Extremality for Gallager’s Reliability Function

\[ E_0 \]

Mine Alsan
Information Theory Laboratory
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: mine.alsan@epfl.ch

Abstract

We describe certain extremalities for Gallager’s \( E_0 \) function evaluated under the uniform input distribution for binary input discrete memoryless channels. The results characterize the extremality of the \( E_0(\rho) \) curves of the binary erasure channel and the binary symmetric channel among all the \( E_0(\rho) \) curves that can be generated by the class of binary discrete memoryless channels whose \( E_0(\rho) \) curves pass through a given point \((\rho_0, e_0)\), for some \( \rho_0 > -1 \).

Index Terms

Channel reliability function, random coding exponent, extremal channels.

I. INTRODUCTION

While the capacity of a memoryless channel \( W \) gives the largest rate for which reliable communication is possible, the reliability function \( E(R, W) \) provides a finer measure on the quality of the channel: for any \( R \) less than the 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 as the blocklength increases — \( E(R, W) \) is the largest possible rate of this decay.

Part of the material in this paper was presented in part at the IEEE International Symposium on Information Theory, Boston, USA, July 2012.
Gallager’s classical treatise [1] 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 [2].

Previously, we investigated in [3] the extremal properties of $E_0(\rho, W)$, for $\rho \in [0, 1]$, for the class of binary input discrete memoryless channels (B-DMC) when the function is evaluated under the uniform input distribution. We have shown that among all 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 the smallest value of $E_0'(\rho_2, W)$ for any $\rho_1, \rho_2 \in [0, 1]$ such that $\rho_2 \geq \rho_1$. As the random coding exponent is obtained by tracing the map $\rho \rightarrow (E_0'(\rho), E_0(\rho) - \rho E_0'(\rho))$, among the simple corollaries of this is the conclusion that of all the symmetric channels with the same capacity, the BEC and the BSC have the largest and the smallest value of $E_r(R, W)$, a result reported in [4].

In this paper, we extend the previous extremality analysis of [3] to both the cases when $\rho > 1$ and when $\rho \in (-1, 0)$. The extremal results for $E_0$ in these regions are motivated by various error exponents such as the list decoding exponent [1], defined for $\rho > 0$, and the exponent which appears in Arimoto’s lower bound for the strong converse of the coding theorem [5], defined for $\rho \in (-1, 0]$. For a concise list of the definitions of various error exponents involving the $E_0$ function, we refer to [6], a recent study which also examined the extremality of $E_0(\rho)$ for $\rho > -1$, but only for the special class of symmetric B-DMCs of the same capacity.

The results of this paper characterize the extremality of the $E_0$ curves of the BEC and the BSC among all the $E_0$ curves that can be generated by the class of B-DMCs whose $E_0$ curves pass through a given point $(\rho_0, e_0)$ for some $\rho_0 > -1$. We prove that when $\rho_0 \in (-1, 1]$, these two channels remain extremal along the $E_0(\rho)$ curves for any $\rho > -1$. We also prove that when $\rho_0 > 1$, while these two channels are extremal along the $E_0(\rho)$ curves for any $\rho \in (-1, 1]$, no extremality beyond $\rho > 1$ can be formulated in general. Furthermore, we show that the conclusion we have mentioned above for $E_0'$ is still valid when $\rho_1 \in (0, 1]$ and $\rho_2 \geq \rho_1$ (even for $\rho_2 > 1$), and also when $\rho_1 \in (-1, 0]$ and $\rho_2 \leq \rho_1$. Using these, we recover the result of [6].
which shows that, for any $\rho > -1$, the BEC and the BSC are $E_0$ extremal among the $E_0(\rho)$ curves of all symmetric channels with the same capacity.

The rest of this paper is organized as follows. Section II starts by giving the preliminary definitions, and then later derives some basic properties of the $E_0$ curves of BECs and BSCs. Subsequently, in Section III, the main results of this paper are stated in Theorem I. The section follows by some convexity lemmas, the proof of the theorem, and a graphical interpretation of the extremality results. Finally, the last section gives the conclusions.

II. Preliminaries

A. Definition of the Random Coding Exponent and $E_0$

Definition 1: [Section 5.6] Given a discrete memoryless channel (DMC) $W$ 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

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

for $R \geq 0$, where

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

with the log denoting the natural logarithm to the base $e$. The random coding exponent of the channel is defined as

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

Throughout this paper, we fix $\mathcal{X}$ to $\{0, 1\}$ and $Q$ to the uniform input distribution. Then, the expression in (2) becomes

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

For symmetric channels, the uniform input distribution corresponds to the distribution which maximizes (3) [I]. The random coding exponent of symmetric channels is then given by

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

Moreover, the right hand side of (5) gives a lower bound to the random coding exponents of B-DMCs which are not symmetric.
The properties of $E_0(\rho, W)$ with respect to the variable $\rho$ are summarized in \[1, \text{Theorem 5.6.3}\]. For $\rho \geq 0$, $E_0(\rho, W)$ is a positive, concave increasing function in $\rho$. By convexity, the maximization in the right hand side of (5) over $\rho \in [0, 1]$ can be described in terms of the following parametric equations:

$$R(\rho, W) = \frac{\partial}{\partial \rho} E_0(\rho, W), \quad (6)$$

$$E_r(\rho, W) = E_0(\rho, W) - \rho \frac{\partial}{\partial \rho} E_0(\rho, W), \quad (7)$$

for $R$ in the range

$$\left. \frac{\partial E_0(\rho, W)}{\partial \rho} \right|_{\rho=1} \leq R \leq \left. \frac{\partial E_0(\rho, W)}{\partial \rho} \right|_{\rho=0}. \quad (8)$$

It is shown in [1] see Figure 5.6.2] that the symmetric capacity of the channel,

$$I(W) = \sum_{y \in Y} \sum_{x \in \{0,1\}} \frac{1}{2} W(y|x) \log \left( \frac{W(y|x)}{2W(y|0) + \frac{1}{2} W(y|1)} \right), \quad (9)$$

is the slope of the $E_0$ curve at $\rho = 0$, i.e.,

$$I(W) = \left. \frac{\partial}{\partial \rho} E_0(\rho, W) \right|_{\rho=0}. \quad (10)$$

Finally, another channel parameter of interest for DMCs, the cut-off rate, can also be derived from $E_0$, see [7] for more information on the significance of this parameter. The cut-off rate of a B-DMC when evaluated under the uniform input distribution is given by $E_0(1, W)$.

\textbf{B. Description of $E_0$ by Rényi’s Entropy Functions}

In this section, we mention an alternative description of $E_0(\rho, W)/\rho$, which also appears in [8] and [2], using the concept of Rényi’s entropy functions. This gives an interpretation to $E_0(\rho, W)/\rho$ as a general measure of information.

Rényi’s entropy function of order $\alpha$ of a discrete random variable $X \sim P(x)$ is defined in [9] as

$$H_\alpha(X) = \frac{\alpha}{1-\alpha} \log \left( \sum_x P(x)^{-\frac{1}{\alpha}} \right)^{\frac{1}{\alpha}}. \quad (11)$$
This definition is extended to the Rényi’s conditional entropy function of order $\alpha$ of a discrete random variable $X$ given $Y$ with joint distribution $P(x,y)$ in [10] as

$$H_\alpha(X \mid Y) = \frac{\alpha}{1 - \alpha} \log \left( \sum_y \left( \sum_x P(x,y)^\alpha \right)^\frac{1}{\alpha} \right),$$

$$= H_\alpha(X) + \frac{\alpha}{1 - \alpha} \log \left( \sum_y \left( \sum_x Q(x) P(y \mid x)^\alpha \right)^\frac{1}{\alpha} \right),$$

where $Q(x) = \frac{P(x)^\alpha}{\sum_x P(x)^\alpha}$ is ‘tilted’ probability distribution. Although different definitions are proposed in the literature for a possible extension of Rényi’s entropy function to a quantity similar to the conditional entropy function, as one suitable for this study, we use the definition in (12).

Taking a uniform input distribution and letting $\alpha = \frac{1}{1 + \rho}$, we get

$$H_{\frac{1}{1+\rho}}(X) = \frac{1}{\rho} \log \left( \sum_x P(x)^{\frac{1}{1+\rho}} \right)^{\frac{1}{1+\rho}},$$

$$H_{\frac{1}{1+\rho}}(X \mid Y) = H_{\frac{1}{1+\rho}}(X) + \frac{1}{\rho} \log \left( \sum_y \left( \sum_x P(x) P(y \mid x)^{\frac{1}{1+\rho}} \right)^{\frac{1}{1+\rho}} \right).$$

Hence, from the definition of $E_0(\rho, W)$ in (4), we deduce

$$\frac{E_0(\rho, W)}{\rho} = H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(X \mid Y).$$

The quantity in the right hand side of (16) is called as the mutual information of order $\frac{1}{1 + \rho}$ in [10]. Moreover, the following properties are proved:

- $\lim_{\alpha \to 1} H_\alpha(X) = H(X)$,
- $H_\alpha(X \mid Y) \leq H_\alpha(X)$, i.e “conditioning reduces entropy” is valid for Rényi’s entropy function, as it is in the Shannon entropy case,
- $\frac{E_0(\rho, W)}{\rho}$ is a decreasing function in $\rho$ with $\lim_{\rho \to 0} \frac{E_0(\rho, W)}{\rho} = I(W)$.

C. An Alternative Representation of $E_0$ for B-DMCs

The extremality results we will prove in Section III will be based neither on the ‘raw definition’ of $E_0(\rho, W)$ in (4), nor on the interpretation in terms of Renyi’s entropy functions of (16). Instead,
we will make use of a description of $E_0(\rho,W)$ introduced by [11] which is more suitable for deriving extremal bounds.

For a given symmetric B-DMC $W : \mathcal{X} \to \mathcal{Y}$ and a fixed $\rho > -1$, [11] shows that there exists a random variable $Z$ taking values in the $[0, 1]$ interval such that

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

(17)

where the function $g(\rho,z)$ is defined as

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

(18)

for $\rho \in \mathbb{R} \setminus \{-1\}$ and $z \in [-1, 1]$. To see this, define

$$W(y) = \frac{W(y | 0) + W(y | 1)}{2},$$

(19)

and

$$\Delta(y) = \frac{W(y | 0) - W(y | 1)}{W(y | 0) + W(y | 1)},$$

(20)

for $y \in \mathcal{Y}$, so that $W(y | 0) = W(y) (1 + \Delta(y))$ and $W(y | 1) = W(y) (1 - \Delta(y))$. Then, one can manipulate (4) to find that $Z = |\Delta(Y)|$ with $Y \sim W(y)$ in (17).

The next lemma gives the first and the second order properties of $g(\rho,z)$ with respect to the variable $z$. The proof is carried in Appendix I.

**Lemma 1:** The function $g(\rho,z)$ defined in (18) is a concave non-increasing function in $z \in [0, 1]$ for $\rho \in (-\infty, -1) \cup [0, \infty)$, and a convex non-decreasing function in $z \in [0, 1]$ for $\rho \in (-1, 0]$. As $g(\rho,z)$ is symmetric around $z = 0$, these properties also determine the function’s behavior for $z \in [-1, 0]$.

We denote by $g^{-1}(\rho,t)$ the inverse of the function $g(\rho,z)$ with respect to its second argument. The variable $t$ always takes values from a subset of the interval $[0, 2]$. More specifically, $t \in [2^{-\rho}, 1]$ when $\rho \geq 0$, and $t \in [1, 2^{-\rho}]$ when $\rho \in (-1, 0)$. For shorthand notation, we denote the range of possible values by $t \in [2^{-\rho}, 1] \cup [1, 2^{-\rho}]$, for $\rho > -1$.

Finally, we note that by using (17), the function $R(\rho,W) = \frac{\partial}{\partial \rho} E_0(\rho,W)$ can be written as

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

(21)

where the second equality follows by the dominated convergence theorem.
D. Fun facts about $E_0$ and $E'_0$ of BECs and BSCs

In this section, we explain some simple facts related to the $E_0$ curves of BECs and BSCs. We will be using some of these facts many times throughout the results section.

Consider first the representation in (17). It is not difficult to see that the BECs and the BSCs are special cases of this representation.

**Fact 1:** The random variable $Z_{BEC}$ of a BEC is $\{0, 1\}$ valued and satisfy $P[Z_{BEC} = 0] = \epsilon$, where $\epsilon \in [0, 1]$ is the erasure probability of the channel. The random variable $Z_{BSC}$ of a BSC is a constant given by $z_{BSC} = 1 - 2x$ assuming that $x \in [0, 0.5]$ is the crossover probability of the channel.

