

# Source Coding and Transmission under Common Knowledge Constraints

Yossef Steinberg

Department of Electrical Engineering  
Technion - Israel Institute of Technology  
Haifa 32000, Israel  
ysteinbe@ee.technion.ac.il

**Abstract**—This work studies problems of source coding under the requirement that the encoder can produce an exact copy of the compressed source constructed by the decoder. This requirement, termed here as a *common knowledge constraint*, is satisfied automatically in rate-distortion theory for single sources. However, in the common formulation of problems of lossy source coding with side information at the decoder (the Wyner-Ziv problem), distributed source coding, and joint source-channel coding for networks, the destination can exploit the information it receives in a manner that cannot be exactly reproduced at the sender side. Some applications, like the transmission of sensitive medical information, may require that both sides – the sender and the receiver – will share a common version of the compressed data, for the purpose of future discussions or consulting. The purpose of this work is to study the implications of common knowledge constraints on the achievable rates in scenarios of lossy source coding. A single letter characterization of the rate distortion function is developed, for the problem of source coding with side information at the decoder, under a common knowledge constraint. Implications of this constraint on problems of joint source channel coding for the degraded broadcast channel are studied. Specifically, it is shown that in this setup, a scheme based on separation achieves optimal distortions.

**Index terms** – Broadcast channel, common knowledge, hierarchical coding, joint source channel coding, source coding with side information, successive refinement, Wyner-Ziv problem.

## I. INTRODUCTION

In the common formulation of lossy source coding problems, the code designer is primarily concerned with reducing the coding rate under a prescribed distortion between the source and the reproduction at the destination. The question of whether the sender is acquainted with the distorted version produced by the decoder is not raised in this context. In lossy compression of single sources, the optimal reproduction  $\hat{X}^n$  is a deterministic function of the original source  $X^n$  and therefore it is readily available to the sender. While in lossy compression of single sources the encoder and decoder can share a common version of the compressed source without sacrificing performance, this is not the situation in more general scenarios. In source coding with side information at the decoder (the Wyner-Ziv problem, [9]), the encoder does not have the realization of the side information, and therefore is not acquainted with the receiver's version of the distorted source. Similar examples abound in multiterminal systems, perhaps most notably the problem of joint source-channel coding for the degraded broadcast channel. It is well

known [4] that for transmission of a Gaussian source over the Gaussian broadcast channel under quadratic distortion and input power constraints, an optimal scheme is not based on separation. Instead, a single-letter code is used and the allowable distortions are incurred by the channel noise. The results of this coding scheme cannot be reproduced at the sender side, since they depend on the channel noise whose realizations are not known.

The requirement to reproduce at the sender side an exact copy of the destination's version may seem superfluous at first: after all, the sender has at hand the source itself – undoubtedly a better version of  $X^n$  than the reproduction  $\hat{X}^n$ . But in some applications, having at the sender a better version than that of the receiver may not be enough. Consider, for example, the transmission of medical information, e.g. MRI results, over a noisy broadcast channel to a number of experts, for the purpose of consultation. The experts have previous data on the patient, that serve as side information in retrieving the new data. Lossy transmission of medical information is undesirable, since the distortion can blur important medical details. Moreover, depending on the coding scheme and the use of side information, the sender may not have control over the distortion pattern, and thus does not know what details were most affected by the distortion. Despite these shortcomings, sometimes there is no choice but to introduce some distortion, due to the limitations of the physical channel over which communication takes place. The lesser of all evils would be then to devise a coding scheme that enables the sender to produce exact copies of the distorted data that arrived to the experts. The role of this common reproduction is twofold. First, it leaves some control at the hands of the sender over the quality of the data at the destination, since the sender can decide to re-transmit in case that important medical details were blurred during transmission. Second, it can serve as a common reference for the consultation, where each side knows exactly what the data at the other side looks like.

As the above example demonstrates, being able to reproduce the receiver's version at the transmitter's side can be beneficial in many practical systems. This requirement is termed throughout the paper as the *common knowledge (CK) constraint*. The purpose of this work is to study the implications of common knowledge constraints on the achievable rates and distortions in source coding and joint source channel coding problems.

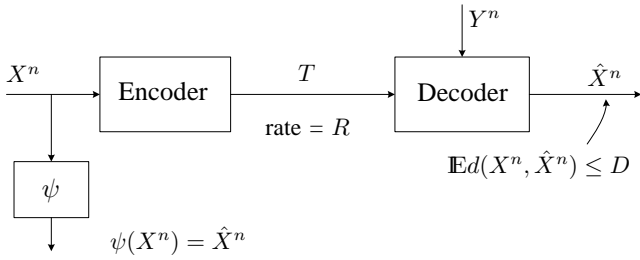


Fig. 1. Source coding with side information under a common knowledge constraint

In Section II, the problem of source coding with decoder side information and common knowledge constraint is posed, and a single letter characterization of the corresponding rate-distortion function is presented. In Section III, a few examples are given, and the relations of the new results to classical rate-distortion functions are discussed. Section IV studies the implications of this constraint on joint source-channel coding for the degraded broadcast channel.

