An Elias Bound on the Bhattacharyya Distance of Codes for Channels with a Zero-Error Capacity

Marco Dalai

Department of Information Engineering
University of Brescia - Italy
Email: marco.dalai@unibs.it

Abstract—In this paper, we propose an upper bound on the minimum Bhattacharyya distance of codes for channels with a zero-error capacity. The bound is obtained by combining an extension of the Elias bound introduced by Blahut, with an extension of a bound previously introduced by the author, which builds upon ideas of Gallager, Lovász and Marton.

I. INTRODUCTION

An intriguing problem in the study of discrete memoryless channels (DMC) is that of determining the asymptotic behavior of the probability of error P_e of optimal codes in the low rate region. In the most general case, the probability of error is precisely zero at rates below the so called zero-error capacity C_0, while for $R > C_0$ it is known to be an exponential function of the block-length n, i.e.

$$\mathsf{P}_e \approx e^{-nE(R)},$$

where $E(R)$ is the so called reliability function of the channel. Both C_0 and $E(R)$ in the proximity of C_0 are unknown in the general case. The most effective upper bounds to C_0 and to $E(R)$ were developed independently and there is not yet a good unified and consistent upper bound to both quantities.

In recent works by this author, a possible approach for unifying bounds to C_0 and $E(R)$ was suggested which attempts at bounding the Bhattacharyya minimum distance of codes at rates $R > \vartheta$, where ϑ is Lovász' upper bound to C_0 [1]. However, the bounds derived in [2], [3] are rather crude and there seems to be room for great improvements. For example, when used for the binary symmetric channel, the bound in [3] gives essentially the simple Plotkin bound for the zero-rate minimum distance of codes. A useful progress with respect to [3] would be a refinement of the ideas to obtain a bound which is both valid in the case of zero-error capacity and not as bad in the case of no zero-error capacity.

In this paper, we make a first step in this direction by proposing an evolution of the idea presented in [3] to bound the Bhattacharyya minimum distance of codes on channels with a zero-error capacity. The obtained bound can be interpreted as an extension of the Elias bound to this setting, and it is based on a combination of ideas introduced by Elias, Blahut [4], Gallager [5], Lovász [1] and Marton [6].

II. ELIAS BOUNDS

Generalizations of the Elias bound to non-binary channels have already appeared in the literature. The main contributions in this direction are those of Berlekamp [7, Ch. 13], Blahut [4], and Piret [8]. All those extensions are based on some notion of distance $d(x, x')$ between symbols x, x', and distance $d(\boldsymbol{x}, \boldsymbol{x}')$ between codewords $\boldsymbol{x}$ and $\boldsymbol{x}'$, and follow a scheme based on two steps. For a given code, one first identifies a subset $\mathcal{T}$ of codewords which are all *packed* in a ball around a properly chosen fixed sequence $\bar{x}$. Then, the Plotkin bound is used to bound the minimum distance of the code in terms of the average distance between pairs of distinct codewords in $\mathcal{T}$ as

$$d_{\min} \leq \frac{1}{|\mathcal{T}|(|\mathcal{T}| - 1)} \sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{T}} d(\boldsymbol{x}, \boldsymbol{x}'). \tag{1}$$

An important point in this scheme is that the distance used for sequences must be based on the additive application of the distance $d(x, x')$ between symbols, which means that for sequences $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$ we have

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

This allows one to rewrite the average in eq. (1) in terms of the componentwise distances as

$$\sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{T}} d(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{T}} \sum_{i} d(x_i, x'_i) \tag{2}$$

$$= \sum_{i} \left(\sum_{\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{T}} d(x_i, x'_i) \right). \tag{3}$$

Then, the constraints on the compositions of the sequences $\boldsymbol{x}, \boldsymbol{x}'$ (and possibly $\bar{x}$) are used to derive the final bound in slightly different ways in the different approaches [7], [4], [8].

