Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels

Marco Dalai, *Member, IEEE*, Andreas Winter

Abstract

The sphere packing bound, in the form given by Shannon, Gallager and Berlekamp, was recently extended to classical-quantum channels, and it was shown that this creates a natural setting for combining probabilistic approaches with some combinatorial ones such as the Lovász theta function. In this paper, we extend the study to the case of constant-composition codes. We first extend the sphere packing bound for classical-quantum channels to this case, and we then show that the obtained result is related to a variation of the Lovász theta function studied by Marton. We then propose a further extension to the case of varying channels and codewords with a constant conditional composition given a particular sequence. This extension is finally applied to auxiliary channels to deduce a bound which is useful in the low rate region and which can be interpreted as an extension of the Elias bound.

I. Introduction

In point to point communication, the study of the performance achievable, in terms of probability of error, by optimal coding schemes at increasing block-length has played a relevant role during many decades. One the central results in this context, the so called sphere packing bound, has been recently extended from classical to classical-quantum channels [2], [3, Sec. V] by resorting to the first rigorous proof given for the case of classical discrete memoryless channels (DMC) by Shannon, Gallager and Berlekamp [4]. That resulted in an upper bound on the reliability function of classical-quantum channels, which is the error exponent achievable by means of optimal codes. For the classical case, the proof given in [4] can be considered a rigorous completion of Fano's first efforts toward proving the bound [5, Ch. 9]. However, while Fano's approach led to an asympyotically tight exponent at high rates for codes with specific constant compositions, the proof in [4] focuses on bounding the performance for the best possible code, without composition constraints.

While sphere packing bounds in the forms given in [4] and [3] have the merit of controlling the best possible performance which can be obtained when no constraints on the codes are imposed, they nevertheless leave unanswered the important question of what is achievable when one does impose constraints on the codes. For example,

M. Dalai is with the Department of Information Engineering, University of Brescia, Italy, email: marco.dalai@unibs.it.

A. Winter is with ICREA & Física Teòrica: Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, Spain, email: andreas.winter@uab.cat

Part of the results were first presented in [1].

under the assumption that each symbol used at the encoder has an associated cost, one might ask what performance can be achieved by using codewords which do not exceed an average cost per symbol. In this setting, one notices that it becomes important to derive bounds on the achievable error exponents when only codewords with a specific composition (i.e., empirical symbol frequencies) are allowed.

Shortly after the appearance of [4], Haroutunian [6], [7], proposed a simple yet rigorous proof of the sphere packing bound for classical channels which gives the tight exponent for codes with general (possibly non optimal) constant composition, thus recovering both the rigorous results of [4] and the practically important cases considered in [5]. Furthermore, the bound presented by Haroutunian has the clear advantage of being directly derived using very intuitive arguments expressed in terms of Kullback-Leibler divergence and mutual information, thus re-using tools which already make their appearance when studying channel capacity. However, despite the great advantages offered in the classical case, Haroutunian's proof does not lead to a good bound when greedily extended to classical-quantum channels (see [8, Th. II.20 and page 35]). This motivated the choice made in [2], [3] to follow the approach of [4].

In this paper, we modify the approach in [2], [3] to derive a sphere packing bound for classical-quantum channels with constant-composition codes. In a sense, our results show how the path taken by Fano [5], Shannon, Gallager and Berlekamp [4] can be used not only to derive a quantum version of the sphere packing bound, but to actually derive the tight form even for constant-composition codes with arbitrary composition. One interpretation of the presented results, which is better clarified by what will be discussed in Appendix C, is that the Rényi divergence really represents the fundamental tool which allows one to deal with relatively general cases, that is, both classical and classical-quantum channels, both optimal and arbitrary codeword compositions. In fact, the main difference in the bound derived in our paper with respect to the classical counterpart is in the resulting possible analytical expressions, which do not seem to be expressible, in our case, in terms of Kullback-Leibler divergence and mutual information.

In analogy with the results obtained in [9] [3, Sec. VI], we then discuss the connections of the constant-composition version of the bound with a quantity introduced by Marton [10] as a generalization of the Lovász theta function for bounding the highest rate achievable by zero-error codes with codewords of a given arbitrary asymptotic composition. These results complete the discussion initiated in [3] of how sphere packing bounds for classical-quantum channels relate to some of the most important bounds on the zero error capacity of channels (or graph capacities), showing that the connection is indeed rooted in a common underlying structure.

Finally, we propose an extension of the sphere packing bound for varying channels and codewords with a constant *conditional* composition given a sequence of "channel states". We then apply this bound to auxiliary channels to bound the reliability in the low rate region, showing that this result includes as a special case a recently developed generalization of the Elias bound [11].

II. Definitions

Consider a classical-quantum channel $\mathfrak{C}$ with finite input alphabet $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$ and associated density operators S_x, $x \in \mathcal{X}$, in a finite-dimensional Hilbert space $\mathcal{H}$. The n-fold product channel acts in the tensor-

product space $\boldsymbol{\mathcal{H}} = \mathcal{H}^{\otimes n}$ of n copies of $\mathcal{H}$. To a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ we associate the signal state $\boldsymbol{S}_{\boldsymbol{x}} = S_{x_1} \otimes S_{x_2} \cdots \otimes S_{x_n}$. A block code with M codewords is a mapping from a set of M messages $\{1, \ldots, M\}$ into a set of M codewords $\boldsymbol{\mathcal{C}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ and the rate of the code is $R = (\log M)/n$. The logarithm here and in the rest of the paper is to the base e.

We consider a quantum decision scheme for such a code, which is a Positive-Operator Valued Measure (POVM) composed of a collection of M positive operators $\{\Pi_1, \Pi_2, \ldots, \Pi_M\}$ such that $\sum \Pi_m = \mathbb{1}$, where $\mathbb{1}$ is the identity operator. The probability that message m' is decoded when message m is transmitted is $\mathsf{P}_{m'|m} = \operatorname{Tr} \Pi_{m'} \boldsymbol{S}_{\boldsymbol{x}_m}$ and the probability of error after sending message m is

$$\mathsf{P}_{\mathrm{e}|m} = 1 - \operatorname{Tr} \left(\Pi_m \boldsymbol{S}_{\boldsymbol{x}_m} \right).$$

The maximum error probability of the code is defined as the largest $\mathsf{P}_{\mathrm{e}|m}$, that is,

$$\mathsf{P}_{\text{e,max}} = \max_m \mathsf{P}_{\mathrm{e}|m}.$$

In this paper, we are interested in bounding the probability of error for codes with constant composition (or *type*, see [7]). Using the notation of [7], we define T_P^n to be the set of sequences of length n which contain any symbol x exactly $nP(x)$ times. We define $\mathcal{T}^n = \{P : \mathsf{T}_P^n \neq \emptyset\}$ and we call a distribution $P \in \mathcal{T}^n$ a *composition* for sequences of length n. A constant-composition code is a code such that $\boldsymbol{\mathcal{C}} \subseteq \mathsf{T}_P^n$ for some composition $P \in \mathcal{T}^n$.

Given a composition P_n, we define $\mathsf{P}_{\text{e,max}}^{(n)}(R, P_n)$ to be the smallest maximum error probability among all codes of length n, rate *at least* R, and composition P_n. For a probability distribution P, we define the asymptotic optimal error exponent with distribution P as

$$E(R, P) = \sup \left\{ \limsup_{n \to \infty} -\frac{1}{n} \log \mathsf{P}_{\text{e,max}}^{(n)}(R_n, P_n) \right\}, \tag{1}$$

where the outer supremum is over all sequences of codes with rates R_n and compositions P_n such that $R_n \to R$ and $P_n \to P$ as $n \to \infty$. For channels with a positive zero-error capacity C_0 [12], distributions P_n exist which make $\mathsf{P}_{\text{e,max}}^{(n)}(R_n, P_n) = 0$ for positive, small enough rates R_n. Thus, assuming $-\log 0 = \infty$ in (1), for these channels $E(R, P)$ is infinite for some distributions P and positive rates R. We define the quantity

$$C_0(P) = \sup\{R : E(R, P) = \infty\}, \tag{2}$$

which we can call the zero-error capacity of the channel relative to P. We observe that $C_0(P)$ could be equivalently defined as

$$C_0(P) = \sup \left\{ \limsup_{n \to \infty} R_n \right\} \tag{3}$$

where the outer supremum is over all sequences of codes with rates R_n and compositions P_n such that $\mathsf{P}_{\text{e,max}}^{(n)}(R_n, P_n) = 0$ and $P_n \to P$ as $n \to \infty$ (that is, $C_0(P)$ is the highest rate asymptotically achievable by zero-error codes with compositions tending to P). Indeed, it is clear that the right-hand side of (3) is not larger than the right-hand side of (2). To prove the opposite, we observe that there exists a constant D which only depends on the channel such that if $\mathsf{P}_{\text{e,max}}^{(n)}(R_n, P_n) > 0$ then $\mathsf{P}_{\text{e,max}}^{(n)}(R_n, P_n) \geq e^{-n(D + o(1))}$. In particular, any code with $\mathsf{P}_{\text{e,max}} > 0$ (for an optimal POVM) contains two codewords which are confusable, and $\mathsf{P}_{\text{e,max}}$ is lower bounded by the smallest

possible probability of error in the binary hypothesis test between those two codewords (see [13], [14] and [3, Th. 4 and Th. 12])

$$P_{e,\max} \geq e^{-n(D+o(1))},$$

where D is the largest possible Chernoff distance between non-orthogonal input states

$$D = \max_{\substack{x' \neq x \\ \operatorname{Tr}(S_x S_{x'}) \neq 0}} \left(-\log \min_{0 \leq s \leq 1} \operatorname{Tr} S_x^{1-s} S_{x'}^s \right)$$

(note that $\mathcal{X}$ is finite in our setting. For more general channels this statement would not hold). This implies that for all rates R larger than the right-hand side of (3) $E(R,P)$ is upper bounded by D.

It is important to observe that, as for the zero-error capacity C_0 (see [15], [3]), the value $C_0(P)$ only depends on the so called *confusability graph* G of the channel, defined as a graph with vertex set $\mathcal{X}$ and edges $(x,x') \in \mathcal{X}^2$ for which $\operatorname{Tr} S_x S_{x'} > 0$. In fact, if a code satisfies $\mathsf{P}_{\text{e,max}} = 0$, then for each $m \neq m'$ we must have $\operatorname{Tr}(\Pi_m \boldsymbol{S}_{\boldsymbol{x}_m}) = 1$ and $\operatorname{Tr}(\Pi_m \boldsymbol{S}_{\boldsymbol{x}_{m'}}) = 0$. This is possible if and only if the signals $\boldsymbol{S}_{\boldsymbol{x}_m}$ and $\boldsymbol{S}_{\boldsymbol{x}_{m'}}$ are orthogonal, that is $\operatorname{Tr}(\boldsymbol{S}_{\boldsymbol{x}_m} \boldsymbol{S}_{\boldsymbol{x}_{m'}}) = 0$. Using the property that $\operatorname{Tr}((A \otimes B)(C \otimes D)) = \operatorname{Tr}(AC) \operatorname{Tr}(BD)$, we then have

$$\operatorname{Tr}(\boldsymbol{S}_{\boldsymbol{x}_m} \boldsymbol{S}_{\boldsymbol{x}_{m'}}) = \prod_{i=1}^{n} \operatorname{Tr}(S_{x_{m,i}} S_{x_{m',i}}).$$

This implies that $\operatorname{Tr}(S_{x_{m,i}} S_{x_{m',i}}) = 0$ for at least one value of i, which means that $x_{m,i}$ and $x_{m',i}$ are perfectly distinguishable. So, a code has probability of error precisely equal to zero if and only if for all m and m' we have $\operatorname{Tr}(S_{x_{m,i}} S_{x_{m',i}}) = 0$ for at least one i, which implies that $C_0(P)$ only depends on the confusability graph G defined as above. Furthermore, for any graph G, we can always find a channel with confusability graph G. Thus, we may equivalently refer to $C_0(P)$ as the zero-error capacity relative to P of a channel or to $C(G,P)$ as the capacity relative to P of a graph G [16], [10].

To avoid unnecessary complications, we use a flexible notation in this paper. We keep it simple as far as possible, progressively increasing its complexity by adding arguments to functions as their definitions become more general. The meaning of all quantities should be clear from the context.

III. Sphere Packing Bound for Constant-Composition Codes

Our main result is the following theorem.

Theorem 1: For all positive rates R, distributions P, and strictly positive $\varepsilon < R$, we have the bound

$$E(R,P) \leq E_{\text{sp}}^{\text{cc}}(R-\varepsilon, P),$$

where $E_{\text{sp}}^{\text{cc}}(R,P)$ is defined by the relations

$$E_{\text{sp}}^{\text{cc}}(R,P) = \sup_{\rho > 0} \left[E_0^{\text{cc}}(\rho, P) - \rho R \right], \tag{4}$$

$$E_0^{\text{cc}}(\rho,P) = \min_{F} \left[-(1+\rho) \sum_x P(x) \log \operatorname{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \right]. \tag{5}$$

the minimum being over all density operators F.