## II. SOURCE CODING WITH SIDE INFORMATION AT THE DECODER

Let  $(X, Y)$  be a joint source. We are interested in lossy coding of the source  $X$  where the  $Y$  component serves as side information at the decoder, under the requirement to reproduce the decoder output at the sender side. The model is depicted in Fig. 1. Let the finite set  $\hat{\mathcal{X}}$  stand for the reproduction alphabet, and let  $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$  stand for a single letter distortion measure. Note that since the sets are finite, the definition of the range of  $d$  implies that the distortion measure is bounded. Define

$$d_{max} = \max_{x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x}).$$

The distortion between  $n$  sequences is defined in the usual way:

$$d(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i). \quad (1)$$

*Definition 1:* Let  $\mathcal{T}$  be a set of integers  $\mathcal{T} = \{1, 2, \dots, M\}$ . An  $(n, M, D, \epsilon)$  common knowledge (CK) code for the source  $X$  with decoder side information  $Y$  consists of an encoder map

$$f : \mathcal{X}^n \rightarrow \mathcal{T},$$

a decoder map

$$g : \mathcal{T} \times \mathcal{Y}^n \rightarrow \hat{\mathcal{X}}^n,$$

and a sender reconstruction map

$$\psi : \mathcal{X}^n \rightarrow \hat{\mathcal{X}}^n,$$

such that the average distortion is not larger than  $D$ :

$$\mathbb{E}d(X^n, g(f(X^n), Y^n)) \leq D, \quad (2)$$

and the sender can acquire the decoder's output with high probability, via the sender reconstruction map  $\psi$ :

$$P_{XY}(\psi(X^n) \neq g(f(X^n), Y^n)) \leq \epsilon. \quad (3)$$

In (2),  $\mathbb{E}$  stands for statistical expectation. The rate of the code is  $R = (1/n) \log M$ . A specific CK code is denoted by the three mappings  $(f, g, \psi)$ , where the dimension will be clear from the context. A rate-distortion pair  $(R, D)$  is said to be *CK-achievable* if for every  $\epsilon > 0$  and sufficiently large  $n$ , there exists an  $(n, 2^{n(R+\epsilon)}, D, \epsilon)$  CK code for the source  $X$  with decoder side information  $Y$ . The *CK rate-distortion function*, denoted by  $R_{ck}(D)$ , is the minimal rate  $R$  such that the pair  $(R, D)$  is CK-achievable.

The only difference between this definition and the classical definition of lossy source coding with side information (the Wyner-Ziv problem), is the requirement (3). The model is depicted in Figure 1. Denote by  $R_{WZ}(D)$  the Wyner-Ziv rate distortion function for the source  $X$  with side information  $Y$  [9], and by  $R(D)$  the regular rate-distortion function for the source  $X$  when no side information is present. Clearly

$$R_{WZ}(D) \leq R_{ck}(D) \leq R(D). \quad (4)$$

For a given distortion level  $D$ , define

$$R_{ck}^*(D) = \min[I(\hat{X}; X) - I(\hat{X}; Y)] \quad (5)$$

where the minimum is over all random variables  $\hat{X}$  taking values in the reconstruction alphabet  $\hat{\mathcal{X}}$ , such that the Markov relations

$$\hat{X} \circlearrowleft X \circlearrowright Y \quad (6)$$

hold, and the distortion constraint

$$\mathbb{E}d(X, \hat{X}) \leq D \quad (7)$$

is satisfied. Due to the Markov conditions (6),  $R_{ck}^*(D)$  can be expressed also as a minimum of conditional mutual information

$$R_{ck}^*(D) = \min I(\hat{X}; X|Y), \quad (8)$$

where the minimum is over the same set of reconstruction random variables as in (5). It is easily verified that  $R_{ck}^*(D)$  is a convex function of  $D$ . The next theorem states the main result on the CK rate-distortion function of a joint memoryless source  $(X, Y)$ .