When one considers the case of general DMCs, the first problem is that of choosing a meaningful distance between sequences and symbols. If we are interested in understanding the probability of error of optimal codes, a reasonable approach is to consider the use of the Bhattacharyya distance between symbols[1]

$$d(x, x') = -\log \sum_{y} \sqrt{W(y|x)W(y|x')}, \tag{4}$$

---

[1]We point out that bounds on the Bhattacharyya distance of codes can be immediately mapped to bounds on the reliability function for certain symmetric channels, for example for pairwise reversible channels in the sense of [9] and (with obvious redefinitions of quantities) for classical-quantum pure-state channels. Due to space limitation, we leave the discussion of such applications of our bound to future work.

where $W(y|x)$ is the channel transition probability from input x to output y. Of the three mentioned generalizations, the only one which considers the case of general DMCs is that of Blahut, which actually studies the minimum Bhattacharyya distance of codes.

Blahut focuses on a subset of channels previously studied by Jelinek [10] and he considers the case of no zero-error capacity. There is a strong technical reason for this choice. In fact, if the channel has a zero-error capacity, the Bhattacharyya distance $d(x, x')$ is infinite for some pairs of inputs x, x'. So, optimal codes will in general contain pairs of codewords $\boldsymbol{x}, \boldsymbol{x}'$ with infinite distance, and any attempt to use the Plotking averaging procedure of equation (1) fails, since it gives the trivial bound $d_{\min} \leq \infty$.

In this paper, we propose an extension of the Elias bound for channels with a zero-error capacity by considering a variation of the Plotkin step. In a nutshell, since infinite distances arise from the use of the logarithm, we get rid of the logarithm or, equivalently, rather than averaging the pairwise distances $d(\boldsymbol{x}, \boldsymbol{x}')$, we average an exponential function of those distances. In particular, we use an approach which in a sense corresponds to substituting equation (1) with

$$d_{\min} \leq -\rho \log \left(\max_{\boldsymbol{x} \in \mathcal{T}} \frac{1}{(|\mathcal{T}| - 1)} \sum_{\boldsymbol{x}' \in \mathcal{T} \setminus \{\boldsymbol{x}\}} e^{-d(\boldsymbol{x}, \boldsymbol{x}')/\rho} \right). \quad (5)$$

There is a drawback of course, in that the derivation of the bound must now follow a different route, since it is no longer possible to use eq. (3). We approach the problem by proposing an extension of the umbrella bound originally introduced in [3]. That bound can in fact be interpreted as a variation of the Plotkin bound (1) in the form of equation (5), when there is no constraint on the composition of the codewords $\boldsymbol{x}, \boldsymbol{x}'$. Here, we propose an extension of the method that allows us to handle composition constraints as is usually done with equation (3).

In the next section, we introduce the notation and report the basic result of [3] for the reader's convenience. We then propose a way to deal with composition constraints and present the associated generalization of the Elias bound. We finally discuss how this bound relates to previously known ones.

III. $\vartheta(\rho)$ AND THE BASIC UMBRELLA BOUND

Let $W(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, be the transition probabilities of a discrete memoryless channel W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and a sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$, the probability of observing $\boldsymbol{y}$ at the output of the channel given $\boldsymbol{x}$ at the input is

$$W^{(n)}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i). \quad (6)$$

For a generic input symbol x, consider the unit norm $|\mathcal{Y}|$-dimensional column "state" vector ψ_x with components $\psi_x(y) = \sqrt{W(y|x)}$. In the same way, for an input sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, consider the unit norm $|\mathcal{Y}|^n$-dimensional column vector $\boldsymbol{\psi_x}$ whose components are the

values $\sqrt{W^{(n)}(\boldsymbol{y}|\boldsymbol{x})}$. Then, since the channel is memoryless, we can write

$$\boldsymbol{\psi_x} = \psi_{x_1} \otimes \psi_{x_2} \otimes \cdots \psi_{x_n} \quad (7)$$

where $\otimes$ is the Kronecker product.