It is well known that the set of BECs and BSCs are ordered in terms of their channel capacities: if the chances of an erasure to happen at the output of a BEC model, or similarly of a bit flip at the output of a BSC model is increasing, the transmission capacities shall decrease, see for instance the textbook [II]. Intuitively, we expect this graceful degradation to order as well other measures of channel quality. For that purpose, we start by computing the $E_0$ and $E'_0$ parameters of a BEC and a BSC as a function of the erasure probability and the crossover probability of the channels. Let $BEC$ be a BEC with erasure probability $\epsilon \in [0, 1]$. Then, one can easily derive that

$$E_0(\rho, BEC) = -\log (2^{-\rho}(1 - \epsilon) + \epsilon), \tag{22}$$

and

$$R(\rho, BEC) = \frac{\partial}{\partial \rho} E_0(\rho, BEC) = \frac{2^{-\rho}(1 - \epsilon) \log 2}{2^{-\rho}(1 - \epsilon) + \epsilon}. \tag{23}$$

Let $BSC$ be a BSC with crossover probability $x \in [0, 0.5]$. In this case, we are saved from the trouble by [II Example 1 p.146] which has the derivation of the $E_0$ parameter of a BSC in Equation (5.6.40) and its rate parameter in Equation (5.6.41). Rewriting these equations, we get

$$E_0(\rho, BSC) = \rho - (1 + \rho) \log \left( x^{\frac{1}{1+\rho}} + (1 - x)^{\frac{1}{1+\rho}} \right), \tag{24}$$

and

$$R(\rho, BSC) = 1 - \mathcal{H}(\delta), \tag{25}$$

where $\delta = \frac{x^{\frac{1}{1+\rho}}}{x^{\frac{1}{1+\rho}} + (1 - x)^{\frac{1}{1+\rho}}}.$

Now, we show that these parameters are monotone functions in the erasure/crossover probabilities of the channels.
Lemma 2: For any $\rho \geq 0$, $E_0(\rho, \text{BEC})$ ($E_0(\rho, \text{BSC})$) is decreasing in $\epsilon (x)$. For any $\rho \in (-1, 0]$, $E_0(\rho, \text{BEC})$ ($E_0(\rho, \text{BSC})$) is increasing in $\epsilon (x)$. Moreover, for any $\rho > -1$, $R(\rho, \text{BEC})$ ($R(\rho, \text{BSC})$) is decreasing in $\epsilon (x)$.

Proof:

Taking the first derivative of (22) with respect to $\epsilon$, we get

$$\frac{\partial}{\partial \epsilon} E_0(\rho, \text{BEC}) = -\frac{1 - 2^{-\rho}}{2^{-\rho} (1 - \epsilon) + \epsilon}. \quad (26)$$

One can check that

$$\frac{\partial}{\partial \epsilon} E_0(\rho, \text{BEC}) \begin{cases} > 0, & \text{for } \rho \in (-1, 0) \\ = 0, & \text{for } \rho = 0 \\ < 0, & \text{for } \rho > 0 \end{cases}. \quad (27)$$

As $E_0(0, W) = 0$, the $E_0$ curves of all BECs will be ordered such that while for $\rho > 0$ the $E_0$ curves of BECs with smaller erasure probabilities will be larger, for $\rho \in (-1, 0)$ the opposite will be true.

Now, we show an ordering also holds for the $R$ parameters of BECs. Taking the first derivative of (23) with respect to $\epsilon$, we get

$$\frac{\partial}{\partial \epsilon} R(\rho, \text{BEC}) = -\frac{2^\rho \log 2}{(1 + (-1 + 2^\rho)\epsilon)^2} < 0. \quad (28)$$

Hence, the rate parameters will be decreasing with the erasure probability of the channel for any $\rho > -1$. This completes the proof for the BEC.

Now, we prove the claims for the set of BSCs. First, we note that the term inside the logarithm in (24) satisfies for $x \in [0, 0.5]$

$$\frac{\partial}{\partial x} \left( x^{\frac{1}{1+\rho}} + (1-x)^{\frac{1}{1+\rho}} \right) = \frac{x^{-\frac{\rho}{1+\rho}} - (1-x)^{-\frac{\rho}{1+\rho}}}{1+\rho} \begin{cases} < 0, & \text{for } \rho \in (-1, 0) \\ = 1, & \text{for } \rho = 0 \\ > 0, & \text{for } \rho > 0 \end{cases}. \quad (29)$$

Hence, we also have

$$\frac{\partial}{\partial x} E_0(\rho, \text{BSC}) \begin{cases} > 0, & \text{for } \rho \in (-1, 0) \\ = 0, & \text{for } \rho = 0 \\ < 0, & \text{for } \rho > 0 \end{cases}. \quad (30)$$
which proves the claimed ordering for $E_0(\rho, BSC)$. To prove the claim for $R(\rho, BSC)$, we simply note that in (25), for $x \in [0, 0.5]$, we have $\delta \in [0, 0.5]$ increasing in $x$ and the binary entropy function $H(\delta)$ increasing in $\delta \in [0, 0.5]$. As a result,
\[
\frac{\partial}{\partial x} R(\rho, BSC) < 0,
\] (31)
as claimed.

By this lemma, the second fact is in order:

**Fact 2:** For any $\rho > -1$, the class of BECs and the class of BSCs ($x \in [0, 0.5]$) are strictly ordered in their $E_0(\rho, W)$ parameters, except at $\rho = 0$ where $E_0(0, W) = 0$, and in their $R(\rho, W)$ parameters.

The ordering we have just discussed is not peculiar to BECs and BSCs and can be generalized to more general classes of channels such as degraded ones. However, Lemma 2 will be sufficient for our purpose as the derivations of Section III does not need results of such a generality.

Next, we argue the validity of an assumption we will encounter in the hypothesis of the main theorem.

**Lemma 3:** For any given B-DMC $W$ and any fixed $\rho > -1$, there exist a BEC $BEC$ and a BSC $BSC$ such that
\[
E_0(\rho, W) = E_0(\rho, BEC) = E_0(\rho, BSC).
\] (32)
The erasure probability of $BEC$ and the crossover probability of $BSC$ depend both on the channel $W$ and the parameter $\rho$.

**Proof:** Observe that, by (17), the equality of the $E_0$ functions in (32) is equivalent to the equality of
\[
\mathbb{E} [g(\rho, Z)] = \mathbb{E} [g(\rho, Z_{BEC})] = g(\rho, z_{BSC}),
\] (33)
where $Z$, $Z_{BEC}$ and $z_{BSC}$ correspond to the ‘$Z$’ random variables of the channel $W$, the channel $BEC$, and the channel $BSC$, respectively. Therefore, to show that there exists a BSC and a BEC satisfying (32), it is sufficient to show that there exists $Z_{BEC}$ and $z_{BSC}$ random variables satisfying (33). By the monotonicity results stated in Lemma 5, we know that
\[
g(\rho, z) \in [2^{-\rho}, 1], \quad \text{for } \rho \geq 0,
\] (34)
\[
g(\rho, z) \in [1, 2^{-\rho}], \quad \text{for } \rho \in (-1, 0],
\] (35)
for $z \in [0, 1]$. As a result,

$$E[g(\rho, Z)] \in [2^{-\rho}, 1], \quad \text{for } \rho \geq 0, \quad (36)$$

$$E[g(\rho, Z)] \in [1, 2^{-\rho}], \quad \text{for } \rho \in (-1, 0]. \quad (37)$$

Moreover, $g$ being continuous in $z$ for fixed values of $\rho$ implies that every intermediate value of the corresponding bounded interval will be taken by the function $g(\rho, z)$ for $z \in [0, 1]$, i.e. we can always find a $z^* \in [0, 1]$ such that

$$E[g(\rho, Z)] = g(\rho, z^*). \quad (38)$$

Since, as indicated in Fact 1, the random variable $Z_{BSC}$ of a BSC is a constant $z_{BSC}$, the BSC defined in (33) will be a BSC such that $z_{BSC} = z^*$. From this the crossover probability of the channel can be inferred.

To find a BEC which satisfies (33), we will use the BSC we have just defined with parameter $z^*$. Note that the extreme values of the bounded interval from which $g(\rho, z)$ takes values are given by $2^{-\rho} = g(\rho, 0)$ and $1 = g(\rho, 1)$. Moreover, the function $g$ being continuous in $z \in [0, 1]$ for fixed values of $\rho$, we can weight these two values with a probability distribution $p_0$ and $1 - p_0$ such that

$$g(\rho, z^*) = p_0 g(\rho, 0) + (1 - p_0) g(\rho, 1). \quad (39)$$

Since, as indicated in Fact 1, the random variable $Z_{BEC}$ of a BEC is $\{0, 1\}$ valued, the BEC defined in (33) will be a BEC with erasure probability given by $P(Z_{BEC} = 0) = p_0$. ■

Upon this lemma, another property of BECs and BSCs is due:

**Fact 3:** The set of BECs and the set of BSCs both sweep all the possible values the $E_0$ parameters of B-DMCs can take at any $\rho > -1$.

Suppose now the $E_0$ curves of a BEC and a BSC intersect at a particular $\rho^* > -1$ other than $\rho^* = 0$. We would like to know if there are any other $\rho > -1$ values apart from the trivial $\rho = 0$ such that the $E_0$ curves of these two channels intersect again? The next lemma answer this question.

**Lemma 4:** Suppose a BSC $BSC$, and a BEC $BEC$ satisfy

$$E_0(\rho^*, BEC) = E_0(\rho^*, BSC), \quad (40)$$

10
for some $\rho^* > -1$ such that $\rho^* \neq 0$. Then, if $\rho^* \leq 1$, there is only one other intersection point between the $E_0$ curves of the channels at $\rho = 0$. If $\rho^* > 1$, the only intersection point in the interval $(-1, 1]$ is once more at $\rho = 0$, and for the rest either the $E_0$ curves of the channels are tangent to each others at $\rho^*$, i.e.,

$$R(\rho^*, \text{BEC}) = R(\rho^*, \text{BSC})$$

is satisfied, or there exists a different $\rho' > 1$ such that

$$E_0(\rho', \text{BEC}) = E_0(\rho', \text{BSC}).$$

**Proof:** Let the erasure probability of the channel $\text{BEC}$ be $\epsilon$ and the channel $\text{BSC}$ be such that $z_{\text{BSC}} = z$. By (22) and (24), the condition for equality in (40) translates into

$$g(\rho^*, z) = 2^{-\rho^*}(1 - \epsilon) + \epsilon.$$  

Let the function $h(\rho, z)$ be defined as

$$h(\rho, z) = \frac{g(\rho, z) - 2^{-\rho}}{1 - 2^{-\rho}}.$$  

Observe that $h(\rho^*, z) = \epsilon$ and, in order for (42) to hold, we are looking for another $\rho'$ such that $h(\rho', z) = \epsilon$ holds. To find the answer, we need to study the monotonicity properties of the function $h(\rho, z)$ with respect to $\rho$. Indeed, one can show that the first derivative of $h(\rho, z)$ with respect to $\rho$ changes sign only once at $\rho_{\text{max}}(z) \geq 3$ for every fixed value of $z$, such that $h(\rho, z)$ is increasing for $\rho \in (0, \rho_{\text{max}}(z))$, and decreasing for $\rho > \rho_{\text{max}}(z)$ with $\lim_{\rho \to \infty} h(\rho, z) = h(1, z)$. Consequently, if $\rho^* \in (-1, 0) \cup (0, 1]$, no other $\rho'$ can satisfy (42). On the other hand, if $\rho^* > 1$, but $\rho^* \neq \rho_{\text{max}}(z)$, then the two curves intersect twice. Finally, if $\rho^* = \rho_{\text{max}}(z)$, not only no other $\rho'$ can satisfy (42), but also

$$h(\rho^*, z) = h(\rho_{\text{max}}(z), z) \geq h(\rho, z)$$

holds for all $\rho > -1$. In this case, the $E_0$ curves of the channels will be tangent to each other, so (41) holds as well. As the analysis of the monotonicity property is tedious, we omit the proof.

The previous lemma says that if the $E_0$ curves of a BEC and a BSC intersect somewhere between the interval $(-1, 0) \cup (0, 1]$, they cannot intersect a second time, except trivially at 0, and if otherwise they intersect in the interval $(1, \infty)$, either the two curves are tangent to
each other or they intersect twice in that interval, and the only intersection point in the interval \((-1, 1]\) is again at 0. The significance of this lemma will become clear later when we interpret the extremality results. The lemma will help us to understand why some intervals of \(\rho > -1\) are more interesting in the context of the extremality results presented in the main theorem.