*Theorem 1:* For any memoryless source with decoder side information,

$$R_{ck}(D) = R_{ck}^*(D).$$

It is instructive to examine  $R_{ck}^*(D)$  in light of the Wyner-Ziv rate distortion function. In [9], Wyner and Ziv showed that the rate-distortion function of  $X$  with decoder side information  $Y$  is given by

$$R_{WZ}(D) = \min[I(Z; X) - I(Z; Y)] \quad (9)$$

where  $Z$  is an external random variable taking values in a set  $\mathcal{Z}$ , and the minimum is over all  $Z$  and deterministic functions  $\phi : \mathcal{Z} \times \mathcal{Y} \rightarrow \hat{\mathcal{X}}$  such that the Markov relations

$$Z \circlearrowleft X \circlearrowright Y \quad (10)$$

hold, and the distortion constraint

$$\mathbb{E}d(X, \phi(Z, Y)) \leq D \quad (11)$$

is satisfied. In a typical Wyner-Ziv code, a codebook  $\{z(i)\}$  of rate  $I(Z; X)$  is constructed, from the external random variable  $Z$ . The side information  $Y$  is exploited in two distinct phases – binning and estimation. In the binning phase, the side information is used to reduce the rate by  $I(Z; Y)$ . In the estimation phase, the side information is used by the decoder to enhance the quality of estimation and reduce the distortion – hence the dependence of  $\phi$  on  $Y$ , in (11). Therefore, the codewords  $\{z(i)\}$  do not have to satisfy the distortion constraint by themselves. The estimation is the operation that cannot be mimicked by the encoder. When the common knowledge constraint is present, binning can still be performed, but the decoder cannot use  $Y$  to reduce the distortion. The structure of  $R_{ck}^*(D)$  reflects precisely this fact: the codewords  $\{\hat{x}(i)\}$  consisting the codebook, satisfy the distortion constraint alone, like in lossy source coding without side information. The binning operation, however, can be performed, and the rate is reduced by  $I(\hat{X}; Y)$ .

The result stated in Theorem 1 is, in a sense, complementary to a result of Weissman and El Gamal in [7]<sup>1</sup>. Specifically, Weissman and El Gamal showed that if the decoder is restricted to use the side information in a *causal* manner, then the rate-distortion function is given by [7, Theorem 1]

$$R_{\text{causal si}}(D) = \min I(W; X) \quad (12)$$

where  $W$  is an external random variable with alphabet  $\mathcal{W}$ , and the minimum is over all random variables  $W$  and deterministic functions  $\tilde{\phi} : \mathcal{W} \times \mathcal{Y} \rightarrow \hat{\mathcal{X}}$  satisfying

$$\mathbb{E}d(X, \tilde{\phi}(W, Y)) \leq D. \quad (13)$$

In other words, under the causality constraint, the side information is used for estimation (since  $\tilde{\phi}$  in (13) depends on  $Y$ ), but binning cannot be performed (since there is no  $-I(W; Y)$  term in (12)). This stands as a complementary to Theorem 1 here, where the side information is used for binning, but estimation cannot be performed.

### III. EXAMPLES

In this section the functions  $R_{ck}(D)$  of few sources are presented, and compared to the Wyner-Ziv rate distortion function and to the rate distortion function without side information.

*Example 1:* Binary source, with a symmetric channel side information and Hamming distortion measure. Here  $\mathcal{X} = \hat{\mathcal{X}} = \mathcal{Y} = \{0, 1\}$ ,  $X \sim \text{Bernoulli}(q)$ , and the side information  $Y$  is given by

$$Y = X \oplus Z,$$

where  $Z$  is binary, independent of  $X$ ,  $Z \sim \text{Bernoulli}(p_z)$ , and  $\oplus$  stands for modulo-2 addition. The distortion measure is the Hamming distance

$$d(x, \hat{x}) = \begin{cases} 0 & x = \hat{x} \\ 1 & \text{otherwise.} \end{cases}$$

<sup>1</sup>This observation is due to Tsachy Weissman.

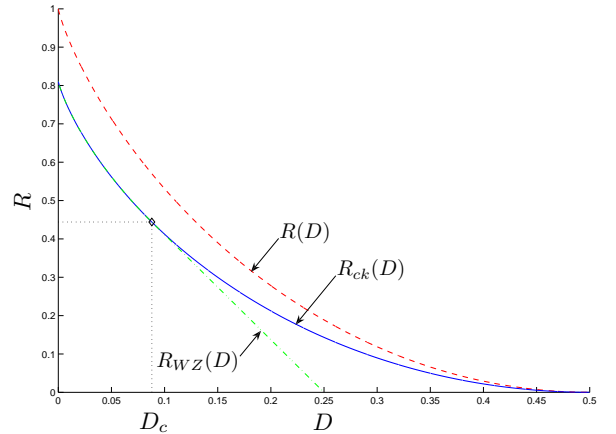


Fig. 2. Rate distortion curves for the doubly symmetric binary source, with cross over probability 0.25.