The function $\vartheta(\rho)$ was derived in [3] as an extension of the Lovász theta function as follows. Consider the inner products between the channel state vectors $\psi_x^\dagger \psi_{x'} \geq 0$. For a fixed $\rho \geq 1$, an *orthonormal representation of degree ρ* of our channel W is a set of "tilted" unit norm vectors $\{\tilde{\psi}_x\}$ in any Hilbert space such that $|\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}| \leq (\psi_x^\dagger \psi_{x'})^{1/\rho}$. Call $\Gamma(\rho)$ the non-empty set of all possible such representations

$$\Gamma(\rho) = \left\{ \{\tilde{\psi}_x\} : |\tilde{\psi}_x^\dagger \tilde{\psi}_{x'}| \leq (\psi_x^\dagger \psi_{x'})^{1/\rho} \right\}, \quad \rho \geq 1. \quad (8)$$

The *value* of an orthonormal representation is the quantity

$$V(\{\tilde{\psi}_x\}) = \min_f \max_x \log \frac{1}{|\tilde{\psi}_x^\dagger f|^2}, \quad (9)$$

where the minimum is over all unit norm vectors f. The optimal choice of the vector f is called the *handle* of the representation. The function $\vartheta(\rho)$ is defined as the minimum value over all representations of degree ρ, that is,

$$\vartheta(\rho) = \min_{\{\tilde{\psi}_x\} \in \Gamma(\rho)} V(\{\tilde{\psi}_x\}). \quad (10)$$

The function $\vartheta(\rho)$ was used in [3] to derive (a stronger form of) the following bound.

Theorem 1: For any code of block-length n with M codewords and any $\rho \geq 1$, we have

$$\max_{m' \neq m} \boldsymbol{\psi}_m^\dagger \boldsymbol{\psi}_{m'} \geq \left(\frac{M e^{-n\vartheta(\rho)} - 1}{M - 1} \right)^\rho.$$

We observe that the proof of Theorem 1 is essentially based on the following result that we prove here for convenience.

Lemma 1: Let $v_1, \ldots, v_M$ and f be unit norm vectors such that $|v_i^\dagger f|^2 \geq c > 0$ for all i. Then

$$\max_{i \neq j} |v_i^\dagger v_j| \geq \frac{Mc - 1}{M - 1}$$

Proof: Let Φ be a matrix whose i-th column is v_i. Then, direct computation shows that

$$f^\dagger \Phi \Phi^\dagger f \geq Mc.$$

Since f is a unit norm vector, $\lambda_{\max}(\Phi \Phi^\dagger) \geq Mc$, where $\lambda_{\max}$ is the largest eigenvalue. This also implies $\lambda_{\max}(\Phi^\dagger \Phi) \geq Mc$. For a matrix A with elements $A_{i,j}$, it is known that

$$\lambda_{\max}(A) \leq \max_i \sum_j |A_{i,j}|. \quad (11)$$

Applying this to $A = \Phi^\dagger \Phi$ we obtain

$$\begin{aligned} Mc &\leq \lambda_{\max}(\Phi^\dagger \Phi) \\ &\leq \max_i \sum_j |v_i^\dagger v_j| \\ &\leq 1 + (M - 1) \max_{i \neq j} |v_i^\dagger v_j| \end{aligned}$$

which implies the statement of the lemma. ∎

Theorem 1 now follows by observing that, setting $\boldsymbol{f} = f^{\otimes n}$, for any sequence $\boldsymbol{x} = (x_1 \ldots, x_n)$ we have

$$|\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^{\dagger} \boldsymbol{f}|^2 = \prod_{i=1}^{n} |\tilde{\psi}_{x_i}^{\dagger} f|^2 \tag{12}$$

$$\geq e^{-n\vartheta(\rho)} \tag{13}$$

and, for any two sequences $\boldsymbol{x}, \boldsymbol{x}'$, $\boldsymbol{\psi}_{\boldsymbol{x}}^{\dagger} \boldsymbol{\psi}_{\boldsymbol{x}'} \geq |\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^{\dagger} \tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}|^{\rho}$.

