A Calculation of the Heegard-Berger Rate-distortion Function for a Binary Source

Chao Tian and Suhas N. Diggavi
Laboratory for Information and Communication Systems (LICOS)
School of Computer and Communication Sciences
EPFL, Lausanne, Switzerland
Email: {chao.tian,suhas.diggavi}@epfl.ch

Abstract — **We provide an explicit calculation of the rate-distortion function for the doubly-symmetric binary source (DSBS), when the side information may be absent at the decoder. The rate-distortion function for general discrete memoryless source was characterized by Heegard and Berger in 1985 [IT-31(6)], who showed that a two-stage coding structure is in fact optimal. However, an explicit characterization of the rate-distortion function for DSBS, and more importantly the optimal forward testing channel structure for this source, was not found despite several attempts. In this work, we resolve this open problem. It is shown that in the two-stage coding structure, the optimal testing channel for the first stage decoder (who does not have side information) is the same as the optimal testing channel for the ordinary symmetric binary source, and this confirms a conjecture made by Fleming and Effros.**

I. Introduction

In the well-known problem of source coding with decoder side information, usually referred to as the Wyner-Ziv problem [1], a source is to be reproduced at the decoder with certain fidelity; the decoder has access to some side information, which is not available to the encoder. Heegard and Berger [2] (see also [3]) extended this problem to the scenario where the side information may be absent at the decoder. More precisely, the problem is defined as follows.

Definition 1 *An (n, M, D_1, D_2) HB code for source X with side information Y with a generic distribution P_{XY} in finite alphabet $\mathcal{X}$ and $\mathcal{Y}$, respectively, consists of an encoding function ϕ, and 2 decoding functions ψ_1 and ψ_2:*

$$\phi : \mathcal{X}^n \to I_M, \quad \psi_1 : I_M \to \hat{\mathcal{X}}^n, \quad \psi_2 : I_M \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n,$$

where $I_k = \{1, 2, ..., k\}$ and $\hat{\mathcal{X}}$ is a finite reconstruction alphabet [1]*, such that*

$$\mathbb{E}d(X^n, \psi_1(\phi(X^n))) \leq D_1 \quad (1)$$

$$\mathbb{E}d(X^n, \psi_2(\phi(X^n), Y^n)) \leq D_2, \quad (2)$$

where $\mathbb{E}$ is the expectation operation, and $d(\cdot, \cdot)$ is the usual per-letter distortion measure.

Definition 2 *Rate R is said to be (D_1, D_2) achievable, if for every $\epsilon > 0$ and sufficient large n there exists an $(n, M, D_1 + \epsilon, D_2 + \epsilon)$ code with $R + \epsilon \geq \frac{1}{n} \log M$.*

The following theorem by Heegard and Berger provides the rate-distortion function for this problem[2].

Theorem 1

$$R_{HB}(D_1, D_2) = \min_{p(D_1,D_2)} [I(X; W_1) + I(X; W_2|W_1, Y)],$$

where $p(D_1, D_2)$ is the set of all random variables $(W_1, W_2) \in \mathcal{W}_1 \times \mathcal{W}_2$ jointly distributed with the generic random variables (X, Y), such that the following conditions are satisfied: (1) $(W_1, W_2) \leftrightarrow X \leftrightarrow Y$ is a Markov string; (2) $|\mathcal{W}_1| \leq |\mathcal{X}| + 2$ and $|\mathcal{W}_2| \leq (|\mathcal{X}| + 1)^2$; (3) there exist deterministic functions f_1 and f_2 such that

$$\mathbb{E}d(X, f(W_1)) \leq D_1, \quad \mathbb{E}d(X, f(W_1, W_2, Y)) \leq D_2.$$

In this work we consider the following doubly-symmetric binary source (DSBS): $X \in \{0, 1\}$ is a discrete memoryless source (DMS) with $P(X = 0) = P(X = 1) = 0.5$. Side information $Y = X \oplus N$, where N is a Bernoulli random variable independent of everything else with $P(N = 1) = p < 0.5$ and $\oplus$ stands for modulo 2 addition; alternatively, Y can be taken as the output of a binary symmetric channel (BSC) with input X, and crossover probability p. The distortion measure is the Hamming distortion $d(x, \hat{x}) = x \oplus \hat{x}$.