### III. Extremality Results

In this section, we study the extremality of the BEC and the BSC with respect to the \(E_0\) channel parameter. In particular, we show in Theorem [III-B](#) that a certain extremality property holds even when the quantities appearing in the parametric form of the random coding error exponent, i.e. \(E_0\) and \(E'_0\), are evaluated at different values of the parameter. The proof of the theorem is carried out in Section [III-B](#).

**Theorem 1:** Given any fixed value of \(\rho_1 > -1\), suppose a B-DMC \(W\), a binary symmetric channel BSC, and a binary erasure channel BEC satisfy

\[
E_0(\rho_1, BSC) \stackrel{(a)}{\leq} E_0(\rho_1, W) \stackrel{(a')}{\leq} E_0(\rho_1, BEC),
\]

for \(\rho_1 \neq 0\), or

\[
\lim_{\rho \to 0} \frac{E_0(\rho, BSC)}{\rho} \stackrel{(a_0)}{\leq} \lim_{\rho \to 0} \frac{E_0(\rho, W)}{\rho} \stackrel{(a_0')}{\leq} \lim_{\rho \to 0} \frac{E_0(\rho, BEC)}{\rho},
\]

for \(\rho_1 = 0\).

**Part 1** If \(\rho_1 \in [0, 3]\), then

\[
R(\rho_2, BSC) \stackrel{(b)}{\leq} R(\rho_2, W) \stackrel{(b')}{\leq} R(\rho_2, BEC),
\]

\[
E_0(\rho_2, BSC) \stackrel{(c)}{\leq} E_0(\rho_2, W) \stackrel{(c')}{\leq} E_0(\rho_2, BEC),
\]

for any \(\rho_2 \in [\rho_1, 3]\).

**Part 2** If \(\rho_1 \in (-1, 0]\), then

\[
R(\rho_2, BEC) \stackrel{(d)}{\leq} R(\rho_2, W) \stackrel{(d')}{\leq} R(\rho_2, BSC),
\]

\[
E_0(\rho_2, BSC) \stackrel{(e)}{\leq} E_0(\rho_2, W) \stackrel{(e')}{\leq} E_0(\rho_2, BEC),
\]

for any \(\rho_2 \in (-1, \rho_1]\),

**Part 3** If \(\rho_1 \in (-1, 0]\), then

\[
E_0(\rho_2, BSC) \stackrel{(f)}{\leq} E_0(\rho_2, W) \stackrel{(f')}{\leq} E_0(\rho_2, BEC),
\]

for any \(\rho_2 \in (-1, \rho_1]\).
for any $\rho_2 \geq 0$.
If $\rho_1 \in [0, 1]$, then
$$E_0(\rho_2, BSC) \leq E_0(\rho_2, W) \leq E_0(\rho_2, BEC), \tag{53}$$
for any $\rho_2 \geq \rho_1$.
If $\rho_1 > 1$, then
$$E_0(\rho_2, BEC) \leq E_0(\rho_2, W) \leq E_0(\rho_2, BSC), \tag{54}$$
for any $\rho_2 \in [0, 1]$.
If $\rho_1 > 1$, then
$$E_0(\rho_2, BSC) \leq E_0(\rho_2, W) \leq E_0(\rho_2, BEC), \tag{55}$$
for any $\rho_2 \in (-1, 0]$.
Moreover, the extremalities hold with strict inequalities, except for $\rho_2 = 0$, whenever (a) and (a’) in (46) are strict for $\rho_1 \neq 0$, or (a0) and (a0’) in (47) are strict for $\rho_1 = 0$.

**Remark 1:** In Theorem 1, the inequalities (a)-(a0) imply the inequalities (b), (c), (d), (e), (f), (g), (h), and (i). Similarly, the inequalities (a’)-(a0’) imply the inequalities (b’) through (i’).

**Remark 2:** The value of “3” that appears in the interval in Part 1 of the theorem is a conservative estimate. The reader who follows the proof of Lemma 5, which is stated in Section III-A and proved in Appendix II, will notice that this “3” may be replaced by a $\rho^*(W)$ that depends on the channel $W$. In the proof of Lemma 5, it is shown that $\rho^*(W) \geq 3$ for any $W$, but the lower bound is not necessarily tight. We chose the value 3 so as to not further complicate the statement of the theorem.

For the special case where $\rho_1 = \rho_2 = \rho$, for $\rho \in [0, 1]$, we recover in the next corollary, a result obtained in [11].

**Corollary 1 ([11]):** Given a symmetric B-DMC $W$, for any fixed value of $\rho \in [0, 1]$, find a binary symmetric channel $BSC$, and a binary erasure channel $BEC$ through the equality
$$R(\rho, W) = R(\rho, BEC) = R(\rho, BSC). \tag{56}$$
Then,
$$E_0(\rho, BEC) \leq E_0(\rho, W) \leq E_0(\rho, BSC), \tag{57}$$
$$E_r(\rho, BEC) \leq E_r(\rho, W) \leq E_r(\rho, BSC). \tag{58}$$
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 (56). Taking $\rho_1 = \rho_2 = \rho$, (57) holds by Theorem 1. To see this, observe that had the channels on the contrary satisfied

$$E_0(\rho, BSC) < E_0(\rho, W) < E_0(\rho, BEC),$$

the results in Part 1 of the theorem would imply

$$R(\rho, BSC) < R(\rho, W) < R(\rho, BEC),$$

contradicting the assumption (56) of the corollary.

Another particular case of Theorem 1 when $\rho_1 = 0$ recovers the result in [4]: amongst all symmetric B-DMCs of the same capacity, the BEC and the BSC are extremal with respect to the random coding exponent.

Corollary 2 (Theorem 2.3 [4]): Given a symmetric B-DMC $W$ of capacity $I(W)$, we define a binary symmetric channel $BSC$, and a binary erasure channel $BEC$ of the same capacity through the equality

$$I(W) = I(BEC) = I(BSC).$$

Then, the random coding error exponent of the channels satisfy

$$E_r(R, BSC) \leq E_r(R, W) \leq E_r(R, BEC).$$

Proof: The equality of capacities is equivalent to

$$\lim_{\rho \to 0} \frac{E_0(\rho, W)}{\rho} = \lim_{\rho \to 0} \frac{E_0(\rho, BEC)}{\rho} = \lim_{\rho \to 0} \frac{E_0(\rho, BSC)}{\rho}.$$

But in this case, we know by Part 1 in Theorem 1 that we have

$$E_0(\rho_2, BSC) \leq E_0(\rho_2, W) \leq E_0(\rho_2, BEC),$$

for any $\rho_2 \in [0, 1]$. This, in turn, implies the inequality for the random coding exponent.

Finally, note that in [6] the above result of [4] was extended to the region where $\rho > -1$. Namely, amongst all symmetric B-DMCs of the same capacity, the BEC and the BSC are extremal with

$$E_0(\rho, BSC) \leq E_0(\rho, W) \leq E_0(\rho, BEC),$$

for all $\rho > -1$. In particular, [6, Theorem 1] can also be recovered from Theorem 1.
A. Convexity Lemmas

The proof of Theorem 1 rests on the next two lemmas. The lemmas are proved in the Appendix.

**Lemma 5:** For fixed values of \( \rho_1, \rho_2 \in \mathbb{R} \setminus \{-1\} \), we define the function \( \tilde{f}_{\rho_1, \rho_2}(t) \) by

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

for \( t \in [2^{-\rho}, 1) \cup (1, 2^{1-\rho}] \). Let \( \tilde{f}_\rho(t) \) denotes the function when \( \rho_1 = \rho_2 = \rho \). Then, \( \tilde{f}_\rho(t) \) is a concave function in \( t \) when \( \rho \in (0, 3) \), convex when \( \rho = (-1, 0] \) and \( \rho \in (-\infty, -1) \). Moreover, the function \( \tilde{f}_{\rho_1, \rho_2}(t) \) is concave when \( \rho_1, \rho_2 \in [0, 1] \) such that \( \rho_2 \geq \rho_1 \).

**Lemma 6:** For fixed values of \( \rho_1, \rho_2 \in \mathbb{R} \setminus \{-1\} \), the function \( f_{\rho_1, \rho_2}(t) \) defined as

\[
f_{\rho_1, \rho_2}(t) = g(\rho_2, g^{-1}(\rho_1, t)),
\]

for \( t \in [2^{-\rho}, 1) \cup (1, 2^{1-\rho}] \), is concave in \( t \) when \( \rho_1 \in (-1, 0] \) and \( \rho_2 \geq 0 \), when \( \rho_1 \in [0, 1] \) and \( \rho_2 \geq \rho_1 \), and when \( \rho_1 > 1 \) and \( \rho_2 \in (-1, 0) \), and the function is convex when \( \rho_1 > 1 \) and \( \rho_2 \in (0, 1) \).

B. Proof of Theorem 1

Before we start proving the theorem’s statement in its most general form, we will prove two particular cases of the theorem in the next two lemmas assuming \( \rho_1 = \rho_2 = \rho \).

**Lemma 7:** Given any fixed value of \( \rho \in (0, 3) \), suppose a B-DMC \( W \), a binary symmetric channel \( BSC \), and a binary erasure channel \( BEC \) satisfy the equality

\[
E_0(\rho, BSC) \leq E_0(\rho, W) \leq E_0(\rho, BEC).
\]

Then, the following holds:

\[
R(\rho, BSC) \leq R(\rho, W) \leq R(\rho, BEC),
\]

where the inequalities are strict if the inequalities in (66) are strict.

**Proof:** Let us define another binary erasure channel \( BEC^* \) and another binary symmetric channel \( BSC^* \) through the following equality:

\[
E_0(\rho, BSC^*) = E_0(\rho, W) = E_0(\rho, BEC^*).
\]
Observe that by (17), the equality condition in Equation (68) is equivalent to the equality of

\[ \mathbb{E}[g(\rho, Z)] = \mathbb{E}[g(\rho, Z_{\text{BEC}})] = g(\rho, z_{\text{BSC}}). \]  \hspace{1cm} (69)

Hence, the denominator in

\[ R(\rho, W) = \frac{\partial}{\partial \rho} E_0(\rho, W) = \frac{\mathbb{E}[-\partial g(\rho, Z)/\partial \rho]}{\mathbb{E}[g(\rho, Z)]} \]  \hspace{1cm} (70)

is the same for the three channels. Then, the proof can be completed using the concavity of the function \( \tilde{f}_\rho(t) \) in \( t \) for \( \rho \in (0, 3] \), which was shown in Lemma 5, and the special structure of the \( Z \) random variable of a BEC and a BSC. To see this, let us define the random variable \( T = g(\rho, Z) \in [2^{-\rho}, 1] \). Then, we note that \( \tilde{f}_\rho(T) = \partial g(\rho, Z)/\partial \rho \), and \( E[T] \) gives (69). So,

\[ R(\rho, W) = \frac{\mathbb{E}[\tilde{f}_\rho(T)]}{E[T]}, \quad R(\rho, \text{BSC}^*) = \frac{\tilde{f}_\rho(E[T])}{E[T]}. \]  \hspace{1cm} (71)

To derive the expression for \( R(\rho, \text{BEC}^*) \), recall by Fact 1 that \( Z_{\text{BEC}^*} = \{0, 1\} \). Using \( E[T] = \mathbb{E}[g(\rho, Z_{\text{BEC}^*})] \), we get

\[ P(Z_{\text{BEC}^*} = 0) = \frac{E[T] - 1}{2^{-\rho} - 1}. \]  \hspace{1cm} (72)

Hence,

\[ R(\rho, \text{BEC}^*) = \frac{\tilde{f}_\rho(2^{-\rho}) P(Z_{\text{BEC}^*} = 0) + \tilde{f}_\rho(1) P(Z_{\text{BEC}^*} = 1)}{E[T]}. \]  \hspace{1cm} (73)

Now, by the two sides of the Jensen’s inequality for concave functions we have

\[ \tilde{f}_\rho(1) + \frac{\tilde{f}_\rho(1) - \tilde{f}_\rho(2^{-\rho})}{1 - 2^{-\rho}} (E[T] - 1) \leq \mathbb{E}[\tilde{f}_\rho(T)] \leq \tilde{f}_\rho(E[T]). \]  \hspace{1cm} (74)

Dividing all sides by \( E[T] > 0 \) and negating the expressions in (74), we get

\[ R(\rho, \text{BSC}^*) \leq R(\rho, W) \leq R(\rho, \text{BEC}^*). \]  \hspace{1cm} (75)

The final step of the proof is to show (75) implies (67). For that purpose, recall that by Fact 2 that the set of BSCs and the set of BECs are strictly ordered in their \( E_0 \) and \( R \) parameters for \( \rho \in (0, 3] \). As we have

\[ E_0(\rho, \text{BSC}) \leq E_0(\rho, \text{BSC}^*), \]  \hspace{1cm} (76)

\[ E_0(\rho, \text{BEC}^*) \leq E_0(\rho, \text{BEC}), \]  \hspace{1cm} (77)

16
we conclude by Lemma 2 that

\begin{align}
R(\rho, BSC) &\leq R(\rho, BSC^*), \\
R(\rho, BEC^*) &\leq R(\rho, BEC)
\end{align}

(78)

(79)

holds for \( \rho \geq 0 \). From this (67) follows. Moreover, if the inequalities in (66) are strict than the ones in (76) and (77), and thus, (78) and (79) are strict as well. Consequently, the inequalities in (67) hold strictly as claimed. 