IV. EXTENSION OF THE BOUND

A. Constant Composition Codes

The first step that we need to consider, for the development of a bound along the Elias scheme, is the extension of the umbrella bound to codes with a constant composition. For the bound derived in the previous section, the main property of the function $\vartheta(\rho)$ that we used is the property expressed in (13). There we see the reason for the definition of $\vartheta(\rho)$. We built a set of vectors $\{\tilde{\psi}_x\}$ associated to symbols, and a vector f such that f is "close" to all possible $\tilde{\psi}_x$. If we are interested in sequences with a particular composition, however, it can be preferable to pick f so that $|\tilde{\psi}_x^{\dagger} f|$ is larger for the symbols x which are used more frequently. This leads to a variation of $\vartheta(\rho)$ which is the analogue of the variation of the Lovász theta function introduced by Marton in [6].

For a distribution P and for $\rho \geq 1$, we define

$$\vartheta(\rho, P) = \min_{\{\tilde{\psi}_x\} \in \Gamma(\rho), f} \sum_x P(x) \log \frac{1}{|\tilde{\psi}_x^{\dagger} f|^2}. \tag{14}$$

With this definition, if $\boldsymbol{x}$ is a sequence with composition P, and $\{\tilde{\psi}_x\}$ is a representation with handle f achieving $\vartheta(\rho, P)$, we have

$$\begin{aligned}
|\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^{\dagger} \boldsymbol{f}|^2 &= \prod_{i=1}^{n} |\tilde{\psi}_{x_i}^{\dagger} f|^2 \tag{15} \\
&= \prod_x |\tilde{\psi}_x^{\dagger} f|^{2nP(x)} \tag{16} \\
&= e^{n \sum_x P(x) \log |\tilde{\psi}_x^{\dagger} f|^2} \tag{17} \\
&= e^{-n\vartheta(\rho, P)}. \tag{18}
\end{aligned}$$

Then, assume we have a code with M codewords $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ of composition P. We can apply Lemma 1 to the vectors $\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}_i}$ and then the inequality $\boldsymbol{\psi}_{\boldsymbol{x}}^{\dagger} \boldsymbol{\psi}_{\boldsymbol{x}'} \geq |\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}}^{\dagger} \tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}|^{\rho}$ to deduce that

$$\max_{m \neq m'} \boldsymbol{\psi}_{\boldsymbol{x}_m}^{\dagger} \boldsymbol{\psi}_{\boldsymbol{x}_{m'}} \geq \left(\frac{M e^{-n\vartheta(\rho, P)} - 1}{M - 1} \right)^{\rho} \tag{19}$$

$$\geq \left(e^{-n\vartheta(\rho, P)} - M^{-1} \right)^{\rho}. \tag{20}$$

Now, we see that if $R > \vartheta(\rho, P)$, as $n \to \infty$ the above quantity goes to zero as $e^{-n\rho\vartheta(\rho, P)}$.

Define then the asymptotic minimum distance

$$d(R, P) = \limsup_{n \to \infty} \max_{\mathcal{C}} \left[-\frac{1}{n} \log \max_{m \neq m'} \boldsymbol{\psi}_m^{\dagger} \boldsymbol{\psi}_{m'} \right] \tag{21}$$

where the maximum is over all codes of length n, rate at least R and compositions tending to P as $n \to \infty$. We have the following result.

Theorem 2: For any $\rho \geq 1$, if $R > \vartheta(\rho, P)$, then $d(R, P) \leq \rho\vartheta(\rho, P)$.

It is obvious from the definitions that $\vartheta(\rho, P) \leq \vartheta(\rho)$. Hence, even after optimization of the distribution P, the bound derived here is at least as good as the one that we can derive from Theorem 1. The variation introduced here is however also useful in the case of cost constraints.

