A Lower Bound on the Sum Rate of Multiple Description Coding With Symmetric Distortion Constraints

Lin Song, Shuo Shao, and Jun Chen, *Member, IEEE*

Abstract—We derive a single-letter lower bound on the minimum sum rate of multiple description coding with symmetric distortion constraints. For the binary uniform source with the erasure distortion measure or Hamming distortion measure, this lower bound can be evaluated with the aid of certain minimax theorems. A similar minimax theorem is established in the quadratic Gaussian setting, which is further leveraged to analyze the special case where the minimum sum rate subject to two levels of distortion constraints (with the second level imposed on the complete set of descriptions) is attained; in particular, we determine the minimum achievable distortions at the intermediate levels.

Index Terms—Erasure distortion, Hamming distortion, mean squared error, minimax theorem, multiple description coding, saddle point.

I. INTRODUCTION

IN MULTIPLE description coding a source is encoded into several (say, L) descriptions such that every subset of these descriptions can be used to reconstruct the source (though the reconstruction distortion in general depends on which subset of descriptions is used). Many coding schemes have been proposed for this problem over the past three decades. Notable examples include the classical El Gamal-Cover [1] and Zhang-Berger [2] schemes for the two-description case as well as their extension to the general L-description case by Venkataramani, Kramer, and Goyal [3]. Special attention [4]–[7] has been paid to the case where the distortion constraints are symmetric, i.e., the distortion constraints imposed on the reconstructions from different subsets of descriptions of the same cardinality are identical (see Fig. 1 for an illustration of the three-description case). In particular,

Pradhan, Puri, and Ramchandran [4], [5] developed a symmetric multiple description coding scheme via an ingenious application of the binning technique; further improvements based on structured codes and the splitting method can be found in [6].

In contrast, the converse results for the multiple description problem are relatively limited. This is partly due to technical difficulties in handling dependencies among different descriptions. In fact, it is already a highly sophisticated task to obtain a tight single-letter bound even when the descriptions are asymptotically independent, as evident from Ahlswede's remarkable work on the characterization of the rate-distortion region of two-description coding with no excess sum rate [8].

We shall show that non-trivial converse results for the multiple description problem can be obtained by augmenting the probability space through the introduction of certain auxiliary remote sources. It is worth emphasizing that the use of auxiliary remote sources is by no means a new idea. Indeed, this idea made its first appearance in the seminal work of Ozarow on the solution of the Gaussian two-description problem [9]; since then, remote sources have played an essential role in the derivation of several conclusive results on the information-theoretic limits of multiple description coding [10]–[12]. However, although this idea has been widely used with great success, to the best of the authors' knowledge, in the context of multiple description coding[1] only a special class of remote sources (specifically, only those that can be generated by the given source via additive Gaussian noise channels) have been exploited. A possible reason for this situation is as follows: if this special class of remote sources are used, then one can derive explicit bounds on the relevant multi-letter expressions by invoking certain extremal inequalities (e.g., the worst additive noise lemma [19], [20] and certain variants of the entropy power inequality) that hinge upon the properties of the Gaussian distribution. However, the use of such remote sources and the associated extremal inequalities impose severe restrictions on the applicability of this idea, rendering it essentially only useful for the quadratic Gaussian case[2]. It will be seen that such extremal inequalities

L. Song was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada. She is now with the Institute of Network Coding, Chinese Univeristy of Hong Kong, Hong Kong (e-mail: linsong@inc.cuhk.edu.hk).

S. Shao was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada. He is now with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: shaoshuo@tamu.edu).

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

---

[1]The idea of using auxiliary remote sources in the converse arguments has also found applications in multiterminal source coding (see [13]–[16]), joint source-channel coding (see [17], [18]), and other network information theory problems.

[2]Strictly speaking, this special class of remote sources and the associated extremal inequalities can be used in the non-Gaussian setting, particularly when the mean squared error distortion measure is adopted (see [21 Th. 5.3]); however, they incur an intrinsic loss in the non-Gaussian setting and consequently the resulting bound is in general strictly suboptimal.
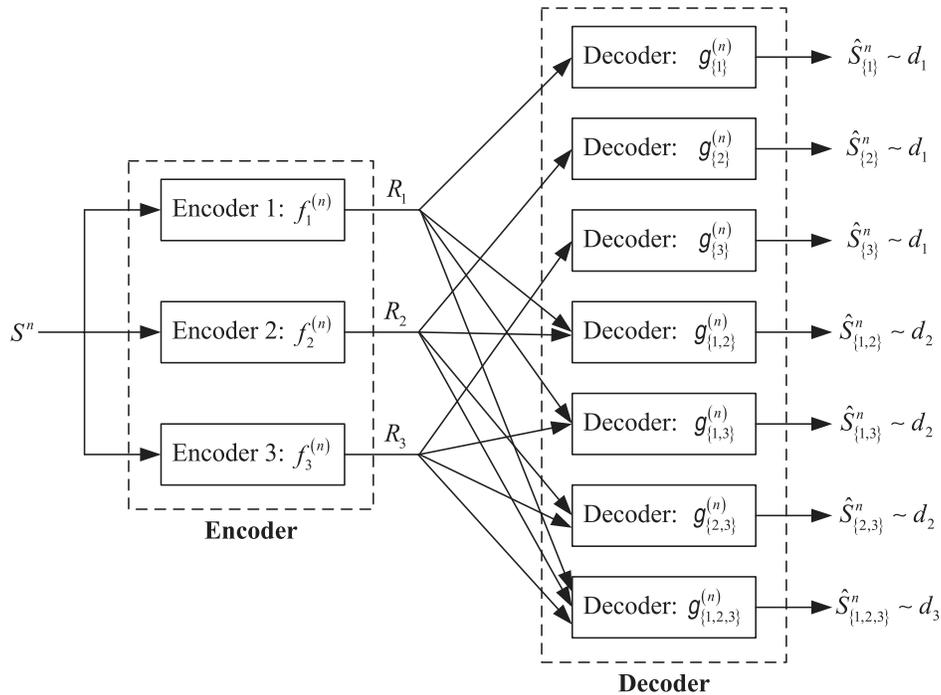
Fig. 1.   System diagram for three-description coding with symmetric distortion constraints.

are in fact not needed for reducing multi-letter bounds to single-letter bounds (though they can be useful for evaluating single-letter bounds in the quadratic Gaussian case) and there is greater flexibility in choosing remote sources. As a consequence, this remote-source idea can be readily applied in the non-Gaussian setting as well. It is worth mentioning that the converse results in [22] and [23] also involve certain auxiliary random objects. However, those objects do not appear to have a remote-source interpretation[3] and their relationship with the remote-source construction initiated by Ozarow remains elusive.

In this work we derive a single-letter lower bound on the minimum sum rate of multiple description coding with symmetric distortion constraints by exploiting the aforementioned remote-source idea. It will be seen that our bounding technique is in fact applicable to the asymmetric case as well. We choose to focus on the symmetric case mostly because the resulting bound has a more compact expression; additionally, from a practical perspective, it often suffices to consider symmetric distortion constraints. Furthermore, we prove several minimax theorems, which are of interest in their own right, and leverage them to evaluate this lower bound in some special settings. Interestingly, the minimax theorem established in the quadratic Gaussian case also enables us to obtain a new conclusive result on the information-theoretic limits of Gaussian multiple description coding.

The rest of this paper is organized as follows. A single-letter lower bound on the minimum sum rate of multiple description

coding with symmetric distortion constraints is presented in Section II. We show in Section III that, for the binary uniform source with the erasure distortion measure or the Hamming distortion measure, this lower bound can be evaluated with the aid of certain minimax theorems. Section IV contains a similar minimax theorem in the quadratic Gaussian setting, which is used to analyze the special case where the minimum sum rate subject to two levels of distortion constraints (with the second level imposed on the complete set of descriptions) is attained; in particular, we determine the minimum achievable distortions at the intermediate levels. We conclude the paper in Section V.

For any nonempty set $\mathcal{A}$, we define $2^{\mathcal{A}}_+ = \{\mathcal{B} : \mathcal{B} \subseteq \mathcal{A}, |\mathcal{B}| > 0\}$, where $|\mathcal{B}|$ is the cardinality of $\mathcal{B}$. We write $(X(1), \ldots, X(n))$ as X^n for any positive integer n and set $X^0 = 0$. Moreover, $\oplus_M$ and $\ominus_M$ are used to denote modulo-M addition and subtraction, respectively, for any integer $M \geq 2$. Let $f(x, y)$ be an arbitrary real-valued function with $x \in \mathcal{X}$ and $y \in \mathcal{Y}$; we define

$$\arg\max_{x \in \mathcal{X}} f(x, y^*) = \{x \in \mathcal{X} : f(x, y^*) \geq f(\tilde{x}, y^*) \quad \text{for all } \tilde{x} \in \mathcal{X}\},$$

$$\arg\min_{y \in \mathcal{Y}} f(x^*, y) = \{y \in \mathcal{Y} : f(x^*, y) \leq f(x^*, \tilde{y}) \quad \text{for all } \tilde{y} \in \mathcal{Y}\}$$

for any $x^* \in \mathcal{X}$ and $y^* \in \mathcal{Y}$. Unless specified otherwise, we adopt the following convention:

$$\sum_{i=1}^{N} \alpha_i \log \infty = \begin{cases} -\infty, & \sum_{i=1}^{N} \alpha_i < 0 \\ 0, & \sum_{i=1}^{N} \alpha_i = 0 \\ \infty, & \sum_{i=1}^{N} \alpha_i > 0. \end{cases}$$

---

[3]Actually they are better interpreted as duplicate copies of the source.

The logarithm function is to base e throughout this paper.

II. A SINGLE-LETTER LOWER BOUND

Let $\{S(t)\}_{t=1}^{\infty}$ be i.i.d. copies of a generic source random variable S with distribution p_S. Let $m : \mathcal{S} \times \hat{\mathcal{S}} \to [0, \infty]$ be a distortion measure, where $\mathcal{S}$ and $\hat{\mathcal{S}}$ are, respectively, the source alphabet and the reconstruction alphabet.

Definition 1: A sum rate R is said to be achievable subject to distortion constraints $[d] \triangleq (d_{\mathcal{A}}, \mathcal{A} \in 2_+^{\mathcal{L}})$ if there exist encoding functions $f_i^{(n)} : \mathcal{S}^n \to \mathcal{C}_i$, $i \in \mathcal{L}$, and decoding functions $g_{\mathcal{A}}^{(n)} : \prod_{i \in \mathcal{A}} \mathcal{C}_i \to \hat{\mathcal{S}}^n$, $\mathcal{A} \in 2_+^{\mathcal{L}}$, such that

$$\frac{1}{n} \sum_{i=1}^{L} \log |\mathcal{C}_i| \leq R,$$

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[m(S(t), \hat{S}_{\mathcal{A}}(t))] \leq d_{\mathcal{A}}, \quad \mathcal{A} \in 2_+^{\mathcal{L}}, \quad (1)$$

where $\mathcal{L} = \{1, \ldots, L\}$ and $\hat{S}_{\mathcal{A}}^n = g_{\mathcal{A}}^{(n)}(f_i^{(n)}(S^n), i \in \mathcal{A})$, $\mathcal{A} \in 2_+^{\mathcal{L}}$. The infimum over all such achievable sum rates is denoted by $R([d])$. When the distortion constraints are symmetric, i.e., there exists $\underline{d} \triangleq (d_1, \ldots, d_L)$ such that $d_{\mathcal{A}} = d_{|\mathcal{A}|}$ for all $\mathcal{A} \in 2_+^{\mathcal{L}}$, we shall denote $R([d])$ by $R(\underline{d})$.

Note that $R([d])$ is the classical rate-distortion function when $L = 1$. Therefore, we shall only consider the case $L \geq 2$ in the rest of this paper.

The following result provides a single-letter lower bound on $R([d])$. Let $\mathcal{P}$ denote the set of conditional distributions $p_{\underline{Z}|S}$ with $\underline{Z} = (Z_0, Z_1, \ldots, Z_L)$ such that $p_{Z_{k-1}|S}$ is physically degraded with respect to $p_{Z_k|S}$, $k = 1, \ldots, L$. Let $\mathcal{P}([d])$ denote the set of conditional distributions $p_{[\hat{S}]|S}$ with $[\hat{S}] = (\hat{S}_{\mathcal{A}}, \mathcal{A} \in 2_+^{\mathcal{L}})$ such that the induced joint distribution $p_{S,[\hat{S}]} \triangleq p_S p_{[\hat{S}]|S}$ satisfies $\mathbb{E}[m(S, \hat{S}_{\mathcal{A}})] \leq d_{\mathcal{A}}$, $\mathcal{A} \in 2_+^{\mathcal{L}}$. Define

$$r([d]) = \inf_{p_{[\hat{S}]|S} \in \mathcal{P}([d])} \sup_{p_{\underline{Z}|S} \in \mathcal{P}}$$

$$\sum_{k=1}^{L} \frac{L}{k \binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}),$$

$$(2)$$

where it is assumed that $\underline{Z} \leftrightarrow S \leftrightarrow [\hat{S}]$ form a Markov chain. For the purpose of evaluating $r([d])$, it suffices to consider $|\mathcal{Z}_k| \leq |\mathcal{S}|^k$, $k = 0, 1, \ldots, L-1$, and $Z_L = S$ (see Appendix A).

Theorem 1: $R([d]) \geq r([d])$.

Proof: Our proof is partly based on the bounding technique developed in [21], which is in turn inspired by the ideas in Ozarow's celebrated work [9] as well as [10] and [11]. To illustrate the key points in the argument (particularly the role of remote sources), we shall first give a sketch of the proof for the case $L = 2$. The detailed proof for the general L-description case is relegated to Appendix B.

Note that

$$\log |\mathcal{C}_1| + \log |\mathcal{C}_2|$$
$$\geq H(f_1^{(n)}(S^n)) + H(f_2^{(n)}(S^n))$$
$$= H(f_1^{(n)}(S^n), f_2^{(n)}(S^n)) + I(f_1^{(n)}(S^n); f_2^{(n)}(S^n))$$
$$= I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) + I(f_1^{(n)}(S^n); f_2^{(n)}(S^n)). \quad (3)$$

If the two terms in (3) are treated separately, then one will encounter difficulties in obtaining a non-trivial single-letter lower bound on the second term, i.e., $I(f_1^{(n)}(S^n); f_2^{(n)}(S^n))$. To address this issue, we introduce remote sources (Z_0, Z_1, Z_2) jointly distributed with the generic source random variable S such that $Z_0 \leftrightarrow Z_1 \leftrightarrow Z_2 \leftrightarrow S$ form a Markov chain, and define $\{Z_0(t), Z_1(t), Z_2(t)\}_{t=1}^{\infty}$ correspondingly. It can be verified that

$$I(f_1^{(n)}(S^n); f_2^{(n)}(S^n))$$
$$= I(Z_1^n; f_1^{(n)}(S^n)) + I(Z_1^n; f_2^{(n)}(S^n))$$
$$- I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) + I(f_1^{(n)}(S^n); f_2^{(n)}(S^n)|Z_1^n).$$

As a consequence, we have

$$\log |\mathcal{C}_1| + \log |\mathcal{C}_2|$$
$$\geq I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) + I(Z_1^n; f_1^{(n)}(S^n))$$
$$+ I(Z_1^n; f_2^{(n)}(S^n)) - I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n))$$
$$+ I(f_1^{(n)}(S^n); f_2^{(n)}(S^n)|Z_1^n). \quad (4)$$

Dropping the last term in (4) yields the following lower bound

$$\log |\mathcal{C}_1| + \log |\mathcal{C}_2|$$
$$\geq I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) + I(Z_1^n; f_1^{(n)}(S^n))$$
$$+ I(Z_1^n; f_2^{(n)}(S^n)) - I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)). \quad (5)$$

The rationale here is that $I(f_1^{(n)}(S^n); f_2^{(n)}(S^n)|Z_1^n)$ might be smaller than $I(f_1^{(n)}(S^n); f_2^{(n)}(S^n))$ for certain choices of Z_1^n, and consequently the resulting bound is tighter than the one obtained by simply dropping $I(f_1^{(n)}(S^n); f_2^{(n)}(S^n))$. Let T be uniformly distributed over $\{1, \ldots, n\}$ and independent of $(S^n, Z_0^n, Z_1^n, Z_2^n)$. It can be readily shown that

$$I(Z_1^n; f_i^{(n)}(S^n))$$
$$\geq I(Z_1^n; \hat{S}_{\{i\}}^n)$$
$$\geq n I(Z_1(T); \hat{S}_{\{i\}}(T))$$
$$\geq n I(Z_1(T); \hat{S}_{\{i\}}(T)|Z_0(T)), \quad i = 1, 2. \quad (6)$$

In contrast, it appears difficult to single-letterize $I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n))$ and $I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n))$ in (5) simultaneously due to the opposite signs in front of them. If S^n and Z^n are jointly Gaussian, then one can overcome this difficulty by invoking the entropy power inequality or the worst additive noise lemma; however, such specialized methods are not suitable in the general setting. It will be seen that this difficulty is actually not intrinsic and can be resolved through simple algebraic manipulations. Indeed, we have

$$I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) - I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n))$$
$$= I(Z_1^n, S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n))$$
$$- I(Z_1^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)) \quad (7)$$
$$= I(S^n; f_1^{(n)}(S^n), f_2^{(n)}(S^n)|Z_1^n)$$
$$\geq I(S^n; \hat{S}_{\{1\}}^n, \hat{S}_{\{2\}}^n, \hat{S}_{\{1,2\}}^n|Z_1^n)$$
$$= \sum_{t=1}^{n} I(S(t); \hat{S}_{\{1\}}^n, \hat{S}_{\{2\}}^n, \hat{S}_{\{1,2\}}^n|Z_1^n, S^{t-1})$$

$$= \sum_{t=1}^{n} I(S(t); \hat{S}_{\{1\}}^{n}, \hat{S}_{\{2\}}^{n}, \hat{S}_{\{1,2\}}^{n}, Z_1^n, S^{t-1}|Z_1(t))$$

$$\geq \sum_{t=1}^{n} I(S(t); \hat{S}_{\{1\}}(t), \hat{S}_{\{2\}}(t), \hat{S}_{\{1,2\}}(t)|Z_1(t))$$

$$\geq n I(S(T); \hat{S}_{\{1\}}(T), \hat{S}_{\{2\}}(T), \hat{S}_{\{1,2\}}(T)|Z_1(T))$$

$$\geq n I(Z_2(T); \hat{S}_{\{1\}}(T), \hat{S}_{\{2\}}(T), \hat{S}_{\{1,2\}}(T)|Z_1(T)), \quad (8)$$

where (7) is due to the fact that $Z_1^n \leftrightarrow S^n \leftrightarrow (f_1^{(n)}(S^n), f_2^{(n)}(S^n))$ form a Markov chain. One can readily obtain the desired result by substituting (6) and (8) into (5). ∎

When the distortion constraints are symmetric, we have the following simplified (albeit potentially weakened) lower bound on the minimum sum rate. Let $\mathcal{P}(\underline{d})$ denote the set of conditional distributions $p_{\underline{\hat{S}}|S}$ with $\underline{\hat{S}} = (\hat{S}_1, \dots, \hat{S}_L)$ such that the induced joint distribution $p_{S,\underline{\hat{S}}}$ satisfies $\mathbb{E}[m(S, \hat{S}_k)] \leq d_k$, $k = 1, \dots, L$. Define

$$r(\underline{d}) = \inf_{p_{\underline{\hat{S}}|S} \in \mathcal{P}(\underline{d})} \sup_{p_{\underline{Z}|S} \in \mathcal{P}} \sum_{k=1}^{L} \frac{L}{k} I(Z_k; \hat{S}_k|Z_{k-1}),$$

where it is assumed that $\underline{Z} \leftrightarrow S \leftrightarrow \underline{\hat{S}}$ form a Markov chain.