Proof: See Appendix A. See Lemma 12 in Appendix D for the existence of the minimum over F. ∎

Remark 2: The quantity $E_{\text{sp}}^{\text{cc}}(R,P)$ is convex function of R, being the supremum of a set of linear functions. Thus, it is continuous on any open interval over which it is finite. Hence, the role of $\varepsilon > 0$ is not really important there, and it only affects the statement of the bound at the smallest point where $E_{\text{sp}}^{\text{cc}}(R,P)$ is finite, called $R_\infty(P)$ later on, since $E(R,P)$ might be infinite there if $C_0 = R_\infty(P)$ (see Theorem 6 below for the boundedness of $E_{\text{sp}}^{\text{cc}}(R,P)$ at this point).

The bound is written here in terms of Rényi divergences. For commuting states, that is, classical channels, the bound can be written in the more usual form in terms of Kullback-Leibler divergences and mutual information as in [7]. In fact, assuming that the states S_x commute, let for notational convenience $W(y|x)$ be their eigenvalues, which we interpret as classical probability distributions, indexing in y the output space. Then we can write (see [7, Sec. II.5, Prob. 23])

$$E_0^{\text{cc}}(\rho,P) = \min_F \left[-(1+\rho) \sum_x P(x) \log \operatorname{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \right] \tag{6}$$

$$= \min_Q \left[-(1+\rho) \sum_x P(x) \log \sum_y W(y|x)^{\frac{1}{1+\rho}} Q(y)^{\frac{\rho}{1+\rho}}) \right] \tag{7}$$

$$= \min_{V,Q} \sum_{x,y} P(x)V(y|x) \left[\log \frac{V(y|x)}{W(y|x)} + \rho \log \frac{V(y|x)}{Q(y)} \right] \tag{8}$$

$$= \min_V \left[D(V\|W|P) + \rho I(P,V) \right], \tag{9}$$

where the $V(\cdot|x)$ and Q run over probability distributions on y, $I(P,V)$ is the mutual information with the notation of [7]

$$I(P,V) = \sum_{x,y} P(x)V(y|x) \log \frac{V(y|x)}{\sum_{x'} P(x')V(y|x')}, \tag{10}$$

and $D(V\|W|P)$ is the conditional information divergence

$$D(V\|W|P) = \sum_x P(x) \sum_y V(y|x) \log \frac{V(y|x)}{W(y|x)}. \tag{11}$$

Hence, for classical channels, we have the more familiar form of the bound (see [7])

$$E_{\text{sp}}^{\text{cc}}(R,P) = \sup_{\rho>0} \left[\min_V \left(D(V\|W|P) + \rho I(P,V) \right) - \rho R \right] \tag{12}$$

$$= \min_{V: I(P,V) \le R} D(V\|W|P). \tag{13}$$

This form of the bound emerges naturally in Haroutunian's proof [6], [7], which is very simple and gives a very intuitive interpretation of the resulting expression. For a given rate R, one considers auxiliary channels V such that $I(P,V) < R$. Given codes with rate R and composition P, by the strong converse to the coding theorem, the probability of error over channel V for at least one codeword is nearly one. For that same codeword, the probability of error over channel W can be lower bounded in terms of the Kullback-Leibler divergence $D(V\|W|P)$, and this leads to the sphere packing bound.

It is interesting to consider what happens in the case of non-commuting states. A reasoning similar to the one described in the last paragraph can be applied to derive a bound which is the formal analog of the classical one in the form given using equation (13), namely (see [8, Th. II.20])

$$E(R,P) \leq \min_{V:I(P,V)\leq R} D(V\|S|P) \tag{14}$$

where now the minimum is over all sets of density operators $\{V_x\}_{x\in\mathcal{X}}$,

$$I(P,V) = H\left(\sum_x P(x)V_x\right) - \sum_x P(x)H(V_x), \quad \text{with } H(\rho) = -\operatorname{Tr}\rho\log\rho, \tag{15}$$

and

$$D(V\|S|P) = \sum_x P(x)\operatorname{Tr}V_x(\log V_x - \log S_x). \tag{16}$$

The main difference with respect to the classical case, however, is that this bound does not have good properties in the more general classical-quantum setting. For example, note that - as in the classical case - the bound is finite only when the V_x can be chosen so that $\operatorname{supp}(V_x) \subseteq \operatorname{supp}(S_x)$. As a consequence, for pure-state channels, that is when all operators S_x have rank one, the bound is infinite for rates $R < I(P,S)$, which means that the bound is essentially trivial in this case. The reason for this unexpected behavior can be traced back to a fundamental difference in the study of error exponents in the classical and quantum binary hypothesis testing (see for example [14, Sec. 4.8]). A more detailed discussion of this issue requires an inspection of the proof of the sphere packing bound and is thus deferred to Appendix C.

Now it is not difficult to show that after optimization of the composition we recover the original bound of [2], [3]. In order to do this, note that

$$\max_P E_{\text{sp}}^{\text{cc}}(R,P) = \sup_{\rho>0}\left[\max_P E_0^{\text{cc}}(\rho,P) - \rho R\right].$$

Here, the existence of the maximum over P can be motivated by noticing that the minimum over F in (5) can be replaced by an infimum over invertible density operators. Then, $E_0^{\text{cc}}(\rho,P)$ as the infimum of continuous functions in P is upper semicontinuous in P and hence achieves its supremum on the probability simplex, as the latter is compact (see [17] for similar considerations in the classical case).

Then,

$$\begin{aligned}
&\max_P E_0^{\text{cc}}(\rho,P) \\
&\quad = \max_P \min_F \left[-(1+\rho)\sum_x P(x)\log\operatorname{Tr}(S_x^{\frac{1}{1+\rho}}F^{\frac{\rho}{1+\rho}})\right]. \\
&\quad = \min_F \max_P \left[-(1+\rho)\sum_x P(x)\log\operatorname{Tr}(S_x^{\frac{1}{1+\rho}}F^{\frac{\rho}{1+\rho}})\right] \\
&\quad = \min_F \max_x \left[-(1+\rho)\log\operatorname{Tr}(S_x^{\frac{1}{1+\rho}}F^{\frac{\rho}{1+\rho}})\right],
\end{aligned}$$

where the minimum and the maximum can be exchanged by invoking an appropriate form of the minimax theorem (see the proof of [17, Prop. 1], with the obvious modifications for density operators, for a detailed discussion of the technicalities. A similar reasoning is also used in the proof of Theorem 5). The resulting expression is in fact

the coefficient $E_0(\rho)$ which defines the sphere packing bound as proved in [3, Th. 6]. Hence, this procedure allows us to recover the results of [2], [3] by noticing that[1]

$$E(R) = \sup_P E(R,P) \tag{17}$$

$$\leq \sup_P E_{\text{sp}}^{\text{cc}}(R-\varepsilon, P) \tag{18}$$

$$= E_{\text{sp}}(R-\varepsilon). \tag{19}$$

Theorem 1 constitutes thus the most general form of the sphere packing bound, from which all other forms can be derived.

Remark 3: The version of the bound derived in [2] and [3] is provably tight in the high rate regime for pure-state channels, in the sense that it matches the achievability result derived in [18], while there is no evidence of tightness for general non classical mixed-state channels, since a matching achievability has not been derived yet.

To the best of our knowledge, the bound of Theorem 1 has no matching achievability counterpart even for pure-state channels. We observe that it is not easy to envision a modification of the proof derived in [18] to obtain a matching achievability for constant-composition codes, since the bound derived in [18] uses in a substantial way the properties of random codes generated with i.i.d. symbols.

IV. Connections with Marton's function

The bound $E_{\text{sp}}^{\text{cc}}(R,P)$ obtained in the previous section can be used as an upper bound for the zero-error capacity of the channel relative to P. Whenever the function $E_{\text{sp}}^{\text{cc}}(R-\varepsilon,P)$ is finite, in fact, then the probability of error at rate R is non-zero. We focus then on the values of R for which $E_{\text{sp}}^{\text{cc}}(R,P)$ is finite. Since the function $E_{\text{sp}}^{\text{cc}}(R,P)$ is the upper envelope of the family of lines $E_0^{\text{cc}}(\rho,P)-\rho R$, $\rho \geq 0$, it is not difficult to observe that the infimum (actually the minimum, see Theorem 6) of the rates at which $E_{\text{sp}}^{\text{cc}}(R,P)$ is finite, call it $R_\infty(P)$, can be evaluated as

$$\begin{aligned} R_\infty(P) &= \lim_{\rho\to\infty} \frac{E_0^{\text{cc}}(\rho,P)}{\rho} \\ &= \lim_{\rho\to\infty} \min_F \left[-\frac{1+\rho}{\rho} \sum_x P(x) \log \text{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \right]. \end{aligned}$$

For fixed F, the quantity $\log \text{Tr}(S_x^{1-s}F^s)$ is a non-positive convex function of $s \in (0,1)$, as seen by computing derivatives (see for example [3, eq. (33)]), and hence the quantity

$$-\frac{1+\rho}{\rho} \sum_x P(x) \log \text{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \tag{20}$$

[1]The identity $E(R) = \sup_P E(R,P)$ follows from standard considerations on constant-composition codes [7]. The inequality $E(R,P) \leq E(R)$ is obvious, while the converse follows from the fact that the number of possible compositions is only polynomial in n. This implies that any code with rate R contains a constant-composition sub-code of rate at least $R' = R - O(\log(n)/n)$, and hence $E(R)$ can be approached arbitrarily with constant-composition codes, which implies that $\max_P E(R,P) = E(R)$.

is non increasing in ρ for fixed F. This implies that the limit over ρ and the minimum over F can be interchanged, obtaining

$$\begin{aligned} R_\infty(P) &= \min_F \lim_{\rho\to\infty} \left[-\frac{1+\rho}{\rho} \sum_x P(x) \log \mathrm{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \right] \\ &= \min_F \left[-\sum_x P(x) \log \mathrm{Tr}(S_x^0 F) \right], \end{aligned} \tag{21}$$

where S_x^0 is the projection onto the range of S_x. Thus, since we have the bound $C_0(P) \leq R_\infty(P)$, equation (21) gives us an upper bound on $C_0(P)$. We observe that a corresponding upper bound on C_0, already discussed in [3], can be derived in consideration of the fact that $C_0 = \max_P C_0(P)$ (see footnote 1). We deduce then that $C_0 \leq R_\infty$, where

$$\begin{aligned} R_\infty &= \max_P R_\infty(P) \\ &= \max_P \min_F \left[-\sum_x P(x) \log \mathrm{Tr}(S_x^0 F) \right] \\ &= \min_F \max_P \left[-\sum_x P(x) \log \mathrm{Tr}(S_x^0 F) \right] \\ &= \min_F \max_x \log \frac{1}{\mathrm{Tr}(S_x^0 F)}. \end{aligned}$$

Here, the minimum over F and maximum over P can be again interchanged by an appropriate version of the minimax theorem (see again [17, proof. of Prop. 1]).

It was observed in [9] and [3, Sec. VI] that R_∞ is related to the Lovász number ϑ [19]. Here, we observe that, in complete analogy, the value $R_\infty(P)$ is related to a variation of the ϑ function introduced by Marton in [10] as an upper bound to $C(G, P)$. Given a (confusability) graph G, Marton introduces the following quantity[2]:

$$\vartheta(G, P) = \min_{\{u_x\}, f} \sum_x P(x) \log \frac{1}{|\langle u_x | f \rangle|^2}, \tag{22}$$

where the minimum is over all sets of unit norm vectors $\{u_x\}_{x\in\mathcal{X}}$ in some Hilbert space such that $\langle u_x | u_{x'} \rangle = 0$ whenever x and x' are not confusable (usually called *representations* of G), and over all unit norm vectors f in the same Hilbert space. She then shows that $C(G, P) \leq \vartheta(G, P)$.

Let us now compare this bound with the best bound on $C(G, P)$ that we can deduce from the sphere packing bound using $R_\infty(P)$. We enforce the notation writing $R_\infty(\{S_x\}, P)$ to point out the dependence of $R_\infty(P)$ on the channel states S_x. For a given confusability graph G, the best upper bound to $C(G, P)$ is obtained by minimizing $R_\infty(\{S_x\}, P)$ over all possible channels with confusability graph G. We may then define

$$\vartheta_{\mathrm{sp}}(G, P) = \inf_{\{S_x\}} R_\infty(\{S_x\}, P) \tag{23}$$

$$= \inf_{\{U_x\}, F} \sum_x P(x) \log \frac{1}{\mathrm{Tr}(U_x F)}, \tag{24}$$

[2]We use the notation $\vartheta(G, P)$ in place of Marton's $\lambda(G, P)$ to preserve a higher coherence with the context of this paper. For the same reason, in what follows we also use, as in [3], a logarithmic version of the ordinary Lovász ϑ function, that is, our ϑ corresponds to $\log \vartheta$ in Lovász' notation.

where $\{U_x\}$ now runs over all sets of projectors in some Hilbert space such that $\operatorname{Tr} U_x U_{x'} = 0$ if x and x' are not adjacent in G, which we might consider a more general form of *representation* for G. Then we have the bound $C(G,P) \leq \vartheta_{\text{sp}}(G,P)$.

