

# Extremality Properties for Gallager's Random Coding Exponent

Mine Alsan  
LTHI, I&C, EPFL  
Lausanne, Switzerland  
Email: mine.alsan@epfl.ch

**Abstract**—We describe certain extremality properties for Gallager's reliability function  $E_0$  for binary input symmetric DMCs. In particular, we show that amongst such DMC's whose  $E_0(\rho_1)$  has a given value for a given  $\rho_1$ , the BEC and BSC have the largest and smallest value of the derivative of  $E_0(\rho_2)$  for any  $\rho_2 \geq \rho_1$ . As the random coding exponent is obtained by tracing the map  $\rho \rightarrow (E'_0(\rho), E_0(\rho) - \rho E'_0(\rho))$  this conclusion includes as a special case the results of [1]. Furthermore, we show that amongst channels  $W$  with a given value of  $E_0(\rho)$  for a given  $\rho$  the BEC and BSC are the most and least polarizing under Arkan's polar transformations in the sense that their polar transforms  $W^+$  and  $W^-$  has the largest and smallest difference in their  $E_0$  values.

**Index Terms**—Channel reliability function, random coding exponent, channel polarization.

## I. INTRODUCTION

While the capacity of a memoryless channel  $W$  gives the largest rate that may be communicated reliably across it, the reliability function  $E(R, W)$  provides a finer measure on the quality of the channel: for any rate  $R$  less than channel capacity, it is possible to find a sequence of codes of increasing blocklength, each of which of rate at least  $R$ , and whose block error probability decays exponentially to zero in the blocklength —  $E(R, W)$  is the largest possible rate of this decay.

Gallager classical treatise [2] gives a lower bound to  $E(R, W)$ , the random coding exponent  $E_r(R, W)$  in the form  $E_r(R, W) = \max_{\rho \in [0,1]} E_0(\rho, W) - \rho R$ . Remarkably, this lower bound is tight for rates above the critical rate  $E'_0(1, W)$ . The function  $E_0(\rho, W)$  that appears as an auxiliary function on the road to deriving  $E_r(R, W)$  turns out to be of independent interest in its own right. In particular,  $E_0(\rho, W)/\rho$  is the largest rate for which a sequential decoder can operate while keeping the  $\rho$ -th moment of the decoder's computation effort per symbol bounded.

In this paper we investigate the extremal properties of  $E_0(\rho, W)$  for the class of binary input symmetric channels. We show that among all such channels with a given value of  $E_0(\rho_1, W)$  the binary erasure channel (BEC) and the binary symmetric channel (BSC) distinguish themselves in certain ways: they have, respectively, the largest and smallest value of  $E'_0(\rho_2, W)$  for any  $\rho_2 \geq \rho_1$ . Among the simple corollaries of this is the conclusion that of all the channels with the same capacity, the BEC and BSC have the largest and smallest value of  $E_r(R, W)$ , a result reported in [1] last year.

The BEC and BSC also exhibit themselves as being extremal for Arkan's polarization transforms. In his award winning paper [3], Arkan describes two synthetic channels  $W^+$ , and  $W^-$  which can be obtained from two independent copies of  $W$ . It is well known (proved as a corollary to extremes of information combining) that among all channels  $W$  with a given symmetric capacity  $I(W)$ , the BEC and BSC polarize most and least in the sense of having the largest and smallest difference between  $I(W^+)$  and  $I(W^-)$ . We report a more general conclusion: amongst all channels  $W$  with a given value of  $E_0(\rho, W)$ , the BEC and BSC polarize most and least in the sense of having the largest difference between  $E_0(\rho, W^+)$  and  $E_0(\rho, W^-)$ .

## II. EXTREMALITY OF THE RANDOM CODING EXPONENT

### A. Definitions and Properties

**Definition 1:** [2] Given a discrete memoryless channel  $W$  with input alphabet  $\mathcal{X}$  and output alphabet  $\mathcal{Y}$ , fix a distribution  $Q$  on its input alphabet.