Corollary 1: $R(\underline{d}) \geq r(\underline{d})$.

Proof: Let $(S, \underline{Z}, [\hat{S}])$ be jointly distributed according to $p_{S,\underline{Z},[\hat{S}]} \triangleq p_S p_{\underline{Z}|S} p_{[\hat{S}]|S}$ with $p_{\underline{Z}|S} \in \mathcal{P}$ and $p_{[\hat{S}]|S} \in \mathcal{P}([d])$, where $d_{\mathcal{A}} = d_{|\mathcal{A}|}$ for all $\mathcal{A} \in 2_+^{\mathcal{L}}$. Note that

$$\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1})$$

$$\geq \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{A}}|Z_{k-1}). \quad (9)$$

Let Q_k be uniformly distributed over $\{\mathcal{A} : \mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}| = k\}$, $k = 1, \dots, L$; furthermore, it is assumed that $(Q_1, \dots, Q_L)$ and $(S, \underline{Z}, [\hat{S}])$ are mutually independent. We have, for any $k \in \mathcal{L}$,

$$\sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{A}}|Z_{k-1})$$

$$= \binom{L}{k} I(Z_k; \hat{S}_{Q_k}|Z_{k-1}, Q_k)$$

$$= \binom{L}{k} I(Z_k; \hat{S}_{Q_k}, Q_k|Z_{k-1}) \quad (10)$$

$$\geq \binom{L}{k} I(Z_k; \hat{S}_{Q_k}|Z_{k-1}), \quad (11)$$

where (10) is due to the fact that $Z_k \leftrightarrow Z_{k-1} \leftrightarrow Q_k$ form a Markov chain. Substituting (11) into (9) gives

$$\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1})$$

$$\geq \sum_{k=1}^{L} \frac{L}{k} I(Z_k; \hat{S}_k|Z_{k-1}),$$

where $\hat{S}_k = \hat{S}_{Q_k}$, $k = 1, \dots, L$. In view of the fact that $\underline{Z} \leftrightarrow S \leftrightarrow \underline{\hat{S}}$ form a Markov chain and the fact that $\mathbb{E}[m(S, \hat{S}_k)] \leq d_k$, $k = 1, \dots, L$, one can readily complete the proof by invoking Theorem 1. ∎

III. THE BINARY UNIFORM SOURCE

In this section we shall evaluate $r(\underline{d})$ for the binary uniform source (i.e., $\mathcal{S} = \{0, 1\}$ and $p_S(0) = p_S(1) = \frac{1}{2}$) with two different distortion measures:

- $m = m_E$, where m_E is the erasure distortion measure with $\hat{\mathcal{S}} = \{0, 1, e\}$ and

$$m_E(s, \hat{s}) = \begin{cases} 0, & s = \hat{s} \\ 1, & \hat{s} = e \\ \infty, & (s, \hat{s}) = (0, 1) \text{ or } (s, \hat{s}) = (1, 0); \end{cases}$$

- $m = m_H$, where m_H is the Hamming distortion measure with $\hat{\mathcal{S}} = \{0, 1\}$ and $m_H(s, \hat{s}) = s \oplus_2 \hat{s}$ for $(s, \hat{s}) \in \mathcal{S} \times \hat{\mathcal{S}}$.

To this end, we need the following two technical lemmas. Their proofs are relegated to Appendix C and Appendix D, respectively.

Let $\mathcal{X} = \{0, 1, \dots, M - 1\}$ for some integer $M \geq 2$, and let $\mathcal{Y}$ be an arbitrary (finite)[4] set. A channel $p_{Y|X} : \mathcal{X} \to \mathcal{Y}$ is said to be circularly symmetric [26] if there exists a bijective function $\mu : \mathcal{Y} \to \mathcal{Y}$ such that $\mu^M(y) = y$ and $p_{Y|X}(\mu^x(y)|x) = p_{Y|X}(y|0)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where μ^k denotes the k-times self-composition of μ (with μ^0 being the identity function). Similarly, a distortion measure $m : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ is said to circularly symmetric if there exists a bijective function $\mu : \mathcal{Y} \to \mathcal{Y}$ such that $\mu^M(y) = y$ and $m(x, \mu^x(y)) = m(0, y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Note that the binary erasure channel with erasure probability δ (i.e., BEC(δ)) and the erasure distortion measure m_E are circularly symmetric with $\mu : \{0, 1, e\} \to \{0, 1, e\}$ given by $\mu(0) = 1$, $\mu(1) = 0$, and $\mu(e) = e$; the binary symmetric channel with crossover probability q (i.e., BSC(q)) and the Hamming distortion measure m_H are also circularly symmetric, and the associated $\mu : \{0, 1\} \to \{0, 1\}$ is given by $\mu(0) = 1$ and $\mu(1) = 0$.

Lemma 1: Let $p_{Y|X} : \mathcal{X} \to \mathcal{Y}$ and $p_{\tilde{Y}|X} : \mathcal{X} \to \tilde{\mathcal{Y}}$ be two circularly symmetric channels. Moreover, let Y and $\tilde{Y}$ be the channel outputs induced by the uniform input X via $p_{Y|X}$ and $p_{\tilde{Y}|X}$, respectively. For any real numbers α and $\tilde{\alpha}$, the maximum value of the following optimization problem

$$\max_{p_{U|X}} -\alpha H(Y|U) + \tilde{\alpha} H(\tilde{Y}|U) \quad (12)$$

is attained by some circularly symmetric channel $p_{U|X} : \mathcal{X} \to \mathcal{U}$ with $\mathcal{U} = \{0, 1, \dots, M - 1\}$ such that $p_{U|X}(u|x)$ depends on (x, u) only through $u \ominus_M x$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$. Here it is assumed that $U \leftrightarrow X \leftrightarrow (Y, \tilde{Y})$ form a Markov chain.

Lemma 2: Let $m : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ be a circularly symmetric distortion measure associated with some bijective function μ. Let $p_{U|X} : \mathcal{X} \to \mathcal{U}$ and $p_{\tilde{U}|X} : \mathcal{X} \to \tilde{\mathcal{U}}$ be two circularly symmetric channels with $\mathcal{U} = \tilde{\mathcal{U}} = \{0, 1, \dots, M - 1\}$ such that $p_{U|X}(u|x)$ and $p_{\tilde{U}|X}(\tilde{u}|x)$ depend on

[4]For simplicity, we implicitly assume $\mathcal{Y}$ is a finite set; however, it will be clear that such an assumption is inessential.

(x, u) and $(x, \tilde{u})$ only through $u \ominus_M x$ and $\tilde{u} \ominus_M x$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$ and $(x, \tilde{u}) \in \mathcal{X} \times \tilde{\mathcal{U}}$; moreover, $p_{\tilde{U}|X}$ is stochastically degraded with respect to $p_{U|X}$. Let U and $\tilde{U}$ be the channel outputs induced by the uniform input X via $p_{U|X}$ and $p_{\tilde{U}|X}$, respectively. For any real numbers α and $\tilde{\alpha}$ such that $\alpha \geq \tilde{\alpha}$ and $\alpha \geq 0$, the minimum value of the following optimization problem[5]

$$\min_{p_{Y|X}:\mathbb{E}[m(X,Y)]\leq d} -\alpha H(U|Y) + \tilde{\alpha} H(\tilde{U}|Y) \tag{13}$$

is attained by some circularly symmetric channel $p_{Y|X} : \mathcal{X} \to \mathcal{Y}$ with the property that $p_{Y|X}(\mu^x(y)|x) = p_{Y|X}(y|0)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Here it is assumed that $(U, \tilde{U}) \leftrightarrow X \leftrightarrow Y$ form a Markov chain.

A. The Erasure Distortion Measure

This subsection is devoted to the evaluation of $r(\underline{d})$ for the binary uniform source with the erasure distortion measure. Without loss of generality, we assume $d_k \in [0, 1]$, $k = 1, \ldots, L$. Let $\underline{\alpha} = (\alpha_1, \ldots, \alpha_L)$, $\underline{q} = (q_0, q_1, \ldots, q_L)$, and $\underline{\delta} = (\delta_1, \ldots, \delta_L)$. Define

$$\kappa(\underline{\alpha}, \underline{q}, \underline{\delta}) = \sum_{k=1}^{L} \alpha_k (1 - \delta_k)[H_b(q_{k-1}) - H_b(q_k)],$$

where $H_b(q) = -q \log q - (1-q) \log(1-q)$. Moreover, define

$$\varphi_*(\underline{\alpha}, \underline{d}) = \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta}),$$

$$\varphi^*(\underline{\alpha}, \underline{d}) = \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta}),$$

where $\mathcal{D}(\underline{d}) = [0, d_1] \times \ldots \times [0, d_L]$.

Theorem 2: There exists a saddle-point solution $(\underline{q}^*, \underline{\delta}^*)$ in the sense that

$$\underline{q}^* \in \arg \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta}^*), \tag{14}$$

$$\underline{\delta}^* \in \arg \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}^*, \underline{\delta}), \tag{15}$$

where $\underline{q}^* = (q_0^*, q_1^*, \ldots, q_L^*)$ and $\underline{\delta}^* = (\delta_1^*, \ldots, \delta_L^*)$. If $\alpha_1 \geq \ldots \geq \alpha_L \geq 0$, then there exists a saddle-point solution $(\underline{q}^*, \underline{\delta}^*)$ such that $q_0^* \geq q_1^* \geq \ldots \geq q_L^*$, $\delta_1^* \geq \ldots \geq \delta_L^*$, and $q_{k-1}^* = q_k^*$ whenever $d_k = 1$, $k = 1, \ldots, L$. Furthermore, if $\alpha_1 > \ldots > \alpha_L > 0$, then every saddle-point solution $(\underline{q}^*, \underline{\delta}^*)$ has the property that $q_0^* \geq q_1^* \geq \ldots \geq q_L^*$.

Proof: See Appendix E. ∎

The following result is a direct consequence of Theorem 2.

Corollary 2: $\varphi_*(\underline{\alpha}, \underline{d}) = \varphi^*(\underline{\alpha}, \underline{d})$.

In view of Corollary 2, we shall simply denote $\varphi_*(\underline{\alpha}, \underline{d})$ and $\varphi^*(\underline{\alpha}, \underline{d})$ by $\varphi(\underline{\alpha}, \underline{d})$. Note that

$$\varphi(\underline{\alpha}, \underline{d}) = \kappa(\underline{\alpha}, \underline{q}^*, \underline{\delta}^*)$$

for any $(\underline{q}^*, \underline{\delta}^*)$ satisfying (14) and (15).

Corollary 3: If $\alpha_1 \geq \ldots \alpha_L \geq 0$, then

$$\varphi(\underline{\alpha}, \underline{d}) = \max_{k \in \mathcal{L}} \alpha_k (1 - d_k) \log 2.$$

Proof: In view of Theorem 2, we have

$$\varphi(\underline{\alpha}, \underline{d}) = \max_{\underline{q}: \frac{1}{2} \geq q_0 \geq q_1 \geq \ldots q_L \geq 0} \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta})$$

$$= \max_{\underline{q}: \frac{1}{2} \geq q_0 \geq q_1 \geq \ldots q_L \geq 0} \kappa(\underline{\alpha}, \underline{q}, \underline{d}).$$

Note that

$$\kappa(\underline{\alpha}, \underline{q}, \underline{d}) = \alpha_1(1 - d_1)H_b(q_0) - \alpha_L(1 - d_L)H_b(q_L)$$

$$+ \sum_{k=1}^{L-1}[\alpha_{k+1}(1 - d_{k+1}) - \alpha_k(1 - d_k)]H_b(q_k).$$

Therefore, to maximize $\kappa(\underline{\alpha}, \underline{q}, \underline{d})$ over $\underline{q}$ subject to the constraint $\frac{1}{2} \geq q_0 \geq q_1 \geq \ldots q_L \geq 0$, one can safely set $q_0 = \frac{1}{2}$ and $q_L = 0$; furthermore, one can eliminate q_1 by setting

$$q_1 = \begin{cases} \frac{1}{2}, & \alpha_1(1 - d_1) < \alpha_2(1 - d_2) \\ q_2, & \text{otherwise.} \end{cases}$$

Note that the resulting expression is a linear combination of $H_b(q_k)$, $k = 2, \ldots, L - 1$, plus a constant term. One can further eliminate q_2 by setting $q_2 = \frac{1}{2}$ if the coefficient in front of $H(q_2)$ is positive and setting $q_2 = q_3$ otherwise. By repeating this process, we will eventually obtain a vector $\underline{q}$ with the property that

$$q_k = \begin{cases} \frac{1}{2}, & k < i \\ 0, & \text{otherwise} \end{cases}$$

for some $i \in \mathcal{L}$. Maximizing $\kappa(\underline{\alpha}, \underline{q}, \underline{d})$ over all such $\underline{q}$ yields the desired result. ∎

Recall that $\mathcal{P}(\underline{d})$ is the set of conditional distributions $p_{\hat{\underline{S}}|S}$ with $\hat{\underline{S}} = (\hat{S}_1, \ldots, \hat{S}_L)$ such that the induced joint distribution $p_{S,\hat{\underline{S}}}$ satisfies $\mathbb{E}[m(S, \hat{S}_k)] \leq d_k$, $k = 1, \ldots, L$. Let $\mathcal{Q}$ denote the set of all possible conditional distributions $p_{\underline{Z}|S}$, where $\underline{Z} = (Z_0, Z_1, \ldots, Z_L)$. Define

$$\Gamma_*(\underline{\alpha}, \underline{d}) = \sup_{p_{\underline{Z}|S} \in \mathcal{Q}} \inf_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}|S}),$$

$$\Gamma^*(\underline{\alpha}, \underline{d}) = \inf_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \sup_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}|S}),$$

where

$$\gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}|S}) = \sum_{k=1}^{L} \alpha_k [I(Z_k; \hat{S}_k) - I(Z_{k-1}; \hat{S}_k)]$$

and it is assumed that $\underline{Z} \leftrightarrow S \leftrightarrow \hat{\underline{S}}$ form a Markov chain. Our main result in this subsection is the following minimax theorem.

Theorem 3: For the binary uniform source with the erasure distortion measure,

$$\Gamma_*(\underline{\alpha}, \underline{d}) = \Gamma^*(\underline{\alpha}, \underline{d}) = \varphi(\underline{\alpha}, \underline{d}).$$

Furthermore, every $(p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S})$ with the property that $p_{Z_k^*|S}$ is a BSC(q_k^*), $k = 0, 1, \ldots, L$, and $p_{\hat{S}_k^*|S}$ is a BEC(δ_k^*), $k = 1, \ldots, L$, for some $(\underline{q}^*, \underline{\delta}^*)$ satisfying (14) and (15) is a saddle-point solution in the sense that

$$p_{\underline{Z}^*|S} \in \arg \max_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S}),$$

$$p_{\hat{\underline{S}}^*|S} \in \arg \min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S}).$$

Proof: Given any $(p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S})$ with the property that $p_{Z_k^*|S}$ is a BSC(q_k^*), $k = 0, 1, \ldots, L$, and $p_{\hat{S}_k^*|S}$ is a BEC(δ_k^*), $k = 1, \ldots, L$, for some $(\underline{q}^*, \underline{\delta}^*)$ satisfying (14) and (15), let us consider the following optimization problems

$$\max_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S}), \tag{16}$$

$$\min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S}). \tag{17}$$

Note that $\gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S})$ depends on $p_{\underline{Z}|S}$ only through $p_{Z_k|S}$, $k = 0, 1, \ldots, L$. As a consequence, (16) can be decomposed into the following sub-problems

$$\max_{p_{Z_0|S}} -\alpha_1 I(Z_0; \hat{S}_1^*),$$

$$\max_{p_{Z_k|S}} \alpha_k I(Z_k; \hat{S}_k^*) - \alpha_{k+1} I(Z_k; \hat{S}_{k+1}^*), \quad k = 1, \ldots, L-1,$$

$$\max_{p_{Z_L|S}} \alpha_L I(Z_L; \hat{S}_L^*),$$

which, in light of the fact that $H(\hat{S}_k^*)$, $k = 1, \ldots, L$, do not depend on $p_{\underline{Z}|S}$, are equivalent to

$$\max_{p_{Z_0|S}} \alpha_1 H(\hat{S}_1^*|Z_0), \tag{18}$$

$$\max_{p_{Z_k|S}} -\alpha_k H(\hat{S}_k^*|Z_k) + \alpha_{k+1} H(\hat{S}_{k+1}^*|Z_k),$$
$$k = 1, \ldots, L-1, \tag{19}$$

$$\max_{p_{Z_L|S}} -\alpha_L H(\hat{S}_L^*|Z_L). \tag{20}$$

By Lemma 1, the maximum values of (18), (19), and (20) are attained by $p_{Z_k|S}$, $k = 0, 1, \ldots, L$, with the property that $p_{Z_k|S}$ is a BSC(q_k) for some $q_k \in [0, \frac{1}{2}]$, $k = 0, 1, \ldots, L$. For such $p_{Z_k|S}$, $k = 0, 1, \ldots, L$, it can be verified that

$$\sum_{k=1}^{L} \alpha_k [I(Z_k; \hat{S}_k^*) - I(Z_{k-1}; \hat{S}_k^*)]$$

$$= \sum_{k=1}^{L} \alpha_k [H(Z_{k-1}|\hat{S}_k^*) - H(Z_k|\hat{S}_k^*)] \tag{21}$$

$$= \sum_{k=1}^{L} \alpha_k \sum_{\hat{s} \in \{0,1,e\}} p_{\hat{S}_k^*}(\hat{s})[H(Z_{k-1}|\hat{S}_k^* = \hat{s}) - H(Z_k|\hat{S}_k^* = \hat{s})]$$

$$= \sum_{k=1}^{L} \alpha_k \sum_{\hat{s} \in \{0,1\}} p_{\hat{S}_k^*}(\hat{s})[H(Z_{k-1}|\hat{S}_k^* = \hat{s}) - H(Z_k|\hat{S}_k^* = \hat{s})] \tag{22}$$

$$= \sum_{k=1}^{L} \alpha_k \sum_{\hat{s} \in \{0,1\}} p_{\hat{S}_k^*}(\hat{s})[H(Z_{k-1}|S = \hat{s}) - H(Z_k|S = \hat{s})] \tag{23}$$

$$= \sum_{k=1}^{L} \alpha_k \sum_{\hat{s} \in \{0,1\}} p_{\hat{S}_k^*}(\hat{s})[H_b(q_{k-1}) - H_b(q_k)]$$

$$= \sum_{k=1}^{L} \alpha_k (1 - \delta_k^*)[H_b(q_{k-1}) - H_b(q_k)],$$

where (21) is because $H(Z_{k-1}) = H(Z_k) = \log 2$, (22) is because $H(Z_{k-1}|\hat{S}_k^* = e) = H(Z_k|\hat{S}_k^* = e) = \log 2$, and (23)

is due to the fact that $(Z_{k-1}, Z_k) \leftrightarrow S \leftrightarrow \hat{S}_k^*$ form a Markov chain and $p_{S|\hat{S}_k^*}(0|0) = p_{S|\hat{S}_k^*}(1|1) = 1$, $k = 1, \ldots, L$. Therefore, we have

$$\max_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S})$$
$$= \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta}^*)$$
$$= \kappa(\underline{\alpha}, \underline{q}^*, \underline{\delta}^*)$$
$$= \gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S}).$$

Similarly, since $\gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S})$ depends on $p_{\hat{\underline{S}}|S}$ only through $p_{\hat{S}_k|S}$, $k = 1, \ldots, L$., we can decompose (17) into

$$\min_{p_{\hat{S}_k|S}: \mathbb{E}[m_E(S, \hat{S}_k)] \leq d_k} \alpha_k I(Z_k^*; \hat{S}_k) - \alpha_k I(Z_{k-1}^*; \hat{S}_k)$$

for $k = 1, \ldots, L$, which, in light of the fact that $H(Z_k^*)$, $k = 0, 1, \ldots, L$, do not depend on $p_{\hat{\underline{S}}|S}$, are equivalent to

$$\min_{p_{\hat{S}_k|S}: \mathbb{E}[m_E(S, \hat{S}_k)] \leq d_k} -\alpha_k H(Z_k^*|\hat{S}_k) + \alpha_k H(Z_{k-1}^*|\hat{S}_k) \tag{24}$$

for $k = 1, \ldots, L$. We shall consider the following cases.