The quantity $\vartheta_{\text{sp}}(G,P)$ is the constant-composition analog of the formal quantity $\vartheta_{\text{sp}}(G)$ defined in [3, Sec. VI]. In that case it was observed by Schrijver and by Duan and Winter [20] that in fact $\vartheta_{\text{sp}}(G) = \vartheta(G)$ (with our logarithmic definition of ϑ, see footnote 2). We have the analogous result for constant compositions, which essentially means that in equation (24) it is enough to consider rank-one projectors U_x and state F, the infimum being thus the same as the minimum in equation (22).

Theorem 4: For any graph G and composition P, $\vartheta_{\text{sp}}(G,P) = \vartheta(G,P)$, and the infimum in (24) can be replaced by minimum.

Proof: It is obvious that $\vartheta_{\text{sp}}(G,P) \leq \vartheta(G,P)$, since the right-hand side of (22) is obtained by restricting the operators in the right-hand side of (24) to have rank one.

We now prove the converse inequality and the second part of the theorem (cf. [20]) by showing that the infimum in (24) can be achieved using rank-one operators. Let $\{U_x\}$ be a representation of G in our more general sense and let F be a state. Let first $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}'$ be a purification of F obtained using an auxiliary space $\mathcal{H}'$, so that $\operatorname{Tr}(U_x F) = \operatorname{Tr}(U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\langle\psi|)$. Let then

$$|w_x\rangle = \frac{U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle}{\|U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\|}. \tag{25}$$

When x and x' are not confusable, we have

$$\begin{aligned}
\langle w_x | w_{x'} \rangle &= \frac{\langle\psi|(U_x \otimes \mathbb{1}_{\mathcal{H}'})(U_{x'} \otimes \mathbb{1}_{\mathcal{H}'})|\psi\rangle}{\|U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\| \|U_{x'} \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\|} \\
&= \frac{\langle\psi|(U_x U_{x'}) \otimes (\mathbb{1}_{\mathcal{H}'} \mathbb{1}_{\mathcal{H}'})|\psi\rangle}{\|U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\| \|U_{x'} \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\|} \\
&= 0
\end{aligned}$$

since U_x and $U_{x'}$ are orthogonal. So, $\{w_x\}$ is a standard vector-valued orthonormal representation of G and, furthermore, we have $\operatorname{Tr}(U_x F) = \operatorname{Tr}(U_x \otimes \mathbb{1}_{\mathcal{H}'} |\psi\rangle\langle\psi|) = |\langle w_x|\psi\rangle|^2$, for all x. Hence, the orthonormal representation $\{w_x\}$ and the unit norm vector ψ satisfy

$$\sum_x P(x) \log \frac{1}{\operatorname{Tr}(U_x F)} = \sum_x P(x) \log \frac{1}{|\langle w_x|\psi\rangle|^2}. \tag{26}$$

That is, any representation $\{U_x\}$ with handle F is equivalent to a rank-one representation $\{w_x\}$ with handle ψ in a space of dimension at most $|\mathcal{X}|$ (namely, the span of the w_x). This implies that $\vartheta(G,P) \leq \vartheta_{\text{sp}}(G,P)$ and, in particular, that the infimum can be replaced by a minimum since the domain is compact and the logarithm is a continuous extended-real valued function from $[0,\infty]$ to $[-\infty,\infty]$ once we set $\log(0) = -\infty$, $\log(\infty) = \infty$ and $1/0 = \infty$. ∎

We can now discuss another interesting issue about the use of the quantity $\vartheta(G,P)$. When we are interested in bounding C_0, we can use the bound $C_0 \leq \vartheta(G)$ or we can also use the bound $C_0 \leq \max_P \vartheta(G,P)$. Marton [10]

states that this does not make a difference since - "as is easily seen" - $\max_P \vartheta(G,P) = \vartheta(G)$. However, a proof of this statement does not seem to follow easily from the definitions. It can in fact be written as

$$\max_P \min_{\{u_x\},f} \sum_x P(x) \log \frac{1}{|\langle u_x|f\rangle|^2} = \min_{\{u_x\},f} \max_x \log \frac{1}{|\langle u_x|f\rangle|^2} \tag{27}$$

$$= \min_{\{u_x\},f} \max_P \sum_x P(x) \log \frac{1}{|\langle u_x|f\rangle|^2} \tag{28}$$

and, in order to prove the equality, we would need to exchange the maximization over P with the minimization over representations and handles. It is not clear in Marton's paper what argument she used to motivate it. We use Theorem 4 to prove this statement.

Theorem 5: For any graph G, $\max_P \vartheta(G,P) = \vartheta(G)$.

Proof: For any representation $\{U_x\}$ of G and density operator F, define the function $f(x) = \operatorname{Tr} U_x F$, and denote the set of all functions f obtained in this way by $\mathrm{OR}(G)$. Let then $\mathrm{OR}^+(G) \subseteq \mathrm{OR}(G)$ be the subset of such f functions which are strictly positive, that is $f(x) > 0\,, \forall x$. The proof of Theorem 4 shows that any $f \in \mathrm{OR}(G)$ can be realized by rank-one projections $U_x = |u_x\rangle\langle u_x|$ and a pure state $F = |f\rangle\langle f|$, in a space of dimension at most $|\mathcal{X}|$. In particular, it follows that $\mathrm{OR}(G)$, as image of a continuous function on a compact set, is closed and compact. Furthermore, it is convex: namely, consider $f_i(x) = \operatorname{Tr} U_x^{(i)} F^{(i)}$ for representations $\{U_x^{(i)}\}$ of G and density operators $F^{(i)}$, $i = 1, 2$. Then, for $0 \leq p \leq 1$, let $U_x = U_x^{(1)} \oplus U_x^{(2)}$ and $F = pF^{(1)} \oplus (1-p)F^{(2)}$. The associated $f(x) = \operatorname{Tr} U_x F = pf_1(x) + (1-p)f_2(x)$, i.e. $pf_1 + (1-p)f_2 \in \mathrm{OR}(G)$. This shows that $\mathrm{OR}(G)$ is convex and, for the same reason, $\mathrm{OR}^+(G)$ is also convex.

Now define the quantity

$$J(f,P) = \sum_x P(x) \log \frac{1}{f(x)}, \tag{29}$$

for distributions P and functions $f \in \mathrm{OR}(G)$. The theorem is equivalent to the statement that

$$\max_P \min_{f \in \mathrm{OR}(G)} J(f,P) = \min_{f \in \mathrm{OR}(G)} \max_P J(f,P), \tag{30}$$

since the left-hand side equals $\max_P \vartheta(G,P)$ by Theorem 4, and the right-hand side equals $\vartheta(G)$ by [3, Th. 8]. Equation (30) can now be proved by resorting to an appropriate form of the minimax theorem. The procedure is very similar to the one followed in [17, Prop. 1].

Observe that for any fixed P

$$\vartheta(G,P) := \min_{f \in \mathrm{OR}(G)} J(f,P) \tag{31}$$

$$= \inf_{f \in \mathrm{OR}^+(G)} J(f,P). \tag{32}$$

For any $f \in \mathrm{OR}^+(G)$ the quantity $J(f,P)$ is a finite linear function of P and hence $\vartheta(G,P)$, as the infimum of linear functions of P, is both concave and upper semicontinuous in P. But since it is a concave function on the simplex, it is also lower semicontinuous and hence it is continuous. Furthermore, for any fixed P, the quantity $J(f,P)$ is convex function of f in $\mathrm{OR}^+(G)$. Then, we are in a condition to use the version of the minimax theorem presented in [17, Appendix A] which gives

$$\max_P \inf_{f \in \mathrm{OR}^+(G)} J(f,P) = \inf_{f \in \mathrm{OR}^+(G)} \max_P J(f,P). \tag{33}$$

We then note

$$\inf_{f\in \mathrm{OR}^+(G)} \max_P J(f,P) = \inf_{f\in \mathrm{OR}^+(G)} \max_x \log \frac{1}{f(x)} \tag{34}$$

$$= \min_{f\in \mathrm{OR}(G)} \max_x \log \frac{1}{f(x)} \tag{35}$$

where the last equality is due to continuity and the compactness of $\mathrm{OR}(G)$. ■

We close this section with a result on the boundedness of the sphere packing exponent at the rate $R_\infty(P)$. The first part of the statement will be used in the next section. This is the analogous of [3, Th. 10] for the constant-composition setting.

Theorem 6: For any pure-state channel we have the inequality $E_{\mathrm{sp}}^{\mathrm{cc}}(R_\infty(P),P) \le R_\infty(P)$. For a general (mixed-state) channel we have the bound

$$E_{\mathrm{sp}}^{\mathrm{cc}}(R_\infty(P),P) \le R_\infty(P) + \sum_x P(x) \operatorname{Tr}\left[\frac{S_x^0 F_1}{\operatorname{Tr} S_x^0 F_1}(\log F_1 - \log S_x)\right] \tag{36}$$

$$< +\infty\,, \tag{37}$$

where F_1 achieves the minimum in (21). So, $E_{\mathrm{sp}}^{\mathrm{cc}}(R,P)$ is finite at $R = R_\infty(P)$.

Proof: For a pure-state channel, since $S_x^{\frac{1}{1+\rho}} = S_x = S_x^0$, we have

$$\begin{aligned} E_0^{\mathrm{cc}}(\rho,P) &= \min_F \left[-(1+\rho)\sum_x P(x)\log \operatorname{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}})\right] \\ &= \min_F \left[-(1+\rho)\sum_x P(x)\log \operatorname{Tr}(S_x F^{\frac{\rho}{1+\rho}})\right] \\ &\le \min_F \left[-(1+\rho)\sum_x P(x)\log \operatorname{Tr}(S_x^0 F)\right] \\ &= (1+\rho) R_\infty(P), \end{aligned}$$

from which we easily deduce the first statement by definition of $E_{\mathrm{sp}}^{\mathrm{cc}}(R,P)$.

For general states S_x, let F_1 achieve the minimum in (21), so that

$$R_\infty(P) = -\sum_x P(x)\log \operatorname{Tr}(S_x^0 F_1)\,. \tag{38}$$

Then, using the parameterization $s = \rho/(1+\rho)$ we have

$$\begin{aligned}
&E_{\text{sp}}^{\text{cc}}(R_\infty(P), P) \\
&\leq \sup_{\rho>0} \Bigg[\min_F \left(-(1+\rho) \sum_x P(x) \log \operatorname{Tr}(S_x^{\frac{1}{1+\rho}} F^{\frac{\rho}{1+\rho}}) \right) \\
&\qquad\qquad -\rho R_\infty(P) \Bigg] \\
&\leq \sup_{\rho>0} \left[-(1+\rho) \sum_x P(x) \log \operatorname{Tr}(S_x^{\frac{1}{1+\rho}} F_1^{\frac{\rho}{1+\rho}}) - \rho R_\infty(P) \right] \\
&\leq \sup_{0<s<1} \left[-\frac{1}{1-s} \left(\sum_x P(x) \log \operatorname{Tr}(S_x^{1-s} F_1^s) + R_\infty(P) \right) \right] \\
&\qquad\qquad +R_\infty(P) \\
&\leq \sup_{0<s<1} \left[-\frac{1}{1-s} \sum_x P(x) \left(\mu_x(s) - \mu_x(1) \right) \right] + R_\infty(P) \,,
\end{aligned}$$

where we set $\mu_x(s) = \log \operatorname{Tr} S_x^{1-s} F_1^s$. Since $\mu_x(s)$ is convex in $[0,1]$, we have $\mu_x(s) \geq \mu_x(1) - \mu_x'(1)(1-s)$ where

$$\mu_x'(1) = \operatorname{Tr} \left[\frac{S_x^0 F_1}{\operatorname{Tr} S_x^0 F_1} \left(\log F_1 - \log S_x \right) \right] \tag{39}$$

is a finite quantity for any x. This gives the claimed bound. ■

V. Low Rate Bounds

As in the classical case, and similarly to what was already discussed in [3], Theorem 1 gives the trivial bound $E(R) \leq \infty$ for rates $R \leq R_\infty(P)$. It can be checked from equation (21) that $R_\infty(P)$ can be positive in general, even if the channel states are pairwise non-orthogonal, that is even when $C_0(P) = 0$. In particular, note that $R_\infty(P)$ is positive for all non-trivial pure-state channels. In this Section we derive a bound which, in the same spirit of [3, Sec. VIII], attempts to upper bound the reliability function at low rates by resorting to auxiliary channels. The improvement, with respect to [3, Sec. VIII], is based on the study of constant *conditional composition* codes as inspired by [21]. In a nutshell, given a constant-composition code for a given channel, we will extract a subcode whose codewords have constant conditional composition with respect to some auxiliary sequence of "states", and we will bound the probability of error on the original channel by considering a different auxiliary channel for each state. This approach can be considered as an evolution of the method used in [3, Sec. VIII] along the same lines taken in [11].