The DSBS source is probably the simplest discrete source in the side information scenario, and it provides considerable insight into the Wyner-Ziv problem [1]. Somewhat surprisingly, an explicit calculation of $R_{HB}(D_1, D_2)$ was not found for it. Heegard and Berger postulated a forward test channel in [2], which was later shown to be not optimal by Kerpez [4]. Kerpez provided upper and lower bounds, neither of which are tight. Fleming and Effros [5] also contributed to this problem by considering it as a rate distortion problem with mixed types of side information. An algorithm to compute the rate-distortion function numerically was devised in [6, 7]. However an explicit expression of the rate distortion function for this source, and more importantly

[1]For simplicity we assume the same reconstruction alphabet for the two decoding functions, which is indeed the case for the DSBS problem being considered.

[2]The definitions of D_1 and D_2 are reversed from that in [2, 4]; the convention used here makes the two-stage coding structure more explicit.

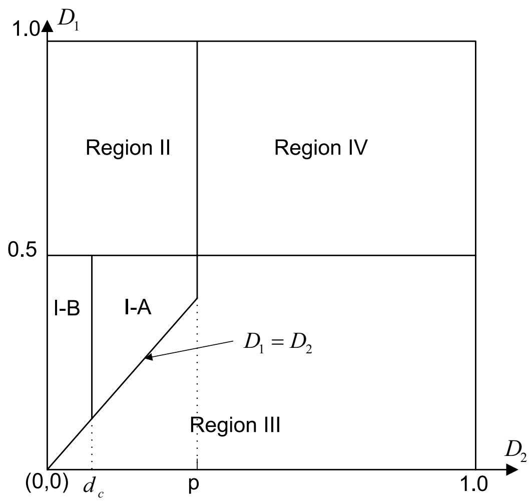

Figure 1: The four parts of the rate-distortion regions. d_c is the critical distortion defined in [1]

the corresponding optimal forward test channel structure have not been given in the literature. In this work, we resolve this open problem. Our interest in this particular source arises from the Wyner-Ziv successive refinement problem recently considered by Steinberg and Merhav [8] (see also [9]).

The rest of the paper is organized as follows. In Section II we state the main result in the form of a theorem and two corollaries. The proofs are given in Section III and Section IV concludes the paper.

II. Main Results

It was shown in [4] that the rate distortion region can be partitioned into four subregions, three of which are degenerate (see Fig. 1).

- Region I: $0 \leq D_1 < 0.5$ and $0 \leq D_2 < \min(D_1, p)$. In this region $R(D_1, D_2)$ is a function of both D_1 and D_2, and it is the only non-degenerate case;
- Region II: $D_1 \geq 0.5$ and $0 \leq D_2 \leq p$. Here the first stage does not have to encode and therefore the problem degenerates to Wyner-Ziv encoding for the second stage;
- Region III: $0 \leq D_1 \leq 0.5$ and $D_2 \geq \min(D_1, p)$. Here the second stage does not have to encode and hence the problem degenerates to the rate-distortion encoding for the first stage;
- Region IV: $D_1 > 0.5$ and $D_2 > p$. The rate is zero since the distortion constraints are trivially met.

We will need the following function from [1], defined on the domain $0 \leq u \leq 1$,

$$G(u) = h(p * u) - h(u), \tag{3}$$

where $h(u)$ is the binary entropy function $h(u) = -u \log u - (1-u)\log(1-u)$ and $u * v$ is the binary convolution for $0 \leq u, v \leq 1$ and $u * v = u(1-v) + v(1-u)$. We will be interested only in the case $0 \leq p < 0.5$. It was shown in [1] that $G(u)$ is (strictly) convex; furthermore, it is easy to show that $G(u)$ is symmetric about 0.5, and is monotonically decreasing for $0 \leq u \leq 0.5$; the minimum of $G(u)$ is zero when $u = 0.5$. It was also shown[3] in [1] that for $0 \leq D < p$ the Wyner-Ziv R-D function is

$$R^*_{X|Y}(D) = \min_{(\beta,\theta): 0 \leq \theta \leq 1, 0 \leq \beta \leq p, D = \theta\beta + (1-\theta)p} [\theta G(\beta)]. \tag{4}$$

We next define the following function

$$S_{D_1}(\alpha, \beta, \theta, \theta_1) = \\ 1 - h(D_1 * p) + (\theta - \theta_1)G(\alpha) + \theta_1 G(\beta) + (1-\theta)G(\gamma)$$

where

$$\gamma = \begin{cases} \frac{D_1 - (\theta-\theta_1)(1-\alpha) - \theta_1\beta}{1-\theta} & \theta \neq 1 \\ 0.5 & \theta = 1 \end{cases}$$

on the domain

$$0 \leq \theta_1 \leq \theta \leq 1, \quad 0 \leq \alpha, \beta \leq p, \quad p \leq \gamma \leq 1-p.$$

Notice that $S_{D_1}(\cdot)$ is continuous at $\theta = 1$.