\textbf{Remark 3:} Note that Lemma 7 and Corollary 1 are of the same flavor. Indeed, one can easily derive one from the other using the degradation argument discussed in Fact 2. So, the result of [11] could also have been used to characterize the behavior of the \( E_0 \) curves for the \( \rho \in (0, 1] \) interval. However, the proofs of the lemma and the corollary are different as they involve different convexity analysis.

\textbf{Lemma 8:} Given any fixed value of \( \rho \in (-1, 0) \), suppose a B-DMC \( W \), a binary symmetric channel \( BSC \), and a binary erasure channel \( BEC \) satisfy the condition (66) of Lemma 7. Then, the following holds:

\begin{align}
R(\rho, BEC) &\leq R(\rho, W) \leq R(\rho, BSC),
\end{align}

(80)

where the inequalities are strict if the inequalities in (66) are strict.

\textbf{Proof:} Let \( BEC^* \) and \( BSC^* \) be as defined in the proof of Lemma 7. Once again, the equality condition in (68) implies the denominator in (70) is the same for the three channels. Then, the inequalities

\begin{align}
R(\rho, BEC^*) &\leq R(\rho, W) \leq R(\rho, BSC^*)
\end{align}

(81)

follow using the convexity of the function \( \tilde{f}_\rho(t) \) in \( t \) when \( \rho \in (-1, 0] \), which was shown in Lemma 5 and applying Jensen’s inequalities. Finally, since \( E_0(\rho, BSC) \leq E_0(\rho, BSC^*) \) and \( E_0(\rho, BEC^*) \leq E_0(\rho, BEC) \), we know by Fact 2 that these BSCs and BECs are ordered by degradation, and we conclude by Lemma 2 that we have \( R(\rho, BSC^*) \leq R(\rho, BSC) \) and \( R(\rho, BEC) \leq R(\rho, BEC^*) \), for \( \rho \in (-1, 0] \). From this (80) follows. The claim about the strictness of the inequalities can be proved similarly as in the proof of Lemma 7.

Now, we are ready to prove the theorem.

\textbf{Proof of Theorem 1}:
We will first prove the claims for \( \rho_1 \in (-1, 0) \cup (0, \infty) \), leaving the case \( \rho_1 = 0 \) to the last. In fact, we will show that the results proved for \( \rho_1 \in (-1, 0) \cup (0, \infty) \) will immediately extend to \( \rho_1 = 0 \) by the continuity of \( E_0 \) in its arguments.

We start by proving the inequalities (48) and (49) in Part 1 for the case \( \rho_1 \in (0, 3] \). By Lemma 7, we know that (48) holds for \( \rho_2 = \rho_1 \). So, we only need to prove the theorem for \( \rho_2 \in (0, 3] \) such that \( \rho_2 > \rho_1 \). By the continuity of \( E_0(\rho, \text{BEC}) \) and \( E_0(\rho, \text{BSC}) \) in the channels’ erasure and crossover probabilities, respectively, it suffices to show that

\[
E_0(\rho_1, \text{BSC}) < E_0(\rho_1, W) < E_0(\rho_1, \text{BEC})
\]  
implies

\[
E_0(\rho_2, \text{BSC}) < E_0(\rho_2, W) < E_0(\rho_2, \text{BEC}).
\]  

Then, Lemma 7 will imply

\[
R(\rho_2, \text{BSC}) < R(\rho_2, W) < R(\rho_2, \text{BEC}).
\]

We define \( D(\rho) = E_0(\rho, W) - E_0(\rho, \text{BEC}) \). Let \( D'(\rho) \) denotes the first derivative of \( D(\rho) \) with respect to \( \rho \). Noting that \( R(\rho, W) = \frac{\partial}{\partial \rho} E_0(\rho, W) \), the inequality in (83) is implied by the following statement:

\[
D(\rho_1) < 0 \quad \text{and by Lemma 7} \quad (D(\rho) < 0 \Rightarrow D'(\rho) < 0) \quad \Rightarrow \quad D(\rho_2) < 0.
\]

But this is true by elementary considerations on differential equations. Indeed, suppose to the contrary that

\[
D(\rho_1) < 0, \quad \text{and} \quad (D(\rho) < 0 \Rightarrow D'(\rho) < 0), \quad \text{but} \quad D(\rho_2) \geq 0.
\]

Then, there exists \( \rho_1 < \rho_3 \leq \rho_2 \) such that \( D(\rho) < 0 \), for \( \forall \rho \in [\rho_1, \rho_3) \), and \( D(\rho_3) = 0 \). But then there exists \( \rho_1 < \rho_4 < \rho_3 \) such that

\[
D'(\rho_4) = \frac{D(\rho_3) - D(\rho_1)}{\rho_3 - \rho_1} > 0,
\]
and \( D(\rho_4) < 0 \), contradicting the assumption.

The inequality for the BSC can be obtained similarly by letting \( D(\rho) = E_0(\rho, \text{BSC}) - E_0(\rho, W) \) and applying the above argument once more.

We continue with the proof of the inequalities in (50) and (51) in Part 2 for the case \( \rho_1 \in (-1, 0) \). The proof follows along the same lines of the previous part. By Lemma 8, we know
that the inequalities in (50) hold for $\rho_2 = \rho_1$. So, we only need to prove the theorem for $\rho_2 < \rho_1$. By the continuity of $E_0(\rho, BEC)$ and $E_0(\rho, BSC)$ in the channels’ erasure and crossover probabilities, respectively, it suffices to show that

$$E_0(\rho_1, BSC) < E_0(\rho_1, W) < E_0(\rho_1, BEC)$$

implies

$$E_0(\rho_2, BSC) < E_0(\rho_2, W) < E_0(\rho_2, BEC).$$

Then, Lemma 8 will imply

$$R(\rho_2, BEC) < R(\rho_2, W) < R(\rho_2, BSC).$$

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

$$D(\rho_1) < 0 \quad \text{and by Lemma 8} \quad (D(\rho) < 0 \Rightarrow D'(\rho) > 0) \quad \Rightarrow \quad D(\rho_2) < 0.$$

But this is true by an analogous reasoning as before.

The inequality for the BSC can be obtained similarly by letting $D(\rho) = E_0(\rho, BSC) - E_0(\rho, W)$ and applying the above argument once more. This concludes the proof of Part 2.

For Part 3, we will only do the proof of (52) for the case $\rho_1 \in (-1, 0)$ and $\rho_2 \geq 0$ as all the other claims can be proved in the same way using the convexity properties of the function $f_{\rho_1, \rho_2}(t)$ discussed in Lemma 6.

Let $T = g(\rho_1, Z)$. We know that the condition in (46) is equivalent to

$$\mathbb{E}[g(\rho_1, Z_{BEC})] \leq \mathbb{E}[g(\rho_1, Z)] \leq g(\rho_1, Z_{BSC}). \quad (88)$$

Define the BEC $BEC^*$ and the BSC $BSC^*$ through the equality

$$\mathbb{E}[g(\rho_1, Z)] = \mathbb{E}[g(\rho_1, Z_{BEC^*})] = g(\rho_1, Z_{BSC^*}). \quad (89)$$

As by Lemma 6 we know the function $f_{\rho_1, \rho_2}(t)$ is concave in $t$ when $\rho_1 \in (-1, 0)$ and $\rho_2 \geq 0$, we can apply the two sides of Jensen’s inequality to obtain

$$\mathbb{E}[f_{\rho_1, \rho_2}(g(\rho_1, Z_{BEC^*}))] \leq \mathbb{E}[f_{\rho_1, \rho_2}(T)] \leq f_{\rho_1, \rho_2}(g(\rho_1, Z_{BSC^*})), \quad (90)$$

which is equivalent to

$$\mathbb{E}[g(\rho_2, Z_{BEC^*})] \leq \mathbb{E}[g(\rho_2, Z)] \leq g(\rho_2, Z_{BSC^*}). \quad (91)$$
To get the claimed inequalities in (52), we simply need to use the ordering argument based on Fact 2 for the two BECs and the two BSCs. As we have illustrated this argument before in the proof of Lemma 7, we do not repeat it here.

The last step is to prove the theorem for the case $\rho_1 = 0$. We will only present the proof extension for the inequalities $(b')$ and $(c')$ in Part 1 as the same argument can be used to extend all the remaining results. Moreover, once again by the continuity of $E_0(\rho, BEC)$ in the channels’ erasure probability, it suffices to show the results assuming $(a'_0)$ in (47) holds with strict inequality.

So, we assume the given channels $W$ and $BEC$ satisfy $I(W) < I(BEC)$. Then,

$$\lim_{\rho \to 0^+} \frac{E_0(\rho, W) - E_0(\rho, BEC)}{\rho} = I(W) - I(BEC) < 0. \quad (92)$$

(We assumed $\rho \to 0^+$ for simplicity as the above limit for $\rho \to 0$ is well defined). Hence, for any sufficiently small $\rho > 0$, we have

$$E_0(\rho, W) < E_0(\rho, BEC). \quad (93)$$

Moreover, we already proved that this implies

$$E_0(\rho_2, W) \leq E_0(\rho_2, BEC). \quad (94)$$

for all $\rho_2 \in [\rho, 3]$. As $\rho > 0$ is arbitrary, we conclude the result should hold for all $\rho_2 \in [0, 3]$.

Now, we can carry the proof as follows. First, we let $\epsilon \in [0, 1]$ be the erasure probability of the BEC $BEC_\epsilon$ which satisfies $I(W_\epsilon) = I(W)$. Then, we take a sequence of BECs $BEC_{\epsilon_n}$ of erasure probabilities $\epsilon_n \in [0, 1]$ such that the sequence $\epsilon_n$ is increasing to $\epsilon$. In this case, we know that

$$I(W) < I(BEC_{\epsilon_n}). \quad (95)$$

By the previous argument, we conclude that for all the channels $BEC_{\epsilon_n}$,

$$E_0(\rho_2, W) \leq E_0(\rho_2, BEC_{\epsilon_n}) \quad (96)$$

holds for all $\rho_2 \in [0, 3]$. Taking the limit for the sequence $\epsilon_n$, we conclude by continuity that the result also holds for the channel $BEC_\epsilon$, i.e.,

$$E_0(\rho_2, W) \leq E_0(\rho_2, BEC_\epsilon) \quad (97)$$

holds for $\rho_2 \in [0, 3]$. As the ordering $E_0(\rho_2, BEC_\epsilon) \leq E_0(\rho_2, BEC)$ holds, the inequality $(c')$ in (49) is proved. By Lemma 7, the inequality $(b')$ follows. $\blacksquare$
C. Extremality of Rényi Entropies

In this section, we show how the results of Theorem 1 can be translated into extremalities for Rényi entropies using the definition given in (16).

Observe that the assumption in (46) of Theorem 1 can be equivalently stated as

\[
\frac{E_0(\rho, \text{BSC})}{\rho_1} \leq \frac{E_0(\rho, \text{W})}{\rho_1} \leq \frac{E_0(\rho, \text{BEC})}{\rho_1},
\]

for \( \rho_1 > 0 \), and

\[
\frac{E_0(\rho, \text{BEC})}{\rho_1} \leq \frac{E_0(\rho, \text{W})}{\rho_1} \leq \frac{E_0(\rho, \text{BSC})}{\rho_1},
\]

for \( \rho_1 \in (-1, 0) \). Note that by Lemma 2 while for \( \rho_1 > 0 \) a worst BEC and a worst BSC has a smaller \( E_0 \) parameter, for \( \rho \in (-1, 0) \) the opposite is true. Consequently, all the results obtained for the parameter \( E_0(\rho, \text{W}) \) can be restated in terms of Rényi entropies via (16). For the sake of brevity, we will only restate in the next corollary the result given in (53) in Part 3 of the theorem in terms of Rényi entropies.