A. Conditional Sphere Packing Bound

We now develop an extension of the sphere packing to handle the case of varying channels with a *conditional composition* constraint on the codewords. Although this setting can appear artificial, the bound will prove useful when applied to auxiliary channels in combination with a construction originally introduced by Elias, and further developed by Blahut [21], to bound the minimum distance of codes.

Here we assume that we have a finite set $\mathcal{A}$ of possible states and a different channel $\mathfrak{C}_a$, for each state $a \in \mathcal{A}$. The communication is governed by a sequence of states $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}^n$ (known to both encoder and decoder) with composition Q_n, which determines the channels to use. In particular, channel $\mathfrak{C}_{a_i}$ is used at time instant i and hence we can consider the global channel $\{\mathfrak{C}_{a_i}\}_{i\in\mathbb{N}}$, as a time-varying memoryless channel. The composition constraint in this case is that all codewords have *conditional* composition $\hat{P}_n$ given $\boldsymbol{a}$ in the sense that any codeword has a symbol x in a fraction $\hat{P}_n(x|a)$ of the $nQ_n(a)$ positions where $a_i = a$. We then assume that, as $n \to \infty$, $Q_n \to Q$ and $\hat{P}_n \to \hat{P}$.

Remark 7: Note that this general scenario includes the ordinary constant-composition situation considered before, which is obtained for example when $Q(a) = 1$ for some a and $\boldsymbol{a} = (a, a, \ldots, a)$. Note that it also includes the study of the parallel use of $K > 1$ channels, which can be recovered by setting $Q(a) = 1/K, \forall a$, and normalizing the block lengths by a factor K.

For a given Q and $\hat{P}$, let now $E(\{\mathfrak{C}_a\}, R, \hat{P}|Q)$ be the optimal asymptotic error exponent achievable by codes with asymptotic conditional composition $\hat{P}$ with respect to a sequence with asymptotic composition Q using the set of channels $\{\mathfrak{C}_a\}$, $a \in \mathcal{A}$. Then we have the following result.

Theorem 8: We have the inequality

$$E(\{\mathfrak{C}_a\}, R, \hat{P}|Q) \leq E_{\text{sp}}^{\text{cc}}(\{\mathfrak{C}_a\}, R - \varepsilon, \hat{P}|Q), \tag{40}$$

where $E_{\text{sp}}^{\text{cc}}(\{\mathfrak{C}_a\}, R, \hat{P}|Q)$ is defined by

$$E_{\text{sp}}^{\text{cc}}(\{\mathfrak{C}_a\}, R, \hat{P}|Q) = \sup_{\rho>0} \left[E_0^{\text{cc}}(\{\mathfrak{C}_a\}, \rho, \hat{P}|Q) - \rho R \right], \tag{41}$$

$$E_0^{\text{cc}}(\{\mathfrak{C}_a\}, \rho, \hat{P}|Q) = \sum_a Q(a) E_0^{\text{cc}}(\mathfrak{C}_a, \rho, \hat{P}(\cdot|a)), \tag{42}$$

and $E_0^{\text{cc}}(\mathfrak{C}_a, \rho, \hat{P}(\cdot|a))$ is the coefficient E_0^{cc} of the sphere packing bound for channel $\mathfrak{C}_a$ with composition $P(\cdot|a)$, as defined in (5).

Proof: See Appendix B. ∎

We observe that the function $E_{\text{sp}}^{\text{cc}}(\{\mathfrak{C}_a\}, R, \hat{P}|Q)$ is finite for all rates $R \geq R_\infty(\{\mathfrak{C}_a\}, \hat{P}|Q)$ where

$$R_\infty(\{\mathfrak{C}_a\}, \hat{P}|Q) = \lim_{\rho\to\infty} \frac{E_0^{\text{cc}}(\{\mathfrak{C}_a\}, \rho, \hat{P}|Q)}{\rho} \tag{43}$$

$$= \lim_{\rho\to\infty} \sum_a Q(a) \frac{E_0^{\text{cc}}(\mathfrak{C}_a, \rho, \hat{P}(\cdot|a))}{\rho} \tag{44}$$

$$= \sum_a Q(a) R_\infty(\mathfrak{C}_a, \hat{P}(\cdot|a)). \tag{45}$$

Furthermore, it is not difficult to show, using the same procedure used in Theorem 6, that for pure-state channels we have the inequality

$$E_{\text{sp}}^{\text{cc}}(\{\mathfrak{C}_a\}, R_\infty(\{\mathfrak{C}_a\}, \hat{P}|Q), \hat{P}|Q) \leq R_\infty(\{\mathfrak{C}_a\}, \hat{P}|Q). \tag{46}$$

B. Improvement of the Sphere-Packed Umbrella Bound

We can now combine the bound derived above with the ideas presented in [22], [3] and [23], much in the same way as done in [11] [24], to obtain a bound on the reliability of a channel $\mathfrak{C}$ using auxiliary classical-quantum

channels $\{\tilde{\mathfrak{C}}_a\}$. We limit here the discussion to the case of a pure-state channel with states $S_x = |\psi_x\rangle\langle\psi_x|$ and pure-state auxiliary channels $\{\tilde{\mathfrak{C}}_a\}$. The general case will become clear in the next section where we reformulate this bound in terms of code *distances*, reinterpreting it as a generalization of the Elias bound.

For a $\rho \geq 1$, we define the set $\Gamma(\rho)$ of admissible pure-state auxiliary channels $\tilde{\mathfrak{C}}$ with states $\tilde{S}_x = |\tilde{\psi}_x\rangle\langle\tilde{\psi}_x|$ such that

$$|\langle\tilde{\psi}_x|\tilde{\psi}_{x'}\rangle| \leq |\langle\psi_x|\psi_{x'}\rangle|^{1/\rho}, \quad \forall x, x' \in \mathcal{X}. \tag{47}$$

For any $a \in \mathcal{A}$ we choose an auxiliary pure-state channel $\tilde{\mathfrak{C}}_a \in \Gamma(\rho)$ with states $\tilde{S}_{a,x} = |\tilde{\psi}_{a,x}\rangle\langle\tilde{\psi}_{a,x}|$. Given a sequence $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}^n$ and a sequence $\boldsymbol{x} = (x_1 \ldots, x_n) \in \mathcal{X}^n$, let

$$\tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}} = \tilde{\psi}_{a_1,x_1} \otimes \cdots \otimes \tilde{\psi}_{a_n,x_n}. \tag{48}$$

Now, given two sequences $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$, we can use these auxiliary channels to bound the overlap $|\langle\boldsymbol{\psi}_{\boldsymbol{x}}|\boldsymbol{\psi}_{\boldsymbol{x}'}\rangle|^2$ as

$$|\langle\boldsymbol{\psi}_{\boldsymbol{x}}|\boldsymbol{\psi}_{\boldsymbol{x}'}\rangle|^2 \geq |\langle\tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}}|\tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}'}\rangle|^{2\rho}. \tag{49}$$

This will allow us to bound $E(R, P)$ for the original channel using the bound (see for example [3, Th. 12])

$$E(R, P) \leq -\frac{1}{n} \log \max_{m \neq m'} |\langle\boldsymbol{\psi}_{\boldsymbol{x}_m}|\boldsymbol{\psi}_{\boldsymbol{x}_{m'}}\rangle|^2 + o(1) \tag{50}$$

$$\leq -\frac{\rho}{n} \log \max_{m \neq m'} |\langle\tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_m}|\tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_{m'}}\rangle|^2 + o(1). \tag{51}$$

We could use the extension of the sphere packing bound considered in this section to upper bound the right-hand side of the last equation as done in [3, Sec. VIII] if all codewords $\boldsymbol{x}_m$ had the same conditional composition given the sequence $\boldsymbol{a}$. Since the sequence $\boldsymbol{a}$ is arbitrary, we choose it so that this condition is met by at least a large enough subset $\mathcal{T}$ of codewords, and we only apply the sphere packing bound to this subset $\mathcal{T}$. In order to do this, we adopt an idea proposed by Blahut [23] in a generalization of the Elias bound and already considered for a further generalization in [11], [24].

Fix a $\delta > 0$. Given a code with $M = e^{nR_n}$ codewords of composition P_n, assume that there exists a set finite $\mathcal{A}$ and a conditional composition $\hat{Q}_n(a|x)$, $x \in \mathcal{X}, a \in \mathcal{A}$ (i.e., $\hat{Q}_n(\cdot|\cdot)$ is a stochastic matrix and $nP_n(x)\hat{Q}_n(a|x)$ is an integer) such that

$$R_n > I(P_n, \hat{Q}_n) + \delta, \tag{52}$$

where $I(P_n, \hat{Q}_n)$ is the mutual information as defined in (10). Define then

$$Q_n(a) = \sum_x P_n(x)\hat{Q}_n(a|x) \tag{53}$$

(that we will write as $P_n\hat{Q}_n = Q_n$) and and let $\hat{P}_n(x|a) = P_n(x)\hat{Q}_n(a|x)/Q_n(a)$, so that $Q_n\hat{P}_n = P_n$. Note that $I(P_n, \hat{Q}_n) = I(Q_n, \hat{P}_n)$.

Then, (see [23, proof of Th. 8], or [24, Lemma 3]) there is at least one sequence $\boldsymbol{a}$ of composition Q_n such that there is a subset $\mathcal{T}_n$ of at least $|\mathcal{T}_n| = e^{n(R_n - I(Q_n,\hat{P}_n) + o(1))} > e^{n(\delta + o(1))}$ codewords with conditional composition

$\hat{P}_n$ given $\boldsymbol{a}$. For these codewords used over this varying channel, there is a decision rule such that ([25], [3, Sec. VIII])

$$\tilde{\mathsf{P}}_{\text{e,max}} \leq (|\mathcal{T}_n| - 1) \max_{m \neq m' \in \mathcal{T}_n} |\langle \tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_m} | \tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_{m'}} \rangle|^2 \tag{54}$$

$$\leq e^{n(R_n - I(Q_n,\hat{P}_n) + o(1))} \max_{m \neq m' \in \mathcal{T}} |\langle \tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_m} | \tilde{\boldsymbol{\psi}}_{\boldsymbol{a},\boldsymbol{x}_{m'}} \rangle|^2. \tag{55}$$

Now, we can use the conditional sphere packing bound introduced in this section to bound the probability of error of the subcode $\mathcal{T}_n$ of rate $\tilde{R}_n = R_n - I(Q_n, \hat{P}_n) + o(1) > \delta + o(1)$ used over the varying channel $\tilde{\mathfrak{C}}_{a_1}, \cdots, \tilde{\mathfrak{C}}_{a_n}$. Since we are interested in the limit as $n \to \infty$, we directly work with the asymptotic rates R and $\tilde{R} = R - I(Q, \hat{P}) \geq \delta$, compositions P and Q and matrix $\hat{Q}$, and we neglect the constraint that $nP_n(x)$, $nP_n(x)\hat{Q}_n(a|x)$ etc. are integers. Then, as $n \to \infty$, Theorem 8 with rate $\tilde{R}$ gives

$$-\frac{1}{n} \log \tilde{\mathsf{P}}_{\text{e,max}} \leq E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, \tilde{R} - \varepsilon, \hat{P}|Q) + o(1) \tag{56}$$

$$\leq E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, R - I(Q, \hat{P}) - \varepsilon, \hat{P}|Q) + o(1). \tag{57}$$

Putting together equations (51), (55) and (57), we obtain

$$E(R, P) \leq \rho[E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, R - I(Q, \hat{P}) - \varepsilon, \hat{P}|Q) + R - I(Q, \hat{P})]. \tag{58}$$

Since $\delta > 0$ is arbitrarily small, since the choice of ρ, of the channels $\{\tilde{\mathfrak{C}}_a\} \in \Gamma(\rho)$ and of the distributions $Q, \hat{P}$ can be optimized, we have, in analogy with [3, Th. 11],

Theorem 9: For a pure-state channel, the reliability function with constant composition P satisfies $E(R, P) \leq E_{\text{spu}}^{\text{cc}}(R, P)$ where

$$E_{\text{spu}}^{\text{cc}}(R, P) = \inf \rho[E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, R - I(Q, \hat{P}) - \varepsilon, \hat{P}|Q) + R - I(Q, \hat{P})], \tag{59}$$

the infimum being over $\varepsilon > 0$, $\rho \geq 1$, auxiliary pure-state channels $\tilde{\mathfrak{C}}_a \in \Gamma(\rho)$, and auxiliary distributions Q and $\hat{P}$ such that $Q\hat{P} = P$ and $R - I(Q, \hat{P}) \geq 0$.