The following theorem characterizes the rate distortion function $R_{HB}(D_1, D_2)$ in Region I.

Theorem 2 *For distortion pairs (D_1, D_2) in Region I:*

$$R_{HB}(D_1, D_2) = \min S_{D_1}(\alpha, \beta, \theta, \theta_1) \triangleq S^*(D_1, D_2),$$

where the minimization is over the domain of $S_{D_1}(\alpha, \beta, \theta, \theta_1)$, subject to the constraint

$$(\theta - \theta_1)\alpha + \theta_1\beta + (1-\theta)p = D_2.$$

This theorem is proved in the next section. One notable consequence in the proof of the forward part of this theorem, is that W_1 can always be taken as the output of a BSC with crossover probability D_1 and input X. This confirms the conjecture made by Fleming and Effros that the optimal forward test channel indeed should have this structure [7].

The following two corollaries are useful, and are straightforward given Theorem 2. The proofs for them, which can be found in [9], are not included here due to limited space. The first corollary provides a lower bound for $R_{HB}(D_1, D_2)$, which is easy to compute and usually tighter than the one given in [4].

Corollary 1 *For distortion pairs (D_1, D_2) in Region I:*

$$R_{HB}(D_1, D_2) \geq 1 - h(D_1 * p) + R^*_{X|Y}(D_2).$$

[3]In [1], the minimization was given instead as an infimum with the feasible range of $0 \leq \beta' < p$, but it can be shown that for $D_2 < p$, these two forms are equivalent.

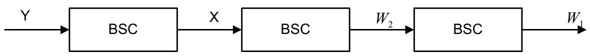

Figure 2: The optimal forward test channel in Region I-B. The crossover probability for the BSC between X and W_2 is D_2, while the crossover probability η for the BSC between W_2 and W_1 is such that $D_2 * \eta = D_1$.

Next recall the definition of the critical distortion d_c in the Wyner-Ziv problem for the DSBS source, where

$$\frac{G(d_c)}{d_c - p} = G'(d_c).$$

We have the following corollary which specifies a simple forward test channel structure for the case $D_2 \leq d_c$.

Corollary 2 *For distortion pairs* (D_1, D_2) *such that* $D_1 \leq 0.5$ *and* $D_2 \leq \min(d_c, D_1)$ *(i.e., Region I-B),*

$$R_{HB}(D_1, D_2) = 1 - h(D_1 * p) + G(D_2).$$

This result suggests that the optimal forward test channel for this case is in fact cascade of two BSC channels depicted in Fig. 2; this agrees with intuition, however is quite difficult to prove without Theorem 2.

III. The Proof Details

We will need the following lemma from [4] to simplify the calculation.

Lemma 1 *[4] For* $(W_1, W_2) \in p(D_1, D_2)$

$$I(X; W_1) + I(X; W_2|YW_1) = H(X) - H(Y|W_1) + H(Y|W_1W_2) - H(X|W_1W_2).$$

The lower bound:

Let $(W_1, W_2) \in P(D_1, D_2)$ define a joint distribution with (X, Y). Furthermore, assume the functions f_1 and f_2 are optimal for these random variables, i.e., there do not exist f_1' (or f_2'), such that $\mathbb{E}d(X, f_1'(W_1)) < \mathbb{E}d(X, f_1(W_1))$ (or $\mathbb{E}d(X, f_2'(W_1, W_2, Y)) < \mathbb{E}d(X, f_2(W_1, W_2, Y))$), because otherwise we can consider the alternative functions f_1' (or f_2') without loss of optimality. Our goal is to show that $I(X; W_1) + I(X; W_2|YW_1) \geq S^*(D_1, D_2)$, then invoke the rate distortion theorem, by which the lower bound can be established.