Consider the function  $E_r(R, Q, W)$  defined as

$$E_r(R, Q, W) = \max_{\rho \in [0,1]} \{E_0(\rho, Q, W) - \rho R\} \quad (1)$$

for  $R \geq 0$ , with

$$E_0(\rho, Q, W) = -\log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathcal{X}} Q(x) W(y | x)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (2)$$

The random coding exponent is defined as

$$E_r(R, W) = \max_Q E_r(R, Q, W). \quad (3)$$

We restrict our attention to symmetric binary input channels (B-DMCs). In this case the uniform input distribution maximizes Equation (3). Accordingly, we drop the variable  $Q$  from the notations. The expression in (2) becomes

$$E_0(\rho, W) = -\log \sum_{y \in \mathcal{Y}} \left[ \frac{1}{2} W(y | 0)^{\frac{1}{1+\rho}} + \frac{1}{2} W(y | 1)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (4)$$

Theorem 5.6.3 in [2] summarizes the properties of  $E_0(\rho, W)$  with respect to the variable  $\rho$ . For  $\rho \geq 0$ ,  $E_0(\rho, W)$  is a positive, concave increasing function in  $\rho$ . Moreover, the symmetric capacity  $I(W)$  and the Bhattacharyya parameter  $Z(W)$  of the channel can be derived from  $E_0(\rho, W)$  by

- $\frac{\partial}{\partial \rho} E_0(\rho, W) \Big|_{\rho=0} = \frac{E_0(\rho, W)}{\rho} \Big|_{\rho=0} = I(W)$
- $E_0(1, W) = \log \frac{2}{1+Z(W)}$ .

The maximization in the right hand side of Equation (1) over  $\rho \in [0, 1]$  can be described in terms of the following parametric equations:

$$\begin{aligned} R(\rho, W) &= \frac{\partial}{\partial \rho} E_0(\rho, W) \\ E_r(\rho, W) &= E_0(\rho, W) - \rho \frac{\partial}{\partial \rho} E_0(\rho, W) \end{aligned} \quad (5)$$

for  $R$  in the range  $\frac{\partial E_0(\rho, W)}{\partial \rho} \Big|_{\rho=1} \leq R \leq \frac{\partial E_0(\rho, W)}{\partial \rho} \Big|_{\rho=0}$ .

### B. Extremality of the Random Coding Exponent

We show in this section that the binary erasure channel and the binary symmetric channel are extremal among all symmetric B-DMCs with respect to the random coding exponent. In particular, we show in Theorem 1 a certain extremality property holds even when the quantities appearing in the parametric form of the error exponent are allowed to be evaluated at different values of the parameter.

*Theorem 1:* Given a symmetric B-DMC  $W$ , for any fixed value of  $\rho_1 \in [0, 1]$ , we define a binary symmetric channel  $W_{\text{BSC}}$ , and a binary erasure channel  $W_{\text{BEC}}$  through the equality

$$E_0(\rho_1, W) = E_0(\rho_1, W_{\text{BSC}}) = E_0(\rho_1, W_{\text{BEC}}). \quad (6)$$

Then, for any  $\rho_2 \in [\rho_1, 1]$ , we have

$$R(\rho_2, W_{\text{BSC}}) \leq R(\rho_2, W) \leq R(\rho_2, W_{\text{BEC}}). \quad (7)$$

The proof of the theorem is given in Appendix VI-A.

*Remark 1:* The erasure probability of  $W_{\text{BEC}}$  and the crossover probability of  $W_{\text{BSC}}$  depend both on the channel  $W$  and the parameter  $\rho_1$ .

For the special case when  $\rho_1 = \rho_2$ , we recover in the next Corollary a result obtained in [4].

*Corollary 1 ([4]):* Given a symmetric B-DMC  $W$ , for any fixed value of  $\rho \in [0, 1]$ , we define a binary symmetric channel  $\tilde{W}_{\text{BSC}}$ , and a binary erasure channel  $\tilde{W}_{\text{BEC}}$  through the equality

$$R(\rho, W) = R(\rho, \tilde{W}_{\text{BEC}}) = R(\rho, \tilde{W}_{\text{BSC}}). \quad (8)$$

Then,

$$\begin{aligned} E_0(\rho, \tilde{W}_{\text{BEC}}) &\leq E_0(\rho, W) \leq E_0(\rho, \tilde{W}_{\text{BSC}}) \\ E_r(\rho, \tilde{W}_{\text{BEC}}) &\leq E_r(\rho, W) \leq E_r(\rho, \tilde{W}_{\text{BSC}}). \end{aligned}$$