Remark 10: Note that for the choice $\mathcal{A} = \mathcal{X}$, $\hat{Q}(a|x) = P(a)$, $\forall a$, we have $I(P, \hat{Q}) = 0$. We can also notice that the optimization of the channels $\tilde{\mathfrak{C}}_a$ will give $\tilde{\mathfrak{C}}_a = \tilde{\mathfrak{C}}$, $\forall a$, for an optimal $\tilde{\mathfrak{C}}$. With this constraints the bound $E(R, P)$ is weakened to

$$\inf \rho[E_{\text{sp}}^{\text{cc}}(\tilde{\mathfrak{C}}, R - \varepsilon, P) + R], \tag{60}$$

where the infimum is now only over $\rho \geq 1$ and $\tilde{\mathfrak{C}} \in \Gamma(\rho)$. This is a constant-composition version of the bound in [3, Th. 11].

C. Connection with the Elias Bound

In the same way as [3, Th. 11] generalizes the results of [3, Sec. III], it possible to reinterpret the idea used to obtain Theorem 9 as a generalization of the Elias bound presented in [11] and [24]. For this purpose, it is useful to introduce a notion of distance between symbols and distance between sequences, and then restate our bound as a bound on the minimum distance of codes. Finally, bounds on the reliability function can be obtained by relating the minimum distance to the probability of error (see [24, Sec. VI] for details).

Let d be a function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+ \cup \{\infty\}$ such that

$$d(x, x') \geq 0$$
$$d(x, x') = d(x', x) \quad \forall x, x'$$
$$d(x, x) = 0.$$

We call this function d a "distance" although, as seen above, we do not really require all the properties of a distance. We stress that d is allowed to take value ∞ for some pairs of symbols, a case which is of practical interest in our context. We extend the distance to sequences of symbols defining, for $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$,

$$d(\boldsymbol{x}, \boldsymbol{x}') := \sum_{i=1}^{n} d(x_i, x'_i). \tag{61}$$

Note in particular that $d(\boldsymbol{x}, \boldsymbol{x}') = \infty$ iff $d(x_i, x'_i) = \infty$ for at least one i.

For a given code $\boldsymbol{\mathcal{C}}$, we define its minimum distance as

$$d_{\min}(\boldsymbol{\mathcal{C}}) := \min_{\boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{\mathcal{C}},\, \boldsymbol{x} \neq \boldsymbol{x}'} d(\boldsymbol{x}, \boldsymbol{x}'). \tag{62}$$

For an integer $n \geq 1$, rate R and a composition $P \in \mathcal{T}_n$, we define

$$d(R, n, P) := \max_{\boldsymbol{\mathcal{C}}} d_{\min}(\boldsymbol{\mathcal{C}}), \tag{63}$$

where the maximum is over all codes of length n, rate at least R, and composition P. For a fixed R and distribution P, we then define

$$\delta^*(R, P) := \sup \left\{ \limsup_{n \to \infty} \frac{1}{n} d(R_n, n, P_n) \right\}, \tag{64}$$

where the outer supremum is over all sequences of rates $R_n \to R$ and compositions $P_n \to P$ as $n \to \infty$.

Note that we can drop the constant-composition constraint defining

$$d(R, n) := \max_{\boldsymbol{\mathcal{C}}} d_{\min}(\boldsymbol{\mathcal{C}}), \tag{65}$$

and, correspondingly,

$$\delta^*(R) := \limsup_{n \to \infty} \frac{1}{n} d(R, n). \tag{66}$$

Then we have

$$\delta^*(R) := \max_{P} \delta^*(R, P). \tag{67}$$

We want to use our results to bound the quantity $\delta^*(R, P)$. In order to do this we proceed in a similar way as done in Section V-B. Note that this corresponds to what was done in [24] with two variations; 1) we use general auxiliary classical-quantum channels in place of the so called representations composed of vectors, and 2) we replace the Lovász-like trick of [24, Lemma 2] with the sphere packing bound.

Given the distance d and a $\rho \geq 1$, we define the set $\Gamma(\rho)$ of admissible auxiliary channels $\tilde{\mathfrak{C}}$ with states $\tilde{S}_x$ such that

$$\operatorname{Tr} \sqrt{\tilde{S}_x} \sqrt{\tilde{S}_{x'}} \leq e^{-d(x,x')/\rho}. \tag{68}$$

We then consider again as in Section V-B the subcode $\mathcal{T}_n$ of codewords with composition P_n all with the same conditional composition $\hat{P}_n$ given the sequence $\boldsymbol{a}$. For any $a \in \mathcal{A}$ we choose an auxiliary channel $\tilde{\mathfrak{C}}_a \in \Gamma(\rho)$ with states $\tilde{S}_{a,x}$ and for an $\boldsymbol{x} \in \mathcal{T}$ we define

$$\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x}} = \tilde{S}_{a_1,x_1} \otimes \cdots \otimes \tilde{S}_{a_n,x_n}. \tag{69}$$

Note that this implies that for two sequences $\boldsymbol{x}$ and $\boldsymbol{x'}$,

$$\operatorname{Tr} \sqrt{\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x}}} \sqrt{\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x'}}} \leq e^{-d(\boldsymbol{x},\boldsymbol{x'})/\rho}. \tag{70}$$

Consider now an optimal decision scheme for the states associated to the subcode $\mathcal{T}_n$, that is, $\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x}}$, $\boldsymbol{x} \in \mathcal{T}_n$. The extension of (55) [25] says that for such a set of states, there exists a measurement such that

$$\tilde{\mathsf{P}}_{\text{e,max}} \leq e^{n(R-I(Q,\hat{P})+o(1))} \max_{m \neq m' \in \mathcal{T}} \operatorname{Tr} \sqrt{\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x}_m}} \sqrt{\tilde{\boldsymbol{S}}_{\boldsymbol{a},\boldsymbol{x}_{m'}}}. \tag{71}$$

But, again, we can use the conditional sphere packing bound to lower bound the probability of error of the subcode $\mathcal{T}_n$ as

$$-\frac{1}{n} \log \tilde{\mathsf{P}}_{\text{e,max}} \leq E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, R - I(Q,\hat{P}) - \varepsilon, \hat{P}|Q) + o(1). \tag{72}$$

Combining equations (70), (71) and (72) we obtain

$$\frac{1}{n} \min_{m \neq m'} d(\boldsymbol{x}_m, \boldsymbol{x}_{m'}) \leq \rho(E_{\text{sp}}^{\text{cc}}(\{\tilde{\mathfrak{C}}_a\}, R - I(Q,\hat{P}) - \varepsilon, \hat{P}|Q) + R - I(Q,\hat{P})) + o(1), \tag{73}$$

which asymptotically gives the following result.

Theorem 11: For a distance d and assuming the above definitions, we have the inequality

$$\delta^*(R,P) \leq E_{\text{spu}}^{\text{cc}}(R,P), \tag{74}$$

where $E_{\text{spu}}^{\text{cc}}(R,P)$ is defined in (59).

As mentioned, this bound is an extension of [24, Th. 6]. To see this, we can consider the particular case in which we restrict the attention to pure-state auxiliary channels with states $\tilde{S}_{a,x} = |\tilde{\psi}_{a,x}\rangle\langle\tilde{\psi}_{a,x}|$ and then study the smallest rate for which the bound $E_{\text{spu}}^{\text{cc}}(R,P)$ (with this additional constraint) is finite. First note that for fixed channels $\{\tilde{\mathfrak{C}}_a\}$, distributions Q and $\hat{P}$, and ε sufficiently small, the quantity on the right-hand side of equation (59) is finite for $R > R_\infty(\{\tilde{\mathfrak{C}}_a\}, \hat{P}|Q) + I(Q,\hat{P})$. Furthermore, when R approaches this value from the right, using equation (46), the right-hand side of equation (59) is upper bounded by $2\rho R_\infty(\{\tilde{\mathfrak{C}}_a\}, \hat{P}|Q)$. So, for $R > R_\infty(\{\tilde{\mathfrak{C}}_a\}, \hat{P}|Q) + I(Q,\hat{P})$ we have the bound

$$\delta^*(R,P) \leq 2\rho R_\infty(\{\tilde{\mathfrak{C}}_a\}, \hat{P}|Q). \tag{75}$$

For pure-state auxiliary channels we can write

$$R_\infty(\{\mathfrak{C}_a\},\hat{P}|Q) = \sum_a Q(a) R_\infty(\mathfrak{C}_a,\hat{P}(\cdot|a)) \tag{76}$$

$$= \sum_{a\in\mathcal{X}} Q(a) \min_{F_a}\left[-\sum_x \hat{P}(x|a)\log \mathrm{Tr}(\tilde{S}^0_{a,x}F_a)\right] \tag{77}$$

$$= \min_{\{F_a\}} \sum_{a,x\in\mathcal{X}} Q(a)\hat{P}(x|a)\log\frac{1}{\langle\tilde{\psi}_{a,x}|F_a|\tilde{\psi}_{a,x}\rangle} \tag{78}$$

$$\leq \min_{\{f_a\}} \sum_{a,x\in\mathcal{X}} Q(a)\hat{P}(x|a)\log\frac{1}{|\langle\tilde{\psi}_{a,x}|f_a\rangle|^2}, \tag{79}$$

where the last step we have enforced minimization over rank-one operators $F_a = |f_a\rangle\langle f_a|$. Optimizing now over ρ, Q and $\hat{P}$ such that $Q\hat{P} = P$, and the auxiliary vectors $\{\tilde{\psi}_{a,x}\}$, and comparing with the definition of $\vartheta(\rho,\hat{P}|Q)$ used in [24], we deduce that the bound of Theorem 11 includes, as a particular case, the bound presented in [24, Th. 6] as a generalization of the Elias bound for general, possibly infinite, distances[3]. Hence, it includes in particular all previously known extensions as discussed in [24].

VI. Acknowledgments

The author(s) would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme "Mathematical Challenges in Quantum Information" where work on this paper was undertaken. AW was supported by the European Commission (STREP "RAQUEL"), the ERC (Advanced Grant "IRQUAT"), the Spanish MINECO (grant FIS2013-40627-P) with the support of FEDER funds, and by the Generalitat de Catalunya CIRIT, project 2014-SGR-966. MD was supported by the Italian Ministry of Education, University and Research (MIUR) under grant PRIN 2015 D72F1600079000.

Appendix A
Proof of Theorem 1

The structure of the proof is the same as in [4], and [3, Th. 5] with some technical changes which are required for dealing with general compositions. While introducing these changes, we also considerably simplify some of the technicalities with respect to [3, Th. 5] in order to give a simpler and more transparent proof of both this and the original theorem.

From the definition of $E(R,P)$, there exists a sequence of codes of block-lengths $n = 1,2,\ldots$ with rates $R_n \to R$, compositions $P_n \to P$ and with probabilities of error $\mathsf{P}^{(n)}_{\text{e,max}}$ such that

$$E(R,P) = \limsup_{n\to\infty} -\frac{1}{n}\log \mathsf{P}^{(n)}_{\text{e,max}}.$$

We first observe that we can just focus on the subset of input symbols with $P(x) > 0$ and assume without loss of generality that $P_n(x) = 0$ if $P(x) = 0$. This technicality is needed shortly after equation (89) below and can be motivated as follows. Let $\mathcal{X}_0$ be the subset of $\mathcal{X}$ such that $P(x) = 0$ if and only if $x \in \mathcal{X}_0$. Then, for

[3]Note that the definition of $\Gamma(\rho)$ in [24] is slightly different than here, so that the parameter ρ here corresponds to the parameter $\rho/2$ there.

any sequence of compositions $P_n \to P$, for all $x \in \mathcal{X}_0$ we can write that $P_n(x) \leq \varepsilon_n/|\mathcal{X}_0|$, where $\varepsilon_n \to 0$ as $n \to \infty$. Any codeword with composition P_n will contain symbols from $\mathcal{X}_0$ in at most $n\varepsilon_n$ positions. There are only nearly $e^{nH(\varepsilon_n)}$ choices for these positions and, for each such choice there are only at most $|\mathcal{X}_0|^{n\varepsilon_n}$ possible combinations of symbols in $\mathcal{X}_0$. Hence, from a code with rate R_n and composition P_n we can extract a subcode with rate $R'_n = R_n - H(\varepsilon_n) - \varepsilon_n \log |\mathcal{X}_0|$ such that each symbol in $\mathcal{X}_0$ appears precisely in the same positions in all codewords. We can then bound $E(R, P)$ by bounding the probability of error for this subcode since, given that $\varepsilon_n \to 0$, we have $(R'_n - R_n) \to 0$. However, in the chosen subcode each symbol in $\mathcal{X}_0$ appears in the same positions in all codewords, and can thus be replaced with any symbol in $\mathcal{X}\backslash\mathcal{X}_0$ without affecting the probability of error.