Corollary 3: Given a binary uniform random variable \( X \), among all jointly distributed random variables \((X, Y)\) of equal Rényi equivocation \( H_\alpha(X \mid Y) \) of order \( \alpha \in (0, 1/2] \), the Rényi equivocation of order \( \beta \geq 0 \) such that \( \beta \geq \alpha \) is maximized when \( X \) and \( Y \) are coupled by a BEC, and minimized when coupled by a BSC. For \( \beta \leq \alpha \) values, the maximizing and minimizing distributions are reversed.

Proof: Recall that \( \alpha = 1/(1 + \rho) \). So for \( \alpha \in (0, 1/2] \), we have \( \rho \in [0, 1] \). Moreover, \( \alpha \) is decreasing with \( \rho \). Hence, the inequalities for \( \beta \leq \alpha \) and for \( \beta \geq \alpha \) follow directly from (53) in Part 3 of Theorem 1 using the definition given in equation (16) together with the fact that \( H_\alpha(X) = 1 \) under the uniform distribution.

D. Graphical Interpretation of the Extremality Results

In this section, we provide a graphical interpretation of the theorem and the corollaries through Figures 1 to 5. Suppose that the \( E_0 \) curves of a given B-DMC, a BEC, and a BSC pass through a given point \((\rho_0, e_0)\), for some \( \rho_0 > -1 \).

By the results stated in (51) and (52) of Theorem 1 we know that when \( \rho_0 \in (-1, 0) \), then these curves do not intersect again except at \( \rho = 0 \), and the BEC and BSC always remain extremal even though their extremal behaviour get reversed after the intersection points. Figure 1 illustrates this relation.
Fig. 1. Extremality of $E_0(\rho)$ when the channels intersect at $\rho_0 \in (-1, 0)$. Dashed line: BEC(0.3) & Solid line: BSC(0.1102).

Fig. 2. Extremality of $E_0(\rho)$ when the channels have equal capacity 0.5. Dashed line: BEC(0.5) & Solid line: BSC(0.1102).

Fig. 3. Extremality of $E_0(\rho)$ when the channels have equal cut-off rate. Dashed line: BEC(0.626278) & Solid line: BSC(0.1102).
Fig. 4. Extremality of $E_0(\rho)$ when the channels have equal $E_0(\rho^*)$ and equal rate at $\rho^* > 1$. Dashed line: BEC(0.6777) & Solid line: BSC(0.1102).

Fig. 5. Extremality of $E_0(\rho)$ when the channels intersect at $\rho_0 > 1$. Dashed line: BEC(0.67) & Solid line: BSC(0.1102).

A special case where the $E_0$ curves of the BEC and the BSC remain extremal for the entire $\rho > -1$ region, and with no reversal, corresponds to channels of the same capacity; as discussed after Corollary [2]. Theorem [1] shows that the $E_0$ curves of these channels are upper bounded by the BEC’s curve and lower bounded by the BSC’s one. Figure 2 illustrates this relation.

Another situation where the $E_0$ curves of the BEC and the BSC exhibit extremality for the entire region $\rho > -1$ occurs when $\rho_0 \in (0, 1]$; (49) and (53) of Theorem [1] imply the BEC and the BSC will be $E_0$ extremal, one again with the extremalities reversed after the intersections.
Figure 3 illustrates this relation.

Now, we consider the case when $\rho_0 > 1$. By Theorem 1, we know the curves only intersect at $\rho = 0$ in the interval $\rho \in (-1, 1]$, and the BEC and the BSC are extremal in $(-1, 0)$ and $(0, 1)$ with reversed extremalities. Although the theorem provides a partial result, it is not clear what happens in the interval $\rho > 1$. It turns out that the BEC and the BSC are no longer extremal for $\rho > 1$ in general. We will show this result by studying the intersection points of the $E_0$ curves of a given BSC with different BECs using Lemma 4.

Suppose a BEC $BEC$ and a BSC $BSC$ satisfy

\[ E_0(\rho^*, BEC) = E_0(\rho^*, BSC), \]
\[ R(\rho^*, BEC) = R(\rho^*, BSC), \]

for a particular $\rho^* > 1$. We know by Lemma 4 that this corresponds to the case the $E_0$ curves of these two channels are tangent at $\rho^* > 1$ and do not intersect at any other point except $\rho = 0$. Moreover, by Theorem 1, we know the capacities of the channels are such that $I(BEC) \leq I(BSC)$. Figure 4 illustrates this relation.

Suppose the erasure probability of the BEC channel is increased. By the ordering we discussed in Fact 2, it is not difficult to see that the $E_0$ curves of the BSC and that BEC will not intersect at any point other than $\rho = 0$. On the other hand, assume instead the erasure probability of the channel is decreased such that the capacity of the new BEC is still smaller than the capacity of the BSC. In this case, as long as the cut-off rate of the BSC is larger than the cut-off rate of the BEC, the BSC and the new BECs will intersect twice after $\rho = 0$, first in the interval $(1, \rho^*)$, then after $\rho^*$. Figure 5 illustrates this relation. Once the cut-off rate of the BEC becomes larger than that of the BSC, we are back at the situation where the intersection point falls in the interval $[0, 1]$, and we recover the general extremality result we have already discussed. Then, we can keep decreasing the erasure probability until the BEC and the BSC have the same capacity to recover another special case. Finally, decreasing more the erasure probability, until there is no other intersection anywhere except at $\rho = 0$, will cause the $E_0$ curves of the BSC and the new BECs to intersect in the interval $(-1, 0)$, in which case once more the BSC and the BECs will be $E_0$ extremal for the entire $\rho > -1$ region.

The analysis above shows us that most of the BECs and the BSCs whose $E_0$ curves intersect in the interval $\rho > 1$ have two intersection points in that interval. In such a case, the BEC and
the BSC are no longer extremal as we do not expect a class of B-DMCs \( \mathcal{W} \) which satisfy for all \( W \in \mathcal{W} \) the equality

\[
E_0(\rho_0, W) = E_0(\rho_0, BEC) = E_0(\rho_0, BSC),
\]

for any fixed \( \rho_0 > 1 \), to intersect a second time at the same point where the BEC and the BSC intersect the second time in the interval \((1, \infty)\).

IV. CONCLUSIONS

We have described certain extremalities for B-DMCs when the information measure is Gallager’s \( E_0 \) evaluated under the uniform input distribution. These properties yield in straightforward fashion recent results by Fabregas et al.[4], [6], and also extremal properties for the Rényi entropies.

Finally, it is worth emphasizing that all the conclusions of the paper are valid for arbitrary binary input channels as long as one evaluates all the quantities under the uniform input distribution.

ACKNOWLEDGMENT

The author would like to thank Emre Telatar for helpful discussions. This work was supported by Swiss National Science Foundation under grant number 200021-125347/1.

APPENDICES

The Appendices contain four parts. In the first three of them, we prove Lemma 1, Lemma 5, and Lemma 6 respectively. The final part proves two other lemmas needed in these proofs.

Appendix I

Proof of Lemma 1. Taking the first derivative of (18) with respect to \( z \), we get

\[
\frac{\partial g(\rho, z)}{\partial z} = \left(\frac{1}{2} (1 + z)^{1+\rho} + \frac{1}{2} (1 - z)^{1+\rho} \right)^\rho \left( \frac{1}{2} (1 + z)^{\frac{1}{1+\rho}} - \frac{1}{2} (1 - z)^{\frac{1}{1+\rho}} \right)
\]

\[
\geq 0
\]

\[
\geq 0
\]

(101)
As we have \( \frac{1 - z}{1 + z} \leq 1 \), for \( \forall z \in [0, 1] \), the monotonicity claims follow by noting that when \( \rho \in (-\infty, -1) \cup [0, \infty) \):

\[
\frac{\rho}{1 + \rho} \geq 0 \quad \Rightarrow \quad \left( 1 - \left( \frac{1 - z}{1 + z} \right)^{\frac{\rho}{1+\rho}} \right) \leq 0 \quad \Rightarrow \quad \frac{\partial g(\rho, z)}{\partial z} \leq 0,
\]

and when \( \rho \in (-1, 0] \):

\[
\frac{\rho}{1 + \rho} \leq 0 \quad \Rightarrow \quad \left( 1 - \left( \frac{1 - z}{1 + z} \right)^{\frac{\rho}{1+\rho}} \right) \geq 0 \quad \Rightarrow \quad \frac{\partial g(\rho, z)}{\partial z} \geq 0.
\]

Taking the second derivative with respect to \( z \), we get

\[
\frac{\partial^2 g(\rho, z)}{\partial z^2} = -\frac{\rho}{1 + \rho} (1 - z^2)^{\frac{\rho}{1+\rho}} - 2 \left( \frac{1}{2} (1 + z)^{\frac{1}{1+\rho}} + \frac{1}{2} (1 - z)^{\frac{1}{1+\rho}} \right)^{1+\rho}.
\]

The convexity claims follow once again by inspecting the sign of \( \frac{\rho}{1 + \rho} \) in different intervals, i.e. when \( \rho \in (-\infty, -1) \cup [0, \infty) \):

\[
\frac{\rho}{1 + \rho} \geq 0 \quad \Rightarrow \quad \frac{\partial^2 g(\rho, z)}{\partial z^2} \leq 0,
\]

and when \( \rho \in (-1, 0] \):

\[
\frac{\rho}{1 + \rho} \leq 0 \quad \Rightarrow \quad \frac{\partial^2 g(\rho, z)}{\partial z^2} \geq 0.
\]

\[\blacksquare\]

Appendix II

Proof of Lemma We begin by introducing some definitions to simplify notations. Let

\[
g'(\rho, z) = \frac{\partial g(\rho, z)}{\partial z}.
\]

We define

\[
h(z) = \frac{1 - z}{1 + z},
\]

\[
\alpha(\rho, z) = (1 + h(z)^{\frac{1}{1+\rho}})^\rho,
\]

\[
\beta(\rho, z) = (1 - h(z)^{\frac{1}{1+\rho}}),
\]

26
for \( z \in [0, 1], \rho \in \mathbb{R} \setminus \{ -1 \} \). By equation 101 in Lemma 1, we have

\[
g'(\rho, z) = \left( \frac{1}{2} \right)^{1+\rho} \alpha(\rho, z) \beta(\rho, z). \quad (106)
\]

Taking the first derivative of \( \tilde{f}_{\rho_1, \rho_2}(t) \) with respect to \( t \), we obtain

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

\[
= \frac{\partial}{\partial \rho_2} g(\rho_2, g^{-1}(\rho_1, t))
\]

\[
= \frac{\partial}{\partial \rho_2} g'(\rho_1, g^{-1}(\rho_1, t)). \quad (109)
\]