*Proof:* Since  $E_r(\rho, W) = E_0(\rho, W) - \rho R(\rho, W)$ , it suffices to prove the first set of inequalities in view of (8). For a fixed value of  $\rho$ , we define another binary erasure channel  $\tilde{W}_{\text{BEC}}$  through the equality

$$E_0(\rho, W) = E_0(\rho, \tilde{W}_{\text{BEC}}).$$

By Theorem 1, we know that  $R(\rho, W) \leq R(\rho, W_{\text{BEC}})$ . Via (8), we thus get  $R(\rho, \tilde{W}_{\text{BEC}}) \leq R(\rho, W_{\text{BEC}})$ . As a result,

$$E_0(\rho, \tilde{W}_{\text{BEC}}) \leq E_0(\rho, W_{\text{BEC}}) = E_0(\rho, W).$$

The inequality for the binary symmetric channel can be obtained similarly. ■

Another particular case of Theorem 1 when  $\rho_1 \rightarrow 0$  recovers the result in [1]: among all symmetric B-DMCs of the same capacity, the binary erasure channel and the binary symmetric channel are extremal with respect to the random coding exponent. This is shown in the next corollary.

*Corollary 2 (Theorem 2.3 [1]):* Given a symmetric B-DMC  $W$  of capacity  $I(W)$ , we define a binary symmetric channel  $W_{\text{BSC}}$ , and a binary erasure channel  $W_{\text{BEC}}$  of the same capacity through the equality

$$I(W) = I(W_{\text{BEC}}) = I(W_{\text{BSC}}).$$

Then, the random coding error exponent of the channels satisfy

$$E_r(\rho, W_{\text{BSC}}) \leq E_r(\rho, W) \leq E_r(\rho, W_{\text{BEC}}). \quad (9)$$

*Proof:* Same capacity is equivalent to

$$\frac{E_0(\rho_1, W)}{\rho_1} \Big|_{\rho_1=0} = \frac{E_0(\rho_1, W_{\text{BEC}})}{\rho_1} \Big|_{\rho_1=0} = \frac{E_0(\rho_1, W_{\text{BSC}})}{\rho_1} \Big|_{\rho_1=0}.$$

Moreover, we can reformulate Theorem 1 by replacing the equality condition imposed for the functions  $E_0(\rho, W)$  in Equation (6), by equality among the functions  $\frac{E_0(\rho, W)}{\rho}$ . Since this latter is a continuous function of  $\rho$ , we deduce that the result of the theorem is still valid for  $\rho_1 \rightarrow 0$ . As a consequence, for any  $\rho_2 \in [0, 1]$ , we have

$$R(\rho_2, W_{\text{BSC}}) \leq R(\rho_2, W) \leq R(\rho_2, W_{\text{BEC}}).$$

As  $\frac{\partial}{\partial \rho} E_0(\rho, W) = R(\rho, W)$ , this last inequality implies that

$$E_0(\rho, W_{\text{BSC}}) \leq E_0(\rho, W) \leq E_0(\rho, W_{\text{BEC}})$$

which, in turn, implies the inequality for the random coding exponent. ■

## III. EXTREMALITY FOR THE POLARIZATION TRANSFORMATIONS

In this section, we study the behavior of  $E_0(\rho, W)$  from the aspect of channel polarization. In Theorem 2, we show that the binary erasure channel and the binary symmetric channel are also extremal in the evolution of  $E_0(\rho, W)$  under the polarization transformations.

### A. Basic Polarization Transformations

In [3], a low complexity code construction that achieves the capacity of symmetric B-DMCs is given based on the recursive application of two basic channel transformations. These transforms synthesize two new channels  $W^- : \mathcal{X} \rightarrow \mathcal{Y}^2$  and  $W^+ : \mathcal{X} \rightarrow \mathcal{Y}^2 \times \mathcal{X}$  by combining two copies of the channel  $W$ . The channels are defined by

$$W^-(y_1 y_2 | u_1) = \sum_{u_2 \in \mathcal{X}} \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2) \quad (10)$$

$$W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2). \quad (11)$$

Arıkan shows that the recursive application of these transforms polarize any  $W$  in the sense that almost all synthesized channels are either almost perfect or almost very noisy. Intuitively, one expects that  $E_0(\rho)/\rho$  similarly polarizes. The next proposition provides a simple proof for this argument.

*Proposition 1:* After the recursive application of the polarization transformations, the  $E_0(\rho)/\rho$  parameters of the synthesized channels polarize to the extremal values  $\{0, 1\}$  when a long sequence of steps is applied to the channel.