For every fixed n, the idea is again as in previous proofs to consider a binary hypothesis test between a properly selected code signal $\boldsymbol{S}_{\boldsymbol{x}_m}$ and an auxiliary density operator $\boldsymbol{F} = F^{\otimes n}$. The main difference with respect to [3, Th. 5] is in the choice of F and, as a consequence, in some technical details.

Let n be fixed and let M be the number of codewords, that is $M = e^{nR_n}$. For any $m = 1, \ldots, M$ consider a binary hypothesis test between $\boldsymbol{S}_{\boldsymbol{x}_m}$ and an auxiliary state $\boldsymbol{F} = F^{\otimes n}$. We assume that the supports of the two operators are not orthogonal and, with the notation used in [3], we define the quantity

$$\begin{aligned}\mu(s) &= \mu_{\boldsymbol{S}_{\boldsymbol{x}_m},\boldsymbol{F}}(s)\\ &= \log \operatorname{Tr} \boldsymbol{S}_{\boldsymbol{x}_m}^{1-s} \boldsymbol{F}^s.\end{aligned}$$

Note that, setting

$$\mu_{S_x,F}(s) = \log \left(\operatorname{Tr} S_x^{1-s} F^s \right), \tag{80}$$

we can write

$$\begin{aligned}\mu_{\boldsymbol{S}_{\boldsymbol{x}_m},\boldsymbol{F}}(s) &= \log \prod_{i=1}^{n} \operatorname{Tr} S_{x_{m,i}}^{1-s} F^s\\ &= \log \prod_{x} \left(\operatorname{Tr} S_x^{1-s} F^s \right)^{nP_n(x)}\\ &= n \sum_x P_n(x) \mu_{S_x,F}(s).\end{aligned} \tag{81}$$

Applying [3, Th. 4], we find that for each s in $(0, 1)$, either

$$\operatorname{Tr}\left[(\mathbb{1} - \Pi_m) \boldsymbol{S}_{\boldsymbol{x}_m} \right] > \frac{1}{8} \exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)} \right] \tag{82}$$

or

$$\operatorname{Tr}\left[\Pi_m \boldsymbol{F} \right] > \frac{1}{8} \exp\left[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)} \right]. \tag{83}$$

Following [4], since $\sum_m \Pi_m \leq \mathbb{1}$, for at least one value of m we have $\operatorname{Tr}\left[\Pi_m \boldsymbol{F} \right] \leq 1/M = e^{-nR_n}$ and, for this same m we have $\operatorname{Tr}\left[(\mathbb{1} - \Pi_m) \boldsymbol{S}_{\boldsymbol{x}_m} \right] = \mathsf{P}_{\mathsf{e}|m} \leq \mathsf{P}_{\mathsf{e,max}}^{(n)}$. So, the above two conditions can be converted into a relation between $\mathsf{P}_{\mathsf{e,max}}^{(n)}$ and R_n in the form that either

$$\mathsf{P}_{\mathsf{e,max}}^{(n)} > \frac{1}{8} \exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)} \right] \tag{84}$$

or

$$R_n < -\frac{1}{n}\left[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)} - \log 8\right]. \tag{85}$$

Note that due to (81), the right-hand side of (85) only depends on n, s, P_n, and F. Let this quantity be called $R_n^*(s, P_n, F)$, that is,

$$R_n^*(s, P_n, F) = -\frac{1}{n}\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)} - \log 8\Big). \tag{86}$$

We can use this equation to write $\mu'(s)$ in terms of $R_n^*(s, P_n, F)$. Using (81), we can state our conditions by saying that either

$$R_n < R_n^*(s, P_n, F) \tag{87}$$

or

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\text{e,max}}^{(n)}} < -\frac{1}{1-s}\sum_x P_n(x)\mu_{S_x,F}(s) - \frac{s}{1-s}R_n^*(s, P_n, F) + \frac{1}{n}\left(2s\sqrt{2\mu''(s)} + \frac{\log 8}{1-s}\right). \tag{88}$$

At this point we introduce the variation with respect to [3]. For any F, one of the two conditions above must be satisfied and, in [3], the choice of F was made which, as in [4], guaranteed the smallest possible upper bound on the error exponent valid for all possible choices of the compositions P_n (and, hence, in particular for the best possible choice). Here, instead, the compositions P_n are forced to tend to a given distribution P and we want to choose F accordingly. For a given $s \in (0,1)$, let F_s be the operator defined by

$$F_s = \arg\min_F \left[-\sum_x P(x)\log(\operatorname{Tr} S_x^{1-s}F^s)\right]. \tag{89}$$

It is shown in Appendix D that F_s above is a well defined continuous function of s in the interval $(0,1)$. We note that for all x with $P(x) > 0$, S_x and F_s have non-orthogonal supports. Since we assumed that $P_n(x) = 0$ whenever $P(x) = 0$, the requirement that $\boldsymbol{S}_{\boldsymbol{x}_m}$ and $\boldsymbol{F}$ have non-orthogonal supports is satisfied, when $\boldsymbol{F} = F_s^{\otimes n}$, for all sequences $\boldsymbol{x}_m$ with composition P_n. Hence, with the choice $F = F_s$, $\mu(s)$ is a finite quantity for all $s \in (0,1)$.

We will now relate the choice of s to the rate R and then use F_s in place of F for the chosen s (it must be clear, however, that $\mu'(s)$ and $\mu''(s)$ are computed by holding F fixed). Note that we can write

$$R_n^*(s, P_n, F_s) = -\sum_x P_n(x)\left[\mu_{S_x,F_s}(s) + (1-s)\mu'_{S_x,F_s}(s)\right] + \frac{1}{\sqrt{n}}(1-s)\sqrt{2\sum_x P_n(x)\mu''_{S_x,F_s}(s)} + \frac{1}{n}\log 8. \tag{90}$$

For any fixed s, the last two terms on the right-hand side vanish as $n \to \infty$, and P_n in the first term tends to P. Hence, it is useful to define the quantity

$$R^*(s, P) = \lim_{n\to\infty} R_n^*(s, P_n, F_s) \tag{91}$$

$$= -\sum_x P(x)\left[\mu_{S_x,F_s}(s) + (1-s)\mu'_{S_x,F_s}(s)\right] \tag{92}$$

and compare this quantity to the rate R which we are considering, which is the limit of the R_n's.

We first observe that, for any x and F, $\mu_{S_x,F}(s)$ is a non-positive convex function of s for all $s \in (0,1)$, which implies that for any F we have

$$\mu_{S_x,F}(s) + (1-s)\mu'_{S_x,F}(s) \le \mu_{S_x,F}(1^-) \le 0.$$

Hence, both $R^*(s,P)$ and $R_n^*(s,P_n,F_s)$ are non-negative quantities. Furthermore since, as observed above, F_s is continuous in s in the interval $0<s<1$, so is $R^*(s,P)$. Hence, $R^*(s,P)$ is a continuous non-negative function of s in the interval $0<s<1$, and we can compare this function with the asymptotic rate R. We only have three possible situations:

1) $R > \sup_{s\in(0,1)} R^*(s,P)$;
2) $R \leq \inf_{s\in(0,1)} R^*(s,P)$;
3) $\inf_{s\in(0,1)} R^*(s,P) < R \leq \sup_{s\in(0,1)} R^*(s,P)$.

Assume case 1) is satisfied. Fix an arbitrary $s\in(0,1)$. Since $R_n \to R$ and $R_n^*(s,P_n,F_s)\to R^*(s,P)<R$, $R_n > R_n^*(s,P_n,F_s)$ for all n large enough. Hence, equation (87) is not satisfied and thus equation (88) is. Since s is fixed and $R_n^*(s,P_n,F_s)\geq 0$, as n goes to infinity we find

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\text{e,max}}^{(n)}} < -\frac{1}{1-s}\sum_x P_n(x)\mu_{S_x,F_s}(s) - \frac{s}{1-s}R_n^*(s,P_n,F_s)+o(1) \tag{93}$$

$$\leq -\frac{1}{1-s}\sum_x P_n(x)\mu_{S_x,F_s}(s)+o(1) \tag{94}$$

and in the limit, since $P_n\to P$,

$$E(R,P)\leq E_0^{\text{cc}}\left(\frac{s}{1-s},P\right). \tag{95}$$

Since this holds for arbitrary $s\in(0,1)$, we have

$$\begin{aligned} E(R,P) &\leq \lim_{s\to 0} E_0^{\text{cc}}\left(\frac{s}{1-s},P\right)\\ &= 0, \end{aligned}$$

where the last step is deduced by noticing that $E_0^{\text{cc}}(\rho,P)$ is continuous at $\rho=0$ and that the argument of the minimization in the definition of $E_0^{\text{cc}}(\rho,P)$ is a non-negative quantity which, for $\rho=0$, vanishes for all F with full support. This proves the theorem in case 1) since $E_{\text{sp}}^{\text{cc}}(R-\varepsilon,P)\geq 0$.

Assume now that case 2) is satisfied, which means by definition of $R^*(s,P)$ that, for any $s\in(0,1)$, we have

$$R\leq -\sum_x P(x)\left[\mu_{S_x,F_s}(s)+(1-s)\mu'_{S_x,F_s}(s)\right].$$

Now, since $\mu_{S_x,F}(s)$ is convex and non-positive for all F, it is possible to observe that $\mu_{S_x,F_s}(s)-s\mu'_{S_x,F_s}(s)\leq 0$, which implies that $-\mu'_{S_x,F_s}(s)\leq -\mu_{S_x,F_s}(s)/s$. Thus, for all $s\in(0,1)$,

$$\begin{aligned} R &\leq \sum_x P(x)\left(-\frac{1}{s}\mu_{S_x,F_s}(s)\right)\\ &= \frac{1-s}{s}E_0^{\text{cc}}\left(\frac{s}{1-s},P\right). \end{aligned}$$

Calling now $\rho = s/(1-s)$, we find that for all $\rho>0$

$$R\leq \frac{E_0^{\text{cc}}(\rho,P)}{\rho}.$$

Hence, for any $\varepsilon > 0$, we find

$$\begin{aligned} E_{\text{sp}}^{\text{cc}}\left(R-\varepsilon, P\right) &= \sup_{\rho>0}\left(E_0^{\text{cc}}\left(\rho, P\right) - \rho\left(R-\varepsilon\right)\right) \\ &\geq \sup_{\rho>0}(\rho\,\varepsilon). \end{aligned}$$

This means that $E_{\text{sp}}^{\text{cc}}(R-\varepsilon, P) = +\infty$ for any $\varepsilon > 0$, which obviously implies that $E(R,P) \leq E_{\text{sp}}^{\text{cc}}(R-\varepsilon, P)$ for all positive ε, proving the theorem in this case.

Finally, assume that case 3) above is satisfied. Then, for any $\varepsilon > 0$ small enough, there is a $\bar{s}$ such that $R^*(\bar{s}, P) = R - \varepsilon$. For this fixed value $\bar{s}$, since again $R_n \to R$ and $R_n^*(\bar{s}, P_n, F_{\bar{s}}) \to R^*(\bar{s}, P) = R - \varepsilon$, $R_n > R_n^*(\bar{s}, P_n, F_{\bar{s}})$ for all n large enough. Hence, for $s = \bar{s}$, for all n large enough equation (87) is not satisfied and thus (88) is. This implies that, for all n large enough

$$\frac{1}{n}\log\frac{1}{\mathsf{P}_{\text{e,max}}^{(n)}} < -\frac{1}{1-\bar{s}}\sum_x P_n(x)\mu_{S_x,F_{\bar{s}}}(\bar{s}) - \frac{\bar{s}}{1-\bar{s}}R_n^*(\bar{s}, P_n, F_{\bar{s}}) + \frac{1}{n}\left(2\bar{s}\sqrt{2\mu''(\bar{s})} + \frac{\log 8}{1-\bar{s}}\right). \tag{96}$$

In the limit as $n \to \infty$ the last term vanishes, $R_n^*(\bar{s}, P_n, F_{\bar{s}}) \to R^*(\bar{s}, P) = R - \varepsilon$ and $P_n \to P$. We thus conclude that

$$\begin{aligned} E(R,P) &\leq -\frac{1}{1-\bar{s}}\sum_x P(x)\mu_{S_x,F_{\bar{s}}}(\bar{s}) - \frac{\bar{s}}{1-\bar{s}}(R-\varepsilon) \\ &= E_0^{\text{cc}}\left(\frac{\bar{s}}{1-\bar{s}}, P\right) - \frac{\bar{s}}{1-\bar{s}}(R-\varepsilon) \\ &\leq \sup_{\rho\geq 0}\left(E_0^{\text{cc}}\left(\rho, P\right) - \rho(R-\varepsilon)\right) \\ &= E_{\text{sp}}^{\text{cc}}(R-\varepsilon, P). \end{aligned}$$

This holds for all $\varepsilon > 0$ small enough and hence, since $E_{\text{sp}}^{\text{cc}}(R,P)$ is non increasing in R, it holds for all $\varepsilon \in (0, R)$. This concludes the proof.