- $q_{k-1}^* \geq q_k^*$ *and* $\alpha_k \geq 0$: In this case $p_{Z_{k-1}^*|S}$ is stochastically degraded with respect to $p_{Z_{k-1}^*|S}$. Therefore, it follows from Lemma 2 that the minimum value of (24) is attained by $p_{\hat{S}_k|S}$ with the property that $p_{\hat{S}_k|S}$ is a BEC(δ_k) for some $\delta_k \in [0, d_k]$.
- $q_{k-1}^* < q_k^*$ *and* $\alpha_k \geq 0$: Since the objective function in (24) depends on $p_{Z_{k-1}^*, Z_k^*|S}$ only through $p_{Z_{k-1}^*|S}$ and $p_{Z_k^*|S}$, there is no loss of generality in assuming that $p_{Z_k^*|S}$ is physically degraded with respect to $p_{Z_{k-1}^*|S}$. As a consequence, we have

$$H(Z_{k-1}^*, Z_k^*|\hat{S}_k) = H(Z_{k-1}^*|\hat{S}_k) + H(Z_k^*|Z_{k-1}^*),$$

which, together with the fact that $H(Z_{k-1}^*, Z_k^*|\hat{S}_k) = H(Z_k^*|\hat{S}_k) + H(Z_{k-1}^*|Z_k^*, \hat{S}_k)$, implies

$$-\alpha_k H(Z_k^*|\hat{S}_k) + \alpha_k H(Z_{k-1}^*|\hat{S}_k)$$
$$= \alpha_k H(Z_k^*|Z_{k-1}^*, \hat{S}_k) - \alpha_k H(Z_{k-1}^*|Z_k^*).$$

Since $H(Z_{k-1}^*|Z_k^*)$ does not depend on $p_{\hat{S}_k|S}$, it follows that (24) is equivalent to

$$\min_{p_{\hat{S}_k|S}: \mathbb{E}[m_E(S, \hat{S}_k)] \leq d_k} \alpha_k H(Z_k^*|Z_{k-1}^*, \hat{S}_k). \tag{25}$$

It is obvious that the minimum value of (25) is attained when $p_{\hat{S}_k|S}$ is a BEC(δ_k) with $\delta_k = 0$.
- $\alpha_k < 0$: This case can be converted to the case $\alpha_k > 0$ by switching the roles of Z_{k-1}^* and Z_k^*.

In summary, the minimum values of (24) for $k = 1, \ldots, L$ are attained by $p_{\hat{S}_k|S}$, $k = 1, \ldots, L$, with the property that $p_{\hat{S}_k|S}$ is a BEC(δ_k) for some $\delta_k \in [0, d_k]$, $k = 1, \ldots, L$. For such $p_{\hat{S}_k|S}$, $k = 1, \ldots, L$, it can be verified that

$$\sum_{k=1}^{L} \alpha_k [I(Z_k^*; \hat{S}_k) - I(Z_{k-1}^*; \hat{S}_k)]$$

$$= \sum_{k=1}^{L} \alpha_k (1 - \delta_k)[H_b(q_{k-1}^*) - H_b(q_k^*)].$$

Therefore, we have

$$\min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma \left(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S} \right)$$

$$= \max_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}^*, \underline{\delta})$$

$$= \kappa(\underline{\alpha}, \underline{q}^*, \underline{\delta}^*)$$

$$= \gamma \left(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S} \right).$$

This completes the proof of Theorem 3. ∎

Corollary 4: For the binary uniform source with the erasure distortion measure,

$$r(\underline{d}) = \max_{k \in \mathcal{L}} \frac{L}{k} (1 - d_k) \log 2.$$

Proof: It is easy to see that

$$\sum_{k=1}^{L} \frac{L}{k} I(Z_k; \hat{S}_k | Z_{k-1}) = \gamma \left(\underline{\alpha}^*, p_{\underline{Z}|S}, p_{\hat{\underline{S}}|S} \right)$$

for $p_{\underline{Z}|S} \in \mathcal{P}$, where $\underline{\alpha}^* = (\alpha_1^*, \dots, \alpha_L^*)$ with $\alpha_k^* = \frac{L}{k}$, $k = 1, \dots, L$. Since $\mathcal{P} \subseteq \mathcal{Q}$, we must have $r(\underline{d}) \le \Gamma^*(\underline{\alpha}^*, \underline{d})$, which, together with Theorem 3 and Corollary 3, implies

$$r(\underline{d}) \le \max_{k \in \mathcal{L}} \frac{L}{k} (1 - d_k) \log 2. \qquad (26)$$

Let $(\underline{q}^*, \underline{\delta}^*)$ be an arbitrary saddle-point solution satisfying (14) and (15) with $\underline{\alpha} = \underline{\alpha}^*$. By Theorem 2, such a saddle-point solution exists and has the property that $q_0^* \ge q_1^* \ge \dots \ge q_L^*$. Now construct $(p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S})$ with the property that $p_{Z_k^*|S}$ is a BSC(q_k^*), $k = 0, 1, \dots, L$, and $p_{\hat{S}_k^*|S}$ is a BEC(δ_k^*), $k = 1, \dots, L$; in particular, since $q_0^* \ge q_1^* \ge \dots \ge q_L^*$, we can construct $p_{\underline{Z}^*|S}$ such that $p_{Z_{k-1}^*|S}$ is physically degraded with respect to $p_{Z_k^*|S}$, $k = 1, \dots, L$. Since $p_{\underline{Z}^*|S} \in \mathcal{P}$, it follows that

$$r(\underline{d}) \ge \min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma \left(\underline{\alpha}^*, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S} \right).$$

By Theorem 3 and Corollary 3,

$$\min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma \left(\underline{\alpha}^*, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S} \right)$$

$$= \gamma \left(\underline{\alpha}^*, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S} \right)$$

$$= \phi(\underline{\alpha}^*, \underline{d})$$

$$= \max_{k \in \mathcal{L}} \frac{L}{k} (1 - d_k) \log 2.$$

Therefore, we have

$$r(\underline{d}) \ge \max_{k \in \mathcal{L}} \frac{L}{k} (1 - d_k) \log 2. \qquad (27)$$

Combining (26) and (27) completes the proof of Corollary 4. ∎

Note that $r(\underline{d})$ has the following simple interpretation. Since the reconstruction distortion based every k descriptions is no greater than d_k, according to the rate-distortion function of the binary uniform source with the erasure distortion measure, the total rate of every k descriptions is at least $(1 - d_k) \log 2$, which implies $R(\underline{d}) \ge \frac{L}{k}(1 - d_k) \log 2$; maximizing $\frac{L}{k}(1 - d_k) \log 2$ over $k \in \mathcal{L}$ yields $r(\underline{d})$.

Now consider the case $R(\underline{d}) = \frac{L}{\ell}(1 - d_\ell) \log 2$ for some ℓ. Since $R(\underline{d}) \ge r(\underline{d})$, it follows that $\frac{L}{k}(1 - d_k) \le \frac{L}{\ell}(1 - d_\ell)$, i.e.,

$$d_k \ge \frac{k}{\ell} d_\ell + \frac{\ell - k}{\ell}$$

for all $k \ne \ell$. Particularly, in the no excess sum rate case $R(\underline{d}) = (1 - d_L) \log 2$, we have

$$d_k \ge \frac{k}{L} d_L + \frac{L - k}{L} \qquad (28)$$

for $k = 1, \dots, L - 1$. Interestingly, the lower bounds in (28) are in fact achievable via a simple time-sharing scheme. Specifically, we partition nL source samples into L segments, each of length n; a lossy source code of rate $(1 - d_L) \log 2$, which is optimal for the binary uniform source with the erasure distortion measure, is used to encode these segments separately such that the resulting distortion for each segment is d_L (here optimal encoding is trivial in the sense that one simply keeps $n(1 - d_L)$ source samples[6] for each segment); let description k contain the encoding output for segment k, $k = 1, \dots, L$. Clearly, the sum rate of this scheme is $(1 - d_L) \log 2$; moreover, this scheme has the property that the reconstruction distortion based on every k descriptions is $\frac{k}{L} d_L + \frac{L-k}{L}$, $k = 1, \dots, L$. It is worth mentioning that, in the current setting, this scheme is essentially equivalent to the one described in [7, p. 1331].

We are now in a position to give a more conceptual explanation of the ideas that lead to the proof of Corollary 4. As pointed out earlier, the lower bound in Corollary 4, which is derived through rather sophisticated analysis, is almost a trivial one.[7] In a certain sense, it is not the result but the proof strategy that is important, and we basically use the binary uniform source with the erasure distortion measure as a toy example to illustrate this strategy, which will be used repeatedly in the subsequent part of this paper to obtain more non-trivial results. Note that the definition of $\Gamma^*(\underline{\alpha}^*, \underline{d})$ is almost identical with that of $r(\underline{d})$ except that no Markov ordering is imposed on remote sources for $\Gamma^*(\underline{\alpha}^*, \underline{d})$. This relaxation is crucial since it decouples remote sources and enables us to decompose the problem of maximization over $p_{\underline{Z}|S}$ into sub-problems of maximization over $p_{Z_k|S}$, $k = 0, 1, \dots, L$. Indeed, our results indicate that the relaxed version has the advantage of being amenable to saddle-point analysis and evaluation. Interestingly, it turns out that the Markov ordering is automatically satisfied by the saddle-point solution of the relaxed minimax problem and consequently $\Gamma^*(\underline{\alpha}^*, \underline{d}) = r(\underline{d})$.

B. The Hamming Distortion Measure

In this subsection we shall evaluate $r(\underline{d})$ for the binary uniform source with the Hamming distortion measure. Without loss of generality, we assume $d_k \in [0, \frac{1}{2}]$, $k = 1, \dots, L$. Define

$$\eta(\underline{\alpha}, \underline{q}, \underline{\delta}) = \sum_{k=1}^{L} \alpha_k [H_b(q_{k-1} \odot \delta_k) - H_b(q_k \odot \delta_k)],$$

---

[6] We assume that $n(1 - d_L)$ is an integer.
[7] See [7, Sec. IV] for stronger converse results in certain special cases.

where $q \odot \delta = q(1-\delta) + (1-q)\delta$. Moreover, define

$$\phi_*(\underline{\alpha}, \underline{d}) = \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \eta(\underline{\alpha}, \underline{q}, \underline{\delta}),$$

$$\phi^*(\underline{\alpha}, \underline{d}) = \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \eta(\underline{\alpha}, \underline{q}, \underline{\delta}).$$

We say that $(\underline{q}^*, \underline{\delta}^*)$ is a saddle-point solution if

$$\underline{q}^* \in \arg \max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \eta(\underline{\alpha}, \underline{q}, \underline{\delta}^*), \tag{29}$$

$$\underline{\delta}^* \in \arg \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \eta(\underline{\alpha}, \underline{q}^*, \underline{\delta}). \tag{30}$$

Theorem 4: If $\alpha_1 \geq \ldots \alpha_L \geq 0$, then there exists a saddle-point solution $(\underline{q}^*, \underline{\delta}^*)$ such that $q_0^* \geq q_1^* \geq \ldots \geq q_L^*$, $\delta_1^* \geq \ldots \geq \delta_L^*$, and $q_{k-1}^* = q_k^*$ whenever $d_k = \frac{1}{2}$, $k = 1, \ldots, L$. Furthermore, if $\alpha_1 > \ldots > \alpha_L > 0$, then every saddle-point solution $(\underline{q}^*, \underline{\delta}^*)$ has the property that $q_0^* \geq q_1^* \geq \ldots \geq q_L^*$.

Proof: See Appendix F. ∎

The following result is a direct consequence of Theorem 4.

Corollary 5: $\phi_*(\underline{\alpha}, \underline{d}) = \phi^*(\underline{\alpha}, \underline{d})$ when $\alpha_1 \geq \ldots \alpha_L \geq 0$.

In view of Corollary 5, we shall simply denote $\phi_*(\underline{\alpha}, \underline{d})$ and $\phi^*(\underline{\alpha}, \underline{d})$ by $\phi(\underline{\alpha}, \underline{d})$ when $\alpha_1 \geq \ldots \alpha_L \geq 0$. Note that

$$\phi(\underline{\alpha}, \underline{d}) = \eta(\underline{\alpha}, \underline{q}^*, \underline{\delta}^*)$$

for any $(\underline{q}^*, \underline{\delta}^*)$ satisfying (29) and (30).

Corollary 6: If $\alpha_1 \geq \ldots \alpha_L \geq 0$ and $d_k = \frac{1}{2}$ for all $k \neq i$ and $k \neq j$ (with $i < j$), then

$\phi(\underline{\alpha}, \underline{d})$
$$= \begin{cases} \alpha_i [\log 2 - H_b(d_i)], \\ \quad \alpha_i (1 - 2d_i) \log \left(\frac{1-d_i}{d_i}\right) \geq \alpha_j (1 - 2d_j) \log \left(\frac{1-d_j}{d_j}\right) \\ \alpha_j [\log 2 - H_b(d_j)], \quad \alpha_i (1-2d_i)^2 \leq \alpha_j (1-2d_j)^2 \\ \alpha_i [\log 2 - H_b(q^\dagger \odot d_i)] \\ \quad + \alpha_j [H_b(q^\dagger \odot d_j) - H_b(d_j)], \quad \text{otherwise,} \end{cases}$$

where $q^\dagger$ is the unique solution of

$$\alpha_i (1-2d_i) \log \left(\frac{1-q \odot d_i}{q \odot d_i}\right) = \alpha_j (1-2d_j) \log \left(\frac{1-q \odot d_j}{q \odot d_j}\right)$$

for $q \in (0, \frac{1}{2})$.

Proof: It follows from Theorem 4 that

$$\phi(\underline{\alpha}, \underline{d}) = \max \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \eta(\underline{\alpha}, \underline{q}, \underline{\delta}),$$

where the maximization is taken over those $\underline{q}$ such that $\frac{1}{2} \geq q_0 = \cdots = q_{i-1} \geq q_i = \cdots = q_{j-1} \geq q_j = \cdots = q_L \geq 0$. As a consequence, we have

$\phi(\underline{\alpha}, \underline{d})$
$$= \max_{(q_0, q_i, q_j): \frac{1}{2} \geq q_0 \geq q_i \geq q_j \geq 0} \min_{(\delta_i, \delta_j) \in [0, d_i] \times [0, d_j]} \alpha_i [H_b(q_0 \odot \delta_i) - H_b(q_i \odot \delta_i)]$$
$$\quad + \alpha_j [H_b(q_i \odot \delta_j) - H_b(q_j \odot \delta_j)]$$
$$= \max_{(q_0, q_i, q_j): \frac{1}{2} \geq q_0 \geq q_i \geq q_j \geq 0} \alpha_i [H_b(q_0 \odot d_i) - H_b(q_i \odot d_i)]$$
$$\quad + \alpha_j [H_b(q_i \odot d_j) - H_b(q_j \odot d_j)]$$
$$= \max_{q_i \in [0, \frac{1}{2}]} \alpha_i [\log 2 - H_b(q_i \odot d_i)] + \alpha_j [H_b(q_i \odot \delta_j) - H_b(d_j)].$$

Invoking Lemma 6 in Appendix F completes the proof of Corollary 6. ∎

Our main result in this subsection is the following minimax theorem.

Theorem 5: For the binary uniform source with the Hamming distortion measure,

$$\Gamma_*(\underline{\alpha}, \underline{d}) = \Gamma^*(\underline{\alpha}, \underline{d}) = \phi(\underline{\alpha}, \underline{d})$$

when $\alpha_1 \geq \cdots \alpha_L \geq 0$. Furthermore, every $(p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S})$ with the property that $p_{Z_k^*|S}$ is a BSC(q_k^*), $k = 0, 1, \ldots, L$, and $p_{\hat{S}_k^*|S}$ is a BSC(δ_k^*), $k = 1, \ldots, L$, for some $(\underline{q}^*, \underline{\delta}^*)$ satisfying (29) and (30) is a saddle-point solution in the sense that

$$p_{\underline{Z}^*|S} \in \arg \max_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S}),$$

$$p_{\hat{\underline{S}}^*|S} \in \arg \min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S}).$$

Corollary 7: For the binary uniform source with the Hamming distortion measure,

$$r(\underline{d}) = \phi(\underline{\alpha}^*, \underline{d}),$$

where $\underline{\alpha}^* = (\alpha_1^*, \ldots, \alpha_L^*)$ with $\alpha_k^* = \frac{L}{k}$, $k = 1, \ldots, L$.

The proofs of Theorem 5 and Corollary 7 are omitted due to their similarity to the proofs of Theorem 3 and Corollary 4.

Now consider the case $R(\underline{d}) = \frac{L}{\ell}[\log 2 - H_b(d_\ell)]$ for some ℓ (note that $\ell = 1$ and $\ell = L$ correspond to the no excess marginal rate case and the no excess sum rate case, respectively). By leveraging Corollary 7 and Corollary 6, one can establish the following lower bounds on d_k for $k \neq \ell$:

$$d_k \geq \frac{1}{2} - \frac{1}{2}\sqrt{\frac{k}{\ell}} + \sqrt{\frac{k}{\ell}} d_\ell, \quad k < \ell,$$

$$d_k \geq \min \left\{ d \in [0, \frac{1}{2}] : \frac{L}{k}(1 - 2d) \log \left(\frac{1-d}{d}\right) \right.$$
$$\left. \leq \frac{L}{\ell}(1 - 2d_\ell) \log \left(\frac{1-d_\ell}{d_\ell}\right) \right\}, \quad k > \ell.$$

In particular, we have

$$d_1 \geq \frac{1}{2} - \frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{2} d_2 \tag{31}$$

when $\ell = L = 2$. For the two-description problem, it is known [2] that $R(\underline{d}) = \log 2 - H_b(d_2)$ if and only if

$$d_1 \geq \frac{\sqrt{2}-1}{2} + (2 - \sqrt{2})d_2.$$

Note that, for $d_2 \in [0, \frac{1}{2}]$,

$$\frac{\sqrt{2}-1}{2} + (2 - \sqrt{2})d_2 - \left(\frac{1}{2} - \frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{2} d_2\right)$$
$$= \frac{3\sqrt{2}-4}{4} + \frac{4-3\sqrt{2}}{2} d_2 \geq 0,$$

where the inequality is strict unless $d_2 = \frac{1}{2}$. Therefore, the lower bound in (31) is loose except for the degenerate case $d_2 = \frac{1}{2}$.