B. The Elias Bound

We now extend further the definition of ϑ in order to apply the scheme developed by Blahut as a generalization of the Elias bound. What we need now is to extend the definition of $\vartheta(\rho, P)$ to deal with stochastic matrices. Given a distribution P and a $|\mathcal{X}| \times |\mathcal{X}|$ stochastic matrix $V(x'|x)$, we define

$$\vartheta(\rho, P, V) = \sum_x P(x) \vartheta(\rho, V(\cdot|x)) \tag{22}$$

$$= \min \sum_{x, x'} P(x) V(x'|x) \log \frac{1}{|\tilde{\psi}_{x, x'}^{\dagger} f_x|^2} \tag{23}$$

where the minimum is over all *sequences* of representations $\{\psi_{x,1}, \ldots, \psi_{x,|\mathcal{X}|}\} \in \Gamma(\rho)$, $x \in \mathcal{X}$ (one representation for each x) and over all sets of unit norm vectors $\{f_x\}$, $x \in \mathcal{X}$ (a different handle for each x).

Consider now the set of optimal representations and optimal handles which achieve $\vartheta(\rho, P, V)$. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be a sequence with composition P and define

$$\boldsymbol{f} = f_{x_1} \otimes f_{x_2} \cdots \otimes f_{x_n} \tag{24}$$

For a sequence $x' = (x_1', x_2', \ldots, x_n')$ which has a conditional composition V given the sequence x, consider the vector

$$\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'} = \tilde{\psi}_{x_1, x_1'} \otimes \tilde{\psi}_{x_2, x_2'} \cdots \otimes \tilde{\psi}_{x_n, x_n'} \tag{25}$$

Then, we have

$$\begin{aligned}
|\tilde{\boldsymbol{\psi}}_{\boldsymbol{x}'}^{\dagger} \boldsymbol{f}|^2 &= \prod_{i=1}^{n} |\tilde{\psi}_{x_i, x_i'}^{\dagger} f_{x_i}|^2 \tag{26} \\
&= \prod_{x, x'} |\tilde{\psi}_{x, x'}^{\dagger} f_x|^{2nP(x)V(x'|x)} \tag{27} \\
&= e^{n \sum_{x, x'} P(x)V(x'|x) \log |\tilde{\psi}_{x, x'}^{\dagger} f_x|^2} \tag{28} \\
&= e^{-n\vartheta(\rho, P, V)}. \tag{29}
\end{aligned}$$

Proceeding as we did in our previous bounds, if we have a set of M codewords all with a conditional composition V from a fixed sequence $\boldsymbol{x}$ with composition P, then

$$\max_{m \neq m'} \boldsymbol{\psi}_{\boldsymbol{x}_m}^{\dagger} \boldsymbol{\psi}_{\boldsymbol{x}_{m'}} \geq \left(\frac{M e^{-n\vartheta(\rho, P, V)} - 1}{M - 1} \right)^{\rho}. \tag{30}$$

In order to use this inequality for a given code, it is now necessary to consider the possible joint compositions of a subset of codewords with some given fixed sequence $\bar{\boldsymbol{x}}$. Given a code with $M = e^{nR}$ codewords of composition P, for a $\rho \geq 1$ and $\varepsilon > 0$, assume that there exists a stochastic matrix $V(x'|x)$ such that $nP(x)V(x'|x)$ is an integer,

$$\sum_x P(x) V(x'|x) = P(x') \tag{31}$$

(that we will write as $PV = P$), and

$$R \geq I(P,V) + \vartheta(\rho,P,V) + \varepsilon, \qquad (32)$$

where $I(P,V)$ is the mutual information with the notation of [11]. Then, (see [4], proof of Th. 8) there is at least one sequence $\bar{x}$ of composition P (not necessarily a codeword) such that there are at least $T = e^{n(\vartheta(\rho,P,V)+\varepsilon-o(1))}$ codewords with conditional composition V from $\bar{x}$. Let $\mathcal{T}$ be the set of such codewords, which plays the same role as in Section I. Then, for these codewords we can apply the bound of equation (30) with T in place of M. Considering the first order exponent, we then deduce that

$$-\frac{1}{n} \log \max_{m \neq m'} \boldsymbol{\psi}_m^\dagger \boldsymbol{\psi}_{m'} \leq \rho\vartheta(\rho,P,V) + o(1). \qquad (33)$$

For fixed n, the choice of V is constrained to satisfy the usual type constraints, but asymptotically as $n \to \infty$ this constraints can be neglected. As a consequence, we have the following theorem.

