

# Hypothesis testing and quasi-perfect codes

**Conference Paper**

**Author(s):**

Vazquez-Vilar, Gonzalo; Guillén i Fàbregas, Albert; Verdú, Sergio

**Publication date:**

2016

**Permanent link:**

<https://doi.org/10.3929/ethz-a-010645813>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

# Hypothesis Testing and Quasi-Perfect Codes

Gonzalo Vazquez-Vilar  
 Universidad Carlos III de Madrid  
 gvazquez@ieee.org

Albert Guillén i Fàbregas  
 ICREA & Universitat Pompeu Fabra  
 University of Cambridge  
 guillen@ieee.org

Sergio Verdú  
 Princeton University  
 verdu@princeton.edu

**Abstract**—Hypothesis testing lower bounds to the channel coding error probability are studied. For a family of symmetric channels, block lengths and coding rates, the error probability of the best code is shown to coincide with that of a binary hypothesis test with certain parameters. The points in which they coincide, are precisely the points at which perfect or quasi-perfect codes exist. General conditions are given for a code to attain minimum error probability.

## I. INTRODUCTION

Consider the channel coding problem of transmitting a set of messages over a binary symmetric channel (BSC). The sphere-packing bound [1, Eq. (5.8.19)] establishes a lower bound on the block error probability of a code with a given rate and blocklength. This bound follows from counting the maximum number of non-overlapping Hamming spheres that can be packed in the output space. In certain cases the sphere-packing bound is achievable. A binary code is said to be *perfect* if non-overlapping Hamming spheres of radius  $t$  centered on the codewords exactly fill out the space. Perfect codes are a subset of the class of quasi-perfect codes. A *quasi-perfect* code is defined as a code in which Hamming spheres of radius  $t$  centered on the codewords are non-overlapping and Hamming spheres of radius  $t+1$  cover the space, possibly with overlaps. Since quasi-perfect codes attain the sphere-packing bound for a BSC, they achieve the minimum error probability among all the codes with the same block length and rate [1, Sec. 5.8]. However, these codes are rare. For each rate  $R$ ,  $0 < R < 1$ , there exists a block length beyond which neither perfect nor quasi-perfect codes exist [2], [3].

A generalization of the definition of perfect and quasi-perfect codes beyond the Hamming space was proposed by Hamada in [4]. Using a variation of the Fano metric, Hamada derived a lower bound to the channel coding error probability. This bound is achievable by perfect and quasi-perfect codes (defined with respect to the new metric), whenever they exist. This result applies for a class of symmetric discrete memoryless channels.

Binary hypothesis testing has been shown instrumental in the derivation of converse bounds (see e.g. [5], [6]), one prominent recent example being the the meta-converse bound

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the Spanish Ministry of Economy and Competitiveness under grants TEC2012-38800-C03-03, TEC2013-41718-R and FPDI-2013-18602, by the US National Science Foundation under Grant CCF-1016625, and by the Center for Science of Information, an NSF Science and Technology Center under Grant CCF-0939370.

by Polyanskiy *et al.* [7, Th. 27]. Particularized for the BSC, the meta-converse bound recovers the sphere-packing bound [1, Eq. (5.8.19)] (see [7, Sec. III.H] for details). As a result, when perfect or quasi-perfect codes exist, the the meta-converse bound gives the minimum error probability in the BSC.

In this work, we generalize the definitions of perfect and quasi-perfect codes for a class of symmetric channels and we establish a connection between hypothesis testing lower bounds and perfect or quasi-perfect codes. The results of this paper are general enough to recover Hamada's condition for achieving minimum error probability [4, Th. 3].

