{87}$$

Now, combining equation (87) with (83) we can set the following lower bound,

$$p(\bar{n}_c > 0) \geq \frac{1}{2} \left(1 - \max_{j \neq 1} \operatorname{Tr}(\sigma_1 \sigma_j)^{L_{U,min}} \right) \geq \frac{1}{2} \left(1 - \max_{j,i; j \neq i} \operatorname{Tr}(\sigma_i \sigma_j)^{L_{U,min}} \right) \tag{88}$$

Parameter $p_+ := \frac{1}{2} (1 - \max_{j,i; j \neq i} \operatorname{Tr}(\sigma_i \sigma_j)^{L_{U,min}})$, is defined as this lower bound. Now,

$$\mathbb{E}[\bar{n}_c] \geq p(\bar{n}_c > 0) \geq p_+ \tag{89}$$

Let us now define $\hat{N}_c = \hat{N}_{c-1} + V$, where $\hat{N}_0 := 0$, and $V \sim \operatorname{Bern}(p_+)$. Clearly, r.v. \hat{N}_c needs on average more steps than \bar{N}_c to reach the value $N - 1$. Therefore, the random variable $Z \sim p_Z(z)$, where $p_Z(z)$ is the probability that $\hat{N}_z = N - 1$ and $\hat{N}_{z-1} = N - 2$, fulfills

$$\mathbb{E}[C_e] \leq \mathbb{E}[Z] \tag{90}$$

To conclude the proof we only have to compute $\mathbb{E}[Z]$. Since $\hat{N}_c \sim \operatorname{B}(c, p_+)$ ⁴ for a fixed value of c , the probability $p_Z(z)$ is given by

$$p_Z(z) = \binom{z-1}{N-2} p_+^{N-2} (1-p_+)^{z-N+1} p_+ \tag{91}$$

where the first term of the product is $p(\hat{N}_{z-1} = N - 2)$, and the second $p(v = 1) = p_+$. Hence, the expected value of Z is

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{z=N-1}^{\infty} z p_Z(z) \\ &= \sum_{z=N-1}^{\infty} z \binom{z-1}{N-2} p_+^{N-1} (1-p_+)^{z-N+1} \end{aligned} \tag{92}$$

Substituting the binomial coefficient $\binom{z-1}{N-2}$ by $\frac{(z-1)!}{(N-2)!(z-N+1)!}$ and arranging terms, the previous expression becomes

$$\frac{p_+^{N-1}}{(N-2)!} \sum_{z=N-1}^{\infty} \prod_{j=0}^{N-2} (z-j) (1-p_+)^{z-N+1} \tag{93}$$

The sum can be rewritten as a derivative of the geometric series, $\sum_{z=0}^{\infty} (1-p_+)^z$, this is

$$\frac{p_+^{N-1}}{(N-2)!} (-1)^{N-1} \frac{d^{N-1}}{dp_+^{N-1}} \sum_{z=0}^{\infty} (1-p_+)^z \tag{94}$$

Finally, since $\sum_{z=0}^{\infty} (1-p_+)^z = \frac{1}{p_+}$ the expected value of Z can be expressed as

$$\mathbb{E}[Z] = \frac{N-1}{p_+} \tag{95}$$

Therefore, $\mathbb{E}[C_e]$ is bounded $\mathbb{E}[C_e] \leq \frac{N-1}{p_+}$.

REFERENCES

- [1] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Phys. Rev. Lett.*, vol. 108, May 2012, Art. no. 200501.
- [2] A. Acín, J. Bae, E. Bagan, M. Baig, L. Masanes, and R. Muñoz-Tapia, "Secrecy properties of quantum channels," *Phys. Rev. A, Gen. Phys.*, vol. 73, no. 1, Jan. 2006, Art. no. 012327.
- [3] C. W. Helstrom, *Quantum Detection and Estimation*. New York, NY, USA: Academic, 1976.
- [4] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*. Pisa, Italy: Edizioni della Normale, 1979.
- [5] K. M. R. Audenaert, "Discriminating states: The quantum Chernoff bound," *Phys. Rev. Lett.*, vol. 98, Apr. 2007, Art. no. 160501.
- [6] M. Nussbaum and A. Szkola, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Statist.*, vol. 37, no. 2, pp. 1040-1057, Apr. 2009.

⁴ $\operatorname{B}(n, p)$ denotes a binomial distribution with parameters $n \in \mathbb{N}$, and $p \in (0, 1)$.