IV. THE QUADRATIC GAUSSIAN CASE

In this section we shall consider the quadratic Gaussian case, where p_S is a Gaussian distribution (with mean zero and variance λ) and $m(s, \hat{s}) = (s - \hat{s})^2$ for $(s, \hat{s}) \in \mathcal{R} \times \mathcal{R}$, and evaluate $r(\underline{d})$ in this setting. Without loss of generality, we assume $d_k \in (0, \lambda]$, $k = 1, \ldots, L$.

First recall the definition of $\underline{\alpha}$, $\underline{\delta}$, $\mathcal{D}(\underline{d})$, $\gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}|S})$, $\Gamma_*(\underline{\alpha}, \underline{d})$, and $\Gamma^*(\underline{\alpha}, \underline{d})$ in Section III-A. Let

$$\omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}) = \sum_{k=1}^{L} \frac{\alpha_k}{2} \log \left(\frac{\lambda \theta_{k-1} + \lambda \delta_k - \theta_{k-1} \delta_k}{\lambda \theta_k + \lambda \delta_k - \theta_k \delta_k} \right),$$

where $\underline{\theta} = (\theta_0, \theta_1, \ldots, \theta_L)$. Moreover, define

$$\psi_*(\underline{\alpha}, \underline{d}) = \max_{\underline{\theta} \in [0, \lambda]^{L+1}} \min_{\underline{\delta} \in \mathcal{D}(d)} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}),$$

$$\psi^*(\underline{\alpha}, \underline{d}) = \min_{\underline{\delta} \in \mathcal{D}(d)} \max_{\underline{\theta} \in [0, \lambda]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}).$$

We say that $(\underline{\theta}^*, \underline{\delta}^*)$ is a saddle-point solution if

$$\underline{\theta}^* \in \arg \max_{\underline{\theta} \in [0, \lambda]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}^*), \quad (32)$$

$$\underline{\delta}^* \in \arg \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}^*, \underline{\delta}). \quad (33)$$

Theorem 6: If $\alpha_1 \geq \cdots \alpha_L \geq 0$, then there exists a saddle-point solution $(\underline{\theta}^*, \underline{\delta}^*)$ such that $\theta_0^* \geq \theta_1^* \geq \cdots \geq \theta_L^*$, $\delta_1^* \geq \cdots \geq \delta_L^*$, and $\theta_{k-1}^* = \theta_k^*$ whenever $d_k = \lambda$, $k = 1, \ldots, L$. Furthermore, if $\alpha_1 \geq \ldots \geq \alpha_L > 0$, then every saddle-point solution $(\underline{\theta}^*, \underline{\delta}^*)$ has the property that $\theta_0^* \geq \theta_1^* \geq \cdots \geq \theta_L^*$.

Proof: See Appendix G. ∎

The following result is a direct consequence of Theorem 6.

Corollary 8: $\psi_*(\underline{\alpha}, \underline{d}) = \psi^*(\underline{\alpha}, \underline{d})$ when $\alpha_1 \geq \ldots \alpha_L \geq 0$.

In view of Corollary 8, we shall simply denote $\psi_*(\underline{\alpha}, \underline{d})$ and $\psi^*(\underline{\alpha}, \underline{d})$ by $\psi(\underline{\alpha}, \underline{d})$ when $\alpha_1 \geq \ldots \alpha_L \geq 0$. Note that

$$\psi(\underline{\alpha}, \underline{d}) = \omega(\underline{\alpha}, \underline{\theta}^*, \underline{\delta}^*)$$

for any $(\underline{\theta}^*, \underline{\delta}^*)$ satisfying (32) and (33).

Corollary 9: If $\alpha_1 \geq \cdots \alpha_L \geq 0$ and $d_k = \lambda$ for all $k \neq i$ and $k \neq j$ (with $i < j$), then

$$\psi(\underline{\alpha}, \underline{d}) = \begin{cases} \frac{\alpha_i}{2} \log \left(\frac{\lambda}{d_i} \right), & \alpha_i (\lambda - d_i) d_j \geq \alpha_j (\lambda - d_j) d_i \\ \frac{\alpha_j}{2} \log \left(\frac{\lambda}{d_j} \right), & \alpha_i (\lambda - d_i) \leq \alpha_j (\lambda - d_j) \\ \frac{\alpha_i}{2} \log \left[\frac{(\alpha_i - \alpha_j)(\lambda - d_j)}{\alpha_i (d_i - d_j)} \right] \\ \quad + \frac{\alpha_j}{2} \log \left[\frac{\alpha_j \lambda (d_i - d_j)}{(\alpha_i - \alpha_j)(\lambda - d_i) d_j} \right], & \text{otherwise.} \end{cases}$$

Proof: See Appendix H. ∎

We say that $p_{Y|X} : \mathcal{R} \to \mathcal{R}$ is an additive Gaussian noise channel with parameter θ (i.e., AGNC(θ)) if $Y = \frac{\lambda - \theta}{\lambda} X + \sqrt{\frac{\theta(\lambda - \theta)}{\lambda}} N$, where $\theta \in [0, \lambda]$ and N is a zero-mean unit-variance Gaussian random variable independent of X. It can be readily verified that, for any θ and $\tilde{\theta}$ with $\lambda \geq \theta \geq \tilde{\theta} \geq 0$, there exist additive Gaussian noise channels $p_{Y|X}$ and $p_{\tilde{Y}|X}$ with parameters θ and $\tilde{\theta}$, respectively, such that $p_{Y|X}$ is physically degraded with respect to $p_{\tilde{Y}|X}$.

The following result is essentially the scalar version of [30, Th. 8].

Lemma 3: Let X be a Gaussian random variable with mean zero and variance λ. Moreover, let $p_{Y|X} : \mathcal{R} \to \mathcal{R}$ and $p_{\tilde{Y}|X} : \mathcal{R} \to \mathcal{R}$ be two additive Gaussian noise channels with parameters θ and $\tilde{\theta}$, respectively. For any real numbers α, $\tilde{\alpha}$, and d such that $\alpha \geq \tilde{\alpha} \geq 0$ and $d \in (0, \lambda]$, the maximum value of the following optimization problem

$$\max_{p_{U|X} : \mathbb{E}[(X - \mathbb{E}[X|U])^2] \leq d} \alpha I(U; Y) - \tilde{\alpha} I(U; \tilde{Y})$$

is attained by some additive Gaussian noise channel $p_{U|X} : \mathcal{R} \to \mathcal{R}$ with parameter $\delta \in [0, d]$. Here it is assumed that $U \leftrightarrow X \leftrightarrow (Y, \tilde{Y})$ form a Markov chain.

The following minimax theorem is the counterpart of Theorem 3 in Section III-A and Theorem 5 in Section III-B.

Theorem 7: For the quadratic Gaussian case,

$$\Gamma_*(\underline{\alpha}, \underline{d}) = \Gamma^*(\underline{\alpha}, \underline{d}) = \psi(\underline{\alpha}, \underline{d})$$

when $\alpha_1 \geq \cdots \alpha_L \geq 0$. Furthermore, every $(p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}^*|S})$ with the property that $p_{Z_k^*|S}$ is an AGNC(θ_k^*), $k = 0, 1, \ldots, L$, and $p_{\hat{S}_k^*|S}$ is an AGNC(δ_k^*), $k = 1, \ldots, L$, for some $(\underline{\theta}^*, \underline{\delta}^*)$ satisfying (32) and (33) is a saddle-point solution in the sense that

$$p_{\underline{Z}^*|S} \in \arg \max_{p_{\underline{Z}|S} \in \mathcal{Q}} \gamma(\underline{\alpha}, p_{\underline{Z}|S}, p_{\hat{\underline{S}}^*|S}),$$

$$p_{\hat{\underline{S}}^*|S} \in \arg \min_{p_{\hat{\underline{S}}|S} \in \mathcal{P}(\underline{d})} \gamma(\underline{\alpha}, p_{\underline{Z}^*|S}, p_{\hat{\underline{S}}|S}).$$

Proof: The proof is similar to that of Theorem 3 with Lemma 3 playing the roles of Lemma 1 and Lemma 2. The details are left to the interested reader. ∎

The following result is a simple consequence of Theorem 6 and Theorem 7. Its proof is similar to that of Corollary 4 and thus is omitted.

Corollary 10: For the quadratic Gaussian case,

$$r(\underline{d}) = \psi(\underline{\alpha}^*, \underline{d}),$$

where $\underline{\alpha}^* = (\alpha_1^*, \ldots, \alpha_L^*)$ with $\alpha_k^* = \frac{L}{k}$, $k = 1, \ldots, L$.

In view of Corollary 9 and Corollary 10, one can obtain an explicit expression of $r(\underline{d})$ when only two levels of distortion constraints are imposed.

Corollary 11: If $d_k = \lambda$ for all $k \neq i$ and $k \neq j$ (with $i < j$), then $r(\underline{d}) = R_{i,j}(d_i, d_j)$, where

$$R_{i,j}(d_i, d_j) = \begin{cases} \frac{L}{2i} \log \left(\frac{\lambda}{d_i} \right), & d_j \geq \left(\frac{j}{i} d_i^{-1} - \frac{j - i}{i} \lambda^{-1} \right)^{-1} \\ \frac{L}{2j} \log \left(\frac{\lambda}{d_j} \right), & d_j \leq \frac{j}{i} d_i - \frac{j - i}{i} \lambda \\ \frac{L}{2i} \log \left[\frac{(j - i)(\lambda - d_j)}{j(d_i - d_j)} \right] \\ \quad + \frac{L}{2j} \log \left[\frac{i \lambda (d_i - d_j)}{(j - i)(\lambda - d_i) d_j} \right], & \text{otherwise.} \end{cases}$$

The following theorem can be deduced from the results in [4, Sec. IV].

Theorem 8: 1) If $d_k = \lambda$ for all $k < \ell$ and $d_k \geq (\frac{k}{\ell} d_\ell^{-1} - \frac{k - \ell}{\ell} \lambda^{-1})^{-1}$ for all $k > \ell$ (with $\ell < L$), then

$$R(\underline{d}) \leq \frac{L}{2\ell} \log \left(\frac{\lambda}{d_\ell} \right). \quad (34)$$

2) If $d_k \geq \frac{k}{L}d_L + \frac{L-k}{L}\lambda$ for all $k < L$, then

$$R(\underline{d}) \leq \frac{1}{2}\log\left(\frac{\lambda}{d_L}\right). \qquad (35)$$

3) If $\frac{L}{\ell}d_\ell - \frac{L-\ell}{\ell}\lambda < d_L < \left(\frac{L}{\ell}d_\ell^{-1} - \frac{L-\ell}{\ell}\lambda^{-1}\right)^{-1}$ for some $\ell < L$ and

$$d_k = \lambda, \quad k < \ell,$$
$$d_k \geq \frac{L(k-\ell)(\lambda - d_\ell)d_L + \ell(L-k)(\lambda - d_L)d_\ell}{k(L-\ell)\lambda - L(k-\ell)d_\ell - \ell(L-k)d_L},$$
$$\ell < k < L,$$

then

$$R(\underline{d}) \leq \frac{L}{2\ell}\log\left[\frac{(L-\ell)(\lambda - d_L)}{L(d_\ell - d_L)}\right]$$
$$+ \frac{1}{2}\log\left[\frac{\ell\lambda(d_\ell - d_L)}{(L-\ell)(\lambda - d_\ell)d_L}\right]. \qquad (36)$$

Remark: One can write (34), (35), and (36) compactly as $R(\underline{d}) \leq R_{\ell,L}(d_\ell, d_L)$.

Combining Corollary 10, Corollary 11, and Theorem 8 yields the scalar version of [11, Th. 4], which determines the minimum sum rate of symmetric Gaussian multiple description coding subject to two levels of distortion constraints (with the second level imposed on the complete set of descriptions).

Corollary 12: If $d_k = \lambda$ for all $k \neq \ell$ and $k \neq L$ (with $\ell < L$), then $R(\underline{d}) = R_{\ell,L}(d_\ell, d_L)$.

The following result provides a partial converse for Theorem 8. In particular, it is shown that, when $R(\underline{d}) = R_{\ell,L}(d_\ell, d_L)$, the achievable distortions d_k, $\ell < k < L$, indicated in Theorem 8 are in fact the best possible.

Theorem 9: 1) If $R(\underline{d}) = \frac{L}{2\ell}\log\left(\frac{\lambda}{d_\ell}\right)$ for some $\ell < L$, then

$$d_k \geq \left(\frac{k}{\ell}d_\ell^{-1} - \frac{k-\ell}{\ell}\lambda^{-1}\right)^{-1}, \quad k > \ell.$$

2) If $R(\underline{d}) = \frac{1}{2}\log\left(\frac{\lambda}{d_L}\right)$, then

$$d_k \geq \frac{k}{L}d_L + \frac{L-k}{L}\lambda, \quad k < L.$$

3) If $R(\underline{d}) = R_{\ell,L}(d_\ell, d_L)$ for some $\ell < L$ and $\frac{L}{\ell}d_\ell - \frac{L-\ell}{\ell}\lambda < d_L < \left(\frac{L}{\ell}d_\ell^{-1} - \frac{L-\ell}{\ell}\lambda^{-1}\right)^{-1}$, then

$$d_k \geq \frac{L(k-\ell)(\lambda - d_\ell)d_L + \ell(L-k)(\lambda - d_L)d_\ell}{k(L-\ell)\lambda - L(k-\ell)d_\ell - \ell(L-k)d_L},$$
$$\ell < k < L.$$

Proof: See Appendix I. ∎

Remark: Case 2) in Theorem 9 corresponds to [7, Th. 11]. Furthermore, in this case the lower bounds on d_k, $k < L$, are achievable via a time-sharing scheme similar to the one described in Section III-A.

V. CONCLUSION

We have derived a single-letter lower bound on the sum rate of multiple description coding subject to symmetric distortion constraints. Furthermore, this lower bound is evaluated in several special cases with the aid of certain minimax theorems. It is worth noting that the minimax theorems established in

this paper depend critically on the properties of the binary erasure channel, the binary symmetric channel, and the additive Gaussian noise channel. An important feature shared by all these channels is that they can all be specified by a single parameter (which induces a natural Markov ordering). An interesting direction for future research is to investigate, to what extent, the minimax theorems in the present work can be extended to the scenarios where the relevant channels are more complex.

APPENDIX A
CARDINALITY BOUNDS

Note that

$$\sup_{p_{\underline{Z}|S} \in \mathcal{P}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1})$$

$$= \sup_{p_{\underline{Z}|S} \in \mathcal{P}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}}$$
$$\times \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0, \ldots, Z_{k-1})$$

$$\leq \sup_{p_{\underline{Z}|S}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}}$$
$$\times \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0, \ldots, Z_{k-1}).$$

On the other hand, setting $Z_k' = (Z_0, \ldots, Z_k)$, $k = 0, 1, \ldots, L$, we have

$$\sup_{p_{\underline{Z}|S}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0, \ldots, Z_{k-1})$$

$$= \sup_{p_{\underline{Z}|S}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k'; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}')$$

$$\leq \sup_{p_{\underline{Z}|S} \in \mathcal{P}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}),$$
$$(37)$$

where (37) is due to the fact that $p_{Z_{k-1}'|S}$ is physically degraded with respect to $p_{Z_k'|S}$, $k = 1, \ldots, L$. Therefore, $r([d])$ can be expressed alternatively as

$$r([d]) = \inf_{p_{[\hat{S}]|S} \in \mathcal{P}([d])} \sup_{p_{\underline{Z}|S}} \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}}$$
$$\times \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0, \ldots, Z_{k-1}).$$

For this expression of $r([d])$, there is no loss of optimality in setting $|\mathcal{Z}_0| = 1$ and $Z_L = S$; moreover, it suffices to have $|\mathcal{Z}_k| \leq |\mathcal{S}|$ in view of the fact [24, p. 631] that, for every $(z_0, \ldots, z_{k-1})$, one can find a conditional distribution of $(S, Z_k, \ldots, Z_{L-1})$ given $(Z_0, \ldots, Z_{k-1}) = (z_0, \ldots, z_{k-1})$

with $|\mathcal{Z}_k| \leq |\mathcal{S}|$ that preserves $p_{S|Z_0,\ldots,Z_{k-1}}(\cdot|z_0,\ldots,z_{k-1})$ and

$$\frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0 = z_0, \ldots,$$
$$Z_{k-1} = z_{k-1})$$
$$+\sum_{j=k+1}^{L} \frac{L}{j\binom{L}{j}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=j} I(Z_j; \hat{S}_{\mathcal{B}}, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_0 = z_0, \ldots,$$
$$Z_{k-1} = z_{k-1}, Z_k, \ldots, Z_{j-1}),$$
$$k = 1, \ldots, L-1.$$

The desired cardinality bounds for the original form of $r([d])$ in (2) follow immediately.