Similar as in [1][4], define the following set

$$A = \{(w_1, w_2) : f_2(w_1, w_2, 0) = f_2(w_1, w_2, 1)\}, \tag{5}$$

which defines its complement as,

$$A^c = \{(w_1, w_2) : f_2(w_1, w_2, 0) \neq f_2(w_1, w_2, 1)\}. \tag{6}$$

For each $w_1 \in \mathcal{W}_1$, define the following two sets

$$B(w_1) = \{w_2 : (w_1, w_2) \in A,\ f_1(w_1) = f_2(w_1, w_2, 0)\},$$
$$B^*(w_1) = \{w_2 : (w_1, w_2) \in A,\ f_1(w_1) \neq f_2(w_1, w_2, 0)\}.$$

Notice that for each fixed $w_1^* \in \mathcal{W}_1$, we have $\mathcal{W}_2 = B(w_1^*) \cup B^*(w_1^*) \cup \{w_2 : (w_1^*, w_2) \in A^c\}$, and the three sets are disjoint. To simplify the notations, write $P\{(W_1W_2) = (w_1w_2)\}$ as $P_{w_1w_2}$, and $P\{W_1 = w_1\}$ as P_{w_1}. Define the following quantity for each $w_1 \in \mathcal{W}_1$

$$D_{1,w_1} \triangleq \mathbb{E}[d(X, \hat{X}_1)|W_1 = w_1] = P\{X \neq f_1(w_1)|W_1 = w_1\}$$

and define the following quantity for each $(w_1, w_2) \in A$,

$$\begin{aligned} D_{2,w_1w_2} &\triangleq \mathbb{E}[d(X, \hat{X}_2)|(W_1, W_2) = (w_1, w_2)] \\ &= P\{X \neq f_2(w_1, w_2, 0)|(W_1, W_2) = (w_1, w_2)\}. \end{aligned}$$

By the Markov string $Y \leftrightarrow X \leftrightarrow (W_1, W_2)$, it follows that for each $w_1 \in \mathcal{W}_1$

$$\begin{aligned} H(X|W_1 = w_1) &= h(D_{1,w_1}) \\ H(Y|W_1 = w_1) &= h(p * D_{1,w_1}), \end{aligned} \tag{7}$$

where as before $u * v \triangleq u(1-v) + v(1-u)$. For each $(w_1, w_2) \in A$, we have

$$\begin{aligned} H[X|(W_1, W_2 = w_1, w_2)] &= h(D_{2,w_1w_2}) \\ H[Y|(W_1, W_2) = (w_1, w_2)] &= h(p * D_{2,w_1w_2}). \end{aligned} \tag{8}$$

And furthermore, for each $(w_1, w_2) \in A^c$, we have

$$\begin{aligned} &H[X|(W_1, W_2 = w_1, w_2)] = \\ &\quad h(P\{X \neq f_1(w_1)|W_1 = w_1, W_2 = w_2\}) \\ &H[Y|(W_1, W_2) = (w_1, w_2)] = \\ &\quad h(p * P\{X \neq f_1(w_1)|W_1 = w_1, W_2 = w_2\}). \end{aligned} \tag{9}$$

We will also need the following quantities

$$\begin{aligned} \theta &\triangleq P\{(W_1, W_2) \in A\} \\ \theta_1 &\triangleq P\{(W_1, W_2) \in \{(w_1, w_2) : w_2 \in B(w_1)\}\}. \end{aligned} \tag{10}$$

Clearly, we have

$$\begin{aligned} H(X) - H(Y|W_1) &= 1 - \sum_{w_1 \in \mathcal{W}_1} P_{w_1} H(Y|W_1 = w_1) \\ &= 1 - \sum_{w_1 \in \mathcal{W}_1} P_{w_1} h(p * D_{1,w_1}) \\ &\geq 1 - h(p * D_1') \end{aligned} \tag{11}$$

where we have used the concavity of function $h(p * u)$ in the last step and

$$D_1' \triangleq \sum_{w_1 \in \mathcal{W}_1} P_{w_1} D_{1,w_1}.$$

Furthermore we have

$$\begin{aligned} &H(Y|W_1W_2) - H(X|W_1W_2) \\ = &\sum_{(w_1,w_2) \in A} P_{w_1,w_2}[H(Y|(W_1, W_2) = (w_1, w_2)) \\ &\qquad - H(X|(W_1, W_2) = (w_1, w_2))] \\ + &\sum_{(w_1,w_2) \in A^c} P_{w_1,w_2}[H(Y|(W_1, W_2) = (w_1, w_2)) \\ &\qquad - H(X|(W_1, W_2) = (w_1, w_2))] \end{aligned}$$