*Proof:* In [5], it is shown that  $E_0(\rho, W)/\rho$  is a decreasing function in  $\rho$ . Therefore, we can squeeze the function between

$$-\log\left(\frac{1+Z(W)}{2}\right) \leq \frac{E_0(\rho, W)}{\rho} \leq I(W).$$

The proof follows based on the fact that the quantities  $I(W)$ , and  $Z(W)$  both polarize, i.e.,  $I(W) \rightarrow 0 \iff Z(W) \rightarrow 1$ , and  $I(W) \rightarrow 1 \iff Z(W) \rightarrow 0$  [3]. ■

*Remark 2:* Based on numerical experiments, we conjecture the channels  $W$ ,  $W^-$ , and  $W^+$  satisfy the next relationship:

$$E_0(\rho, W^-) + E_0(\rho, W^+) \geq 2E_0(\rho, W).$$

*Remark 3:* [6] For a symmetric  $W$  the channels  $W^-$ ,  $W$ , and  $W^+$  are ordered by degradation. Consequently,

$$E_0(\rho, W^-) \leq E_0(\rho, W) \leq E_0(\rho, W^+).$$

#### B. Extremality of the Basic Channel Transformations

*Theorem 2:* Given a symmetric B-DMC  $W$ , for any fixed value of  $\rho \in [0, 1]$ , we define (as in Theorem 1) a binary symmetric channel  $W_{\text{BSC}}$ , and a binary erasure channel  $W_{\text{BEC}}$  through the equality

$$E_0(\rho, W) = E_0(\rho, W_{\text{BEC}}) = E_0(\rho, W_{\text{BSC}}). \quad (12)$$

Then

$$E_0(\rho, W_{\text{BEC}}^-) \leq E_0(\rho, W^-) \leq E_0(\rho, W_{\text{BSC}}^-) \quad (13)$$

$$E_0(\rho, W_{\text{BSC}}^+) \leq E_0(\rho, W^+) \leq E_0(\rho, W_{\text{BEC}}^+). \quad (14)$$

The proof is given in Appendix VI-B.

In Theorem 2, we have shown that among all B-DMC's  $W$  of fixed  $E_0(\rho, W)$ , the binary erasure channel  $W^-$  transformation results in a lower bound to any  $E_0(\rho, W^-)$  and the binary symmetric channel's one in an upper bound to any  $E_0(\rho, W^-)$ . For the  $W^+$  transformation, a similar extremality property holds except the difference that the binary erasure channel  $W^+$  transformation provides an upper bound and the binary symmetric channel's one a lower bound to any  $E_0(\rho, W^+)$ . This shows that the binary erasure and binary symmetric channels appear on reversed sides of the inequalities for  $E_0(\rho, W^-)$  and  $E_0(\rho, W^+)$ .

*Corollary 3:* Under the same assumptions as Theorem 2

$$E_0(\rho, W_{\text{BSC}}^+) - E_0(\rho, W_{\text{BSC}}^-) \leq E_0(\rho, W^+) - E_0(\rho, W^-) \\ \leq E_0(\rho, W_{\text{BEC}}^+) - E_0(\rho, W_{\text{BEC}}^-).$$

*Remark 4:* Dividing all sides of the inequality above by  $\rho$  and taking the limit as  $\rho \rightarrow 0$ , we see that among channels

of a given symmetric capacity the BEC and BSC are extremal with respect to the polarization transformations, in the sense that

$$I(W_{\text{BSC}}^+) - I(W_{\text{BSC}}^-) \leq I(W^+) - I(W^-) \\ \leq I(W_{\text{BEC}}^+) - I(W_{\text{BEC}}^-).$$

These inequalities can also be obtained by the results on the extremes of information combining [7], together with the fact that symmetric capacity is preserved under the polarization transformations [3]. Note also that, when combined with the preservation property, only one of the Equations (13) and (14) would be sufficient to get the result.

#### IV. CONCLUSIONS

We have described some extremality properties for binary input channels when the information measure is Gallager's  $E_0$ . These properties yield in straightforward fashion already known extremality results.

The extremality of BEC and BSC for polar transforms can be interpreted in the context of information combining. Theorem 2 shows that even if we change the measure of information from the customary mutual information to  $E_0$  the channels BEC and BSC still remain extremal.