Let \( z = g^{-1}(\rho_1, t) \). As \( g(\rho, z) \) is a monotone function in \( z \) by Lemma 1 in Appendix I, so is \( z = g^{-1}(\rho, t) \) in \( t \). Hence, we can check the convexity of \( \tilde{f}_{\rho_1, \rho_2}(t) \) with respect to \( t \) from the monotonicity with respect to \( z \) of the following expression:

\[
\frac{\partial}{\partial \rho_2} \frac{g'(\rho_2, z)}{g'(\rho_1, z)} = \frac{\partial}{\partial \rho_2} \left( 2^{\rho_1 - \rho_2} \frac{\alpha(\rho_2, z) \beta(\rho_2, z)}{\alpha(\rho_1, z) \beta(\rho_1, z)} \right)
\]

\[
= \frac{2^{-\rho_2} \alpha(\rho_2, z) \beta(\rho_2, z)}{2^{-\rho_1} \alpha(\rho_1, z) \beta(\rho_1, z)} \left( \frac{2^{-\rho_2} \alpha(\rho_2, z)}{2^{-\rho_2} \alpha(\rho_2, z)} + \frac{\partial \beta(\rho_2, z)}{\partial \rho_2} \beta(\rho_2, z) \right) \quad (110)
\]

where

\[
\frac{\partial 2^{-\rho_2} \alpha(\rho_2, z)}{\partial \rho_2} = \frac{\partial}{\partial \rho_2} \left( \frac{1}{2} + \frac{1}{2} h(z) \frac{1}{1+\rho_2} \right)^{\rho_2}
\]

\[
= \left( \frac{1}{2} + \frac{1}{2} h(z) \frac{1}{1+\rho_2} \right)^{\rho_2} \left( \log \left( \frac{1}{2} + \frac{1}{2} h(z) \frac{1}{1+\rho_2} \right) + \rho_2 \frac{1}{2} h(z) \frac{1}{1+\rho_2} \log h(z) \right)
\]

\[
= 2^{-\rho_2} \alpha(\rho_2, z) \left( \log \left( \frac{1}{2} + \frac{1}{2} h(z) \frac{1}{1+\rho_2} \right) - \frac{\rho_2 h(z) \frac{1}{1+\rho_2} \log h(z)}{(1+\rho_2)^2 \left( 1 + h(z) \frac{1}{1+\rho_2} \right)} \right), \quad (113)
\]

and

\[
\frac{\partial \beta(\rho_2, z)}{\partial \rho_2} = \frac{\partial}{\partial \rho_2} \left( 1 - h(z) \frac{\rho_2}{1+\rho_2} \right)
\]

\[
= \frac{1}{(1+\rho_2)^2} h(z) \frac{\rho_2}{1+\rho_2} \log h(z). \quad (115)
\]
Hence, the expression inside the parenthesis in (143) equals
\[
\log \left( \frac{1}{2} + \frac{1}{2} h(z)^{1+\rho_2} \right) - \frac{\rho_2 h(z)^{1+\rho_2} \log h(z)}{(1 + \rho_2)^2 (1 + h(z)^{1+\rho_2})} + \frac{h(z)^{1+\rho_2} \log h(z)}{(1 + \rho_2)^2 (1 - h(z)^{1+\rho_2})} \quad (116)
\]
\[
= \log \left( \frac{1}{2} + \frac{1}{2} h(z)^{1+\rho_2} \right) - \frac{\rho_2 h(z)^{1+\rho_2} \log h(z)}{(1 + \rho_2)^2 (1 + h(z)^{1+\rho_2})} + \frac{\log h(z)}{(1 + \rho_2)^2 (h(z)^{1+\rho_2} - 1)} \quad (117)
\]
\[
(118)
\]
To simplify derivations we define
\[
\Phi(k, \rho_1, \rho_2) = \left( \frac{1}{2} + \frac{1}{2} k^{1+\rho_2} \right)^{\rho_2} \left( 1 - k^{-\rho_1} \right)^{\rho_1} \quad (119)
\]
\[
\Psi(k, \rho_2) = \log \left( \frac{1}{2} + \frac{1}{2} k^{1+\rho_2} \right) + \frac{\log k}{(1 + \rho_2)^2} \left( -\frac{\rho_2 k^{-\rho_1}}{1 + k^{1+\rho_2}} + \frac{1}{k^{\rho_2} - 1} \right) \quad (120)
\]
\[
= \log \left( \frac{1}{2} + \frac{1}{2} k^{1+\rho_2} \right) + \frac{\left( 1 + k^{1+\rho_2} - \rho_2 \left( k - k^{1+\rho_2} \right) \right) \log k}{(1 + \rho_2)^2 \gamma(k, \rho_2)}
\]
where
\[
\gamma(k, \rho_2) = \left( 1 + k^{1+\rho_2} \right) \left( k^{\rho_2} - 1 \right). \quad (121)
\]
Then, equation (143) equals to the product
\[
\frac{\partial}{\partial \rho_2} g'(\rho_2, z) = \Phi(h(z), \rho_1, \rho_2) \Psi(h(z), \rho_2). \quad (122)
\]
Let \( k = h(z) \in [0, 1] \). As \( k = h(z) \) is decreasing in \( z \), to check the monotonicity of the above expression with respect to \( z \), we can equivalently check the monotonicity with respect to \( k \) of the following expression:
\[
\Phi(k, \rho_1, \rho_2) \Psi(k, \rho_2). \quad (123)
\]
Taking the derivative with respect to \( k \) gives
\[
\frac{\partial \Phi(k, \rho_1, \rho_2) \Psi(k, \rho_2)}{\partial k} = \Phi'(k, \rho_1, \rho_2) \Psi(k, \rho_2) + \Phi(k, \rho_1, \rho_2) \Psi'(k, \rho_2)
\]
\[
= \Phi(k, \rho_1, \rho_2) \Psi(k, \rho_2) \left( \frac{\partial \log \Phi(k, \rho_1, \rho_2)}{\partial k} + \frac{\Psi'(k, \rho_2)}{\Psi(k, \rho_2)} \right) \quad (124)
\]
where \( \Phi'(k, \rho_1, \rho_2) = \frac{\partial \Phi(k, \rho_1, \rho_2)}{\partial k}, \) and \( \Psi'(k, \rho) = \frac{\partial \Psi(k, \rho)}{\partial k} \).
Now, we derive the expressions in Equation (124):

\[
\log \Phi(k, \rho_1, \rho_2) = \rho_2 \log \left( \frac{1}{2} + \frac{1}{2} k^{\frac{\rho_2}{1 + \rho_2}} \right) + \log \left( 1 - k^{\frac{-\rho_2}{1 + \rho_2}} \right) - \rho_1 \log \left( \frac{1}{2} + \frac{1}{2} k^{\frac{1}{1 + \rho_1}} \right) - \log \left( 1 - k^{\frac{-\rho_1}{1 + \rho_1}} \right)
\]

\[
\frac{\partial \log \Phi(k, \rho_1, \rho_2)}{\partial k} = \frac{\rho_2}{1 + \rho_2} \frac{k^{\frac{\rho_2}{1 + \rho_2}}}{1 + k^{\frac{1}{1 + \rho_2}}} + \frac{\rho_2}{1 + \rho_2} \frac{k^{\frac{-\rho_2}{1 + \rho_2}} - 1}{1 + k^{\frac{-\rho_2}{1 + \rho_2}}} - \frac{\rho_1}{1 + \rho_1} \frac{k^{\frac{\rho_1}{1 + \rho_1}}}{1 + k^{\frac{1}{1 + \rho_1}}} - \frac{\rho_1}{1 + \rho_1} \frac{k^{\frac{-\rho_1}{1 + \rho_1}} - 1}{1 + k^{\frac{-\rho_1}{1 + \rho_1}}}
\]

\[
= F(k, \rho_2) - F(k, \rho_1)
\]

where

\[
F(k, \rho) = \frac{\rho}{1 + \rho} \frac{1 + k}{k} \gamma(k, \rho),
\]

and

\[
\Psi'(k, \rho_2) = \frac{\partial}{\partial k} \left( \log \left( \frac{1}{2} + \frac{1}{2} k^{\frac{1}{1 + \rho_2}} \right) + \log k \left( \frac{1}{1 + \rho_2} \right)^2 \left( - \frac{\rho_2 k^{\frac{-\rho_2}{1 + \rho_2}}}{1 + k^{\frac{1}{1 + \rho_2}}} + \frac{1}{k^{\frac{\rho_2}{1 + \rho_2}} - 1} \right) \right)
\]

\[
= \frac{k^{\frac{\rho_2}{1 + \rho_2}}}{(1 + \rho_2)(1 + k^{\frac{1}{1 + \rho_2}})} + \frac{1}{(1 + \rho_2)^2} \left( - \frac{\rho_2 k^{\frac{-\rho_2}{1 + \rho_2}}}{1 + k^{\frac{1}{1 + \rho_2}}} + \frac{1}{k^{\frac{\rho_2}{1 + \rho_2}} - 1} \right)
\]

\[
+ \frac{\log k}{(1 + \rho_2)^2} \left( - \frac{\rho_2 k^{\frac{-\rho_2}{1 + \rho_2}}}{1 + k^{\frac{1}{1 + \rho_2}}} \right)^2 - \frac{\rho_2 k^{\frac{-\rho_2}{1 + \rho_2}}}{(1 + \rho_2) \left( 1 + k^{\frac{1}{1 + \rho_2}} \right)^2} - \frac{\rho_2 k^{\frac{-\rho_2}{1 + \rho_2}}}{(1 + \rho_2) \left( 1 + k^{\frac{1}{1 + \rho_2}} \right)^2}
\]

\[
= \frac{k + 1}{(1 + \rho_2)^2} k \gamma(k, \rho_2) - \frac{\rho_2 (k + 1)}{(1 + \rho_2)^3} k \gamma^2(k, \rho_2)
\]

where \(\gamma(k, \rho)\) is defined in Equation (121).

To summarize the steps so far, we have shown that the second derivative of \(\tilde{f}_{\rho_1, \rho_2}(t)\) with
respect to $t$ is given by

$$
\frac{\partial^2 \tilde{f}_{\rho_1, \rho_2}(t)}{\partial t^2} = \frac{\partial}{\partial t} \frac{\partial}{\partial \rho_2} g'(\rho_2, g^{-1}(\rho_1, t))
$$

(127)

$$
= \frac{\partial}{\partial z} \left( \frac{\partial}{\partial \rho_2} g'(\rho_2, z) \right) \frac{\partial z}{\partial t}
$$

(128)

$$
= \frac{\partial \Phi(k, \rho_1, \rho_2)}{\partial k} \frac{\partial z}{\partial t} \frac{\partial k}{\partial z}
$$

(129)

where $z = g^{-1}(\rho_1, t)$, $k = h(z)$ with $h(z)$ defined in (103), $\Phi(k, \rho_1, \rho_2)$ given by (119), and $\Psi(k, \rho_2)$ given by (120).

We first prove the claims of the lemma for $\rho_1 = \rho_2 = \rho$. Coming back to Equation (124),

$$
\frac{\partial \Phi(k, \rho_1, \rho_2)}{\partial k} \Psi(k, \rho_2) \frac{\partial z}{\partial t} \frac{\partial k}{\partial z}
$$

(130)

as $\Phi(k, \rho, \rho) = 1$, and $\frac{\partial \log \Phi(k, \rho, \rho)}{\partial k} = 0$. Hence to prove the convexity claims, we need to investigate the sign of $\Psi'(k, \rho)$ we derived in Equation (126).

Note that the factor in front of the parenthesis in Equation (126) is always positive for $k \in [0, 1]$, $\rho_2 \in \mathbb{R} \setminus \{-1\}$, and the term inside the parenthesis equals the function $m(k, \rho_2)$ defined in Lemma 9 in Appendix IV. So the sign of $\Psi'(k, \rho_2)$ is determined by the sign of $m(k, \rho_2)$. By Lemma 9, we have

$$
\Psi'(k, \rho_2) \geq 0, \quad \forall \rho_2 < -1
$$

(131)

$$
\Psi'(k, \rho_2) \leq 0, \quad \forall \rho_2 \in (-1, 0),
$$

(132)

$$
\Psi'(k, 0) = 0,
$$

(133)

$$
\Psi'(k, \rho_2) \leq 0, \quad \forall \rho_2 \in (0, \rho^*(k)),
$$

(134)

$$
\Psi'(k, \rho^*(k)) = 0,
$$

(135)

$$
\Psi'(k, \rho_2) \geq 0, \quad \forall \rho_2 \geq \rho^*(k).
$$

(136)

where $\rho^*(k) \geq 3$ is a constant which depends on $k \in [0, 1]$.

As $k$ is decreasing in $z$, which is non-increasing in $t$ when $\rho \geq 0$ by Lemma 1, we have

$$
\frac{\partial^2 \tilde{f}_\rho(t)}{\partial t^2} = \Psi'(k, \rho) \left. \frac{\partial k}{\partial z} \right|_{\rho \geq 0} \left. \frac{\partial z}{\partial t} \right|_{\rho < 0} \leq 0
$$

(137)
for \( \rho \in [0, \rho^*(k)] \), and
\[
\frac{\partial^2 \tilde{f}_\rho(t)}{\partial t^2} = \Psi'(k, \rho) \frac{\partial k}{\partial z} \frac{\partial z}{\partial t} \geq 0
\]
(138)

for \( \rho \geq \rho^*(k) \). Hence, the function \( \tilde{f}_\rho(t) \) is concave in \( t \) when \( \rho \in (0, 3] \) as claimed.

On the other hand, we know by Lemma [1] that \( z \) is non-decreasing in \( t \) when \( \rho \in (-1, 0) \). Hence, the function \( \tilde{f}_\rho(t) \) is convex in \( t \) whenever \( \rho \in (-1, 0) \).