## II. GENERALIZED QUASI-PERFECT CODES

Consider the one-shot channel coding problem, where an equiprobable message  $v \in \{1, \dots, M\}$  is to be transmitted over a random transformation  $P_{Y|X}$ ,  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$  with  $\mathcal{X}$  and  $\mathcal{Y}$  discrete alphabets. A channel code  $\mathcal{C}$  is defined as the set of  $M$  codewords  $\mathcal{C} = \{x_1, \dots, x_M\}$  assigned to each of the messages. We assume that the maximum likelihood (ML) rule is used to choose the decoded message  $\hat{v} \in \{1, \dots, M\}$ . The error probability is given by

$$\epsilon(\mathcal{C}) = \Pr[\hat{V} \neq V] \quad (1)$$

$$= 1 - \frac{1}{M} \sum_y \max_{x \in \mathcal{C}} P_{Y|X}(y|x). \quad (2)$$

*Definition 1:* A discrete channel is *symmetric* if the rows of the transition matrix of the channel (with inputs as rows and outputs as columns), i. e.,  $P_{Y|X}(\cdot|x)$ , are permutations of each other.

This definition of symmetric channels coincides with that of uniformly dispersive channels of Massey [8, Sec. 4.2] and is less restrictive than those of Cover and Thomas [9] and Gallager [1]. The definition in [9, Sec. 7.2] additionally requires that the columns of the channel transition matrix be permutations of each other, i.e., uniformly focusing according to [8, Sec. 4.2]. The definition in [1, p. 94] requires the channel transition matrix to be partitioned in submatrices such that each submatrix fulfills the condition in [9, Sec. 7.2]. Relations among these definitions are investigated in [10, Sec. VI.B].

We define  $\mathcal{S}_x(\theta)$  to be the set of output sequences  $y$  with a likelihood given input  $x$  of at least  $\theta \in [0, 1]$ , i. e.,

$$\mathcal{S}_x(\theta) \triangleq \left\{ y \in \mathcal{Y} \mid P_{Y|X}(y|x) \geq \theta \right\}. \quad (3)$$

We denote the interior and the shell of  $\mathcal{S}_x(\theta)$ , respectively, as (9), given by

$$\mathcal{S}_x^\bullet(\theta) \triangleq \left\{ y \in \mathcal{Y} \mid P_{Y|X}(y|x) > \theta \right\}, \quad (4)$$

$$\mathcal{S}_x^\circ(\theta) \triangleq \left\{ y \in \mathcal{Y} \mid P_{Y|X}(y|x) = \theta \right\}. \quad (5)$$

Although we are not assuming that the input and output alphabets are identical and  $P_{Y|X}(y|x)$  (or the related Fano metric  $\sim \log P_{Y|X}(y|x)$ ) do not fulfill the properties of a mathematical distance in general, we refer to  $\mathcal{S}_x(\theta)$  as a sphere of radius  $\theta$  centered on  $x$ . For specific channels, such as the binary symmetric channel,  $\log P_{Y|X}(y|x)$  is an affine function of the Hamming distance between  $x$  and  $y$  and hence  $\mathcal{S}_x(\theta)$  becomes a sphere with respect to that distance.

*Proposition 1:* Let  $P_{Y|X}(y|x)$  be a symmetric channel defined over input and output alphabets  $\mathcal{X}, \mathcal{Y}$ . Then, cardinalities (or “volumes”)  $|\mathcal{S}_x(\theta)|, |\mathcal{S}_x^\bullet(\theta)|, |\mathcal{S}_x^\circ(\theta)|$  are independent of  $x$ .

Then, for any symmetric channel, we define  $S(\theta) \triangleq |\mathcal{S}_x(\theta)|$ ,  $S_\bullet(\theta) \triangleq |\mathcal{S}_x^\bullet(\theta)|$ ,  $S_\circ(\theta) \triangleq |\mathcal{S}_x^\circ(\theta)|$ . Obviously,  $S(\theta) = S_\bullet(\theta) + S_\circ(\theta)$ .

*Definition 2:* A code is *perfect* if there exists  $\theta \in [0, 1]$  such that

$$\bigcup_{x \in \mathcal{C}} \mathcal{S}_x(\theta) = \mathcal{Y}, \quad (6)$$

where the union is disjoint. More generally, a code is *quasi-perfect* if there exists  $\theta \in [0, 1]$  such that (6) is satisfied and the codeword-centered spheres  $\{\mathcal{S}_x(\theta), x \in \mathcal{C}\}$  are disjoint.