The assumed symmetry of the channel  $W$  is not a major limitation: as long as we are interested only in the case when the input distribution  $Q$  is uniform, the  $E_0$  of a channel  $W$  is the same as an associated symmetric channel  $\tilde{W}$  as defined in Lemma 1.4 of [8]. Since the symmetrizing operator  $W \rightarrow \tilde{W}$  commutes with the polar transforms, all the conclusions of the paper are valid for arbitrary binary input channels as long as one evaluates all quantities under the uniform input distribution.

#### V. ACKNOWLEDGMENT

The author would like to thank Emre Telatar for helpful discussions.

#### REFERENCES

- [1] A. Guillen i Fabregas, I. Land, and A. Martinez. Extremes of random coding error exponents. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2896–2898, 31 2011-aug. 5 2011.
- [2] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.
- [3] E. Arıkan. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theor.*, 55(7):3051–3073, 2009.
- [4] E. Arıkan and E. Telatar. BEC and BSC are  $E_0$  extremal. Unpublished manuscript. July 2008.
- [5] S. Arimoto. Information measures and capacity of order  $\alpha$  for discrete memoryless channels. In *Topics in information theory*, I.Csiszar and P. Elias, editors, Amsterdam, The Netherlands, 1977. North-Holland Publishing Co.
- [6] E. Telatar, Private Communications.
- [7] I. Sutskever, S. Shamai, and J. Ziv. Extremes of information combining. *Information Theory, IEEE Transactions on*, 51(4):1313 – 1325, April 2005.
- [8] S.B. Korada, "Polar codes for channel and source coding," PhD Thesis, EPFL, Switzerland, 2009.
- [9] M. Alsan. Channel polarization and polar codes. Technical Report. EPFL, Switzerland, 2012. <http://infoscience.epfl.ch/record/176515>
- [10] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press Inc, 1994.

## VI. APPENDICES

### A. Appendix A: Proof of Theorem 1

We start with a lemma proved in [4] that expresses the function  $E_0(\rho, W)$  in a more suitable form for the proof.

*Lemma 1:* [4] Given a symmetric B-DMC  $W$ , and a fixed  $\rho \in [0, 1]$ , there exist a random variable  $Z$  taking values in the  $[0, 1]$  interval such that

$$E_0(\rho, W) = -\log \mathbb{E}[g(\rho, Z)] \quad (15)$$

where

$$g(\rho, z) = \left( \frac{1}{2}(1+z)^{\frac{1}{1+\rho}} + \frac{1}{2}(1-z)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (16)$$

Moreover, the random variable  $Z_{\text{BEC}}$  of a binary erasure channel is  $\{0, 1\}$  valued. The random variable  $Z_{\text{BSC}}$  of a binary symmetric channel is a constant  $z_{\text{BSC}}$ .

The proof of the lemma can be found in Lemma 4.1 [9].

Based on this representation, the function  $R(\rho, W)$  equals

$$R(\rho, W) = \frac{\mathbb{E}[-\partial g(\rho, Z)/\partial \rho]}{\mathbb{E}[g(\rho, Z)]}. \quad (17)$$

To prove the theorem we need to introduce a set of lemmas, and corollaries to them.

*Lemma 2:* For a fixed value of  $\rho$ , the function  $g(\rho, z)$  defined in Equation (16) is a concave decreasing function in the variable  $z$ .

The proof of the lemma is carried in Appendix A of [9].

*Lemma 3:* For fixed values of  $\rho_1, \rho_2 \in [0, 1]$  such that  $\rho_1 \leq \rho_2$ , the function  $\tilde{f}_{\rho_1, \rho_2}(t)$  defined as

$$\tilde{f}_{\rho_1, \rho_2}(t) = \frac{\partial}{\partial \rho_2} g(\rho_2, g^{-1}(\rho_1, t))$$

is a concave function in the variable  $t$ .

The convexity analysis is tedious, and we refer the readers to Appendix D of [9] for a proof. We first prove Theorem 1 for the case  $\rho_1 = \rho_2$  in the next corollary.