- [7] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, 1952.
- [8] K. Li, "Discriminating quantum states: The multiple chernoff distance," *Ann. Statist.*, vol. 44, no. 4, pp. 1661–1679, 2016.
- [9] A. Acín, E. Bagan, M. Baig, L. Masanes, and R. M. Tapia, "Multiple-copy two-state discrimination with individual measurements," *Phys. Rev. A, Gen. Phys.*, vol. 71, Mar. 2005, Art. no. 032338.
- [10] J. Calsamiglia, J. I. de Vicente, R. Muñoz-Tapia, and E. Bagan, "Local discrimination of mixed states," *Phys. Rev. Lett.*, vol. 105, no. 8, Aug. 2010, Art. no. 080504.
- [11] E. Martínez Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz-Tapia, and J. Calsamiglia, "Quantum sequential hypothesis testing," *Phys. Rev. Lett.*, vol. 126, no. 18, May 2021, Art. no. 180502.
- [12] Y. Li, F. Tan, and M. Tomamichel, "Optimal adaptive strategies for sequential quantum hypothesis testing," 2021, *arXiv:2104.14706*.
- [13] H. Nagaoka, "The converse part of the theorem for quantum Hoeffding bound," 2006, *arXiv:quant-ph/0611289*.
- [14] T. Ogawa and M. Hayashi, "On error exponents in quantum hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1368–1372, Jun. 2004.
- [15] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acín, and E. Bagan, "Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states," *Phys. Rev. A, Gen. Phys.*, vol. 77, no. 3, Mar. 2008, Art. no. 032311.
- [16] K. M. R. Audenaert and M. Mosonyi, "Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination," *J. Math. Phys.*, vol. 55, no. 10, Oct. 2014, Art. no. 102201.
- [17] I. D. Ivanovic, "How to differentiate between non-orthogonal states," *Phys. Rev. A, Gen. Phys.*, vol. 123, no. 6, pp. 257–259, Jan. 1987.
- [18] D. Dieks, "Overlap and distinguishability of quantum states," *Phys. Lett. A*, vol. 126, nos. 5–6, pp. 303–306, Jan. 1988.
- [19] A. Peres, "How to differentiate between non-orthogonal states," *Phys. Lett. A*, vol. 128, nos. 1–2, p. 19, Mar. 1988.



ALBA PAGÈS-ZAMORA (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Universitat Politècnica de Catalunya (UPC), Spain, in 1992 and 1996, respectively.

In 1992, she joined the Department of Signal Theory and Communications, UPC, and became an Associate Professor, in 2001. Since April 2019, she has been the Vice-Dean of Academic Planning and Quality of the Electrical Engineering School ETSETB, UPC. She has coauthored one patent, two book chapters, 20 articles in international periodic journals, and about 60 papers in international conferences. She has been UPC's principal investigator of five national and European research projects, and has participated in 15 more. Her current research interests include signal processing on graphs, ensemble learning, and quantum signal processing.

Dr. Pagès-Zamora is a member of the IEEE Signal Processing Theory and Methods Technical Committee. Since February 2018, she has been a member of the Editorial Board of *EURASIP Signal Processing* journal. From February 2012 to 2015, she served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. She regularly participates as an evaluator of research projects for the European Commission and for national and international agencies.



JAVIER RODRÍGUEZ FONOLLOSA (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, USA, in 1992.

In 1993, he joined the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), where he became an Associate Professor, in 1996, a Professor, in 2003, and the Department Head, from 2006 to 2010. He was the Head of the Signal Theory and Communications Department, UPC, from October 2006 to January 2010. In 1995, he led UPC's participation in the European Commission funded ACTS Mobile projects TSUNAMI (II) and SUNBEAM that included the analysis of adaptive antennas in 2nd and 3rd generation cellular mobile communication systems. From January 2000 to 2003, he was a Technical and a Project Co-ordinator of the IST projects METRA and I-METRA dedicated to the introduction of multi-antenna terminals in UMTS and Systems beyond 3G. From January 2006 to December 2008, he co-ordinated the Sixth Framework Program IST project SURFACE which evaluated the performance of a generalized air interface with self-configuration capabilities. In November 2006, he initiated the co-ordination of the 5 year Type C-Consolider project Fundamental bounds in Network Information Theory of the National Research Plan of Spain. From December 2008 to December 2014, he was the Co-ordinator of the CONSOLIDER-INGENIO 2010 Foundations and Methodologies for Future Communication and Sensor Networks (COMONSENS), a six year € 3.5 Million effort of 135 researchers belonging to ten universities and research centers in Spain. The project continued under his co-ordination as a research network Red COMONSENS, since December 2015, for seven more years. In 2009, he co-ordinated the ERASMUS MUNDUS 2009–2013 Master of Science in Research on Information and Communication Technologies (MERIT) EMMC Joint Master Program. From February 2010 to July 2014, he was a manager of the communications and electronic technologies (TEC) area of the National Research Plan of Spain.

Dr. Fonollosa has been a member of the Editorial Board of the *EURASIP Signal Processing* journal, since May 2005. In June 1995 and September 2001, he was the Co-Chairperson and an Organizer of the IEEE Signal Processing/ATHOS Workshop on Higher-Order Statistics held in Begur, Girona, Spain, and of the IST Mobile Communications Summit 2001 held in Sitges, Barcelona, Spain. He was elected as a member of the Signal Processing for Communications (SPCOM) Technical Committee of the IEEE Signal Processing Society, in February 1998 and January 1999, respectively.

• • •



JORDI PÉREZ-GUIJARRO received the B.S. degree in telecommunications engineering from Miguel Hernández University, Elche, Spain, in 2018, and the M.S. degree in telecommunications engineering from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 2020, where he is currently pursuing the Ph.D. degree.