This definition of perfect codes coincides with that in [4, Def. 1] when the channel fulfills the Properties 1-4 in [4]. Definition 2 applies however to any symmetric channel according to 1 (which corresponds to Property 4 in [4]). Also, the definition of quasi-perfect code in Definition 2 includes both perfect and quasi-perfect codes from [4, Def. 1].

### III. THE META-CONVERSE BOUND

Let  $\hat{H} \in \{0, 1\}$  be the random variable associated to the output of a binary hypothesis test discriminating between distributions  $P$  (hypothesis 0) and  $Q$  (hypothesis 1). Then, the test can be described by the conditional distribution  $P_{\hat{H}|Y}$ . Let  $\pi_{j|i}$  denote the probability of deciding  $j$  when  $i$  is the true hypothesis. More precisely, we define

$$\pi_{0|1} \triangleq \sum_y Q(y) P_{\hat{H}|Y}(0|y), \quad (7)$$

$$\pi_{1|0} \triangleq \sum_y P(y) P_{\hat{H}|Y}(1|y). \quad (8)$$

Let  $\alpha_\beta(P, Q)$  denote the minimum error probability  $\pi_{1|0}$  among all tests  $T \triangleq P_{\hat{H}|Y}$  with  $\pi_{0|1}$  at most  $\beta$ , that is

$$\alpha_\beta(P, Q) \triangleq \inf_{T: \pi_{0|1} \leq \beta} \pi_{1|0}. \quad (9)$$

In [11], Neyman and Pearson derived the explicit form of a (possibly randomized) test  $T$  achieving the optimum trade-off

$$T_{\text{NP}}(0|y) = \begin{cases} 1, & \text{if } \frac{P(y)}{Q(y)} > \gamma, \\ p, & \text{if } \frac{P(y)}{Q(y)} = \gamma, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where  $\gamma \geq 0$  and  $p \in [0, 1]$  are parameters chosen such that  $\pi_{0|1} = \beta$ .

Let  $P_X^C$  denote the channel input distribution induced by the codebook  $\mathcal{C} = \{x_1, \dots, x_M\}$ , i. e.,

$$P_X^C(x) \triangleq \frac{1}{M} \sum_{m=1}^M \mathbb{1}\{x = x_m\}, \quad (11)$$

where  $\mathbb{1}\{\cdot\}$  denotes the indicator function.

It has been shown in [12, Th. 1] that the exact error probability  $\epsilon(\mathcal{C})$  in (2) can be expressed as the best type-0 error probability of an induced binary hypothesis test discriminating between the original distribution  $P_X^C \times P_{Y|X}$  and an alternative product distribution  $P_X^C \times Q_Y$  with type-1-error equal to  $\frac{1}{M}$ , i. e.,

$$\epsilon(\mathcal{C}) = \max_{Q_Y} \left\{ \alpha_{\frac{1}{M}} \left( P_X^C \times P_{Y|X}, P_X^C \times Q_Y \right) \right\}. \quad (12)$$

The right hand side of Eq. (12) is precisely the meta-converse bound [7, Th. 26] after optimization over the auxiliary distribution  $Q_Y$ . By choosing the auxiliary output distribution  $\bar{Q}_Y(y) = |\mathcal{Y}|^{-1}$  and minimizing over all distributions defined over the input alphabet  $\mathcal{X}$ , identity (12) can be weakened to obtain

$$\epsilon(\mathcal{C}) \geq \inf_{P_X} \left\{ \alpha_{\frac{1}{M}} \left( P_X \times P_{Y|X}, P_X \times \bar{Q}_Y \right) \right\}. \quad (13)$$

For the class of symmetric channels considered in Definition 1, we resort to the Neyman-Pearson lemma to find an alternative expression for right-hand side of (13). This expression will be then shown to coincide with the exact error probability  $\epsilon(\mathcal{C})$  when  $\mathcal{C}$  is a quasi-perfect code according to Definition 2.