*Corollary 4:* Under the same assumptions of Theorem 1, the following extremality property holds:

$$R(\rho_1, W_{\text{BSC}}) \leq R(\rho_1, W) \leq R(\rho_1, W_{\text{BEC}}). \quad (18)$$

*Proof:* By the equality conditions in Equation (6), the denominator in Equation (17) is the same for all the channels. Then, the proof follows directly by the concavity of the function  $\tilde{f}_{\rho_1}(t)$  in  $t$ , and the special structure of the  $Z$  random variable corresponding to the channels  $W_{\text{BSC}}$ , and  $W_{\text{BEC}}$ . We define the random variable  $T = g(\rho, Z)$ . By the two sides of the Jensen's inequality for concave functions, i.e.

$$\begin{aligned} \tilde{f}_{\rho_1}(1) + \frac{\tilde{f}_{\rho_1}(1) - \tilde{f}_{\rho_1}(2^{-\rho_1})}{1 - 2^{-\rho_1}} (\mathbb{E}[T] - 1) &\leq \mathbb{E}[\tilde{f}_{\rho_1}(T)] \\ &\leq \tilde{f}_{\rho_1}(\mathbb{E}[T]) \end{aligned}$$

Equation (18) follows ( $-\tilde{f}$  is convex).

*Corollary 5:* Under the same assumptions of Theorem 1, we have for  $\rho_2 \geq \rho_1$

$$E_0(\rho_2, W_{\text{BSC}}) \leq E_0(\rho_2, W) \leq E_0(\rho_2, W_{\text{BEC}}). \quad (19)$$

*Proof:* By the continuity of  $E_0(\rho, W)$  in the channel, it suffices to show that

$$E_0(\rho_1, W_{\text{BSC}}) < E_0(\rho_1, W) < E_0(\rho_1, W_{\text{BEC}})$$

implies

$$E_0(\rho_2, W_{\text{BSC}}) < E_0(\rho_2, W) < E_0(\rho_2, W_{\text{BEC}}).$$

From Theorem 1,

$$E_0(\rho, W_{\text{BSC}}) < E_0(\rho, W) < E_0(\rho, W_{\text{BEC}})$$

implies

$$R(\rho, W_{\text{BSC}}) < R(\rho, W) < R(\rho, W_{\text{BEC}}).$$

We define  $F(\rho) = E_0(\rho, W) - E_0(\rho, W_{\text{BEC}})$ , or  $F(\rho) = E_0(\rho, W_{\text{BSC}}) - E_0(\rho, W)$ . Noting that  $R(\rho) = \frac{\partial}{\partial \rho} E_0(\rho)$ , the corollary is implied by the following statement:

$$F(\rho_1) < 0 \text{ and } (F(\rho) < 0 \Rightarrow F'(\rho) < 0) \Rightarrow F(\rho_2) < 0.$$

But this is true by elementary considerations on differential equations. ■

Now, we proceed with the proof of the theorem for the case  $\rho_2 > \rho_1$ . By Lemma 3, the function  $\tilde{f}_{\rho_1, \rho_2}(t)$  is concave in  $t$ . So, we can apply the two sides of Jensen's inequality to obtain

$$\begin{aligned} \mathbb{E}[-\tilde{f}_{\rho_1, \rho_2}(g(\rho_1, Z_{\text{BSC}}))] &\leq \mathbb{E}[-\tilde{f}_{\rho_1, \rho_2}(g(\rho_1, Z))] \\ &\leq \mathbb{E}[-\tilde{f}_{\rho_1, \rho_2}(g(\rho_1, Z_{\text{BEC}}))]. \end{aligned}$$

On the other hand, as  $E_0(\rho, W) = -\log \mathbb{E}[g(\rho, Z)]$  by Lemma 1, we know by Corollary 5 that

$$\mathbb{E}[g(\rho_2, Z_{\text{BEC}})] \leq \mathbb{E}[g(\rho_2, Z)] \leq \mathbb{E}[g(\rho_2, Z_{\text{BSC}})].$$

Taking the ratios of the last two set of inequalities proves the extremality property stated in Equation (7).

### B. Appendix B: Proof of Theorem 2

As we did in Appendix A, we start with two lemmas to express the functions  $E_0(\rho, W^+)$ , and  $E_0(\rho, W^-)$  in a similar form as in Lemma 1.

*Lemma 4:* Given a B-DMC  $W$  and a fixed  $\rho \in [0, 1]$ , there exist i.i.d. random variables  $Z_1$  and  $Z_2$  taking values in the  $[0, 1]$  interval such that

$$E_0(\rho, W) = -\log \mathbb{E}[g(\rho, Z)]$$

and

$$E_0(\rho, W^-) = -\log \mathbb{E}[g(\rho, Z_1 Z_2)]$$

■ where  $g(\rho, z)$  is given by (16).