Theorem 3: For given R, P and $\rho \geq 1$, let V be a $|\mathcal{X}| \times |\mathcal{X}|$ stochastic matrix such that $PV = P$. If $R > I(P,V) + \vartheta(\rho,P,V)$, then $d(R,P) \leq \rho\vartheta(\rho,P,V)$.

Remark 1: We observe that with the choice $V(x'|x) = P(x')$ we have $PV = P$, $I(P,V) = 0$ and $\vartheta(\rho,P,V) = \vartheta(\rho,P)$. Hence, if $R > \vartheta(\rho,P)$ for a given ρ, the particular choice $V(x'|x) = P(x')$ gives the same bound of Theorem 2, which is thus included as a particular case in Theorem 3.

V. AN ANALYSIS OF THE BOUND

A. Binary Channels

Since most readers are probably familiar with the original Elias bound, we first give evidence that the proposed bound is a generalization by showing in detail how the original bound for binary channels is recovered as a special case. This shows that, even in the binary case, there is no loss in the use of equation (5) with the approach based on ϑ with respect to the standard use of the Plotkin bound (1) under composition constraints. In particular, the original bound for binary channels is obtained in the limit $\rho \to \infty$.

For a binary channel, let $Z = -\log \psi_0^\dagger \psi_1$ be the Battacharyya distance between the two inputs (0 and 1). Then, for any ρ it is not difficult to see that one can always take as an optimal representation of degree ρ the two-dimensional vectors

$$\tilde{\psi}_0 = [\cos(\alpha), \sin(\alpha)]^\dagger$$
$$\tilde{\psi}_1 = [\cos(\alpha), -\sin(\alpha)]^\dagger$$

where α satisfies $\cos(2\alpha) = e^{-Z/\rho}$. For a given distribution Q, let the optimal handle which achieves $\vartheta(\rho,Q)$ be

$$f = [\cos(\beta), \sin(\beta)]^\dagger.$$

Then

$$\vartheta(\rho,Q) = -2Q(0)\log\cos(\alpha-\beta) - 2Q(1)\log\cos(\alpha+\beta). \qquad (34)$$

where the value of β can be determined by minimizing this expression. Upon differentiation and a little of algebra we find

$$\sin(2\beta) = (Q(0) - Q(1))\sin(2\alpha). \qquad (35)$$

The value of $\vartheta(\rho,Q)$ can now be computed analytically by using this relation in (34). The resulting expression is complicated and not very useful here. So, we only study the bound of Theorem 3 asymptotically obtained by letting $\rho \to \infty$. We also only study the bound obtained for the uniform composition P, since we already know that this is the interesting case for the original Elias bound.

First note that, for any V, $\vartheta(\rho,P,V) \to 0$ as $\rho \to \infty$, which means that we can obtain a bound for any R by choosing V such that $I(P,V) < R$. Let us then choose V such that $V(1|0) = V(0|1) = \lambda$, with λ such that $I(P,V) = 1 - h(\lambda) < R$, where $h(\cdot)$ is the binary entropy function. If we set $Q = V(\cdot|0)$, then by symmetry we have $\vartheta(\rho,P,V) = \vartheta(\rho,Q)$. In the limit $\rho \to \infty$, since $\cos(2\alpha) = e^{-Z/\rho}$, we have $\alpha \to 0$, and from equation (35) we deduce that $\beta \approx \alpha(1-2\lambda)$. The expression for $\vartheta(\rho,Q)$ is then asymptotically

$$\vartheta(\rho,Q) \approx -2(1-\lambda)\log\cos(2\lambda\alpha) - 2\lambda\log\cos(2(1-\lambda)\alpha)$$
$$\approx (1-\lambda)(4\lambda^2\alpha^2) + \lambda(4(1-\lambda)^2\alpha^2)$$
$$= 4\lambda(1-\lambda)\alpha^2.$$

Using again the relation $e^{-Z/\rho} = \cos(2\alpha)$ we deduce that

$$\rho = \frac{-Z}{\log\cos(2\alpha)} \qquad (36)$$
$$\approx \frac{Z}{2\alpha^2}. \qquad (37)$$