Finally, when \( \rho < -1 \), \( z \) is non-increasing in \( t \) by Lemma [1] so that \( \tilde{f}_\rho(t) \) is convex in \( t \).

To prove the last claim of the lemma concerned with the case \( \rho_1, \rho_2 \in (0, 1] \) such that \( \rho_1 < \rho_2 \), we need to determine the sign of \( \Psi(k, \rho_2) \). Note that, \( \Psi'(k, \rho) \leq 0 \) for \( \rho \in (0, 3] \) implies
\[
\Psi(k, \rho) \geq \lim_{k \to 1} \Psi(1, \rho) = \frac{2}{(1+\rho)^2} \lim_{k \to 1} \frac{\log k}{\gamma(k, \rho)} = \frac{1}{\rho(1+\rho)} \geq 0
\]

since
\[
\lim_{k \to 1} \frac{\log k}{\gamma(k, \rho)} = 0 \lim_{k \to 1} \frac{\partial \log k/\partial k}{\partial \gamma(k, \rho)/\partial k} = \lim_{k \to 1} \frac{k + \rho k}{k(1 + \rho k - k^{1/\rho} + gk^{\varphi_{1/\rho}})} = \frac{1 + \rho}{2\rho}.
\]

As a result, \( \Psi(k, \rho_2) \geq 0 \) whenever \( \rho_2 \in (0, 3] \).

Recall that we are interested in the sign of the following expression
\[
\frac{\partial \Phi(k, \rho_1, \rho_2)\Psi(k, \rho_2)}{\partial k} = \Phi(k, \rho_1, \rho_2)\Psi(k, \rho_2) \left( \frac{\Psi'(k, \rho_2)}{\Psi(k, \rho_2)} + F(k, \rho_2) - F(k, \rho_1) \right).
\]
(139)

Lemma [10] in Appendix IV shows that the function \( F(k, \rho) \) is decreasing in \( \rho \in (0, 1] \). Moreover, we have just shown \( \frac{\Psi'(k, \rho_2)}{\Psi(k, \rho_2)} \leq 0 \), for \( \rho_2 \in (0, 3] \). Consequently, when \( \rho_1, \rho_2 \in (0, 1] \) such that \( \rho_1 \leq \rho_2 \)
\[
\frac{\Psi'(k, \rho_2)}{\Psi(k, \rho_2)} + F(k, \rho_2) - F(k, \rho_1) \leq 0
\]
(140)

holds, and the product \( \Phi(k, \rho_1, \rho_2)\Psi(k, \rho_2) \) is non-increasing in \( k \). As \( k \) is decreasing in \( z \), which is in turn non-increasing in \( t \) when \( \rho \geq 0 \) by Lemma [1], the expression in equation (143), is decreasing in \( z \) whenever \( \rho_1, \rho_2 \in (0, 1] \) such that \( \rho_1 \leq \rho_2 \). In this case,
\[
\frac{\partial^2 \tilde{f}_{\rho_1, \rho_2}(t)}{\partial t^2} = \frac{\partial \Phi(k, \rho_1, \rho_2)\Psi(k, \rho_2)}{\partial k} \frac{\partial k}{\partial z} \frac{\partial z}{\partial t} \leq 0,
\]
(141)

whence the function \( \tilde{f}_{\rho_1, \rho_2}(t) \) is concave in \( t \) as claimed.
Appendix III

Proof of Lemma 6: Taking the first derivative of \( f_{\rho_1,\rho_2}(t) \) with respect to \( t \), we obtain

\[
\frac{\partial f_{\rho_1,\rho_2}(t)}{\partial t} = \frac{\partial g(\rho_2, g^{-1}(\rho_1, t))}{\partial t}
\]

(142)

\[
= g'(\rho_2, g^{-1}(\rho_1, t)) \frac{g'(\rho_1, g^{-1}(\rho_1, t))}{g'_{(1,1)}}
\]

(143)

Let \( z = g^{-1}(\rho_1, t) \). As \( g(\rho, z) \) is a monotone function in \( z \) by Lemma 1, so is \( z = g^{-1}(\rho, t) \) in \( t \). Hence we can check the convexity of \( f_{\rho_1,\rho_2}(t) \) with respect to \( t \), from the monotonicity with respect to \( z \) of the following expression:

\[
g'(\rho_2, z) g'(\rho_1, z)
\]

(144)

Taking the derivative with respect to \( z \), we get

\[
\frac{\partial}{\partial z} g'(\rho_2, z) g'(\rho_1, z) = \frac{\partial}{\partial z} 2^{\rho_1-\rho_2} \frac{\alpha(\rho_2, z) \beta(\rho_2, z)}{\alpha(\rho_1, z) \beta(\rho_1, z)}
\]

\[
= 2^{\rho_1-\rho_2} \frac{\alpha(\rho_2, z) \beta(\rho_2, z)}{\alpha(\rho_1, z) \beta(\rho_1, z)} (\ell(\rho_2, z) - \ell(\rho_1, z))
\]

where

\[
\ell(\rho, z) = \frac{\partial \alpha(\rho, z) / \partial z}{\alpha(\rho, z)} + \frac{\partial \beta(\rho, z) / \partial z}{\beta(\rho, z)}.
\]

(145)

One can easily check that the function

\[
\alpha(\rho, z) \geq 0,
\]

(146)

for any \( \rho > -1 \), and while the function

\[
\beta(\rho, z) \geq 0,
\]

(147)

for \( \rho \in (-1, 0) \), we have

\[
\beta(\rho, z) \leq 0,
\]

(148)

for \( \rho \geq 0 \).

Moreover, we claim that

\[
\ell(\rho_2, z) - \ell(\rho_1, z) \geq 0
\]

(149)

when \( \rho_1 \in (-1, 0) \), and \( \rho_2 \geq 0 \), or when \( \rho_1 \in (0, 1] \), and \( \rho_2 \geq \rho_1 \), and that

\[
\ell(\rho_2, z) - \ell(\rho_1, z) \leq 0
\]

(150)
when \( \rho_1 > 1 \), and \( \rho_2 \in (-1, 0) \), or when \( \rho_1 > 1 \), and \( \rho_2 \in (0, 1] \).

Therefore, if \( \rho_1 \in (-1, 0] \), and \( \rho_2 \geq 0 \), we have

\[
\frac{\partial}{\partial z} g'(\rho_2, z) = \frac{2^{-\rho_2} \alpha(\rho_2, z) \beta(\rho_2, z)}{2^{-\rho_2} \alpha(\rho_1, z) \beta(\rho_1, z)} (\ell(\rho_2, z) - \ell(\rho_1, z)) \leq 0,
\]

and if \( \rho_1 \in [0, 1] \), and \( \rho_2 \geq \rho_1 \), we have

\[
\frac{\partial}{\partial z} g'(\rho_2, z) = \frac{2^{-\rho_2} \alpha(\rho_2, z) \beta(\rho_2, z)}{2^{-\rho_1} \alpha(\rho_1, z) \beta(\rho_1, z)} (\ell(\rho_2, z) - \ell(\rho_1, z)) \geq 0.
\]

On the other hand, if \( \rho_1 > 1 \), and \( \rho_2 \in (-1, 0) \), we have

\[
\frac{\partial}{\partial z} g'(\rho_2, z) = \frac{2^{-\rho_2} \alpha(\rho_2, z) \beta(\rho_2, z)}{2^{-\rho_1} \alpha(\rho_1, z) \beta(\rho_1, z)} (\ell(\rho_2, z) - \ell(\rho_1, z)) \geq 0,
\]

and if \( \rho_1 > 1 \), and \( \rho_2 \in (0, 1] \), we have

\[
\frac{\partial}{\partial z} g'(\rho_2, z) = \frac{2^{-\rho_2} \alpha(\rho_2, z) \beta(\rho_2, z)}{2^{-\rho_1} \alpha(\rho_1, z) \beta(\rho_1, z)} (\ell(\rho_2, z) - \ell(\rho_1, z)) \leq 0.
\]

Recall that we are interested in the sign of the second derivative of \( f_{\rho_1, \rho_2} \) with respect to \( t \) given by

\[
\frac{\partial^2 f_{\rho_1, \rho_2}}{\partial t^2} = \frac{\partial}{\partial t} \left( \frac{\partial g(\rho_2, g^{-1}(\rho_1, t))}{\partial t} \right) = \frac{\partial}{\partial z} g'(\rho_2, z) \frac{\partial}{\partial t}.
\]

As \( z \) is non-decreasing in \( t \) for \( \rho_1 \in (-1, 0) \), and non-increasing for \( \rho_1 \geq 0 \) by Lemma \[1\] the function \( f_{\rho_1, \rho_2}(t) \) is concave in \( t \) when \( \rho_1 \in (-1, 0] \), and \( \rho_2 \geq 0 \), or when \( \rho_1 \in [0, 1] \), and \( \rho_2 \geq \rho_1 \), or when \( \rho_1 > 1 \), and \( \rho_2 \in (-1, 0) \), and convex when \( \rho_1 > 1 \), and \( \rho_2 \in (0, 1] \).

Now, we prove the claim. For that purpose, we show that the function \( \ell(\rho, z) \) is non-decreasing in \( \rho \) for the interval \( \rho \in (-1, 3) \), and \( \frac{\partial \ell(\rho, z)}{\partial \rho} \) changes sign only once after \( \rho \geq 3 \). As

\[
\lim_{\rho \to 1} \ell(\rho, z) = \lim_{\rho \to \infty} \ell(\rho, z) = \frac{1}{z - z^3}
\]

holds, we conclude that

\[
\ell(\rho, z) \geq \ell(1, z), \quad \text{when } \rho \geq 1,
\]

\[
\ell(\rho, z) \leq \ell(1, z), \quad \text{when } \rho \in (-1, 1].
\]
The above inequalities ensure \( \ell(\rho_2, z) - \ell(\rho_1, z) \geq 0 \) when \( \rho_1 \in (-1, 0) \), and \( \rho_2 \geq 0 \), or when \( \rho_1 \in [0, 1] \), and \( \rho_2 \geq \rho_1 \). Similarly, the previous arguments ensure that \( \ell(\rho_2, z) - \ell(\rho_1, z) \leq 0 \) when \( \rho_1 > 1 \), and \( \rho_2 \in (-1, 1) \).

Note that

\[
\ell(\rho, z) = \frac{\partial}{\partial z} \left( \log \left( 2^{-\rho} \alpha(\rho, z) \right) + \log \left( \beta(\rho, z) \right) \right). \tag{160}
\]

Hence,

\[
\frac{\partial \ell(\rho, z)}{\partial \rho} = \frac{\partial}{\partial z} \left( \frac{\partial 2^{-\rho} \alpha(\rho, z) / \partial \rho}{2^{-\rho} \alpha(\rho, z)} + \frac{\partial \beta(\rho, z) / \partial \rho}{\beta(\rho, z)} \right), \tag{161}
\]

\[
= \frac{\partial \Psi(k, \rho)}{\partial k} \frac{\partial k}{\partial z}, \tag{162}
\]

\[
= \Psi'(k, \rho) \frac{\partial k}{\partial z}, \tag{163}
\]

where \( k = h(z) \) is defined in Equation (103), and \( \Psi'(k, \rho) \) is defined in Equation (126). Luckily, we have already investigated the sign of \( \Psi'(k, \rho) \) in the proof of Lemma 5 we previously stated. Indeed, we have shown that \( \Psi'(k, \rho) \leq 0 \), for \( \rho \in (-1, 3) \), and the function changes sign only once after \( \rho \geq 3 \). As \( k \) is decreasing in \( z \), the sign of \( \frac{\partial \ell(\rho, z)}{\partial \rho} \) is exactly the opposite of \( \Psi'(k, \rho) \).

This concludes the proof.