### IV. OPTIMAL CODE STRUCTURE

We particularize the Neyman-Pearson test (10) with  $P \leftarrow P_X \times P_{Y|X}$  and  $Q \leftarrow P_X \times Q_Y$ ,

$$T_{\text{NP}}(0|x, y) = \begin{cases} 1, & \text{if } y \in \mathcal{S}_x^\bullet(\theta), \\ p, & \text{if } y \in \mathcal{S}_x^\circ(\theta), \\ 0, & \text{otherwise,} \end{cases} \quad (14)$$

where  $\theta = \gamma|\mathcal{Y}|^{-1}$  and  $p \in [0, 1]$  are parameters that allow to balance  $\pi_{1|0}$  and  $\pi_{0|1}$ . We proceed to analyze the two error types.

Substituting (14) in (7) we obtain

$$\pi_{0|1} = \sum_{x, y} P_X(x) \bar{Q}_Y(y) T_{\text{NP}}(0|x, y) \quad (15)$$

$$= |\mathcal{Y}|^{-1} \sum_x P_X(x) \left( |\mathcal{S}_x^\bullet(\theta)| + p |\mathcal{S}_x^\circ(\theta)| \right) \quad (16)$$

$$= |\mathcal{Y}|^{-1} \left( S_\bullet(\theta) + p S_\circ(\theta) \right). \quad (17)$$

Given the constraint on  $\pi_{0|1}$  imposed by (13), and the structure of the Neyman-Pearson test, the parameters  $p, \theta \in [0, 1]$  are chosen such that  $\pi_{0|1} = \frac{1}{M}$ , i.e.,

$$S_{\bullet}(\theta) + pS_{\circ}(\theta) = \frac{|\mathcal{Y}|}{M}. \quad (18)$$

Substituting (14) in (8) we obtain

$$\begin{aligned} \pi_{1|0} &= 1 - \sum_{x,y} P_X(x)P_{Y|X}(y|x)T_{\text{NP}}(0|x,y) \\ &= 1 - \sum_x P_X(x) \left( \sum_{y \in \mathcal{S}_{\bullet}^{\circ}(\theta)} P_{Y|X}(y|x) \right. \\ &\quad \left. + p \sum_{y \in \mathcal{S}_{\bullet}^{\circ}(\theta)} P_{Y|X}(y|x) \right). \end{aligned} \quad (19)$$

For an arbitrary  $x$ , let  $P_{Y|X}(y_i|x)$ ,  $i = 1, \dots, |\mathcal{Y}|$ , denote the output likelihoods indexed in decreasing order. Given the symmetry condition in Definition 1, the vector  $(P_{Y|X}(y_1|x), \dots, P_{Y|X}(y_{|\mathcal{Y}}|x))$  does not depend on the specific value of  $x$ . Then, for any  $x$ , we define  $\psi_i \triangleq P_{Y|X}(y_i|x)$ ,  $i = 1, \dots, |\mathcal{Y}|$ , and rewrite (20) as

$$\pi_{1|0} = 1 - \left( \sum_{i=1}^{S_{\bullet}(\theta)} \psi_i + p \sum_{i=1}^{S_{\circ}(\theta)} \psi_{i+S_{\bullet}(\theta)} \right). \quad (21)$$

Using (18) and (21), it follows that the lower bound (13) can be rewritten as

$$\epsilon(\mathcal{C}) \geq 1 - \left( \sum_{i=1}^{S_{\bullet}(\theta)} \psi_i + p \sum_{i=1}^{S_{\circ}(\theta)} \psi_{i+S_{\bullet}(\theta)} \right), \quad (22)$$

where  $p, \theta \in [0, 1]$  are such that  $S_{\bullet}(\theta) + pS_{\circ}(\theta) = \frac{|\mathcal{Y}|}{M}$ .