APPENDIX B
PROOF OF THEOREM 8

The proof is obtained by introducing a variation in the proof of theorem 1 presented in Appendix A. In particular, we use a different operator $\boldsymbol{F}$ which we choose so as to take into account the state-dependent structure of the communication process.

From the hypotheses, the communication is governed by the sequence of states $\boldsymbol{a} = (a_1, \ldots, a_n)$ with composition Q_n, where $Q_n \to Q$, and codes are considered with conditional compositions $\hat{P}_n$ given $\boldsymbol{a}$, where $\hat{P}_n \to \hat{P}$. Here again, as in the other proof, we can assume that $\hat{P}_n(x|a) = 0$ if $Q(a) = 0$ or $\hat{P}(x|a) = 0$. The structure of the proof remains unchanged with the only difference that, instead of building $\boldsymbol{F}$ using n identical copies of a single density operator F, we can use $|\mathcal{A}|$ different operators F_a, $a \in \mathcal{A}$ to build $\boldsymbol{F}$ as

$$\boldsymbol{F} = F_{a_1} \otimes F_{a_2} \otimes \cdots \otimes F_{a_n}. \tag{97}$$

Then we can still use the two equations (84) and (85) to bound the probability of error as a function of the rate, with the difference that the function $\mu(s)$ now reads

$$\mu_{\boldsymbol{S}_{\boldsymbol{x}_m},\boldsymbol{F}}(s) = n \sum_{a,x} Q_n(a) \hat{P}_n(x|a) \mu_{S_{a,x},F_a}(s). \tag{98}$$

For a given $a \in \mathcal{A}$ and fixed s, we then choose

$$F_{a,s} = \arg\min_{F} - \sum_{x} \hat{P}(x|a) \log(\operatorname{Tr} S_{a,x}^{1-s} F^s), \tag{99}$$

again ensuring that $\mu_{\boldsymbol{S}_{\boldsymbol{x}_m},\boldsymbol{F}}(s)$ is finite. The rest of the proof follows essentially identical with the obvious differences due to the use of quantities $E_0^{\mathrm{cc}}(\mathfrak{C}_a, \rho, \hat{P}(\cdot|a))$ in place of $E_0^{\mathrm{cc}}(\rho, P)$ used before.

APPENDIX C
A REMARK ON HAROUTUNIAN'S PROOF OF THE SHERE PACKING BOUND

As mentioned, a greedy extension of Haroutunian's proof of the sphere packing bound to quantum channels, as outlined in equation (14), gives a bound which is in general weak. The reason why this happens in the quantum case and not in the classical one can be traced back to a fundamental difference in the solution of the quantum binary hypothesis testing problem in those two contexts. In fact, as seen from equations (82) and (83), the key ingredient in the proof of the sphere packing bound is a binary hypothesis test to distinguish the state $\boldsymbol{S}_{\boldsymbol{x}_m}$ from the auxiliary state $\boldsymbol{F}$. Here, a fundamental difference with the classical counterpart is related to the roles of the Kullback-Leibler discrimination and Rényi divergence in the expression for the error exponents in binary hypothesis testing. This difference was already observed in [26, Sec. 4, Remark 1] and [14, Sec. 4.8] and leads to the mentioned difference in the expressions for the sphere packing bound. We discuss it here in detail for completeness.

In a binary hypothesis testing between two density operators A and B, based on n independent extractions, the error exponents of the first and second kind can be expressed parametrically as (see [14], [3])

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|A} = -\mu(s) + s\mu'(s) + o(1) \tag{100}$$

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|B} = -\mu(s) - (1-s)\mu'(s) + o(1) \tag{101}$$

where

$$\mu(s) = \log \operatorname{Tr} A^{1-s} B^s. \tag{102}$$

Upon differentiation, one finds

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|A} = -\log \operatorname{Tr}(A^{1-s}B^s) + \operatorname{Tr}\left[\frac{A^{1-s}B^s}{\operatorname{Tr} A^{1-s}B^s} \left(\log B^s - \log A^s\right)\right] + o(1) \tag{103}$$

$$-\frac{1}{n} \log \mathsf{P}_{\mathrm{e}|B} = -\log \operatorname{Tr}(A^{1-s}B^s) + \operatorname{Tr}\left[\frac{A^{1-s}B^s}{\operatorname{Tr} A^{1-s}B^s} \left(\log A^{1-s} - \log B^{1-s}\right)\right] + o(1) \tag{104}$$

In the classical case, A and B commute. We can then define the density operator $V = \frac{A^{1-s}B^s}{\operatorname{Tr} A^{1-s}B^s}$ (we use a simple notation keeping implicit the dependence of V on s) and use the properties $\log B^s - \log A^s = \log A^{1-s}B^s - \log A$ and $\log A^{1-s} - \log B^{1-s} = \log A^{1-s}B^s - \log B$ to obtain

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|A} = \operatorname{Tr} V(\log V - \log A) + o(1) \tag{105}$$

$$= D(V||A) + o(1) \tag{106}$$

and

$$-\frac{1}{n}\log \mathsf{P}_{\mathrm{e}|B} = \operatorname{Tr} V(\log V - \log A) + o(1) \tag{107}$$

$$= D(V||B) + o(1) \tag{108}$$

So, the choice of s induces a new state V which allows us to represent the error exponents as Kullback-Leibler divergences $D(V||A)$ and $D(V||B)$. However, this well-known classical formulation has no quantum counterpart [14, Sec. 4.8]. In the classical case, the property extends to the case of product of non-identical states and, hence, it can be adopted in the binary hypothesis test between the state $\boldsymbol{S}_{\boldsymbol{x}_m}$ and the auxiliary state $\boldsymbol{F}$ used in the sphere packing bound. If we assume that all the S_x operators and F commute, the exponents in equations (82) and (83) can be expressed in terms of conditional Kullback-Leibler divergences. In particular, for a given s, the choice of the operator F_s induces density operators

$$V_x = S_x^{1-s}F_s^s / \operatorname{Tr}(S_x^{1-s}F^s)$$

which allow us to rewrite the conditions in equations (82) and (83) as

$$\operatorname{Tr}\left[(\mathbb{1} - \Pi_m)\,\boldsymbol{S}_{\boldsymbol{x}_m}\right] > e^{-nD(V\|S|P_n)+o(n)} \tag{109}$$

or

$$\operatorname{Tr}\left[\Pi_m \boldsymbol{F}_{\boldsymbol{s}}\right] > e^{-nD(V\|F_s|P_n)+o(n)} \tag{110}$$

It also turns out that the operator F_s defined in equation (89), is such that (see [5, eq. (9.50)], [21, Cor. 3])

$$F_s = \sum_x P(x)V_x \tag{111}$$

so that

$$D(V\|F_s|P) = I(P,V).$$

Since we had the bound $\operatorname{Tr}\left[\Pi_m \boldsymbol{F}\right] \leq e^{-nR_n}$, the choice of the operator F_s can be interpreted as the choice of an auxiliary channel $\{V_x\}$ such that $I(P,V) < R$, which asymptotically gives an upper bound of $D(V\|S|P)$ on the error exponent, as discussed in Section III.

Haroutunian's proof of the sphere packing bound takes this interpretation as the starting point and directly bounds the probability of error for the original channel S in terms of an auxiliary, properly chosen channel V such that $I(P,V) < R$. For this auxiliary channel, the strong converse holds at rate R and this implies that for any decoding

rule, for at least one message the probability of error is $1 - o(1)$. In particular for the given POVM used on the original channel S, for at least one m we have

$$\mathrm{Tr}(I - \mathbf{\Pi}_m)\boldsymbol{V}_{\boldsymbol{x}_m} = 1 - o(1). \tag{112}$$

Using a data processing inequality for the Kullback-Leibler divergence (see [7] and [8]) one then finds that

$$\log \mathrm{Tr}(I - \mathbf{\Pi}_m)\boldsymbol{S}_{\boldsymbol{x}_m} \geq -\frac{nD(V\|S|P) + 1}{1 + o(1)}, \tag{113}$$

and thus

$$\frac{1}{n} \log \frac{1}{\mathsf{P}_{\mathrm{e}|\boldsymbol{S}_{\boldsymbol{x}_m}}} \leq \min_{V: I(P,V) \leq R} D(V\|S|P)(1 + o(1)). \tag{114}$$

For commuting states S_x, that is, classical channels, the resulting bound is equivalent to the one stated in Theorem 1. For non-commuting states, however, we expect some difference since we know that the conditions expressed in equations (106)-(108) do not hold and hence we do not expect expressions of the form (109)-(110) to be deducible from Theorem 1. Indeed, it is easy to see that the bound in (114) is not as good as the one in Theorem 1. In particular, if S is a pure-state channel, any auxiliary channel $V \neq S$ gives $D(V\|S|P) = \infty$, so that the bound is trivial for all pure-state channels. It is important to observe that this is not due to a weakness in the used data processing inequality. In a binary hypothesis test between the the pure state $\boldsymbol{S}_{\boldsymbol{x}_m}$ and a state $\boldsymbol{V}_{\boldsymbol{x}_m}$ built from a different channel V (that is, such that $V_x \neq S_x$ for at least one x with $P(x) > 0$), one can notice that the POVM $\{\tilde{\Pi}, I - \tilde{\Pi}\}$ with $\tilde{\Pi} = \boldsymbol{S}_{\boldsymbol{x}_m}$ satisfies

$$\mathrm{Tr}(I - \tilde{\Pi})\boldsymbol{V}_{\boldsymbol{x}_m} = 1 + o(1), \quad \mathrm{Tr}(I - \tilde{\Pi})\boldsymbol{S}_{\boldsymbol{x}_m} = 0. \tag{115}$$

So, it is really impossible to deduce a positive lower bound for $\mathrm{Tr}(I - \mathbf{\Pi}_m)\boldsymbol{S}_{\boldsymbol{x}_m}$ using only the fact that $\mathrm{Tr}(I - \mathbf{\Pi}_m)\boldsymbol{V}_{\boldsymbol{x}_m} = 1 + o(1)$.

We close by observing that one should not misinterpret the above discussion. The idea of using an auxiliary channel $\{V_x\}$ is general enough to include the approach used in Theorem 1 - call it the MIT approach - as a particular case. This is obtained with the use of a dummy auxiliary channel whose states V_x all equal F. Both the MIT and Haroutunian's proof follow in fact a common general scheme, which has been discussed in great detail in [27]. From this point of view, we observe that the MIT proof uses an auxiliary channel with $I(P, V) = 0$ for which the strong converse takes the simple form $\mathsf{P}_{\mathrm{e}} \geq 1 - e^{-nR}$. It then relates the probability of error under S to the strong converse exponent for V by using bounds on the error exponents in a binary hypothesis test between S and V. Haroutunian's proof chooses an auxiliary channel with $I(P, V) = R^-$ with an associated strong converse of the form $\mathsf{P}_{\mathrm{e}} = 1 - o(1)$, which allows one to compare S with V using a simple data processing inequality. For general classical-quantum channels this is not optimal because the simple strong converse $\mathsf{P}_{\mathrm{e}} = 1 - o(1)$ for V does not imply a lower bound on the error exponent for S as expressed by equation (115). In general, one might consider using some intermediate choice of channel V, its correct strong converse exponent as derived recently in [28], and then using bounds on the error exponents in a binary hypothesis test between S ad V. In the classical case one can show, by invoking result for error exponents with list-decoding, that no choice can do better than those two extremal ones. In the classical-quantum case, instead, the question remains open since no achievability

result matching the sphere packing bound is known for general non-commuting mixed states. Further details on this can be found in [29].

APPENDIX D
DEFINITION AND CONTINUITY OF F_s

We want to prove that, for a fixed distribution P, the operator

$$F_s = \arg\min_{F\in\mathcal{S}(\mathcal{H})} \left[-\sum_x P(x) \log(\operatorname{Tr} S_x^{1-s} F^s) \right]. \tag{116}$$

is a well defined, continuous function of s in the domain $s \in (0,1)$. We thus have to prove that the function

$$g(F,s) = -\sum_x P(x) \log(\operatorname{Tr} S_x^{1-s} F^s) \tag{117}$$

admits a unique minimum in F on $\mathcal{S}(\mathcal{H})$ for any fixed $s \in (0,1)$ and that the minimizing F depends continously on s.