The first term can be bounded as follows

$$\sum_{(w_1,w_2)\in A} P_{w_1,w_2}[H(Y|(W_1,W_2)=(w_1,w_2)) - H(X|(W_1,W_2)=(w_1,w_2))]$$
$$= \sum_{w_1}\sum_{w_2\in B(w_1)} P_{w_1,w_2}[h(p*D_{2,w_1w_2}) - h(D_{2,w_1w_2})]$$
$$+ \sum_{w_1}\sum_{w_2\in B^*(w_1)} P_{w_1,w_2}[h(p*D_{2,w_1w_2}) - h(D_{2,w_1w_2})]$$
$$\geq \theta_1 G(\beta) + (\theta-\theta_1)G(\alpha), \tag{12}$$

where as before $G(u) \triangleq h(p*u) - h(u)$, and

$$\alpha \triangleq \sum_{w_1}\sum_{w_2\in B^*(w_1)} \frac{P_{w_1w_2}}{\theta-\theta_1} D_{2,w_1w_2}$$
$$\beta \triangleq \sum_{w_1}\sum_{w_2\in B(w_1)} \frac{P_{w_1w_2}}{\theta_1} D_{2,w_1w_2}, \tag{13}$$

and the convexity of function $G(u)$ is used in the last step. Next, notice the identity that for each $w_1 \in \mathcal{W}_1$

$$P_{w_1}D_{1,w_1}$$
$$= P\{X \neq f_1(w_1), W_1 = w_1\}$$
$$= \sum_{w_2\in B(w_1)} P\{X \neq f_2(w_1,w_2,0), W_1=w_1, W_2=w_2\}$$
$$+ \sum_{w_2\in B^*(w_1)} P\{X = f_2(w_1,w_2,0), W_1=w_1, W_2=w_2\}$$
$$+ \sum_{w_2:(w_1,w_2)\in A^c} P\{X \neq f_1(w_1), W_1=w_1, W_2=w_2\}$$
$$= \sum_{w_2\in B(w_1)} P_{w_1w_2}D_{2,w_1w_2}$$
$$+ \sum_{w_2\in B^*(w_1)} P_{w_1w_2}(1-D_{2,w_1w_2})$$
$$+ \sum_{w_2:(w_1,w_2)\in A^c} P_{w_1w_2}P\{X \neq f_1(w_1)|W_1=w_1, W_2=w_2\}. \tag{14}$$

It follows that

$$\sum_{(w_1,w_2)\in A^c} P_{w_1,w_2}[H(Y|(W_1,W_2)=(w_1,w_2)) - H(X|(W_1,W_2)=(w_1,w_2))]$$
$$= \sum_{w_1}\sum_{w_2:(w_1,w_2)\in A^c} P_{w_1,w_2}G[P\{X\neq f_1(w_1)|(w_1,w_2)\}]$$
$$\geq (1-\theta)G(\gamma), \tag{15}$$

where again the convexity of function $G(u)$ is used, and because of the identity (14), we have

$$\gamma = \sum_{w_1}\sum_{w_2:(w_1,w_2)\in A^c} \frac{P_{w_1w_2}}{1-\theta} P\{X\neq f_1(w_1)|(w_1,w_2)\}$$
$$= \frac{D_1' - \theta_1\beta - (\theta-\theta_1)(1-\alpha)}{1-\theta}. \tag{16}$$

It was shown in [4], by a straightforward generalization of the argument in [1], that

$$E[d(X,\hat{X}_2)|(W_1,W_2)\in A^c] \geq p. \tag{17}$$

By the hypothesis

$$D_2' \triangleq \theta_1\beta + (\theta-\theta_1)\alpha + (1-\theta)p \leq D_2$$
$$D_1' \leq D_1.$$

Notice that for each $(w_1,w_2)\in A$, $D_{2,w_1w_2} \leq p$, because otherwise for this (w_1,w_2) pair, making $f_2(w_1,w_2,Y) = Y$ will in fact reduce the distortion, which contradicts with the optimality of the decoding function. Thus $0 \leq \alpha, \beta \leq p$. Similarly, $p \leq \gamma \leq 1-p$, because $p \leq P\{X \neq f_1(w_1)|W_1 = w_1, W_2 = w_2\} \leq 1-p$, otherwise we can modify the decoder function f_2 to reduce the distortion. Clearly, $0 \leq \theta_1 \leq \theta \leq 1$ by definition.