The next result shows that for a quasi-perfect code  $\mathcal{C}$ , (22) holds with equality. That is, when they exist, quasi-perfect codes attain the minimum error probability.

*Theorem 1:* Let  $P_{Y|X}$  be a symmetric channel according to Definition 1 and let  $\mathcal{C}$  be a quasi-perfect code according to Definition 2. Then,

$$\epsilon(\mathcal{C}) = 1 - \left( \sum_{i=1}^{S_{\bullet}(\theta)} \psi_i + p \sum_{i=1}^{S_{\circ}(\theta)} \psi_{i+S_{\bullet}(\theta)} \right), \quad (23)$$

where  $p, \theta \in [0, 1]$  are such that  $S_{\bullet}(\theta) + pS_{\circ}(\theta) = \frac{|\mathcal{Y}|}{M}$ .

*Proof:* Before showing that (23) holds with equality for arbitrary quasi-perfect codes, we include the (simpler) proof for the particular case of perfect codes.

*a) Perfect codes:* Consider a perfect code  $\mathcal{C}$  according to Definition 2. Then, the spheres  $\mathcal{S}_x(\theta)$  centered at the codewords are disjoint and their union covers the output space, thus, we have that  $MS(\theta) = |\mathcal{Y}|$ . These spheres are precisely the ML decision regions for each of the codewords. Then, the error probability (2) can be written as

$$\epsilon(\mathcal{C}) = 1 - \frac{1}{M} \sum_{m=1}^M \sum_{y \in \mathcal{S}_{x_m}(\theta)} P_{Y|X}(y|x_m). \quad (24)$$

For symmetric channels, the set  $\{P_{Y|X}(y|x_m) \mid y \in \mathcal{S}_{x_m}(\theta)\}$  does not depend on the specific codeword  $x_m$ . This set coincides with  $\{\psi_1, \dots, \psi_{S(\theta)}\}$ , which are, by definition, the  $S(\theta)$  largest elements in  $\{\psi_1, \dots, \psi_{|\mathcal{Y}|}\}$ . Then, we rewrite (24) as

$$\epsilon(\mathcal{C}) = 1 - \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{S(\theta)} \psi_i \quad (25)$$

$$= 1 - \sum_{i=1}^{S(\theta)} \psi_i. \quad (26)$$

Since  $MS(\theta) = |\mathcal{Y}|$ , according to (18), we must have  $p = 1$ , and (26) coincides with the right-hand side of (23).

*b) Quasi-perfect codes:* Consider now a quasi-perfect code  $\mathcal{C}$  according to Definition 2. The spheres  $\mathcal{S}_x^{\circ}(\theta)$  centered at the codewords are disjoint. However, in general, the sets  $\mathcal{S}_x^{\circ}(\theta)$  centered at each of the codewords do overlap. These overlaps correspond to ML decoding ties, and can be resolved arbitrarily without affecting the error probability.

Let  $\{\mathcal{P}_m\}$ ,  $m = 1, \dots, M$ , be any partition of the output space such that  $\mathcal{P}_m \subseteq \mathcal{S}_{x_m}(\theta)$ ,  $m = 1, \dots, M$ . Let  $P_m^{\circ} \triangleq |\mathcal{P}_m \cap \mathcal{S}_{x_m}^{\circ}(\theta)|$ . Following similar steps as in (25), we obtain

$$\epsilon(\mathcal{C}) = 1 - \frac{1}{M} \sum_{m=1}^M \left( \sum_{i=1}^{S_{\bullet}(\theta)} \psi_i + \sum_{i=1}^{P_m^{\circ}} \psi_{i+S_{\bullet}(\theta)} \right) \quad (27)$$

$$= 1 - \left( \sum_{i=1}^{S_{\bullet}(\theta)} \psi_i + \frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{P_m^{\circ}} \psi_{i+S_{\bullet}(\theta)} \right). \quad (28)$$