**Appendix IV**

**Lemma 9:** For \( k \in [0, 1] \), we define

\[
m(k, \rho) = -1 + k - k^{1/\rho} + k \frac{\rho}{2 \rho + 1} \left( k^{1/\rho} + k^{1/\rho} \right) \log k^{1/\rho}. \tag{164}
\]

Then, for \( \forall k \in [0, 1] \), we have

\[
m(k, \rho) \geq 0, \quad \forall \rho < -1,
\]

\[
m(k, \rho) \leq 0, \quad \forall \rho \in (-1, 0),
\]

\[
m(k, \rho) = 0.
\]

Moreover, \( \exists \rho^*(k) \geq 3 \) which depends on \( k \) such that:

\[
m(k, \rho) \leq 0, \quad \forall \rho \in (-1, \rho^*(k)),
\]

\[
m(k, \rho^*) = 0,
\]

\[
m(k, \rho) \geq 0, \quad \forall \rho \in (\rho^*, \infty).
\]
Proof: We now follow a series of transformations. Let
\[ t = \frac{\rho}{1 + \rho} \]
Then, (164) reduces to
\[ m\left(k, \frac{t}{1-t}\right) = -1 + k - k^{1-t} + k^t - (k^t + k^{1-t}) \log k^t. \]
In addition, let
\[ s = -t \log k. \]
Then,
\[ m\left(k, \frac{-s}{\log k + s}\right) = -1 + k - ke^s + e^{-s} + s(e^{-s} + ke^s). \quad (165) \]
We first note that the function is zero at \( s = 0 \). Taking the first derivative with respect to \( s \), we get
\[
\frac{\partial}{\partial s} m\left(k, \frac{-s}{\log k + s}\right) = -ke^s - e^{-s} + e^{-s} + ke^s + s(-e^{-s} + ke^s) \\
= s(-e^{-s} + ke^s) \\
= t(k^t - k^{1-t}) \log k.
\]
Hence the function \( m\left(k, \frac{-s}{\log k + s}\right) \) is non-increasing in \( s \) for \( t \in [0, 1/2] \), and non-decreasing otherwise.

Moreover, the derivative of \( m\left(k, \frac{t}{1-t}\right) \) with respect to \( t \) is given by
\[
\frac{\partial}{\partial t} m\left(k, \frac{t}{1-t}\right) = \frac{\partial}{\partial s} m\left(k, \frac{-s}{\log k + s}\right) \frac{\partial s}{\partial t}
\]
As \( s \) is non-decreasing in \( t \), we have shown that \( m\left(k, \frac{t}{1-t}\right) \) is non-increasing in \( t \) for \( t \in [0, 1/2] \), and non-decreasing otherwise.

Similarly, the derivative of \( m(k, \rho) \) with respect to \( \rho \) is given by
\[
\frac{\partial m(k, \rho)}{\partial \rho} = \frac{\partial}{\partial t} m\left(k, \frac{t}{1-t}\right) \frac{\partial t}{\partial \rho}
\]
As \( t \) is increasing in \( \rho \) for the intervals \((-\infty, -1)\), and \((-1, \infty)\), \( m(k, \rho) \) will be non-increasing in \( \rho \) for \( t \in [0, 1/2] \), and non-decreasing otherwise. We simply need to map this result to the claims of the lemma in terms of the intervals defined by \( \rho \).
For the interval $t \in [1, \infty)$, we have $\rho < -1$, and $m(k, \rho)$ is non-decreasing in $\rho$. Moreover,

$$\lim_{\rho \to -\infty} m(k, \rho) = (-1 + k - 1 + k) - (k + 1) \log k = -2(1 - k) + (k + 1) \log k \geq 0$$

where the sign follows by noting that at $k = 1$ the expression evaluates to 0, and it is non-increasing in $k$ as

$$\frac{\partial}{\partial k} (-2(1 - k) + (k + 1) \log k) = 1 - \frac{1}{k} + \log \frac{1}{k} \leq 0$$

using $\log x \leq x - 1$ inequality. This shows $m(k, \rho) \geq 0$ for $\rho < -1$.

For the interval $t \in (-\infty, 0]$, we have $\rho \in (-1, 0]$, and $m(k, \rho)$ is non-decreasing in $\rho$. As we have $m(k, 0) = 0$, we conclude $m(k, \rho) \leq 0$ for $\rho \in (-1, 0)$.

For the interval $t \in [0, 1/2]$, we have $\rho \in [0, 1]$, and $m(k, \rho)$ is non-increasing in $\rho$. As we have $m(k, 0) = 0$, we conclude $m(k, \rho) \leq 0$ for $\rho \in (0, 1]$.

For the interval $t \in [1/2, 1]$, we have $\rho \geq 1$, and $m(k, \rho)$ is non-decreasing in $\rho$. As $m(k, 1) \leq 0$, and

$$\lim_{\rho \to -\infty} m(k, \rho) = (-1 + k - 1 + k) - (k + 1) \log k = -2(1 - k) + (k + 1) \log k \geq 0,$$

the function will eventually cross zero. Now, we prove that the crossing point $\rho^*$, i.e. $m(k, \rho^*) = 0$, is such that $\rho^* \geq 3$. For that purpose, we only need to show that $m(k, 3)$ is increasing in $k$ because $m(1, 3) = 0$ holds.

Taking the first derivative with respect to $k$, we get

$$\frac{\partial m(k, 3)}{\partial k} = \frac{4(-1 + k^{3/4}) - 3/4(1 + 3\sqrt{k}) \log k}{4k^{3/4}} \geq 0$$

with equality iff $k = 1$. The sign follows by noting that the denominator is positive, the numerator is decreasing in $k$, and is equal to 0 iff $k = 1$. Indeed, taking the first derivative with respect to $k$ of the numerator, we get

$$\frac{\partial}{\partial k} \left(4(-1 + k^{3/4}) - 3/4(1 + 3\sqrt{k}) \log k\right) = -3(2 + 6\sqrt{k} - 8k^{3/4} + 3\sqrt{k} \log k) \leq 0$$

with equality iff $k = 1$. The sign follows by noting that the denominator is positive, the numerator is increasing in $k$, and is equal to 0 iff $k = 1$. To see this, once more we take the first derivative with respect to $k$ of the numerator. Then, we get

$$\frac{\partial}{\partial k} \left(-3(2 + 6\sqrt{k} - 8k^{3/4} + 3\sqrt{k} \log k) \log k\right) = \frac{-9(4 - 4k^{1/4} + \log k)}{2\sqrt{k}} \geq 0$$
with equality iff $k = 1$. The sign follows by noting that the denominator is positive, the numerator is decreasing in $k$, and is equal to 0 iff $k = 1$. To show this, we need to take the first derivative with respect to $k$ of the numerator one last time. Doing so, we get

$$
\frac{\partial}{\partial k} \left(-9(4 - 4k^{1/4} + \log k)\right) = \frac{9(-1 + k^{1/4})}{k} \leq 0
$$

for $k \in [0, 1]$, and with equality iff $k = 1$. This concludes the proof of the lemma.

Lemma 10 ([12]): The function $F(k, \rho)$ defined in (125) is a decreasing function in $\rho \in [0, 1]$.

Proof: For convenience, we define the function $H(k, \rho) = -k^{1/4}F(k, \rho)$ as

$$
H(k, \rho) = \frac{\rho}{1 + \rho} \left(\frac{1}{1 + k^{1/4}}\right) \left(\frac{1}{1 - k^{1/4}}\right) \geq 0
$$

(166)

where $k \in [0, 1]$. We note that instead of $F(k, \rho)$, we can also check the monotonicity of $H(k, \rho)$ with respect to $\rho$.

We now follow a series of transformations. Let

$$
t = \frac{\rho}{1 + \rho} \quad \text{for} \quad t \in [0, \frac{1}{2}].
$$

Then, (166) reduces to

$$
H(k, \frac{t}{1 - t}) = \frac{t}{(1 - k^4)(1 + k^{1-t})}. \tag{167}
$$

In addition, let

$$
s = -t \ln k \quad \text{for} \quad s \in [0, \frac{1}{2} \ln \frac{1}{k}].
$$

Then,

$$
H(k, \frac{-s}{\log k + s}) = \frac{1}{\log \frac{1}{k}} \left(\frac{s}{1 - e^{-s}}\right) \left(\frac{1}{1 + ke^s}\right) \tag{167}
$$

We note that the first fraction in (167) can be treated as a constant and we ignore it. We define the variable $a = \frac{1}{k} \geq 1$. For simplicity, we consider the function

$$
\frac{1}{H(k, \frac{-s}{\log k + s})} = \frac{\ln a}{\left\langle a \right\rangle} \frac{1 - e^{-s}}{s} (a + e^s).
$$

We first show that $\ln \left(\frac{1 - e^{-s}}{s} (a + e^s)\right)$ is a convex function for all $s \geq 0$. Taking the first derivative with respect to $s$, we obtain

$$
\frac{\partial}{\partial s} \left(-\ln s + \ln \left(\frac{1}{1 - e^{-s}}\right) + \ln \left(\frac{e^s}{a + e^s}\right)\right) = -\frac{1}{s} + \frac{e^s}{a + e^s} + \frac{1}{e^s - 1}. \tag{168}
$$
Taking the second derivative in \( s \), we get
\[
\frac{\partial^2}{\partial s^2} \left( -\ln s + \ln \left( \frac{1}{1-e^{-s}} \right) + \ln \left( \frac{e^s}{a+e^s} \right) \right) \]
\[
= \frac{1}{s^2} + \frac{ae^s}{(a+e^s)^2} - \frac{e^s}{(e^s - 1)^2}
\geq \frac{1}{s^2} - \frac{e^s}{(e^s - 1)^2}
= \frac{1}{s^2} - \left( \frac{1}{e^s + e^{-s}} \right)^2
= \frac{1}{s^2} - \frac{1}{\left( 2 \sinh \frac{s}{2} \right)^2}
\geq 0
\]
where the last inequality follows from \( \sinh x \geq x \), for \( x \geq 0 \). We proved that \( \ln \left( \frac{1-e^{-s}}{s} (a+e^s) \right) \) is a convex function for all \( s \geq 0 \). Therefore the function has only one minimum, and to decide whether the expression is decreasing in \( s \in [0, \frac{1}{2} \ln a] \), it is sufficient to evaluate (168) at \( s = \frac{1}{2} \ln a \).

\[
\frac{\partial}{\partial s} \left( -\ln s + \ln \left( \frac{1}{1-e^{-s}} \right) + \ln \left( \frac{e^s}{a+e^s} \right) \right) \bigg|_{s=\frac{1}{2} \ln a}
\]
\[
= -\frac{1}{\ln \sqrt{a}} + \frac{\sqrt{a}}{a + \sqrt{a}} + \frac{1}{\sqrt{a} - 1}
= -\frac{1}{\ln \sqrt{a}} + \frac{2\sqrt{a}}{a - 1}
\leq 0
\]
since for \( b = \sqrt{a} \geq 1 \), we can show that
\[
\frac{b^2 - 1}{2b} - \ln b \geq 0. \tag{169}
\]
Taking the first derivative of (169) with respect to \( b \), we get
\[
\frac{\partial}{\partial b} \left( \frac{b^2 - 1}{2b} - \ln b \right) = \frac{1}{2} + \frac{1}{2b^2} - \frac{1}{b} = \frac{(b-1)^2}{2b^2} \geq 0.
\]
Therefore, we proved that for each \( k \in [0, 1] \) the function \( \frac{1}{H(k, \log k + \rho)} \) is decreasing in \( s \). By definition, the variable \( t \) is increasing in \( \rho \), and \( s = -t \ln k \) is also increasing in \( t \) for a given \( k \). As a consequence, the function \( F(k, \rho) = -\frac{1+k}{k} H(k, \rho) \) is decreasing in \( \rho \).
REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[2] E. Arıkan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[3] M. Alsan, “Extremality properties for Gallager’s random coding exponent,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2944–2948.

[4] 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*, 2011, pp. 2896–2898.

[5] S. Arimoto, “On the converse to the coding theorem for discrete memoryless channels (corresp.),” *Information Theory, IEEE Transactions on*, vol. 19, no. 3, pp. 357–359, 1973.

[6] A. Guillen i Fabregas, I. Land, and A. Martinez, “Extremes of error exponents,” *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2201–2207, 2013.

[7] E. Arıkan, “Channel combining and splitting for cutoff rate improvement,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 628–639, 2006.

[8] I. Csiszar, “Generalized cutoff rates and renyi’s information measures,” *Information Theory, IEEE Transactions on*, vol. 41, no. 1, pp. 26–34, 1995.

[9] A. Rényi, “On measures of entropy and information,” *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, pp. 547–561, 1961.

[10] S. Arimoto, “Information measures and capacity of order $\alpha$ for discrete memoryless channels,” in *Topics in information theory*, I. Csiszár and P. Elias, Eds., vol. 16. Amsterdam, The Netherlands: North-Holland Publishing Co., 1977, pp. 41–52.

[11] E. Arıkan and E. Telatar, “BEC and BSC are $E_0$ extremal,” Unpublished.

[12] E. Telatar, “Private communication.”