APPENDIX B
PROOF OF THEOREM 1

We augment the probability space by introducing an auxiliary random vector process $\{(Z_0(t), Z_1(t), \ldots, Z_L(t))\}_{t=1}^{\infty}$. It is assumed that $\{(Z_0(t), Z_1(t), \ldots, Z_L(t))\}_{t=1}^{\infty}$ and $\{S(t)\}_{t=1}^{\infty}$ form a jointly stationary and memoryless process with $(S(t), Z_0(t), Z_1(t), \ldots, Z_L(t))$ distributed according to $p_{S,\underline{Z}} \triangleq p_S p_{\underline{Z}|S}$ for every t, where $p_{\underline{Z}|S} \in \mathcal{P}$. Let $f_i^{(n)}: \mathcal{S}^n \rightarrow \mathcal{C}_i$, $i \in \mathcal{L}$, and $g_{\mathcal{A}}^{(n)}: \prod_{i \in \mathcal{A}} \mathcal{C}_i \rightarrow \hat{\mathcal{S}}^n$, $\mathcal{A} \in 2_+^{\mathcal{L}}$, be arbitrary encoding and decoding functions satisfying (1). By Han's subset entropy inequality [25],

$$\frac{1}{(k-1)\binom{L}{k-1}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k-1} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n)$$
$$\geq \frac{1}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n)$$

for $k = 2, \ldots, L$. Therefore, we have

$$\sum_{k=1}^{L-1} \frac{1}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_k^n)$$
$$\geq \sum_{k=2}^{L} \frac{1}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n). \quad (38)$$

Note that

$$\sum_{i=1}^{L} \log |\mathcal{C}_i|$$
$$\geq \sum_{i=1}^{L} H(f_i^{(n)}(S^n))$$
$$\geq \sum_{i=1}^{L} H(f_i^{(n)}(S^n))$$
$$-\sum_{k=1}^{L-1} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_k^n)$$
$$+\sum_{k=2}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n) \quad (39)$$

$$\geq -\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_k^n)$$
$$+\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n)$$
$$= -\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_k^n)$$
$$+\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n)$$
$$+\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A})$$
$$-\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} H(f_i^{(n)}(S^n), i \in \mathcal{A})$$
$$= \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k^n; f_i^{(n)}(S^n), i \in \mathcal{A})$$
$$-\sum_{k=1}^{L} \frac{L}{k\binom{L}{k}} \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_{k-1}^n; f_i^{(n)}(S^n), i \in \mathcal{A}), \quad (40)$$

where (39) is due to (38). It can be verified that, for any $k \in \mathcal{L}$ and $\mathcal{A} \in 2_+^{\mathcal{L}}$,

$$I(Z_k^n; f_i^{(n)}(S^n), i \in \mathcal{A}) - I(Z_{k-1}^n; f_i^{(n)}(S^n), i \in \mathcal{A})$$
$$= I(Z_{k-1}^n, Z_k^n; f_i^{(n)}(S^n), i \in \mathcal{A})$$
$$\quad -I(Z_{k-1}^n; f_i^{(n)}(S^n), i \in \mathcal{A}) \quad (41)$$
$$= I(Z_k^n; f_i^{(n)}(S^n), i \in \mathcal{A}|Z_{k-1}^n)$$
$$\geq I(Z_k^n; \hat{S}_{\mathcal{B}}^n, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}^n)$$
$$= \sum_{t=1}^{n} I(Z_k(t); \hat{S}_{\mathcal{B}}^n, \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}^n, Z_k^{t-1})$$
$$= \sum_{t=1}^{n} I(Z_k(t); \hat{S}_{\mathcal{B}}^n, \mathcal{B} \in 2_+^{\mathcal{A}}, Z_{k-1}^n, Z_k^{t-1}|Z_{k-1}(t)) \quad (42)$$
$$\geq \sum_{t=1}^{n} I(Z_k(t); \hat{S}_{\mathcal{B}}(t), \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}(t)), \quad (43)$$

where (41) is due to the fact that $Z_{k-1}^n \leftrightarrow Z_k^n \leftrightarrow (f_i^{(n)}(S^n), i \in \mathcal{A})$ form a Markov chain, and (42) is due to the fact that $Z_k(t) \leftrightarrow Z_{k-1}(t) \leftrightarrow (Z_{k-1}^n, Z_k^{t-1})$ form a Markov chain. Now let T be uniformly distributed over $\{1, \ldots, n\}$ and independent of $(S^n, Z_0^n, Z_1^n, \ldots, Z_L^n)$. We have, for any $k \in \mathcal{L}$ and $\mathcal{A} \in 2_+^{\mathcal{L}}$,

$$\sum_{t=1}^{n} I(Z_k(t); \hat{S}_{\mathcal{B}}(t), \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}(t))$$
$$= nI(Z_k(T); \hat{S}_{\mathcal{B}}(T), \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}(T), T)$$
$$= nI(Z_k(T); \hat{S}_{\mathcal{B}}(T), \mathcal{B} \in 2_+^{\mathcal{A}}, T|Z_{k-1}(T)) \quad (44)$$
$$\geq nI(Z_k(T); \hat{S}_{\mathcal{B}}(T), \mathcal{B} \in 2_+^{\mathcal{A}}|Z_{k-1}(T)), \quad (45)$$

where (44) is due to the fact that $Z_k(T) \leftrightarrow Z_{k-1}(T) \leftrightarrow T$ form a Markov chain. Combining (40), (43), and (45) gives

$$\frac{1}{n} \sum_{i=1}^{L} \log |\mathcal{C}_i| \geq \sum_{k=1}^{L} \frac{L}{k\binom{L}{k}}$$
$$\times \sum_{\mathcal{A} \in 2_+^{\mathcal{L}}, |\mathcal{A}|=k} I(Z_k(T); \hat{S}_{\mathcal{B}}(T), \mathcal{B} \in 2_+^{\mathcal{A}} | Z_{k-1}(T)). \tag{46}$$

Moreover, one can readily verify the following facts:

- $(S(T), Z_0(T), Z_1(T), \ldots, Z_L(T))$ and $(S, Z_0, Z_1, \ldots, Z_L)$ are identically distributed;
- $(Z_0(T), Z_1(T), \ldots, Z_L(T)) \leftrightarrow S(T) \leftrightarrow (\hat{S}_{\mathcal{A}}(T), \mathcal{A} \in 2_+^{\mathcal{L}})$ form a Markov chain;
- $\mathbb{E}[m(S(T), \hat{S}_{\mathcal{A}}(T))] \leq d_{\mathcal{A}}, \mathcal{A} \in 2_+^{\mathcal{L}}$.

Note that, given encoding and decoding functions, the conditional distribution $p_{[\hat{S}(T)]|S(T)}$ (with $[\hat{S}(T)] = (\hat{S}_{\mathcal{A}}(T), \mathcal{A} \in 2_+^{\mathcal{L}})$) is fixed, and (46) holds for an arbitrary choice of $p_{Z|S}$. This completes the proof of Theorem 1.

Remark: In the proof of Theorem 1, remote sources are incorporated in a specific way to facilitate the use of Han's subset entropy inequality. It is worth mentioning that, when the distortion constraints have a hierarchical structure, one can exploit remote sources in a different manner as shown in [12]. However, neither the method in [12] not the one in the present work seems powerful enough to yield conceptually satisfactory single-letter lower bounds on weighted sum rates subject to general distortion constraints. Establishing such bounds requires a deeper investigation and is beyond the scope of this work.

APPENDIX C
PROOF OF LEMMA 1

Our proof is based on a natural extension of an interesting method developed in [27] (see [27, p. 7, Example 1]).

Let $Y(X')$ and $\tilde{Y}(X')$ be the channel outputs induced by input X' via $p_{Y|X}$ and $p_{\tilde{Y}|X}$, respectively. Moreover, let p_{X^*} be an optimal solution to the following optimization problem

$$\max_{p_{X'}} -\alpha H(Y(X')) + \tilde{\alpha} H(\tilde{Y}(X')).$$

Note that, for any $p_{U|X} : \mathcal{X} \to \mathcal{U}$, we have

$$-\alpha H(Y|U) + \tilde{\alpha} H(\tilde{Y}|U)$$
$$\leq \sup_{u \in \mathcal{U}} -\alpha H(Y|U=u) + \tilde{\alpha} H(\tilde{Y}|U=u)$$
$$\leq -\alpha H(Y^*) + \tilde{\alpha} H(\tilde{Y}^*), \tag{47}$$

where $Y^* = Y(X^*)$ and $\tilde{Y}^* = \tilde{Y}(X^*)$.

Since $p_{Y|X}$ and $p_{\tilde{Y}|X}$ are circularly symmetric, there exist bijective functions $\mu : \mathcal{Y} \to \mathcal{Y}$ and $\tilde{\mu} : \tilde{\mathcal{Y}} \to \tilde{\mathcal{Y}}$ such that $\mu^M(y) = y$, $p_{Y|X}(\mu^x(y)|x) = p_{Y|X}(y|0)$, $\tilde{\mu}^M(\tilde{y}) = \tilde{y}$, and $p_{\tilde{Y}|X}(\tilde{\mu}^x(\tilde{y})|x) = p_{\tilde{Y}|X}(\tilde{y}|0)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $(x, \tilde{y}) \in \mathcal{X} \times \tilde{\mathcal{Y}}$. Define distributions $p_{X^{(k)}}$ over $\mathcal{X}$, $k = 0, 1, \ldots, M - 1$, such that $p_{X^{(k)}}(x \oplus_M k) = p_{X^*}(x)$ for all $x \in \mathcal{X}$. Let $Y^{(k)}$ and $\tilde{Y}^{(k)}$ be the channel outputs

induced by input $X^{(k)}$ via $p_{Y|X}$ and $p_{\tilde{Y}|X}$, respectively, $k = 0, 1, \ldots, M - 1$. Note that

$$p_{Y^{(k)}}(\mu^k(y)) = \sum_{x \in \mathcal{X}} p_{X^{(k)}}(x) p_{Y|X}(\mu^k(y)|x)$$
$$= \sum_{x \in \mathcal{X}} p_{X^{(k)}}(x \oplus_M k) p_{Y|X}(\mu^k(y)|x \oplus_M k)$$
$$= \sum_{x \in \mathcal{X}} p_{X^*}(x) p_{Y|X}(y|x)$$
$$= p_{Y^*}(y)$$

for all $y \in \mathcal{Y}$. Similarly,

$$p_{\tilde{Y}^{(k)}}(\tilde{\mu}^k(\tilde{y})) = p_{\tilde{Y}^*}(\tilde{y})$$

for all $\tilde{y} \in \tilde{\mathcal{Y}}$. Since μ and $\tilde{\mu}$ are bijective functions, it follows that

$$H(Y^{(k)}) = H(Y^*), \tag{48}$$
$$H(\tilde{Y}^{(k)}) = H(\tilde{Y}^*) \tag{49}$$

for $k = 0, 1, \ldots, M - 1$. Now choose $\mathcal{U} = \{0, 1, \ldots, M - 1\}$ and set $p_{U^*}(u) = \frac{1}{M}$ for all $u \in \mathcal{U}$. Moreover, let $p_{X|U^*}(x|u) = p_{X^{(u)}}(x)$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$. One can readily verify that this construction preserves the uniform distribution of X and the induced $p_{U^*|X}(u|x)$ depends on (x, u) only through $u \ominus_M x$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$. Furthermore, we have

$$-\alpha H(Y|U^*) + \tilde{\alpha} H(\tilde{Y}|U^*)$$
$$= \frac{1}{M} \sum_{u \in \mathcal{U}} [-\alpha H(Y|U^*=u) + \tilde{\alpha} H(\tilde{Y}|U^*=u)]$$
$$= \frac{1}{M} \sum_{u \in \mathcal{U}} [-\alpha H(Y^{(u)}) + \tilde{\alpha} H(\tilde{Y}^{(u)})]$$
$$= -\alpha H(Y^*) + \tilde{\alpha} H(\tilde{Y}^*), \tag{50}$$

where (50) is due to (48) and (49). In view of (47) and (50), $p_{U^*|X}$ must be a maximizer of (12). This completes the proof of Lemma 1.

Remark: The proof in fact implies that maximum value of (12) remains the same even if one has to the freedom to optimize over $p_{X,U}$.

APPENDIX D
PROOF OF LEMMA 2

Since $-H(U|Y) + H(\tilde{U}|Y)$ depends on $p_{U,\tilde{U}|X}$ only through $p_{U|X}$ and $p_{\tilde{U}|X}$, there is no loss of generality in assuming that $p_{\tilde{U}|X}$ is physically degraded with respect to $p_{U|X}$. As a consequence,

$$H(U, \tilde{U}|Y) = H(U|Y) + H(\tilde{U}|U).$$

On the other hand,

$$H(U, \tilde{U}|Y) = H(\tilde{U}|Y) + H(U|Y, \tilde{U}).$$

Therefore, we have

$$-H(U|Y) + H(\tilde{U}|Y) = -H(U|Y, \tilde{U}) + H(\tilde{U}|U),$$

which implies

$$-\alpha H(U|Y) + \tilde{\alpha} H(\tilde{U}|Y)$$
$$= -\alpha[H(U|Y, \tilde{U}) - H(\tilde{U}|U)] - (\alpha - \tilde{\alpha}) H(\tilde{U}|Y).$$

In view of the fact that $H(\tilde{U}|U)$ does not depend on $p_{Y|X}$ and the fact that $H(U|Y, \tilde{U})$ and $H(\tilde{U}|Y)$ are concave functions of $p_{Y|X}$, one can readily show that $-\alpha H(U|Y) + \tilde{\alpha} H(\tilde{U}|Y)$ is a convex function of $p_{Y|X}$.

Let $p_{Y^*|X}$ be a minimizer of (13). Define $p_{Y^{(k)}|X} : \mathcal{X} \to \mathcal{Y}$, $k = 0, 1, \ldots, M - 1$, such that $p_{Y^{(k)}|X}(\mu^k(y)|x \oplus_M k) = p_{Y^*|X}(y|x)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. It is easy to see that

$$p_{X,U,Y^{(k)}}(x \oplus_M k, u \oplus_M k, \mu^k(y))$$
$$= p_{X,U}(x \oplus_M k, u \oplus_M k) p_{Y^{(k)}|X}(\mu^k(y)|x \oplus_M k)$$
$$= p_{X,U}(x, u) p_{Y^*|X}(y|x)$$
$$= p_{X,U,Y^*}(x, u, y)$$

for all $(x, u, y) \in \mathcal{X} \times \mathcal{U} \times \mathcal{Y}$ and $k = 0, 1, \ldots, M - 1$, which, together with the fact that μ is a bijective function, implies

$$H(U|Y^*) = H(U|Y^{(k)}) \tag{51}$$

for $k = 0, 1, \ldots, M - 1$. Similarly, we have

$$H(\tilde{U}|Y^*) = H(\tilde{U}|Y^{(k)}) \tag{52}$$

for $k = 0, 1, \ldots, M - 1$.

Now define $p_{\bar{Y}|X} : \mathcal{X} \to \mathcal{Y}$ such that

$$p_{\bar{Y}|X}(y|x) = \frac{1}{M} \sum_{k=0}^{M-1} p_{Y^{(k)}|X}(y|x)$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Note that

$$p_{\bar{Y}|X}(\mu^x(y)|x) = \frac{1}{M} \sum_{k=0}^{M-1} p_{Y^{(k)}|X}(\mu^x(y)|x)$$
$$= \frac{1}{M} \sum_{k=0}^{M-1} p_{Y^*|X}(\mu^{x \ominus_M k}(y)|x \ominus_M k)$$
$$= \frac{1}{M} \sum_{k=0}^{M-1} p_{Y^*|X}(\mu^k(y)|k),$$

which does not depend on x. Therefore,

$$p_{\bar{Y}|X}(\mu^x(y)|x) = p_{\bar{Y}|X}(y|0)$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Moreover, we have

$$\mathbb{E}[m(X, \bar{Y})] = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_X(x) p_{\bar{Y}|X}(y|x) m(x, y)$$
$$= \frac{1}{M} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \sum_{k=0}^{M-1} p_X(x) p_{Y^{(k)}|X}(y|x) m(x, y)$$
$$= \frac{1}{M} \sum_{k=0}^{M-1} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_X(x) p_{Y^{(k)}|X}(y|x) m(x, y)$$
$$= \frac{1}{M} \sum_{k=0}^{M-1} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_X(x \oplus_M k)$$
$$\quad \times p_{Y^{(k)}|X}(\mu^k(y)|x \oplus k) m(x \oplus_M k, \mu^k(y))$$
$$= \frac{1}{M} \sum_{k=0}^{M-1} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p_X(x) p_{Y^*|X}(y|x) m(x, y)$$
$$= \mathbb{E}[m(X, Y^*)]$$
$$\leq d.$$

Since $-\alpha H(U|Y) + \tilde{\alpha} H(\tilde{U}|Y)$ is a convex function of $p_{Y|X}$, it follows that

$$-\alpha H(U|\bar{Y}) + \tilde{\alpha} H(\tilde{U}|\bar{Y})$$
$$\leq \frac{1}{M} \sum_{k=0}^{M-1} [-\alpha H(U|Y^{(k)}) + \tilde{\alpha} H(\tilde{U}|Y^{(k)})]$$
$$= -\alpha H(U|Y^*) + \tilde{\alpha} H(\tilde{U}|Y^*), \tag{53}$$

where (53) is due to (51) and (52). Therefore, $p_{\bar{Y}|X}$ must also be a minimizer of (13). This completes the proof of Lemma 2.

APPENDIX E
PROOF OF THEOREM 2

The following result due to von Neumann [28] (see also [29, Th. 2]) plays a crucial role in establishing the minimax theorems in this paper.

Lemma 4: Let $\mathcal{X}$ and $\mathcal{Y}$ be two bounded closed convex sets in the Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and $\mathcal{X} \times \mathcal{Y}$ be their Cartesian product in $\mathbb{R}^{m+n}$. Let $\mathcal{U}$ and $\mathcal{V}$ be two closed subsets of $\mathcal{X} \times \mathcal{Y}$ such that for any $x \in \mathcal{X}$ the set $\{y \in \mathcal{Y} : (x, y) \in \mathcal{U}\}$ is non-empty, closed, and convex, and such that for any $y \in \mathcal{Y}$ the set $\{x \in \mathcal{X} : (x, y) \in \mathcal{V}\}$ is non-empty, closed, and convex. Under these assumptions, $\mathcal{U}$ and $\mathcal{V}$ have a common point.

The next result is a direct consequence of Lemma 4.

Lemma 5: Let $f(x, y)$ be a continuous real-valued function defined for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are two bounded closed convex sets in the Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. If for any $x \in \mathcal{X}$ the set $\arg\min_{y \in \mathcal{Y}} f(x, y)$ is non-empty, closed, and convex, and for any $y \in \mathcal{Y}$ the set $\arg\max_{x \in \mathcal{X}} f(x, y)$ is non-empty, closed, and convex, then there exists a saddle point (x^*, y^*) in the sense that

$$x^* \in \arg\max_{x \in \mathcal{X}} f(x, y^*), \tag{54}$$
$$y^* \in \arg\min_{y \in \mathcal{Y}} f(x^*, y). \tag{55}$$

Proof: Let $\mathcal{U} = \{(x, y) : x \in \mathcal{X}, y \in \mathcal{U}(x)\}$ and $\mathcal{V} = \{(x, y) : x \in \mathcal{V}(y), y \in \mathcal{Y}\}$, where $\mathcal{U}(x) = \arg\min_{y \in \mathcal{Y}} f(x, y)$ and $\mathcal{V}(y) = \arg\max_{x \in \mathcal{X}} f(x, y)$. Consider a Cauchy sequence $(x^{(n)}, y^{(n)})$, $n = 1, 2 \ldots$, with $(x^{(n)}, y^{(n)}) \in \mathcal{U}$ for every n. Denote the limit of this sequence by (x, y). Note that

$$f(x, y) = \lim_{n \to \infty} f(x^{(n)}, y^{(n)})$$
$$\leq \lim_{n \to \infty} f(x^{(n)}, y(x))$$
$$= f(x, y(x)),$$

where $y(x) \in \mathcal{U}(x)$. Hence, we must have $y \in \mathcal{U}(x)$, which implies $(x, y) \in \mathcal{U}$. This proves that $\mathcal{U}$ is closed. Similarly, it can be proved that $\mathcal{V}$ is also closed. Now it follows from Lemma 4 that $\mathcal{U} \cap \mathcal{V} \neq \emptyset$, i.e., there exists (x^*, y^*) such that (54) and (55) are satisfied. ∎

Now we proceed to prove Theorem 2. Note that the maximization problem

$$\max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta})$$

can be decomposed into the following sub-problems

$$\max_{q_0 \in [0,\frac{1}{2}]} \alpha_1 (1 - \delta_1) H_b(q_0), \tag{56}$$

$$\max_{q_k \in [0,\frac{1}{2}]} [-\alpha_k (1 - \delta_k) + \alpha_{k+1}(1 - \delta_{k+1})] H_b(q_k),$$
$$k = 1, \ldots, L - 1, \tag{57}$$

$$\max_{q_L \in [0,\frac{1}{2}]} -\alpha_L (1 - \delta_L) H_b(q_L). \tag{58}$$

It is clear that the maximizers of (56), (57), (58) are, respectively, given by

$$q_0 = \begin{cases} 0, & \alpha_1(1 - \delta_1) < 0 \\ \text{any number in } [0, \frac{1}{2}], & \alpha_1(1 - \delta_1) = 0 \\ \frac{1}{2}, & \alpha_1(1 - \delta_1) > 0, \end{cases} \tag{59}$$

$$q_k = \begin{cases} 0, & \alpha_k(1 - \delta_k) > \alpha_{k+1}(1 - \delta_{k+1}) \\ \text{any number in } [0, \frac{1}{2}], & \\ & \alpha_k(1 - \delta_k) = \alpha_{k+1}(1 - \delta_{k+1}) \\ \frac{1}{2}, & \alpha_k(1 - \delta_k) < \alpha_{k+1}(1 - \delta_{k+1}), \\ & k = 1, \ldots, L - 1, \end{cases} \tag{60}$$

$$q_L = \begin{cases} 0, & \alpha_L(1 - \delta_L) > 0 \\ \text{any number in } [0, \frac{1}{2}], & \alpha_L(1 - \delta_L) = 0 \\ \frac{1}{2}, & \alpha_L(1 - \delta_L) < 0. \end{cases} \tag{61}$$

Similarly, the minimization problem

$$\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta})$$

can be decomposed into the following sub-problems

$$\min_{\delta_k \in [0, d_k]} \alpha_k(1 - \delta_k)[H_b(q_{k-1}) - H_b(q_k)], \quad k = 1, \ldots, L,$$

and the corresponding minimizers are given by

$$\delta_k = \begin{cases} 0, & \alpha_k[H_b(q_{k-1}) - H_b(q_k)] < 0 \\ \text{any number in } [0, d_k], & \\ & \alpha_k[H_b(q_{k-1}) - H_b(q_k)] = 0 \\ d_k, & \alpha_k[H_b(q_{k-1}) - H_b(q_k)] > 0, \\ & k = 1, \ldots, L. \end{cases} \tag{62}$$

According to (62), the set $\arg\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta})$ is non-empty, closed, and convex for every $\underline{q} \in [0, \frac{1}{2}]^{L+1}$. Moreover, according to (59)-(61), the set $\arg\max_{\underline{q} \in [0,\frac{1}{2}]^{L+1}} \kappa(\underline{\alpha}, \underline{q}, \underline{\delta})$ is non-empty, closed, and convex for every $\underline{\delta} \in \mathcal{D}(\underline{d})$. Therefore, it follows from Lemma 5 that there exists $(\underline{q}^*, \underline{\delta}^*)$ such that (14) and (15) are satisfied.