*Lemma 5:* Given a B-DMC  $W$  and a fixed  $\rho \in [0, 1]$ , there exist i.i.d. random variables  $Z_1$  and  $Z_2$  taking values in the  $[0, 1]$  interval such that

$$E_0(\rho, W) = -\log \mathbb{E}[g(\rho, Z)]$$

and

$$E_0(\rho, W^+) = -\log \mathbb{E} \left[ \frac{1}{2}(1 + Z_1 Z_2) g\left(\rho, \frac{Z_1 + Z_2}{1 + Z_1 Z_2}\right) + \frac{1}{2}(1 - Z_1 Z_2) g\left(\rho, \frac{Z_1 - Z_2}{1 - Z_1 Z_2}\right) \right]$$

where  $g(\rho, z)$  is given by (16).

The proofs can be found in Lemmas 4.2, 4.3 in [9].

The proof of the extremality property for the  $W^-$  transformation, given in Equation (13), relies on the convexity of the function defined in the next lemma.

*Lemma 6:* We define the function  $F_{z,\rho}(t) : [2^{-\rho}, 1] \rightarrow [g(\rho, z), 1]$  as

$$F_{z,\rho}(t) = g(\rho, z g^{-1}(\rho, t)) \quad (20)$$

where both  $z, \rho \in [0, 1]$ . For fixed values of  $\rho$  and  $z$ , the function  $F_{z,\rho}(t)$  is convex with respect to the variable  $t$ .

The convexity analysis of this function is tedious. We refer the readers to Appendix B of [9] for a proof. By Lemmas 1 and 4, we have

$$\begin{aligned} \exp\{-E_0(\rho, W)\} &= \mathbb{E}[g(\rho, Z)] \\ \exp\{-E_0(\rho, W^-)\} &= \mathbb{E}[g(\rho, Z_1 Z_2)] \end{aligned}$$

where  $Z_1$  and  $Z_2$  are independent random variables. Moreover by Lemma 1, we know  $Z_{\text{BSC}} = z_{\text{BSC}}$  and  $Z_{\text{BEC}} \in \{0, 1\}$ . Hence,

$$\exp\{-E_0(\rho, W_{\text{BSC}}^-)\} = g(\rho, z_{\text{BSC}} z_{\text{BSC}}).$$

Let  $\epsilon$  be the erasure probability of  $W_{\text{BEC}}$ . Then, we have  $P(Z_{\text{BEC}} = 0) = \epsilon$ , and

$$\exp\{-E_0(\rho, W_{\text{BEC}})\} = P(Z_{\text{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho}.$$

Since, it is known that the channel  $W^-$  is a BEC with erasure probability  $2\epsilon - \epsilon^2$ , we get

$$\begin{aligned} \exp\{-E_0(\rho, W_{\text{BEC}}^-)\} \\ = [2P(Z_{\text{BEC}} = 0) - P(Z_{\text{BEC}} = 0)^2] (1 - 2^{-\rho}) + 2^{-\rho}. \end{aligned}$$

Moreover, given  $E_0(\rho, W) = E_0(\rho, W_{\text{BSC}})$ , we also have

$$\mathbb{E}[g(\rho, Z)] = g(\rho, z_{\text{BSC}}).$$

Therefore, using Jensen's inequality we obtain

$$\begin{aligned} \exp\{-E_0(\rho, W^-)\} &= \mathbb{E}_{Z_1}[\mathbb{E}_{Z_2}[F_{z_1,\rho}(g(\rho, Z_2)) \mid Z_1 = z_1]] \\ &\geq \mathbb{E}_{Z_1}[F_{Z_1,\rho}(\mathbb{E}_{Z_2}[g(\rho, Z_2)])] \\ &= \mathbb{E}_{Z_1}[F_{Z_1,\rho}(g(\rho, z_{\text{BSC}}))] \\ &\stackrel{(1)}{=} \mathbb{E}_{Z_1}[F_{z_{\text{BSC}},\rho}(g(\rho, Z_1))] \\ &\geq F_{z_{\text{BSC}},\rho}(\mathbb{E}_{Z_1}[g(\rho, Z_1)]) \\ &= \exp\{-E_0(\rho, W_{\text{BSC}}^-)\} \end{aligned}$$

where (1) follows by symmetry of the variables  $Z_1$  and  $z_{\text{BSC}}$ . Similarly, given  $E_0(\rho, W) = E_0(\rho, W_{\text{BEC}})$ , we have

$$\mathbb{E}[g(\rho, Z)] = \mathbb{E}[g(\rho, Z_{\text{BEC}})] = P(Z_{\text{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho}.$$