Since the total number of sequences in the output space is  $|\mathcal{Y}|$ , then it must hold that  $MS_{\bullet}(\theta) + \sum_{m=1}^M P_m^{\circ} = |\mathcal{Y}|$ . Using (18) we obtain

$$pS_{\circ}(\theta) = \frac{1}{M} \sum_{m=1}^M P_m^{\circ}. \quad (29)$$

From the definition of  $\mathcal{S}_{x_m}^{\circ}$ , it follows that  $\psi_i = \theta$  for  $S_{\bullet}(\theta) + 1 \leq i \leq S_{\bullet}(\theta) + S_{\circ}(\theta)$ . Since by definition,  $P_m^{\circ} \leq S_{\circ}(\theta)$ , we have that

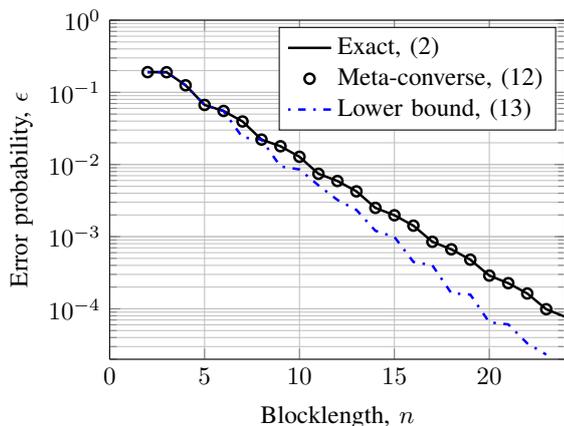
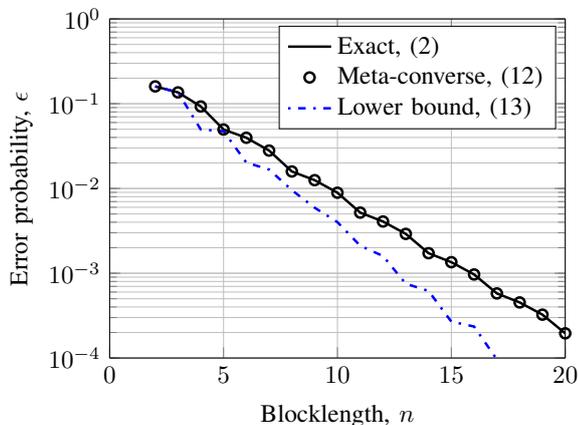
$$\frac{1}{M} \sum_{m=1}^M \sum_{i=1}^{P_m^{\circ}} \psi_{i+S_{\bullet}(\theta)} = \frac{\theta}{M} \sum_{m=1}^M P_m^{\circ} \quad (30)$$

$$= \theta p S_{\circ}(\theta) \quad (31)$$

$$= p \sum_{i=1}^{S_{\circ}(\theta)} \psi_{i+S_{\bullet}(\theta)}, \quad (32)$$

where (31) follows from (29). As a result, the right-hand side of (23) and (28) coincide. ■

Eq. (12) shows that the meta-converse bound, after optimization over the auxiliary distribution  $Q_Y$ , coincides with the exact error probability  $\epsilon(\mathcal{C})$  of any code  $\mathcal{C}$  (see [12] for details). Theorem 1 shows that, for certain symmetric channels, the relaxation (13) also coincides with the minimum error probability for quasi-perfect codes, whenever they exist.


 Fig. 1. Error probability for the BSC with parameters  $\delta = 0.1$ ,  $M = 4$ .

 Fig. 2. Error probability for the BSC with parameters  $\delta = 0.1$ ,  $M = 3$ .

Theorem 1 recovers [4, Th. 3] in the same generality. The hypothesis testing approach reported in this work is conceptually different to that in [4] and allows further extensions. For example, in this work we have restricted ourselves  $Q_Y = \bar{Q}_Y$ , although different  $Q_Y$  are obviously possible.