So, $\rho\vartheta(\rho,Q) \approx 2\lambda(1-\lambda)Z$. The bound of Theorem 3 states that for $R > \vartheta(\rho,P,V) + I(P,V)$ we have $d(R,P) \leq \rho\vartheta(\rho,P,V)$. Since here $\vartheta(\rho,P,V) = \vartheta(\rho,Q) \to 0$ as $\rho \to \infty$, in this limit the theorem says that if $R > 1 - h(\lambda)$ then $d(R) \leq 2\lambda(1-\lambda)Z$. This is an equivalent formulation of the Elias bound. One may wonder whether for finite ρ a better bound can be obtained. Unfortunately, a rigorous analysis seems to be painful, but numerical evaluation shows that this is not the case, the optimal bound is achieved as $\rho \to \infty$.

B. Non-Negative Definite Channels and Euclidean Space Codes

The detailed analysis of the bound obtained for the BSC as $\rho \to \infty$ can be extended to all non-negative definite channels without a zero-error capacity. In this case, the bound obtained as $\rho \to \infty$ is precisely the same as that of Blahut. Due to space limitation, we can only give a sketch of the proof. For a fixed value of x, consider the quantity $\vartheta(\rho,V(\cdot|x))$ which appears in the definition (22). Let for ease of notation $Q = V(\cdot|x)$, so that we can focus on the evaluation of $\vartheta(\rho,Q)$ for a general Q and get rid of x. As mentioned in [3], for these channels, for any $\rho \geq 1$, representations of degree ρ exist which meet the constraints $\tilde{\psi}_{x_1}^\dagger \tilde{\psi}_{x_2} \leq \psi_{x_1}^\dagger \psi_{x_2}^{1/\rho}$ with equality. All these vectors tend to concentrate in a small cap on the unit sphere as $\rho \to \infty$, and $\vartheta(\rho,Q) \to 0$. Using the asymptotic expansions $\sin(2t) \approx$

$2t$ and $\log(\cos^2(t)) \approx -t^2$, valid for $t \to 0$, one finds that the optimal choice of the handle f for achieving $\vartheta(\rho, Q)$ is asymptotically the center of mass of the points $\tilde\psi_x$ (if vectors are interpreted as points and Q as a mass distribution). Then one comes to the conclusion that, as $\rho \to \infty$,

$$\rho\vartheta(\rho, Q) \to - \sum_{x_1,x_2} Q(x_1)Q(x_2) \log \psi_{x_1}^\dagger \psi_{x_2}. \qquad (38)$$

Furthermore, since $\vartheta(\rho, Q) \to 0$, the constraint on the rate becomes $R > I(P, V)$, and the bound on the distance takes the same form as Blahut's one. So, our bound is actually a generalization of Blahut's to general channels possibly with a zero-error capacity.

Finally, we point out that the bound derived by Piret for the squared euclidean distance of codes on the unit circle can also be obtained as a particular case of our bound. This is however due to a rather interesting independent fact, namely that for any choice of points in a euclidean space there exists a set of unit norm vectors in some other space whose pairwise Bhattachryya distances are precisely the squared euclidean distances between the original points. These vectors trivially satisfy the properties required for non-negative definite channels and thus Blahut's bound applies. In fact, Piret's bound can be recast as a special case of Blahut's one and it is thus also included in our method.

C. Complexity

The proposed bound has a non-trivial complexity and the evaluation of the optimal choice of ρ and V for a given channel, given P and R does not seem to be simple. It must be stressed, however, that for any choice of ρ and V we obtain a bound for the rate $R = \vartheta(\rho, P, V) + I(P, V)$. There are two main factors that should be analyzed for a deeper understanding of whether such a high complexity is reasonable or not for these kind of bounds. One reason is that we have no closed form expression for the function $\vartheta(\rho, P, V)$, and this prevents any particularly interesting analytic study of the resulting bound. This is due to the fact that we want to cope with channels with a zero-error capacity and that we chose to build upon the work of Lovász, since it is the most effective in this sense, which also does not lead to closed form expressions for bounds to C_0.