Due to convexity, we also know the following inequality holds:

$$F_{z,\rho}(t) \leq 1 + \frac{g(\rho, z) - 1}{2^{-\rho} - 1}(t - 1).$$

Therefore,

$$\begin{aligned} &\exp\{-E_0(\rho, W^-)\} \quad (21) \\ &= \mathbb{E}_{Z_1}[\mathbb{E}_{Z_2}[F_{z_1,\rho}(g(\rho, Z_2)) \mid Z_1 = z_1]] \\ &\leq \mathbb{E}_{Z_1} \left[ 1 + \frac{g(\rho, Z_1) - 1}{2^{-\rho} - 1} (\mathbb{E}_{Z_2}[g(\rho, Z_2)] - 1) \right] \\ &= 1 + \frac{\mathbb{E}_{Z_1}[g(\rho, Z_1)] - 1}{2^{-\rho} - 1} (\mathbb{E}_{Z_2}[g(\rho, Z_2)] - 1) \\ &= \frac{2^{-\rho} - 1 + [P(Z_{\text{BEC}} = 0)(1 - 2^{-\rho}) + 2^{-\rho} - 1]^2}{2^{-\rho} - 1} \\ &= [2P(Z_{\text{BEC}} = 0) - P(Z_{\text{BEC}} = 0)^2] (1 - 2^{-\rho}) + 2^{-\rho} \\ &= \exp\{-E_0(\rho, W_{\text{BEC}}^-)\}. \end{aligned}$$

This concludes the proof for the  $W^-$  transformation.

Now, we sketch the proof of the extremality property for the  $W^+$  transformation, given in Equation (14).

We define the function  $h_\rho(z_1, z_2)$  as

$$\begin{aligned} h_\rho(z_1, z_2) &= \frac{1}{2}(1 + z_1 z_2) g\left(\rho, \frac{z_1 + z_2}{1 + z_1 z_2}\right) \\ &\quad + \frac{1}{2}(1 - z_1 z_2) g\left(\rho, \frac{z_1 - z_2}{1 - z_1 z_2}\right) \quad (22) \end{aligned}$$

where  $\rho, z_1, z_2 \in [0, 1]$ . Note that  $h_\rho(z_1, z_2)$  is symmetric in the variables  $z_1$ , and  $z_2$ .

*Lemma 7:* Let the function  $H_{z,\rho}(t) : [2^{-\rho}, 1] \rightarrow [2^{-\rho}, g(\rho, z)]$  be defined as

$$H_{z,\rho}(t) = h_\rho(g^{-1}(\rho, t), z)$$

where both  $z, \rho \in [0, 1]$ . For fixed values of  $\rho$  and  $z$ , the function  $H_{z,\rho}(t)$  is concave with respect to the variable  $t$ .

We refer the readers to Appendix F of [9] for a proof.

The proof of the theorem for the  $W^+$  transformation can be completed following similar steps to the  $W^-$  case. By Lemma 5, we have  $\mathbb{E}[h_\rho(Z_1, Z_2)] = \exp\{-E_0(\rho, W^+)\}$ . We define  $T_1 = g(\rho, Z_1)$ , and  $T_2 = g(\rho, Z_2)$ . Then, using the concavity of the function  $H_{z,\rho}(t)$  with respect to  $t$ , and symmetry of  $Z_1$  and  $Z_2$ , we obtain

$$\begin{aligned} \exp\{-E_0(\rho, W^+)\} &= \mathbb{E}[H_{g^{-1}(\rho, T_2), \rho}(T_1)] \\ &\leq h_\rho(z_{\text{BSC}}, z_{\text{BSC}}) = \exp\{-E_0(\rho, W_{\text{BSC}}^+)\} \end{aligned}$$

and

$$\begin{aligned} \exp\{-E_0(\rho, W^+)\} &= \mathbb{E}[H_{g^{-1}(\rho, T_2), \rho}(T_1)] \\ &\geq 2^{-\rho} + P(Z_{\text{BEC}} = 0)^2 (1 - 2^{-\rho}) = \exp\{-E_0(\rho, W_{\text{BEC}}^+)\}. \end{aligned}$$