#### Example: BSC

Figures 1 and 2 depict the minimal error probability for the transmission of  $M$  messages over  $n$  channel uses of a BSC with cross-over probability  $\delta = 0.1$ . We plot the exact error probability (2) and the meta-converse bound (12) computed for the best code [13], compared with the lower bound in (13).

From Fig. 1 we can see that the three curves coincide for  $M = 4$  and  $n = 2, 3, 4, 5, 6, 8$ . According to Theorem 1, a quasi-perfect code can be built for these values of  $n$  as follows. The output sequences belonging to the decision regions of each of the codewords must have the  $\lceil \frac{2^n}{M} \rceil$  or  $\lfloor \frac{2^n}{M} \rfloor$  largest likelihoods in  $\{\psi_i\}$ . For instance, for  $M = 4$  and  $n = 4$ , this implies that the decision regions must include 1 output sequence at Hamming distance 0 to the closest codeword, and 3 output sequences at distance 1. This distance spectrum is achievable, for example, by the code  $\mathcal{C} = \{0000, 0001, 1110, 1111\}$ , that

therefore attains the smallest error probability. Note that this code is not optimum in terms of minimum distance (see [13, Sec. IV] for details).

Similarly, Fig. 2 shows the three curves for  $M = 3$ . We can see that they coincide for  $M = 3$  and  $n = 2, 3, 5$ . For  $n = 4$  the decision regions of a quasi-perfect code should include 1 output sequence at Hamming distance 0 of the corresponding codeword, 4 output sequences at distance 1, and at most 1 output sequence at distance 2. However, there exists no configuration of the codewords such that three of these sets are packed in the output space. Therefore, there exists a strictly positive gap between (12) and (13) and the bound in (13) is not achievable.

#### Example: BEC

Since the binary erasure channel (BEC) is symmetric, quasi-perfect codes according to Definition 2 attain the minimum error probability. Unfortunately, these codes might not exist in general. To see this, consider a BEC with erasure probability  $0 < \delta < \frac{1}{2}$ . For any input  $x \in \mathcal{X}^n$ , the all-erasures sequence is the least probable of the  $2^n$  output sequences with non-zero probability. Therefore, for values of  $\theta$  such that  $S(\theta) < 2^n$ , the all-erasures sequence does not belong to any set  $\mathcal{S}_x(\theta)$ ,  $x \in \mathcal{X}^n$ . Since for any perfect code  $S(\theta) \approx \frac{3^n}{M}$  (see (18)), even moderate values of  $M$  imply that (6) does not hold, and neither perfect nor quasi-perfect codes exist.

#### REFERENCES

- [1] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.
- [2] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Information and Control*, vol. 10, no. 5, pp. 522 – 552, 1967.
- [3] T. Baicheva, I. Bouyukliev, S. Dodunekov, and V. Fack, "Binary and ternary linear quasi-perfect codes with small dimensions," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4335–4339, Sept 2008.
- [4] M. Hamada, "A sufficient condition for a code to achieve the minimum decoding error probability—generalization of perfect and quasi-perfect codes," *IEICE Trans. on Fund. of Electronics, Comm. and Comp. Sciences*, vol. E83-A, no. 10, pp. 1870–1877, Oct. 2000.
- [5] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65 – 103, 1967.
- [6] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, 1974.
- [7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [8] J. L. Massey. (1998) Applied Digital Information Theory I, Lecture Notes. [Online]. Available: <http://www.isi.ee.ethz.ch/research.html>
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. NJ: Wiley-Interscience, 2006.
- [10] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.
- [11] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc. Lond. A*, vol. 231, no. 694-706, p. 289, 1933.
- [12] G. Vazquez-Vilar, A. Tauste Campo, A. Guillén i Fàbregas, and A. Martínez, "Bayesian M-ary hypothesis testing: The meta-converse and Verdú-Han bounds are tight," *preprint arXiv:1411.3292v2*, 2015.
- [13] P.-N. Chen, H.-Y. Lin, and S. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov 2013.