Summarizing the bounds, we have shown that

$$R_{HB}(D_1,D_2) \geq \min_{(\alpha,\beta,\theta,\theta_1,D_1')\in\mathcal{Q}_{\geq}} [1 - h(D_1'*p) + (1-\theta)G(\gamma) + \theta_1 G(\beta) + (\theta-\theta_1)G(\alpha)], \tag{18}$$

where the minimization is within the following set

$$\mathcal{Q}_{\leq} = \{(\alpha,\beta,\theta,\theta_1,D_1') : 0\leq\theta_1\leq\theta\leq 1, \quad 0\leq\alpha,$$
$$(1-\theta)p \leq D_1' - (\theta-\theta_1)(1-\alpha) - \theta_1\beta \leq (1-\theta)(1-p),$$
$$\beta \leq p, (\theta-\theta_1)\alpha + \theta_1\beta + (1-\theta)p \leq D_2, D_1' \leq D_1\}.$$

This is not yet the function given in Theorem 2, because the minimization given there is within the set

$$\mathcal{Q}_{=} = \{(\alpha,\beta,\theta,\theta_1,D_1') : 0\leq\theta_1\leq\theta\leq 1, \quad 0\leq\alpha,$$
$$(1-\theta)p \leq D_1' - (\theta-\theta_1)(1-\alpha) - \theta_1\beta \leq (1-\theta)(1-p),$$
$$\beta \leq p, (\theta-\theta_1)\alpha + \theta_1\beta + (1-\theta)p = D_2, D_1' = D_1\}.$$

This gap will be closed after we give the forward test channel structure. □

The upper bound:

We explicitly construct the random variables with joint pmf given in Table 1. It is straightforward to verify that it is a valid pmf, given the conditions in the definition of $S_{D_1}(\alpha,\beta,\theta,\theta_1)$. Furthermore, the rate $I(X;W_1) + I(X;W_2|W_1Y)$ is exactly $S_{D_1}(\alpha,\beta,\theta,\theta_1)$. The decoding functions are $f_1(W_1) = W_1$ and $f_2(W_1,W_2,Y) = W_2$ if $W_2 \neq 2$, otherwise $f_2(W_1,W_2,Y) = Y$. This establishes the upper bound.

Now we show that the gap aforementioned in the proof of the lower bound can be closed. Suppose that the parameters that minimize the right hand side of (18) are $(\alpha,\beta,\theta,\theta_1,D_1')$, and furthermore $D_1' < D_1$. The set of random variables W_1', W_2' can be constructed as given in Table 1 with D_1' replacing D_1. By the lower bound established above, we have

$$R_{HB}(D_1,D_2) \geq I(X;W_1') + (X;W_2'|W_1'Y). \tag{19}$$

	$w_1=0$		$w_1=1$	
	$x=0$	$x=1$	$x=0$	$x=1$
$w_2=0$	$0.5\theta_1(1-\beta)$	$0.5\theta_1\beta$	$0.5(\theta-\theta_1)(1-\alpha)$	$0.5(\theta-\theta_1)\alpha$
$w_2=1$	$0.5(\theta-\theta_1)\alpha$	$0.5(\theta-\theta_1)(1-\alpha)$	$0.5\theta_1\beta$	$0.5\theta_1(1-\beta)$
$w_2=2$	$0.5(1-\theta)(1-\gamma)$	$0.5(1-\theta)\gamma$	$0.5(1-\theta)\gamma$	$0.5(1-\theta)(1-\gamma)$
$p(x,w_1)$	$0.5(1-D_1)$	$0.5D_1$	$0.5D_1$	$0.5(1-D_1)$

Table 1: Joint distribution $p(x,w_1,w_2)$ and the marginal $p(x,w_1)$.