In what follows, we consider the $\log$ as a continuous function from $[0,\infty]$ to $[-\infty,+\infty]$ where we set $\log(0) = -\infty$, $\log(+\infty) = +\infty$ and $1/0 = +\infty$. Then, for a continuous function f from a set D to $[0,+\infty]$, the function $\log(f(\cdot))$ is a continuous function from D to $[-\infty,+\infty]$.

Lemma 12: The function

$$g(F,s) = -\sum_x P(x) \log(\operatorname{Tr} S_x^{1-s} F^s) \tag{118}$$

is jointly continuous in (F,s) in the domain $\mathcal{S}(\mathcal{H}) \times (0,1)$ and has a minimum in F for any fixed $s \in (0,1)$.

Proof: For fixed $s \in (0,1)$ and $A, B \in \mathcal{S}(\mathcal{H})$, denoting by $\|\cdot\|$ the operator norm, by [30, Th. X.1.1] we have

$$\|A^s - B^s\| \leq \|A - B\|^s, \tag{119}$$

and hence continuity of $\operatorname{Tr} S_x^{1-s} F^s$ in F for any $s \in (0,1)$. Hence, with our assumption $\log(0) = -\infty$, $-\log(\operatorname{Tr} S_x^{1-s} F^s)$ is also continuous in F for any fixed s, and so is $g(F,s)$. Since $\mathcal{S}(\mathcal{H})$ is compact, this implies that for fixed s the function $F \mapsto g(F,s)$ achieves its minimum over $\mathcal{S}(\mathcal{H})$.

Furthermore, for fixed (A, s_1) and $\|A - B\| \leq \epsilon_1$, $|s_1 - s_2| \leq \epsilon_2$ we have

$$\begin{aligned}
\|A^{s_1} - B^{s_2}\| &\leq \|A^{s_1} - A^{s_2}\| + \|A^{s_2} - B^{s_2}\| \\
&\leq \max_i |\lambda_i(A)^{s_1} - \lambda_i(A)^{s_2}| + \|A - B\|^{s_2} \\
&\leq \max_{i:\lambda_i(A)>0} |s_1 - s_2| \cdot |\log(\lambda_i(A))| + \epsilon_1^{s_2} \\
&\leq \epsilon_2 |\log(\lambda_{\min}^+(A))| + \epsilon_1^{s_1-\epsilon_2},
\end{aligned}$$

where $\lambda_{\min}^+(A)$ is the smallest non-zero eigenvalue of A. This implies that $\operatorname{Tr} S_x^{1-s} F^s$ is continuous in (F,s) at any point $(A, s_1) \in \mathcal{S}(\mathcal{H}) \times (0,1)$, and hence $g(F,s)$ is also continuous there. ■

We are now ready to prove that F_s is a well defined continuous function of $s \in (0,1)$. We first show that for fixed $s \in (0,1)$, the minimum of $g(F,s)$ over F is only achieved on the subspace $\mathcal{S}_P(\mathcal{H})$ of operators with support in the subspace spanned by the supports of the S_x such that $P(x) > 0$. This follows from the data processing

inequality for the Rényi divergence ([31], [32]), which asserts that for any completely positive trace preserving (CPTP) map Φ we have

$$-\log \mathrm{Tr}(\Phi(S_x)^{1-s}\Phi(F)^s) \leq -\log \mathrm{Tr}(S_x^{1-s}F^s)\,. \tag{120}$$

Indeed, let V be the projector onto the subspace $\mathcal{H}_P$ spanned by the supports of the S_x with $P(x) > 0$, and let $W = \mathbb{1} - V$ be the projector onto the orthogonal complement of $\mathcal{H}_P$ in $\mathcal{H}$. Consider the CPTP map $\Phi : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ defined by

$$\Phi(A) = V^*AV + W^*AW \tag{121}$$

which reduces A to a block diagonal operator. Then we have $\Phi(S_x) = S_x$ for any S_x with $P(x) > 0$, so that by the data processing inequality (120) we have $g(\Phi(F), s) \leq g(F, s)$. Now note that since $\Phi(F)$ is block diagonal,

$$\Phi(F)^s = (V^*FV)^s + (W^*FW)^s\,, \tag{122}$$

so that for any x with $P(x) > 0$

$$\mathrm{Tr}\, S_x^{1-s}\Phi(F)^s = \mathrm{Tr}\, S_x^{1-s}(V^*FV)^s\,. \tag{123}$$

Now, if $W^*FW \neq 0$ then $\mathrm{Tr}(V^*FV) < 1$ and so there exists[4] $c > 1$ and a density operator $\tilde{F} = cV^*FV$ such that $\mathrm{Tr}(S_x^{1-s}\tilde{F}^s) > \mathrm{Tr}(S_x^{1-s}F^s)$, $\forall x : P(x) > 0$, which implies $g(F', s) < g(F, s)$. We deduce that $g(F, s)$ achieves its minimum over F only in the subspace $\mathcal{S}_P(\mathcal{H})$.

We now prove that the minimum is unique by showing that in $\mathcal{S}_P(\mathcal{H})$ the function $g(F, s)$ is strictly convex in F for fixed $s \in (0, 1)$. For $\alpha \in (0, 1)$ and distinct $A, B \in \mathcal{S}_P(\mathcal{H})$ we have

$$\begin{aligned}
\alpha g(A, s) + (1-\alpha)g(B, s) &= -\sum_x P(x)\left[\alpha \log(\mathrm{Tr}\, S_x^{1-s}A^s) + (1-\alpha)\log(\mathrm{Tr}\, S_x^{1-s}B^s)\right] \\
&\geq -\sum_x P(x) \log \mathrm{Tr}\left[S_x^{1-s}\left(\alpha A^s + (1-\alpha)B^s\right)\right] \\
&\overset{(a)}{>} -\sum_x P(x) \log \mathrm{Tr}\left[S_x^{1-s}\left(\alpha A + (1-\alpha)B\right)^s\right] \\
&= g(\alpha A + (1-\alpha)B, s)
\end{aligned}$$

where the first inequality is due to the concavity of the logarithm and the strict inequality (a) can be motivated as follows. Since the map $F \to F^s$ is operator concave for $0 < s < 1$, we have

$$(\alpha A + (1-\alpha)B)^s - (\alpha A^s + (1-\alpha)B^s) \geq 0\,,$$

so that

$$\forall x \quad \mathrm{Tr}\, S_x^{1-s}\left(\alpha A + (1-\alpha)B\right)^s \geq \mathrm{Tr}\, S_x^{1-s}\left(\alpha A^s + (1-\alpha)B^s\right)\,.$$

Furthermore, since the map $F \to F^s$ is actually *strictly* operator concave

$$(\alpha A + (1-\alpha)B)^s - (\alpha A^s + (1-\alpha)B^s) \neq 0\,.$$

[4]Of course we can assume $V^*FV \neq 0$, for otherwise $g(F, s) = \infty$ and hence F is surely not optimal anyway.

Let now ψ be an eigenvector of the left-hand side operator with positive eigenvalue. Since $A, B \in \mathcal{S}_P(\mathcal{H})$, then for at least one x we have $\langle\psi|S_x|\psi\rangle > 0$ and so, for that x,

$$\mathrm{Tr}\left[S_x^{1-s}\left[(\alpha A + (1-\alpha)B)^s - (\alpha A^s + (1-\alpha)B^s)\right]\right] > 0 \tag{124}$$

This, combined with the strictly increasing nature of the logarithm implies the strict inequality (a) and hence that $g(F, s)$ has a unique minimum in F.

We finally prove that F_s defined above is a continuous function of s. Assume instead that there exists a $\bar{s}$ and a sequence $\{s_i\}_{i=1,\ldots}$ such that $\lim_{i\to\infty} s_i = \bar{s}$ but F_{s_i} does not tend to $F_{\bar{s}}$. Assume also that $\lim_{i\to\infty} F_{s_i} = \tilde{F} \neq F_{\bar{s}}$ (otherwise we take an appropriate converging subsequence). Then by definition of F_{s_i} we have $g(F_{s_i}, s_i) \leq g(F_{\bar{s}}, s_i)$ and, using the continuity of g (Lemma 12) we deduce $g(\tilde{F}, \bar{s}) \leq g(F_{\bar{s}}, \bar{s})$. But by definition of $F_{\bar{s}}$ we have $g(F_{\bar{s}}, \bar{s}) \leq g(\tilde{F}, \bar{s})$ and hence $g(F_{\bar{s}}, \bar{s}) = g(\tilde{F}, \bar{s})$. The assumption that $\tilde{F} \neq F_{\bar{s}}$ now contradicts the fact that $g(F, s)$ has a unique minimizer for each s due to strict convexity on $\mathcal{S}_P(\mathcal{H})$.

REFERENCES

[1] M. Dalai and A. Winter, "Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2014.

[2] M. Dalai, "Sphere Packing Bound for Quantum Channels," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2012, pp. 160 – 164.

[3] ——, "Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 8027 – 8056, 2013.

[4] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[5] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. Wiley, New York, 1961.

[6] E. A. Haroutunian, "Estimates of the error exponents for the semi-continuous memoryless channel," *(in Russian) Probl. Peredachi Inform.*, vol. 4, no. 4, pp. 37–48, 1968.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[8] A. Winter, "Coding Theorems of Quantum Information Theory," *Ph.D. dissertation, Uni Bielefeld, arXiv:quant-ph/9907077*.

[9] M. Dalai, "Lovász's Theta Function, Rényi's Divergence and the Sphere-Packing Bound," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013, pp. 231–235.

[10] K. Marton, "On the Shannon Capacity of Probabilistic Graphs," *Journal of Combinatorial Theory, Series B*, vol. 57, no. 2, pp. 183 – 195, 1993.

[11] M. Dalai, "An Elias Bound on the Bhattacharyya Distance of Codes for Channels with a Zero-Error Capacity," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2014.

[12] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, 1956.

[13] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Statist.*, vol. 37, no. 2, pp. 1040–1057, 2009.

[14] K. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Communications in Mathematical Physics*, vol. 279, pp. 251–283, 2008, 10.1007/s00220-008-0417-5. [Online]. Available: http://dx.doi.org/10.1007/s00220-008-0417-5

[15] R. A. Medeiros, R. Alleaume, G. Cohen, and F. M. De Assis, "Zero-Error Capacity of Quantum Channels and Noiseless Subsystems," in *Proc. Int. Telecomm. Symposium*. Fortaleza, Brazil, 2006, pp. 900–905.

[16] I. Csiszár and J. Körner, "On the Capacity of the Arbitrarily Varying Channel for Maximum Probability of Error," *Zeitschrift für Wahrscheinlichkeitstheorie and Verwandte Gebiete*, vol. 57, no. 1, pp. 87–101, 1981.

[17] I. Csiszár, "Generalized Cutoff Rates and Rényi's Information Measures," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 26 –34, Jan. 1995.

[18] M. V. Burnashev and A. S. Holevo, "On the Reliability Function for a Quantum Communication Channel," *Probl. Peredachi Inform.*, vol. 34, no. 2, pp. 3–15, 1998.

[19] L. Lovász, "On the Shannon Capacity of a Graph," *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[20] R. Duan and A. Winter, "Zero-Error Classical Channel Capacity and Simulation Cost Assisted by Quantum Non-Signalling Correlations," *In preparation*, 2014.

[21] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, 1974.

[22] M. Dalai, "An "Umbrella" Bound of the Lovász-Gallager Type," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013, pp. 3025–3029.

[23] R. Blahut, "Composition bounds for channel block codes," *IEEE Trans. Inform. Theory*, vol. 23, no. 6, pp. 656 – 674, nov 1977.

[24] M. Dalai, "Elias Bound for General Distances and Stable Sets in Edge-Weighted Graphs," *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2335–2350, May 2015.

[25] A. S. Holevo, "Reliability Function of General Classical-Quantum Channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2256 –2261, Sep. 2000.

[26] H. Nagaoka, "The Converse Part of the Theorem for Quantum Hoeffding Bound," *arXiv:quant-ph/0611289v1*.

[27] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[28] M. Mosonyi and T. Ogawa, "Strong converse exponent for classical-quantum channel coding," *arXiv:1409.3562v5 [quant-ph]*, 2014.

[29] M. Dalai, "Classical and Classical-Quantum Sphere Packing Bounds: Rényi vs Kullback and Leibler," in *International Zurich Seminar 2016*, Mar 2016.

[30] R. Bhatia, *Matrix Analysis*, ser. Graduate Texts in Mathematics. Springer New York, 1997.

[31] D. Petz, "Quasi-entropies for finite quantum systems," *Rep. Math. Phys.*, vol. 23, pp. 57–65, 1986.

[32] M. Tomamichel, *Quantum Information Processing with Finite Resources*, ser. SpringerBriefs in Mathematical Physics. Springer, 2015, vol. 5.