Now consider the case $\alpha_1 \geq \cdots \geq \alpha_L \geq 0$. Let $(\underline{q}^*, \underline{\delta}^*)$ be an arbitrary saddle point solution. If $q_{k-1}^* < q_k^*$ for some k, then, in light of (60)-(62), one of the following must be true:

- $\alpha_k = \alpha_{k+1}$ and $\delta_k^* = \delta_{k+1}^* = 0$;
- $\alpha_k = 0$.

Moreover, if $q_{k-1}^* > q_k^*$ and $d_k = 1$ for some k, then, in light of (60)-(62), one of the following must be true:

- $\delta_k^* = \delta_{k+1}^* = 1$;
- $\alpha_{k+1} = 0$ and $\delta_k^* = 1$;

- $k = L$ and $\delta_L^* = 1$;
- $\alpha_k = 0$.

Based on these observations, one can readily show via induction that $(\tilde{\underline{q}}^*, \underline{\delta}^*)$ with $\tilde{\underline{q}}^* = (\tilde{q}_0^*, \tilde{q}_1^*, \ldots, \tilde{q}_L^*)$ is also a saddle-point solution, where $\tilde{q}_0^* = q_0^*$ and

$$\tilde{q}_k^* = \begin{cases} \tilde{q}_{k-1}^*, & \tilde{q}_{k-1}^* < q_k^* \\ \tilde{q}_{k-1}^*, & \tilde{q}_{k-1}^* > q_k^* \text{ and } d_k = 1 \\ q_k^* & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, L$. Note that by construction we have $\tilde{q}_0^* \geq \tilde{q}_1^* \geq \ldots \geq \tilde{q}_L^*$ and $\tilde{q}_{k-1}^* = \tilde{q}_k^*$ whenever $d_k = 1, k = 1, \ldots, L$. In view of (60), it is clear that, if $\delta_k^* < \delta_{k+1}^*$ for some k, then either $\tilde{q}_k^* = 0$ or $\alpha_k = 0$. Based on this observation, one can show via a simple induction that $(\tilde{\underline{q}}^*, \tilde{\underline{\delta}}^*)$ with $\tilde{\underline{\delta}}^* = (\tilde{\delta}_1^*, \ldots, \tilde{\delta}_L^*)$ is still a saddle-point solution, where $\tilde{\delta}_1^* = \delta_1^*$ and

$$\tilde{\delta}_{k+1}^* = \begin{cases} \tilde{\delta}_k^*, & \tilde{\delta}_k^* < \delta_{k+1}^* \\ \delta_{k+1}^*, & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, L - 1$. Note that by construction we have $\tilde{\delta}_1^* \geq \ldots \geq \tilde{\delta}_L^*$. In the case $\alpha_1 > \ldots > \alpha_L > 0$, if $q_{k-1}^* < q_k^*$ for some k, then it follows by (62) that $\delta_k^* = 0$, which, in view of (60) and (61), further implies $q_k^* = 0$, leading to a contradiction with the assumption $q_{k-1}^* < q_k^*$; therefore, we must have $q_0^* \geq q_1^* \geq \cdots \geq q_L^*$. This completes the proof of Theorem 2.

APPENDIX F
PROOF OF THEOREM 4

We first prove the following technical result. Define

$$g(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}) = -\alpha(1 - 2\delta) \log\left(\frac{1 - \delta}{\delta}\right)$$
$$+ \tilde{\alpha}(1 - 2\tilde{\delta}) \log\left(\frac{1 - \tilde{\delta}}{\tilde{\delta}}\right),$$

$$\tilde{g}(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}) = -\alpha(1 - 2\delta)^2 + \tilde{\alpha}(1 - 2\tilde{\delta})^2,$$

where $\alpha \geq \tilde{\alpha} \geq 0$, $\delta \in [0, \frac{1}{2}]$, and $\tilde{\delta} \in [0, \frac{1}{2}]$. When no confusion can arise, we will simply write $g(\alpha, \tilde{\alpha}, \delta, \tilde{\delta})$ and $\tilde{g}(\alpha, \tilde{\alpha}, \delta, \tilde{\delta})$ as g and $\tilde{g}$, respectively.

Lemma 6: The maximizer of the following optimization problem

$$\max_{q \in [0,\frac{1}{2}]} -\alpha H_b(q \odot \delta) + \tilde{\alpha} H_b(q \odot \tilde{\delta}),$$

where $\alpha \geq \tilde{\alpha} \geq 0$, $\delta \in [0, \frac{1}{2}]$, and $\tilde{\delta} \in [0, \frac{1}{2}]$, is given by

$$q = \begin{cases} 0, & g \leq 0 \text{ and } \tilde{g} < 0 \\ q^\dagger, & g > 0 \text{ and } \tilde{g} < 0 \\ \frac{1}{2}, & g > 0 \text{ and } \tilde{g} \geq 0 \\ \text{any number in } [0, \frac{1}{2}], & g = \tilde{g} = 0 \end{cases}$$

with $q^\dagger$ being the unique solution of

$$\alpha(1 - 2\delta) \log\left(\frac{1 - q \odot \delta}{q \odot \delta}\right) = \tilde{\alpha}(1 - 2\tilde{\delta}) \log\left(\frac{1 - q \odot \tilde{\delta}}{q \odot \tilde{\delta}}\right)$$

for $q \in (0, \frac{1}{2})$.

Proof: One can readily verify Lemma 6 for the following degenerate cases: 1) $\tilde{\alpha} = 0$, 2) $\delta = \tilde{\delta}$, 3) $\delta = \frac{1}{2}$, 4) $\tilde{\delta} = \frac{1}{2}$. Therefore, it suffices to consider the case where $\alpha \geq \tilde{\alpha} > 0$, $\delta \in [0, \frac{1}{2})$, $\tilde{\delta} \in [0, \frac{1}{2})$, and $\delta \neq \tilde{\delta}$.

Let $\varsigma(q) = -\alpha H_b(q \odot \delta) + \tilde{\alpha} H_b(q \odot \tilde{\delta})$. Note that

$$\varsigma'(q) = -\alpha(1 - 2\delta) \log \left(\frac{1 - q \odot \delta}{q \odot \delta} \right)$$
$$+ \tilde{\alpha}(1 - 2\tilde{\delta}) \log \left(\frac{1 - q \odot \tilde{\delta}}{q \odot \tilde{\delta}} \right),$$

$$\varsigma''(q) = \frac{\alpha(1 - 2\delta)^2}{(q \odot \delta)(1 - q \odot \delta)} - \frac{\tilde{\alpha}(1 - 2\tilde{\delta})^2}{(q \odot \tilde{\delta})(1 - q \odot \tilde{\delta})}$$

for $q \in [0, \frac{1}{2}]$; in particular,

$$\varsigma'(0) = g(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}),$$
$$\varsigma''(\frac{1}{2}) = -4\tilde{g}(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}).$$

We shall prove the following statements:

a) $\varsigma'(q) = 0$ has no more than one solution in $[0, \frac{1}{2})$;
b) $\varsigma''(q) = 0$ cannot have two different solutions in $[0, \frac{1}{2}]$.

In view of Rolle's theorem and the fact that $\varsigma'(\frac{1}{2}) = 0$, it suffices to prove statement b). Note that

$$\varsigma''(q) = 0 \iff (\alpha - \tilde{\alpha})(1 - 2\delta)^2(1 - 2\tilde{\delta})^2(q^2 - q)$$
$$- \alpha(1 - 2\delta)^2\tilde{\delta}(1 - \tilde{\delta}) + \tilde{\alpha}(1 - 2\tilde{\delta})^2\delta(1 - \delta) = 0$$
(63)

for $q \in [0, \frac{1}{2}]$. If $\alpha = \tilde{\alpha} > 0$, then we further have

$$\varsigma''(q) = 0 \iff \tau(\delta) = \tau(\tilde{\delta}),$$

where $\tau(x) = \frac{x(1-x)}{(1-2x)^2}$; since $\tau(x)$ is a strictly increasing function for $x \in [0, \frac{1}{2})$, it follows that $\varsigma''(q) = 0$ has no solution in $(0, \frac{1}{2})$ (under our assumption $\delta \in [0, \frac{1}{2})$, $\tilde{\delta} \in [0, \frac{1}{2})$, and $\delta \neq \tilde{\delta}$). For the case $\alpha > \tilde{\alpha} > 0$, note that the sum of the two roots of the equation in (63) is equal to one, which implies that $\varsigma''(q) = 0$ cannot have two different solutions in $[0, \frac{1}{2}]$.

Now consider the following cases.

- $g \leq 0$ *and* $\tilde{g} < 0$ (i.e., $\varsigma'(0) \leq 0$ and $\varsigma''(\frac{1}{2}) > 0$): Recall that $\varsigma'(\frac{1}{2}) = 0$. Since $\varsigma''(\frac{1}{2}) > 0$, there exists some $\epsilon \in (0, \frac{1}{2}]$ such that $\varsigma'(q) < 0$ for $q \in [\frac{1}{2} - \epsilon, \frac{1}{2})$; moreover, in view of the fact that $\varsigma'(0) \leq 0$ and the fact that $\varsigma'(q) = 0$ has no more than one solution in $[0, \frac{1}{2})$, we must have $\varsigma'(q) \leq 0$ for $q \in [0, \frac{1}{2} - \epsilon]$; therefore, $q = 0$ is the unique maximizer of $\varsigma(q)$ for $q \in [0, \frac{1}{2}]$.
- $g > 0$ *and* $\tilde{g} < 0$ (i.e., $\varsigma'(0) > 0$ and $\varsigma''(\frac{1}{2}) > 0$): In this case, $q = q^\dagger$ is the unique solution of $\varsigma'(q) = 0$ for $q \in (0, \frac{1}{2})$; furthermore, we have $\varsigma'(q) > 0$ for $q \in [0, q^\dagger)$ and $\varsigma'(q) < 0$ for $q \in (q^\dagger, \frac{1}{2})$. Therefore, $q = q^\dagger$ is the unique maximizer of $\varsigma(q)$ for $q \in [0, \frac{1}{2}]$.
- $g > 0$ *and* $\tilde{g} \geq 0$ (i.e., $\varsigma'(0) > 0$ and $\varsigma''(\frac{1}{2}) \leq 0$): Again recall that $\varsigma'(\frac{1}{2}) = 0$. If $\varsigma''(\frac{1}{2}) < 0$, then there exists some $\epsilon \in (0, \frac{1}{2}]$ such that $\varsigma'(q) > 0$ for $q \in [\frac{1}{2} - \epsilon, \frac{1}{2})$; moreover, in view of the fact that $\varsigma'(0) > 0$ and the fact that $\varsigma'(q) = 0$ has no more than one solution in $[0, \frac{1}{2})$,

we must have $\varsigma'(q) \geq 0$ for $q \in [0, \frac{1}{2} - \epsilon]$; therefore, $q = \frac{1}{2}$ is the unique maximizer of $\varsigma(q)$ for $q \in [0, \frac{1}{2}]$. For the case $\varsigma'(0) > 0$ and $\varsigma''(\frac{1}{2}) = 0$, in view of the fact that $\varsigma'(\frac{1}{2}) = 0$ and the fact that $\varsigma''(q) = 0$ cannot have two different solutions in $[0, \frac{1}{2}]$, we must have $\varsigma'(q) > 0$ for $q \in [0, \frac{1}{2})$; as a consequence, $q = \frac{1}{2}$ is again the unique maximizer of $\varsigma(q)$ for $q \in [0, \frac{1}{2}]$.
- $g \leq 0$ *and* $\tilde{g} \geq 0$: We shall show that this case does not exist under our assumption $\alpha \geq \tilde{\alpha} > 0$, $\delta \in [0, \frac{1}{2})$, $\tilde{\delta} \in [0, \frac{1}{2})$, and $\delta \neq \tilde{\delta}$. Since $g \leq 0$ and $\tilde{g} \geq 0$, it follows that $\rho(\delta) \geq \rho(\tilde{\delta})$, where

$$\rho(x) = \frac{1}{1 - 2x} \log \left(\frac{1 - x}{x} \right).$$

It can be verified that

$$\rho'(x) = \frac{1}{(1 - 2x)^2} \xi(x),$$

where

$$\xi(x) = 2 \log \left(\frac{1 - x}{x} \right) - \frac{1 - 2x}{x(1 - x)}.$$

Moreover,

$$\xi'(x) = -\frac{2}{x(1 - x)} + \frac{2x(1 - x) + (1 - 2x)^2}{x^2(1 - x)^2}$$
$$= \frac{(1 - 2x)^2}{x^2(1 - x)^2} > 0$$

for $x \in (0, \frac{1}{2})$. Therefore, we have $\xi(x) < \xi(\frac{1}{2}) = 0$ for $x \in (0, \frac{1}{2})$, which implies $\rho'(x) < 0$ for $x \in (0, \frac{1}{2})$ and further implies that $\rho(x)$ a strictly decreasing function for $x \in [0, \frac{1}{2})$. As a consequence, $\rho(\delta) \geq \rho(\tilde{\delta})$ if and only if $\delta \leq \tilde{\delta}$. On the other hand, $\tilde{g} \geq 0$ implies $\delta \geq \tilde{\delta}$. Hence, we must have $\delta = \tilde{\delta}$. However, this case is excluded by our assumption $\delta \neq \tilde{\delta}$.

This completes the proof of Lemma 6. ∎

Now we proceed to prove Theorem 4. Note that the maximization problem

$$\max_{\underline{q} \in [0, \frac{1}{2}]^{L+1}} \eta(\underline{\alpha}, \underline{q}, \underline{\delta})$$

can be decomposed into the following sub-problems

$$\max_{q_0 \in [0, \frac{1}{2}]} \alpha_1 H_b(q_0 \odot \delta_1), \quad (64)$$

$$\max_{q_k \in [0, \frac{1}{2}]} -\alpha_k H_b(q_k \odot \delta_k) + \alpha_{k+1} H_b(q_k \odot \delta_{k+1}),$$
$$k = 1, \ldots, L - 1, \quad (65)$$

$$\max_{q_L \in [0, \frac{1}{2}]} -\alpha_L H_b(q_L \odot \delta_L). \quad (66)$$

It is clear that the maximizers of (64) and (66) are, respectively, given by

$$q_0 = \begin{cases} \frac{1}{2}, & \alpha_1 > 0 \text{ and } \delta_1 \in [0, \frac{1}{2}) \\ \text{any number in } [0, \frac{1}{2}], & \alpha_1 = 0 \text{ or } \delta_1 = \frac{1}{2}, \end{cases}$$

$$q_L = \begin{cases} 0, & \alpha_L > 0 \text{ and } \delta_L \in [0, \frac{1}{2}) \\ \text{any number in } [0, \frac{1}{2}], & \alpha_L = 0 \text{ or } \delta_L = \frac{1}{2}. \end{cases}$$

Moreover, it follows from Lemma 6 that the maximizers of (65) are given by

$$
q_k = \begin{cases}
0, & g_k \leq 0 \text{ and } \tilde{g}_k < 0 \\
q_k^\dagger, & g_k > 0 \text{ and } \tilde{g}_k < 0 \\
1, & g_k > 0 \text{ and } \tilde{g}_k \geq 0 \\
\frac{1}{2}, & \\
\text{any number in } [0, \tfrac{1}{2}], & g_k = \tilde{g}_k = 0,
\end{cases}
$$

where $g_k = g(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})$, $\tilde{g}_k = \tilde{g}(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})$, and $q_k^\dagger$ is the unique solution of

$$
\alpha_k(1 - 2\delta) \log\left(\frac{1 - q \odot \delta_k}{q \odot \delta_k}\right)
$$
$$
= \alpha_{k+1}(1 - 2\delta_{k+1}) \log\left(\frac{1 - q \odot \delta_{k+1}}{q \odot \delta_{k+1}}\right)
$$

for $q \in (0, \frac{1}{2})$, $k = 1, \ldots, L - 1$. Similarly, the minimization problem

$$
\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \eta(\underline{\alpha}, \underline{q}, \underline{\delta})
$$

can be decomposed into the following sub-problems

$$
\min_{\delta_k \in [0, d_k]} \alpha_k[H_b(q_{k-1} \odot \delta_k) - H_b(q_k \odot \delta_k)], \quad k = 1, \ldots, L.
\tag{67}
$$

It is easy to verify that the minimizers of (67) are given by

$$
\delta_k = \begin{cases}
0, & \alpha_k > 0 \text{ and } q_{k-1} < q_k \\
d_k, & \alpha_k > 0 \text{ and } q_{k-1} > q_k \\
\text{any number in } [0, d_k], & \alpha_k = 0 \text{ or } q_{k-1} = q_k, \\
& k = 1, \ldots, L.
\end{cases}
$$

The rest of the proof is almost identical with its counterpart in the proof of Theorem 2 (see the steps after Equation (62)) and thus is omitted.

Appendix G
Proof of Theorem 6

The following technical lemma is needed for the proof of Theorem 6. Define

$$
h(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}) = -(\alpha - \tilde{\alpha})(\lambda - \delta)(\lambda - \tilde{\delta}),
$$
$$
\tilde{h}(\alpha, \tilde{\alpha}, \delta, \tilde{\delta}) = -\alpha(\lambda - \delta)\tilde{\delta} + \tilde{\alpha}(\lambda - \delta)\delta,
$$

where $\alpha \geq \tilde{\alpha} \geq 0$, $\delta \in [0, \lambda]$, and $\tilde{\delta} \in [0, \lambda]$. When no confusion can arise, we will simply write $h(\alpha, \tilde{\alpha}, \delta, \tilde{\delta})$ and $\tilde{h}(\alpha, \tilde{\alpha}, \delta, \tilde{\delta})$ as h and $\tilde{h}$, respectively.