Consider a random variable $W_1''=W_1'\oplus N$, where N is a Bernoulli random variable independent of everything else with $P(N=1)=\eta$ such that $\eta * D_1'=D_1=D_1''$, which is valid since $\max\{D_1,D_1'\}\leq\frac{1}{2}$. Let $W_2''=(W_1',W_2')$, and we have $(W_1'',W_2'')\in P(D_1,D_2)$. Clearly, $W_1''\leftrightarrow W_1'\leftrightarrow X\leftrightarrow Y$, and $W_1''\leftrightarrow W_1'\leftrightarrow W_2'$. Thus by the rate distortion theorem for this problem

$$I(X;W_1'')+I(X;W_2''|W_1''Y)\geq R_{HB}(D_1,D_2). \qquad (20)$$

Notice that

$$\begin{aligned}
& I(X;W_1')+I(X;W_2'|W_1'Y)\\
\overset{(a)}{=}\ & I(X;W_1',W_1'')+I(X;W_1'',W_2'|W_1'Y)\\
=\ & I(X;W_1'')+I(X;W_1'|W_1'')+I(X;W_2'|W_1'W_1''Y)\\
\overset{(b)}{=}\ & I(X;W_1'')+I(X;W_1'|W_1'')\\
& \quad +I(X;W_1'W_2'|W_1''Y)-I(X;W_1'|W_1''Y)\\
\overset{(c)}{=}\ & I(X;W_1'')+I(X;W_1'W_2'|W_1''Y)+I(Y;W_1'|W_1'')\\
=\ & I(X;W_1'')+I(X;W_1'W_2'|W_1''Y)\\
& \quad +h(p*D_1'')-h(p*D_1')\\
>\ & I(X;W_1'')+I(X;W_1'W_2'|W_1''Y)
\end{aligned}$$

where (a) and (c) follow because of the Markov chain $W_1''\leftrightarrow W_1'\leftrightarrow X\leftrightarrow Y$, (b) is by applying chain rule to the last term in the previous line, and the last step is because $p<0.5$ and $D_1'<D_1=D_1''\leq 0.5$. However, this implies

$$\begin{aligned}
& I(X;W_1'')+I(X;W_1'W_2'|W_1''Y)\\
& \geq R_{HB}(D_1,D_2)\\
& \geq I(X;W_1')+(X;W_2'|W_1'Y)\\
& > I(X;W_1'')+I(X;W_1'W_2'|W_1''Y),
\end{aligned}$$

which is a contradiction. Thus we conclude that the minimum must be achieved with $D_1'=D_1$.

It can be shown that the constraint $(\theta-\theta_1)\alpha+\theta_1\beta+(1-\theta)p\leq D_2$ can be met with equality without loss of optimality, by analyzing the first and second order derivatives, and the proof is completed. □

IV. Conclusion

We gave an explicit calculation of the Heegard-Berger rate-distortion function for the doubly-symmetric binary source. It reveals that the optimal auxiliary random variable W_1 for the first stage decoder, which does not have side information, can be taken as the output of a BSC with input X and crossover probability D_1. This suggests that for this problem, the first stage is in fact a rate-distortion optimal without side information. In [9] we introduced the notion of generalized successive refinability for the Wyner-Ziv successive refinement problem, and the results given here imply that the DSBS is indeed generalized successive refinable.

References

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

[2] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Information Theory*, vol. 31, pp. 727–734, Nov. 1985.

[3] A. Kaspi, "Rate-distortion when side-information may be present at the decoder," *IEEE Trans. Information Theory*, vol. 40, pp. 2031–2034, Nov. 1994.

[4] K. J. Kerpez, "The rate-distortion function of a binary symmetric source when side information may be absent," *IEEE Trans. Information Theory*, vol. 33, pp. 448–452, May. 1987.

[5] M. Fleming and M. Effros, "Rate-distortion with mixed types of side information," in *Proc. IEEE Symposium Information Theory*, p. 144, Jun.-Jul 2003.

[6] M. Fleming, *On source coding for networks.* PhD thesis, California Institute of Technology, 2004.

[7] M. Fleming and M. Effros, "On rate-distortion with mixed types of side information," *IEEE Trans. Information Theory*, vol. 52, pp. 1698 – 1705, Apr. 2006.

[8] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Information Theory*, vol. 50, pp. 1636–1654, Aug. 2004.

[9] C. Tian and S. Diggavi, "On multistage successive refinement for Wyner-Ziv source coding with degraded side information," in *EPFL Technical report, Jan. 2006. Available at arXiv:cs.IT/0605001.*