Another source of complexity, instead, seems to be intrinsic in all attempts to generalize the Elias bound. We should spend a few words on this. In [7], Berlekamp uses a slightly different approach when compared to [4] and [8] and, since he considers only the case of the Hamming and Lee metrics, he uses a symmetry argument to derive a bound which has a simpler form. Still, it can be checked that the bound does not give a closed form relation between R and the minimum distance, since it involves the inversion of [7, eq. (13.15)], which requires a non trivial computation. Furthermore, in the general case where there is no symmetry in the distances, this method cannot be applied.

The bound proposed by Piret in [8], as well, is only valid for a certain symmetric setting but, as mentioned before, it is much related to Blahut's. Using the symmetry, he obtains a bound that can be expressed in terms of one distribution (his β, in place of our stochastic matrix V), but even there, there is no closed form expression for the optimal distribution to choose for a given rate R, although they conjecture what it could be (see discussion after [8, eq. (41)]).

Finally, the only approach which works for DMCs is the one proposed by Blahut. He gives a complete description of his bound $E_U(R)$ in a form (see his Definition 3) which looks very similar to classical bounds to the reliability function[2] $E(R)$. Unfortunately, however, the computation of this function is in our opinion much more difficult than expected. The problem is that there seems to be a sign error in the proof of Lemma 5 which leads to erroneously consider his function $F(P)$ convex while it is actually concave. So, in our opinion, Theorem 7 is not valid, and in the definition of $E_U(R)$ we have a *minimization* of a concave function over a convex set. Thus, the evaluation of $E_U(R)$ it much more difficult than expected.

We close by pointing out that this concavity issue of Blahut's $F(P)$ mentioned above is essentially the same reason which prevents a closed form expression for Piret's bound. It is actually even present in the original case of binary channels, although with a trivial solution; it is the point where, for a give R, we need to find infimum value of the concave function $2\lambda(1-\lambda)$ under the constraint that $1 - h(\lambda) < R$. On the other hand, this concavity is also a key property which is needed in the original Elias bound and in all above mentioned extensions, see [9, after (2.46)], [7, after (13.48)], [4, Lemma 6] and [8, Lemma 4.2]. Hence, we doubt that a simpler solution could ever been found along these lines. It may just be more effective to find an empirically good selection of V as suggested by Piret for his β.

REFERENCES

[1] L. Lovász, "On the Shannon Capacity of a Graph," *IEEE Trans. Inform. Theory*, vol. 25, no. 1, pp. 1–7, 1979.
[2] M. Dalai, "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.
[3] ——, "An "Umbrella" Bound of the Lovász-Gallager Type," in *Proc. IEEE Intern. Symp. Inform. Theory*, 2013, pp. 3025–3029.
[4] R. Blahut, "Composition bounds for channel block codes," *IEEE Trans. Inform. Theory*, vol. 23, no. 6, pp. 656 – 674, nov 1977.
[5] R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, 1965.
[6] K. Marton, "On the Shannon Capacity of Probabilistic Graphs," *Journal of Combinatorial Theory, Series B*, vol. 57, no. 2, pp. 183 – 195, 1993.
[7] E. Berlekamp, *Algebraic Coding Theory*, ser. McGraw-Hill series in systems science. Aegean Park Press, 1984.
[8] P. Piret, "Bounds for Codes Over the Unit Circle," *Information Theory, IEEE Transactions on*, vol. 32, no. 6, pp. 760–767, 1986.
[9] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding in Discrete Memoryless Channels. II," *Information and Control*, vol. 10, pp. 522–552, 1967.
[10] F. Jelinek, "Evaluation of Expurgated Error Bounds," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 501–505, 1968.
[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

---

[2] We focus on his function $E_U(R)$ only as an upper bound on the minimum Bhattacharyya distance in this paper. The use of $E_U(R)$ as an upper bound to $E(R)$ seems to be valid only for pairwise reversible channels.