Lemma 7: The maximizer of the following optimization problem

$$
\max_{\theta \in [0, \lambda]} -\alpha \log(\lambda\theta + \lambda\delta - \theta\delta) + \tilde{\alpha} \log(\lambda\theta + \lambda\tilde{\delta} - \theta\tilde{\delta}),
$$

where $\alpha \geq \tilde{\alpha} \geq 0$, $\delta \in [0, \lambda]$, and $\tilde{\delta} \in [0, \lambda]$, is given by

$$
\theta = \begin{cases}
0, & h \leq 0, \tilde{h} \leq 0, \text{ and } h + \tilde{h} < 0 \\
\min\left\{-\frac{\tilde{h}}{h}\lambda, \lambda\right\}, & h < 0 \text{ and } \tilde{h} > 0 \\
\lambda, & h = 0 \text{ and } \tilde{h} > 0 \\
\text{any number in } [0, \lambda], & h = \tilde{h} = 0.
\end{cases}
$$

Proof: Note that

$$
\varrho'(\theta) = \frac{h\theta + \tilde{h}\lambda}{(\lambda\theta + \lambda\delta - \theta\delta)(\lambda\theta + \lambda\tilde{\delta} - \theta\tilde{\delta})}
$$

for $\theta > 0$, where

$$
\varrho(\theta) = -\alpha \log(\lambda\theta + \lambda\delta - \theta\delta) + \tilde{\alpha} \log(\lambda\theta + \lambda\tilde{\delta} - \theta\tilde{\delta}).
$$

Now consider the following cases.

- $h \leq 0$, $\tilde{h} \leq 0$, and $h + \tilde{h} < 0$: In this case, $\varrho(\theta)$ is a strictly decreasing function for $\theta \geq 0$; consequently, $\theta = 0$ is the unique maximizer of $\varrho(\theta)$ for $\theta \in [0, \lambda]$.
- $h < 0$ and $\tilde{h} > 0$: In this case, $\varrho(\theta)$ is a strictly increasing function for $\theta \in [0, -\frac{\tilde{h}}{h}\lambda]$ and a strictly decreasing function for $\theta > -\frac{\tilde{h}}{h}\lambda$; consequently, $\theta = \min\{-\frac{\tilde{h}}{h}\lambda, \lambda\}$ is the unique maximizer of $\varrho(\theta)$ for $\theta \in [0, \lambda]$.
- $h = 0$ and $\tilde{h} > 0$: In this case, $\varrho(\theta)$ is a strictly increasing function for $\theta \geq 0$; consequently, $\theta = \lambda$ is the unique maximizer of $\varrho(\theta)$ for $\theta \in [0, \lambda]$.
- $h = \tilde{h} = 0$: In this case, $\varrho(\theta)$ is a constant. ∎

This completes the proof of Lemma 7.

Now we proceed to prove Theorem 6. The main difficulty here is that $\omega(\underline{\alpha}, \cdot, \cdot)$ is not continuous at certain boundary points[8]; as a consequence, Lemma 5 in Appendix E is not applicable anymore. So we shall instead rely on Lemma 4 in Appendix E.

Note that the maximization problem

$$
\max_{\underline{\theta} \in [0, \lambda]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta})
$$

can be decomposed into the following sub-problems

$$
\max_{\theta_0 \in [0, \lambda]} \frac{\alpha_1}{2} \log(\lambda\theta_0 + \lambda\delta_1 - \theta_0\delta_1),
\tag{68}
$$
$$
\max_{\theta_k \in [0, \lambda]} -\frac{\alpha_k}{2} \log(\lambda\theta_k + \lambda\delta_k - \theta_k\delta_k)
$$
$$
+ \frac{\alpha_{k+1}}{2} \log(\lambda\theta_k + \lambda\delta_{k+1} - \theta_k\delta_{k+1}),
$$
$$
k = 1, \ldots, L - 1, \tag{69}
$$
$$
\max_{\theta_L \in [0, \lambda]} -\frac{\alpha_L}{2} \log(\lambda\theta_L + \lambda\delta_L - \theta_L\delta_L).
\tag{70}
$$

It is clear that the maximizers of (68) and (70) are, respectively, given by

$$
\theta_0 = \begin{cases}
\lambda, & \alpha_1 > 0 \text{ and } \delta_1 \in [0, \lambda) \\
\text{any number in } [0, \lambda], & \alpha_1 = 0 \text{ or } \delta_1 = \lambda,
\end{cases}
\tag{71}
$$
$$
\theta_L = \begin{cases}
0, & \alpha_L > 0 \text{ and } \delta_L \in [0, \lambda) \\
\text{any number in } [0, \tfrac{1}{2}], & \alpha_L = 0 \text{ or } \delta_L = \lambda.
\end{cases}
\tag{72}
$$

[8]For example, consider an arbitrary $(\underline{\alpha}, \underline{\theta}, \underline{\delta})$ such that $\alpha_k > 0$ and $\theta_{k-1} = \theta_k = \delta_k = 0$ for some k. It can be shown that, for any $a \in [-\infty, \infty]$, one can find a sequence $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)})$, $n = 1, 2, \ldots$, converging to $(\underline{\theta}, \underline{\delta})$ such that $\lim_{n\to\infty} \omega(\underline{\alpha}, \underline{\theta}^{(n)}, \underline{\delta}^{(n)}) = a$. Similarly, if $\alpha_k > 0$, $\alpha_{k+1} > 0$, and $\theta_k = \delta_k = \delta_{k+1} = 0$ for some k, then, for any $a \in [-\infty, \infty]$, one can find a sequence $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)})$, $n = 1, 2, \ldots$, converging to $(\underline{\theta}, \underline{\delta})$ such that $\lim_{n\to\infty} \omega(\underline{\alpha}, \underline{\theta}^{(n)}, \underline{\delta}^{(n)}) = a$.

Moreover, it follows from Lemma 7 that the maximizers of (69) are given by

$$
\theta_k = \begin{cases}
0, & h_k \le 0, \ \tilde{h}_k \le 0, \ \text{and} \ h_k + \tilde{h}_k < 0 \\
\min\left\{-\frac{\tilde{h}_k}{h_k}\lambda, \lambda\right\}, & h_k < 0 \ \text{and} \ \tilde{h}_k > 0 \\
\lambda, & h_k = 0 \ \text{and} \ \tilde{h}_k > 0 \\
\text{any number in } [0, \lambda], & h_k = \tilde{h}_k = 0,
\end{cases}
$$
$$
k = 1, \ldots, L-1, \qquad (73)
$$

where $h_k = h(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})$ and $\tilde{h}_k = \tilde{h}(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})$. Similarly, the minimization problem

$$
\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta})
$$

can be decomposed into the following sub-problems

$$
\min_{\delta_k \in [0, d_k]} \frac{\alpha_k}{2} \log\left(\frac{\lambda\theta_{k-1} + \lambda\delta_k - \theta_{k-1}\delta_k}{\lambda\theta_k + \lambda\delta_k - \theta_k\delta_k}\right), \quad k = 1, \ldots, L. \quad (74)
$$

It is easy to verify that the minimizers of (74) are given by

$$
\delta_k = \begin{cases}
0, & \alpha_k > 0 \ \text{and} \ \theta_{k-1} < \theta_k \\
d_k, & \alpha_k > 0 \ \text{and} \ \theta_{k-1} > \theta_k \\
\text{any number in } [0, d_k], & \alpha_k = 0 \ \text{or} \ \theta_{k-1} = \theta_k,
\end{cases}
$$
$$
k = 1, \ldots, L. \qquad (75)
$$

For every $\underline{\theta} \in [0, \frac{1}{2}]^{L+1}$, define $\mathcal{U}(\underline{\theta})$ to be the set of $\underline{\delta} \in \mathcal{D}(\underline{d})$ satisfying (75); similarly, for every $\underline{\delta} \in \mathcal{D}(\underline{d})$, define $\mathcal{V}(\underline{\delta})$ to be the set of $\underline{\theta} \in [0, \frac{1}{2}]^{L+1}$ satisfying (71), (72), and (73). Furthermore, define

$$
\mathcal{U} = \{(\underline{\theta}, \underline{\delta}) : \underline{\theta} \in [0, \tfrac{1}{2}]^{L+1}, \underline{\delta} \in \mathcal{U}(\underline{\theta})\},
$$
$$
\mathcal{V} = \{(\underline{\theta}, \underline{\delta}) : \underline{\theta} \in \mathcal{V}(\underline{\delta}), \underline{\delta} \in \mathcal{D}(\underline{d})\}.
$$

Note that[9]

$$
\mathcal{U}(\underline{\theta}) \subseteq \arg\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}), \quad \underline{\theta} \in [0, \frac{1}{2}]^{L+1}, \quad (76)
$$
$$
\mathcal{V}(\underline{\delta}) \subseteq \arg\max_{\underline{\theta} \in [0, \frac{1}{2}]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}), \quad \underline{\delta} \in \mathcal{D}(\underline{d}), \quad (77)
$$

where the equality in (76) holds if

$$
\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}) > -\infty,
$$

and the equality in (77) holds if

$$
\max_{\underline{\theta} \in [0, \frac{1}{2}]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}) < \infty.
$$

In view of (75), the set $\mathcal{U}(\underline{\theta})$ is non-empty, closed, and convex for every $\underline{\theta} \in [0, \frac{1}{2}]^{L+1}$. Moreover, in view of (71)-(73), the set $\mathcal{V}(\underline{\delta})$ is non-empty, closed, and convex for every $\underline{\delta} \in \mathcal{D}(\underline{d})$. Consider a Cauchy sequence $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)})$ with $\underline{\theta}^{(n)} = (\theta_0^{(n)}, \theta_1^{(n)}, \ldots, \theta_L^{(n)})$ and $\underline{\delta}^{(n)} = (\delta_1^{(n)}, \ldots, \delta_L^{(n)})$, $n = 1, 2, \ldots$, such that $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{U}$ for every n. Denote the limit of this sequence by $(\underline{\theta}, \underline{\delta})$. Note that, for any k, if $\alpha_k = 0$

[9] For example, if $\alpha_1 \ge \ldots \ge \alpha_L > 0$, $\theta_0 > \theta_1 = 0$, and $\theta_2 > \ldots > \theta_L$, then $\arg\min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}) = \{\underline{\delta} \in \mathcal{D}(\underline{d}) : \delta_1 > 0, \delta_2 = 0\}$; in contrast, we have $\mathcal{U}(\underline{\theta}) = \{(d_1, 0, d_3, \ldots, d_L)\}$. Note that in this example $\mathcal{U}(\underline{\theta})$ is a closed set whereas $\arg\min_{\underline{\delta} \in \mathcal{D}(\underline{d})}$ is not. Similarly, if $\alpha_1 > \ldots > \alpha_L > 0$, $\delta_1 = 0$, and $\delta_k \in (0, \lambda)$, $k = 2, \ldots, L$, then $\arg\max_{\underline{\theta} \in [0, \frac{1}{2}]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}) = \{\underline{\theta} \in [0, \frac{1}{2}]^{L+1} : \theta_0 > 0, \theta_1 = 0\}$; in contrast, $\mathcal{V}(\underline{\delta})$ contains a single element. Note that in this example $\mathcal{V}(\underline{\delta})$ is a closed set whereas $\arg\max_{\underline{\theta} \in [0, \frac{1}{2}]^{L+1}} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta})$ is not.

or $\theta_{k-1} = \theta_k$, then (75) is clearly satisfied by δ_k. On the other hand, if $\alpha_k > 0$ and $\theta_{k-1} < \theta_k$ ($\theta_{k-1} > \theta_k$), then we must have $\theta_{k-1}^{(n)} < \theta_k^{(n)}$ ($\theta_{k-1}^{(n)} > \theta_k^{(n)}$) for all sufficiently large n, which, together with the fact that $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{U}$ for every n, implies $\delta_k^{(n)} = 0$ ($\delta_k^{(n)} = d_k$) for all sufficiently large n and consequently $\delta_k = 0$ ($\delta_k = d_k$); hence, (75) is still satisfied by δ_k. Therefore, we have $(\underline{\theta}, \underline{\delta}) \in \mathcal{U}$. This proves that $\mathcal{U}$ is a closed set. Similarly, let $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)})$, $n = 1, 2, \ldots$, be a Cauchy sequence with $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{V}$ for every n. Again denote the limit of this sequence by $(\underline{\theta}, \underline{\delta})$. It is easy to verify that θ_0 and θ_L satisfy (71) and (72), respectively. In order to show that (73) is satisfied by θ_k, $k = 1, \ldots, L-1$, we consider the following cases.

- $h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) \le 0$, $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) \le 0$, and they are not equal to zero at the same time: If $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) < 0$, then we have $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) < 0$ for all sufficiently large n, which, together with the fact that $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{V}$ for every n, implies $\theta_k^{(n)} = 0$ for all sufficiently large n and consequently $\theta_k = 0$. On the other hand, if $h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) < 0$ and $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) = 0$, then $h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) < 0$ for all sufficiently large n. Since $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{V}$ for every n, it follows that, when n is sufficiently large, we have $\theta_k^{(n)} = 0$ if $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) \le 0$, and

$$
\theta_k^{(n)} = \min\left\{-\frac{\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}{h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}\lambda, \lambda\right\}
$$

if $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) > 0$. This, together with the fact that

$$
\lim_{n \to \infty} h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) < 0,
$$
$$
\lim_{n \to \infty} \tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) = 0,
$$

implies $\theta_k = 0$. Hence, (73) is satisfied in this case.

- $h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) < 0$ and $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) > 0$: We must have $h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) < 0$ and $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) > 0$ for all sufficiently large n, which, together with the fact that $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{V}$ for every n, implies

$$
\theta_k^{(n)} = \min\left\{-\frac{\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}{h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}\lambda, \lambda\right\}
$$

for all sufficiently large n. Hence,

$$
\theta_k = \lim_{n \to \infty} \theta_k^{(n)} = \min\left\{-\frac{\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})}{h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1})}\lambda, \lambda\right\},
$$

which satisfies (73).

- $h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) = 0$ and $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) > 0$: In this case $\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) > 0$ for all sufficiently large n. Recall that $(\underline{\theta}^{(n)}, \underline{\delta}^{(n)}) \in \mathcal{V}$ for every n. Therefore, when n is sufficiently large, we must have

$$
\theta_k^{(n)} = \min\left\{-\frac{\tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}{h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)})}\lambda, \lambda\right\}
$$

if $h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) < 0$, and $\theta_k^{(n)} = \lambda$ if $h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) = 0$. This, together with the fact that

$$\lim_{n \to \infty} h_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) = 0,$$
$$\lim_{n \to \infty} \tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k^{(n)}, \delta_{k+1}^{(n)}) > 0,$$

implies $\theta_k = \lambda$. Hence, (73) is satisfied in this case.

• $h_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) = \tilde{h}_k(\alpha_k, \alpha_{k+1}, \delta_k, \delta_{k+1}) = 0$: This case is trivial.

This proves $(\underline{\theta}, \underline{\delta}) \in \mathcal{V}$, which further implies that $\mathcal{V}$ is a closed set. Therefore, it follows from Lemma 4 that $\mathcal{U} \cap \mathcal{V} \neq \emptyset$; as a consequence, there exists $(\underline{\theta}^*, \underline{\delta}^*)$ such that (32) and (33) are satisfied.

Now consider the case $\alpha_1 \geq \ldots \geq \alpha_L \geq 0$. Let $(\underline{\theta}^*, \underline{\delta}^*)$ be an arbitrary saddle point solution. It is clear that $\omega(\underline{\alpha}, \underline{\theta}^*, \underline{\delta}^*) \in (-\infty, \infty)$. Therefore, we must have $(\underline{\theta}^*, \underline{\delta}^*) \in \mathcal{U} \cap \mathcal{V}$. Define $\underline{\tilde{\theta}}^* = (\tilde{\theta}_0^*, \tilde{\theta}_1^*, \ldots, \tilde{\theta}_L^*)$, where $\tilde{\theta}_0^* = \theta_0^*$ and

$$\tilde{\theta}_k^* = \begin{cases} \tilde{\theta}_{k-1}^*, & \tilde{\theta}_{k-1}^* < \theta_k^* \\ \tilde{\theta}_{k-1}^*, & \tilde{\theta}_{k-1}^* > \theta_k^* \text{ and } d_k = \lambda \\ \theta_k^* & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, L$; note that by construction we have $\tilde{\theta}_0^* \geq \tilde{\theta}_1^* \geq \ldots \geq \tilde{\theta}_L^*$ and $\tilde{\theta}_{k-1}^* = \tilde{\theta}_k^*$ whenever $d_k = \lambda$, $k = 1, \ldots, L$. Moreover, define $\underline{\tilde{\delta}}^* = (\delta_1^*, \ldots, \delta_L^*)$, where $\tilde{\delta}_1^* = \delta_1^*$ and

$$\tilde{\delta}_{k+1}^* = \begin{cases} \tilde{\delta}_k^*, & \tilde{\delta}_k^* < \delta_{k+1}^* \\ \delta_{k+1}^*, & \text{otherwise} \end{cases}$$

for $k = 1, \ldots, L-1$; note that by construction we have $\tilde{\delta}_1^* \geq \ldots \geq \tilde{\delta}_L^*$. It can be shown that $(\underline{\tilde{\theta}}^*, \underline{\tilde{\delta}}^*)$ is also a saddle-point solution (see the counterpart in the proof of Theorem 2).

Finally consider the case $\alpha_1 \geq \ldots \geq \alpha_L > 0$. If $\theta_{k-1}^* < \theta_k^*$ for some k, then it follows by (75) that $\delta_k^* = 0$. We shall show that it is impossible to have $\theta_k^* > 0$ and $\delta_k^* = 0$ at the same time. Indeed, if $\theta_k^* > 0$ and $\delta_k^* = 0$, then, according to (73), we must have $\delta_{k+1}^* = 0$ (and $\alpha_k = \alpha_{k+1}$), which, in view of (75), further implies $\theta_k^* \leq \theta_{k+1}^*$ (and consequently $\theta_{k+1}^* > 0$); now a simple induction yields $\theta_L^* > 0$ and $\delta_L^* = 0$, leading to a contradiction with (72). Therefore, we must have $\theta_0^* \geq \theta_1^* \geq \ldots \geq \theta_L^*$. This completes the proof of Theorem 6.

APPENDIX H
PROOF OF COROLLARY 9

It follows from Theorem 6 that

$$\psi(\underline{\alpha}, \underline{d}) = \max \min_{\underline{\delta} \in \mathcal{D}(\underline{d})} \omega(\underline{\alpha}, \underline{\theta}, \underline{\delta}),$$

where the maximization is taken over those $\underline{\theta}$ such that $\lambda \geq \theta_0 = \cdots = \theta_{i-1} \geq \theta_i = \cdots = \theta_{j-1} \geq \theta_j = \cdots =$

$\theta_L \geq 0$. As a consequence, we have

$$\psi(\underline{\alpha}, \underline{d}) = \max_{(\theta_0, \theta_i, \theta_j): \lambda \geq \theta_0 \geq \theta_i \geq \theta_j \geq 0} \min_{(\delta_i, \delta_j) \in [0, d_i] \times [0, d_j]}$$
$$\frac{\alpha_i}{2} \log\left(\frac{\lambda\theta_0 + \lambda\delta_i - \theta_0\delta_i}{\lambda\theta_i + \lambda\delta_i - \theta_i\delta_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda\delta_j - \theta_i\delta_j}{\lambda\theta_j + \lambda\delta_j - \theta_j\delta_j}\right)$$
$$= \max_{(\theta_0, \theta_i, \theta_j): \lambda \geq \theta_0 \geq \theta_i \geq \theta_j \geq 0} \frac{\alpha_i}{2} \log\left(\frac{\lambda\theta_0 + \lambda d_i - \theta_0 d_i}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda\theta_j + \lambda d_j - \theta_j d_j}\right)$$
$$= \max_{\theta_i \in [0, \lambda]} \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right).$$

One can readily verify the following statements by invoking Lemma 7 in Appendix G.

• $\tilde{h}(\alpha_i, \alpha_j, d_i, d_j) \leq 0$ (i.e., $\alpha_i(\lambda - d_i)d_j \geq \alpha_j(\lambda - d_j)d_i$): We have

$$\max_{\theta_i \in [0, \lambda]} \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right)$$
$$= \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right)\Bigg|_{\theta_i = 0}$$
$$= \frac{\alpha_i}{2} \log\left(\frac{\lambda}{d_i}\right).$$

• $h(\alpha_i, \alpha_j, d_i, d_j) + \tilde{h}(\alpha_i, \alpha_j, d_i, d_j) \geq 0$ (i.e., $\alpha_i(\lambda - d_i) \leq \alpha_j(\lambda - d_j)$): We have

$$\max_{\theta_i \in [0, \lambda]} \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right)$$
$$= \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right)\Bigg|_{\theta_i = \lambda}$$
$$= \frac{\alpha_j}{2} \log\left(\frac{\lambda}{d_j}\right).$$

• $\tilde{h}(\alpha_i, \alpha_j, d_i, d_j) > 0$ and $h(\alpha_i, \alpha_j, d_i, d_j) + \tilde{h}(\alpha_i, \alpha_j, d_i, d_j) < 0$ (i.e., $\alpha_i(\lambda - d_i)d_j < \alpha_j(\lambda - d_j)d_i$ and $\alpha_i(\lambda - d_i) > \alpha_j(\lambda - d_j)$): We have

$$\max_{\theta_i \in [0, \lambda]} \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta_i + \lambda d_i - \theta_i d_i}\right)$$
$$+ \frac{\alpha_j}{2} \log\left(\frac{\lambda\theta_i + \lambda d_j - \theta_i d_j}{\lambda d_j}\right)$$
$$= \frac{\alpha_i}{2} \log\left(\frac{\lambda^2}{\lambda\theta^\dagger + \lambda d_i - \theta^\dagger d_i}\right)$$

$$+\frac{\alpha_j}{2}\log\Big(\frac{\lambda\theta^\dagger+\lambda d_j-\theta^\dagger d_j}{\lambda d_j}\Big),$$

where

$$\theta^\dagger=-\frac{\tilde{h}(\alpha_i,\alpha_j,d_i,d_j)}{h(\alpha_i,\alpha_j,d_i,d_j)}\lambda$$

$$=\frac{-\alpha_i\lambda(\lambda-d_i)d_j+\alpha_j\lambda(\lambda-d_j)d_i}{(\alpha_i-\alpha_j)(\lambda-d_i)(\lambda-d_j)}.$$

It can be verified that

$$\frac{\alpha_i}{2}\log\Big(\frac{\lambda^2}{\lambda\theta^\dagger+\lambda d_i-\theta^\dagger d_i}\Big)+\frac{\alpha_j}{2}\log\Big(\frac{\lambda\theta^\dagger+\lambda d_j-\theta^\dagger d_j}{\lambda d_j}\Big)$$

$$=\frac{\alpha_i}{2}\log\Big[\frac{(\alpha_i-\alpha_j)(\lambda-d_j)}{\alpha_i(d_i-d_j)}\Big]$$

$$+\frac{\alpha_j}{2}\log\Big[\frac{\alpha_j\lambda(d_i-d_j)}{(\alpha_i-\alpha_j)(\lambda-d_i)d_j}\Big].$$

This completes the proof of Corollary 9.

APPENDIX I
PROOF OF THEOREM 9

We shall treat the three cases separately. It will be seen that Case 3) is the most non-trivial one whereas the other two are simple consequences of Corollary 11.

1) In view of Corollary 11, we must have $R(\underline{d})\geq R_{\ell,k}(d_\ell,d_k)$ and consequently $R_{\ell,k}(d_\ell,d_k)=\frac{L}{2\ell}\log\big(\frac{\lambda}{d_\ell}\big)$ for all $k>\ell$. Note that $R_{\ell,k}(d_\ell,d_k)=\frac{L}{2\ell}\log\big(\frac{\lambda}{d_\ell}\big)$ if and only if

$$d_k\geq\Big(\frac{k}{\ell}d_\ell^{-1}-\frac{k-\ell}{\ell}\lambda^{-1}\Big)^{-1}.$$

2) In view of Corollary 11 (or Corollary 12), we must have $R(\underline{d})\geq R_{k,L}(d_k,d_L)$ and consequently $R_{k,L}(d_k,d_L)=\frac{1}{2}\log\big(\frac{\lambda}{d_L}\big)$ for all $k<L$. Note that $R_{k,L}(d_k,d_L)=\frac{1}{2}\log\big(\frac{\lambda}{d_L}\big)$ if and only if

$$d_L\leq\frac{L}{k}d_k-\frac{L-k}{k}\lambda,$$

i.e.,

$$d_k\geq\frac{k}{L}d_L+\frac{L-k}{L}\lambda.$$

3) For any integer $k\in(\ell,L)$, it follows from Corollary 10 that $R(\underline{d})\geq\psi(\underline{\alpha}^*,\tilde{\underline{d}})$, where $\tilde{\underline{d}}=(\tilde{d}_1,\dots,\tilde{d}_L)$ with $\tilde{d}_\ell=d_\ell$, $\tilde{d}_k=d_k$, $\tilde{d}_L=d_L$, and $\tilde{d}_{k'}=\lambda$, $k'\notin\{\ell,k,L\}$. By Theorem 6, we have

$$\psi(\underline{\alpha}^*,\tilde{\underline{d}})=\max\min_{\underline{\delta}\in\mathcal{D}(\tilde{\underline{d}})}\omega(\underline{\alpha}^*,\underline{\theta},\underline{\delta}),$$

where the maximization is taken over those $\underline{\theta}$ such that $\lambda\geq\theta_0=\cdots=\theta_{\ell-1}\geq\theta_\ell=\cdots=\theta_{k-1}\geq\theta_k=\cdots=$

$\theta_{L-1}\geq\theta_L\geq0$. As a consequence, we have

$$\psi(\underline{\alpha}^*,\tilde{\underline{d}})$$

$$=\max_{(\theta_0,\theta_\ell,\theta_k,\theta_L):\lambda\geq\theta_0\geq\theta_\ell\geq\theta_k\geq\theta_L\geq0}$$
$$\min_{(\delta_\ell,\delta_k,\delta_L)\in[0,d_\ell]\times[0,d_k]\times[0,d_L]}$$
$$\frac{L}{2\ell}\log\Big(\frac{\lambda\theta_0+\lambda\delta_\ell-\theta_0\delta_\ell}{\lambda\theta_\ell+\lambda\delta_\ell-\theta_\ell\delta_\ell}\Big)$$
$$+\frac{L}{2k}\log\Big(\frac{\lambda\theta_\ell+\lambda\delta_k-\theta_\ell\delta_k}{\lambda\theta_k+\lambda\delta_k-\theta_k\delta_k}\Big)$$
$$+\frac{1}{2}\log\Big(\frac{\lambda\theta_k+\lambda\delta_L-\theta_k\delta_L}{\lambda\theta_L+\lambda\delta_L-\theta_L\delta_L}\Big)$$
$$=\max_{(\theta_0,\theta_\ell,\theta_k,\theta_L):\lambda\geq\theta_0\geq\theta_\ell\geq\theta_k\geq\theta_L\geq0}$$
$$\frac{L}{2\ell}\log\Big(\frac{\lambda\theta_0+\lambda d_\ell-\theta_0 d_\ell}{\lambda\theta_\ell+\lambda d_\ell-\theta_\ell d_\ell}\Big)$$
$$+\frac{L}{2k}\log\Big(\frac{\lambda\theta_\ell+\lambda d_k-\theta_\ell d_k}{\lambda\theta_k+\lambda d_k-\theta_k d_k}\Big)$$
$$+\frac{1}{2}\log\Big(\frac{\lambda\theta_k+\lambda d_L-\theta_k d_L}{\lambda\theta_L+\lambda d_L-\theta_L d_L}\Big)$$
$$=\max_{(\theta_\ell,\theta_k):\lambda\geq\theta_\ell\geq\theta_k\geq0}\frac{L}{2\ell}\log\Big(\frac{\lambda^2}{\lambda\theta_\ell+\lambda d_\ell-\theta_\ell d_\ell}\Big)$$
$$+\frac{L}{2k}\log\Big(\frac{\lambda\theta_\ell+\lambda d_k-\theta_\ell d_k}{\lambda\theta_k+\lambda d_k-\theta_k d_k}\Big)$$
$$+\frac{1}{2}\log\Big(\frac{\lambda\theta_k+\lambda d_L-\theta_k d_L}{\lambda d_L}\Big).$$

Now consider the following optimization problems:

$$\max_{\theta_\ell\in[0,\lambda]}-\frac{L}{2\ell}\log(\lambda\theta_\ell+\lambda d_\ell-\theta_\ell d_\ell)$$
$$+\frac{L}{2k}\log(\lambda\theta_\ell+\lambda d_k-\theta_\ell d_k),\qquad(78)$$
$$\max_{\theta_k\in[0,\lambda]}-\frac{L}{2k}\log(\lambda\theta_k+\lambda d_k-\theta_k d_k)$$
$$+\frac{1}{2}\log(\lambda\theta_k+\lambda d_L-\theta_k d_L).\qquad(79)$$

First note that the condition $\frac{L}{\ell}d_\ell-\frac{L-\ell}{\ell}\lambda<d_L<(\frac{L}{\ell}d_\ell^{-1}-\frac{L-\ell}{\ell}\lambda^{-1})^{-1}$, together with the assumption that $d_\ell\in(0,\lambda]$ and $d_L\in(0,\lambda]$, implies $0<d_L<d_\ell<\lambda$, $\tilde{h}(\alpha_\ell^*,\alpha_L^*,d_\ell,d_L)>0$, and $h(\alpha_\ell^*,\alpha_L^*,d_\ell,d_L)+\tilde{h}(\alpha_\ell^*,\alpha_L^*,d_\ell,d_L)<0$, which further implies $\theta^\dagger\in(0,\lambda)$, where

$$\theta^\dagger=-\frac{\tilde{h}(\alpha_\ell^*,\alpha_L^*,d_\ell,d_L)}{h(\alpha_\ell^*,\alpha_L^*,d_\ell,d_L)}\lambda$$

$$=\frac{-L\lambda(\lambda-d_\ell)d_L+\ell\lambda(\lambda-d_L)d_\ell}{(L-\ell)(\lambda-d_\ell)(\lambda-d_L)}.$$

Now define

$$\theta(d_k)=-\frac{\tilde{h}(\alpha_\ell^*,\alpha_k^*,d_\ell,d_k)}{h(\alpha_\ell^*,\alpha_k^*,d_\ell,d_k)}\lambda,$$

$$\tilde{\theta}(d_k)=-\frac{\tilde{h}(\alpha_k^*,\alpha_L^*,d_k,d_L)}{h(\alpha_k^*,\alpha_L^*,d_k,d_L)}\lambda.$$

It is easy to verify that

$$\theta(d_k) = \frac{-k\lambda(\lambda - d_\ell)d_k + \ell\lambda(\lambda - d_k)d_\ell}{(k - \ell)(\lambda - d_\ell)(\lambda - d_k)},$$

$$\tilde{\theta}(d_k) = \frac{-L\lambda(\lambda - d_k)d_L + k\lambda(\lambda - d_L)d_k}{(L - k)(\lambda - d_k)(\lambda - d_L)};$$

moreover, $\theta(d_k)$ is a strictly decreasing function and $\tilde{\theta}(d_k)$ is a strictly increasing function for $d_k \in (0, \lambda)$. Note that

$$
\begin{aligned}
\tilde{\theta}(d_\ell) &= \frac{-L\lambda(\lambda - d_\ell)d_L + k\lambda(\lambda - d_L)d_\ell}{(L - k)(\lambda - d_\ell)(\lambda - d_L)} \\
&< \frac{-L\lambda(\lambda - d_L)d_\ell + k\lambda(\lambda - d_L)d_\ell}{(L - k)(\lambda - d_\ell)(\lambda - d_L)} \\
&= \frac{-\lambda d_\ell}{\lambda - d_\ell} \\
&= \theta(d_\ell),
\end{aligned}
$$

$$
\begin{aligned}
\theta(d_L) &= \frac{-k\lambda(\lambda - d_\ell)d_L + \ell\lambda(\lambda - d_L)d_\ell}{(k - \ell)(\lambda - d_\ell)(\lambda - d_L)} \\
&< \frac{-k\lambda(\lambda - d_\ell)d_L + \ell\lambda(\lambda - d_\ell)d_L}{(k - \ell)(\lambda - d_\ell)(\lambda - d_L)} \\
&= \frac{-\lambda d_L}{\lambda - d_L} \\
&= \tilde{\theta}(d_L).
\end{aligned}
$$

Therefore, the equation $\theta(d_k) = \tilde{\theta}(d_k)$ has a unique solution for $d_k \in (d_\ell, d_L)$, which is given by

$$d_k^* = \frac{L(k - \ell)(\lambda - d_\ell)d_L + \ell(L - k)(\lambda - d_L)d_\ell}{k(L - \ell)\lambda - L(k - \ell)d_\ell - \ell(L - k)d_L}.$$

It is clear that $\tilde{h}(\alpha_\ell^*, \alpha_k^*, d_\ell, d_k^*) > 0$ and $\tilde{h}(\alpha_k^*, \alpha_L^*, d_k^*, d_L) > 0$. Moreover, it can be verified that

$$\theta(d_k^*) = \tilde{\theta}(d_k^*) = \theta^\dagger \in (0, \lambda).$$

Therefore, there exists an $\epsilon \in (0, d_k^*]$ such that

$$
\begin{aligned}
\tilde{h}(\alpha_\ell^*, \alpha_k^*, d_\ell, d_k) &> 0, \\
\tilde{h}(\alpha_k^*, \alpha_L^*, d_k, d_L) &> 0, \\
0 < \tilde{\theta}(d_k) < \theta(d_k) &< \lambda \quad (80)
\end{aligned}
$$

for any $d_k \in (d_k^* - \epsilon, d_k^*)$. It follows from Lemma 7 in Appendix G that $\theta_\ell = \theta(d_k)$ is the unique maximizer of (78) and $\theta_k = \tilde{\theta}(d_k)$ is the unique maximizer of (79) for any $d_k \in (d_k^* - \epsilon, d_k^*)$, which, together with (80), implies

$$
\begin{aligned}
\psi(\underline{\alpha}^*, \tilde{\underline{d}}) &= \frac{L}{2\ell} \log \left(\frac{\lambda^2}{\lambda\theta(d_k) + \lambda d_\ell - \theta(d_k)d_\ell} \right) \\
&+ \frac{L}{2k} \log \left(\frac{\lambda\theta(d_k) + \lambda d_k - \theta(d_k)d_k}{\lambda\tilde{\theta}(d_k) + \lambda d_k - \tilde{\theta}(d_k)d_k} \right) \\
&+ \frac{1}{2} \log \left(\frac{\lambda\tilde{\theta}(d_k) + \lambda d_L - \tilde{\theta}(d_k)d_L}{\lambda d_L} \right) \\
&> \frac{L}{2\ell} \log \left(\frac{\lambda^2}{\lambda\theta^\dagger + \lambda d_\ell - \theta^\dagger d_\ell} \right) \\
&+ \frac{L}{2k} \log \left(\frac{\lambda\theta^\dagger + \lambda d_k - \theta^\dagger d_k}{\lambda\theta^\dagger + \lambda d_k - \theta^\dagger d_k} \right) \\
&+ \frac{1}{2} \log \left(\frac{\lambda\theta^\dagger + \lambda d_L - \theta^\dagger d_L}{\lambda d_L} \right) \\
&= R_{\ell,L}(d_\ell, L)
\end{aligned}
$$

for any $d_k \in (d_k^* - \epsilon, d_k^*)$. Therefore, if $R(\underline{d}) = R_{\ell,L}(d_\ell, d_L)$, then we must have $d_k \geq d_k^*$.

This completes the proof of Theorem 9.

REFERENCES

[1] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[2] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 502–521, Jul. 1987.

[3] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[4] S. S. Pradhan, R. Puri, and K. Ramchandran, "*n*-channel symmetric multiple descriptions—Part I: (n, k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[5] R. Puri, S. S. Pradhan, and K. Ramchandran, "*n*-channel symmetric multiple descriptions—Part II: An achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1377–1392, Apr. 2005.

[6] C. Tian and J. Chen, "New coding schemes for the symmetric *K*-description problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.

[7] E. Ahmed and A. B. Wagner, "Erasure multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1328–1344, Mar. 2012.

[8] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 721–726, Nov. 1985.

[9] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[10] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

[11] H. Wang and P. Viswanath, "Vector Gaussian multiple description with two levels of receivers," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 401–410, Jan. 2009.

[12] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3991–4005, Sep. 2009.

[13] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.

[14] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[15] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3946–3960, Aug. 2010.

[16] J. Wang and J. Chen, "Vector Gaussian two-terminal source coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3693–3708, Jun. 2013.

[17] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3778–3788, Aug. 2006.

[18] L. Song, J. Chen, and C. Tian, "Broadcasting correlated vector Gaussians," *IEEE Trans. Inf. Theory*, to be published.

[19] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Inf. Control*, vol. 37, no. 1, pp. 34–39, Apr. 1978.

[20] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[21] C. Tian, S. Mohajer, and S. N. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3869–3891, Aug. 2009.

[22] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 349–357, Mar. 1995.

[23] F.-W. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012–2021, Jul. 2002.

[24] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[25] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Inf. Control*, vol. 36, no. 2, pp. 133–156, Feb. 1978.

[26] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Finite-dimensional bounds on $\mathbb{Z}_m$ and binary LDPC codes with belief propagation decoders," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 56–81, Jan. 2007.

[27] C. Nair, "Upper concave envelopes and auxiliary random variables," *Int. J. Adv. Eng. Sci. Appl. Math.*, vol. 5, no. 1, pp. 12–20, Mar. 2013.

[28] J. von Neumann, "Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," *Erg. Math. Kolloqu.*, vol. 8, pp. 73–83, 1937.

[29] S. Kakutani, "A generalization of Brouwer's fixed point theorem," *Duke Math. J.*, vol. 8, no. 3, pp. 457–459, 1941.

[30] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

Lin Song received the B.Sc. degree in Electronic Information Engineering, the M.S. degree in Communication and Information Systems from the Harbin Institute of Technology, Harbin, Heilongjiang, China, in 2006 and 2008, respectively, and the Ph.D. degree in Electrical and Computer Engineering from McMaster University, Hamilton, ON, Canada, in 2012. She was a postdoctoral research assistant at McMaster University, Hamilton, ON, Canada from September 2012 to August 2013. She is currently with the Institute of Network Coding at the Chinese University of Hong Kong, Hong Kong, as a postdoctoral fellow. Her research interests include network information theory and coding theory.

Shuo Shao received the B.Sc. degree in Information Engineering from Southeast University, Nanjing, Jiangsu, China, in 2011, and the M.S. degree in Electrical and Computer Engineering from McMaster University, Hamilton, ON, Canada, in 2013. He is currently pursuing his Ph.D. degree in Electrical and Computer Engineering at Texas A&M University, College Station, TX, US. His primary research interests are network information theory and data processing.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from September 2005 to July 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from July 2006 to August 2007. Since September 2007 he has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he is currently an Associate Professor. His research interests include information theory, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010. He is currently serving as an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.