The CK rate distortion function for this source is given by

$$R_{ck}(D) = \begin{cases} R(D) - h(p_z \star q) \\ \quad + h(p_z \star D) & 0 \leq D \leq q \\ 0 & q \leq D, \end{cases} \quad (14)$$

where  $R(D)$  is the rate distortion function of  $X$  without side information

$$R(D) = h(q) - h(D), \quad (15)$$

$h$  is the binary entropy function  $h(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha)$ ,  $\alpha \in [0, 1]$ , and  $\star$  is the cyclic convolution

$$p_z \star q = p_z(1 - q) + (1 - p_z)q.$$

The characterization (14) can be written also in the following form

$$R_{ck}(D) = G(D) + h(q) - h(p_z \star q), \quad 0 \leq D \leq q, \quad (16)$$

where  $G(D)$  is the function defined by Wyner and Ziv in [9]

$$G(D) = h(p_z \star D) - h(D). \quad (17)$$

It is interesting to compare  $R_{ck}(D)$  to the Wyner-Ziv rate distortion function  $R_{WZ}(D)$ . Note that for binary sources, only the case of doubly-symmetric source ( $q = 1/2$ ) is solved. For  $q = 1/2$ , the CK rate distortion function is given by

$$R_{ck}(D) = G(D) \quad 0 \leq D \leq 1/2. \quad (18)$$

As shown in [9], the Wyner-Ziv rate distortion function for this source is the *lower convex envelope* (lce) of  $g(D)$ , a function closely related to  $G(D)$ :

$$g(D) = \begin{cases} G(D) & 0 \leq D \leq q_z \\ 0 & q_z \leq D \end{cases} \quad (19)$$

$$R_{WZ}(D) = \text{lce}(g(D)) = \inf_{\theta, \eta} [\theta (h(p_z \star \eta) - h(\eta))] \quad (20)$$

where the infimum is over all  $\theta, \eta$  such that

$$0 \leq \theta \leq 1, \quad 0 \leq \eta \leq p_z \quad (21)$$

and

$$D = \theta\eta + (1 - \theta)p_z. \quad (22)$$

Since  $G(D)$  is a convex function of  $D$ , the Wyner-Ziv rate distortion function coincides with  $g(D)$  in the region  $0 \leq D \leq D_c$  for  $D_c$  that solves the equation

$$\frac{G(D_c)}{D_c - p_z} = G'(D_c), \quad (23)$$

and for  $D_c \leq D \leq q_z$  it is given by the straight line connecting the points  $(G(D_c), D_c)$  and  $(p_z, 0)$ . Hence, in the region  $0 \leq D \leq D_c$ , the functions  $R_{WZ}(D)$  and  $R_{ck}(D)$  coincide. This is stated in the next corollary.

*Corollary 1:* For the binary doubly symmetric source with Hamming distortion measure, no penalty is incurred due to the common knowledge constraint in the region  $0 \leq D \leq D_c$ .

Figure 2 shows the curves  $R_{WZ}(D)$ ,  $R_{ck}(D)$ , and  $R(D)$ , for the doubly symmetric binary source.

*Example 2:* Gaussian source and square error distortion measure. As in [8], the result stated in Theorem 1, which is proved for finite alphabets, can be extended to infinite (discrete or continuous) alphabets. Note that also in [8], the distortion measure is bounded on bounded sets, and this is satisfied by the square error distortion measure.

In the current example,  $X \sim \mathcal{N}(0, \sigma_X^2)$ , the side information  $Y$  is given by

$$Y = X + Z \quad (24)$$

where  $Z \sim \mathcal{N}(0, \sigma_Z^2)$  and is independent of  $X$ . The distortion measure is the square error,  $d(x, \hat{x}) = (x - \hat{x})^2$ . The CK rate-distortion function of this source is

$$R_{ck}(D) = \frac{1}{2} \log \left( \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Z^2} \cdot \frac{D + \sigma_Z^2}{D} \right). \quad (25)$$

As shown in [8], the Wyner-Ziv rate distortion function for the Gaussian source coincides with the rate distortion function when  $Y$  is known at both sides, encoder and decoder, and is given by

$$R_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2} \log \left( \frac{\sigma_X^2 \sigma_Z^2}{(\sigma_X^2 + \sigma_Z^2)D} \right). \quad (26)$$

Figure 3 shows the curves  $R_{WZ}(D)$ ,  $R_{ck}(D)$ , and  $R(D)$  for the Gaussian source.

#### IV. JOINT SOURCE-CHANNEL CODING FOR THE DEGRADED BROADCAST CHANNEL

It is well known that Shannon's source-channel separation theorem does not hold in networks. The two most studied examples are the transmission of sources over the degraded broadcast channel (DBC), and over the multiple access channel (MAC). For these two problems, counter examples are known that show that separation is not an optimal strategy [1], [2], [4]. In this section the implications of the CK constraint on the problem of transmission of sources over the degraded broadcast channel are examined. In particular, it is shown that under the common knowledge constraint, a given distortion pair  $(D_1, D_2)$  can be achieved if and only if the successive

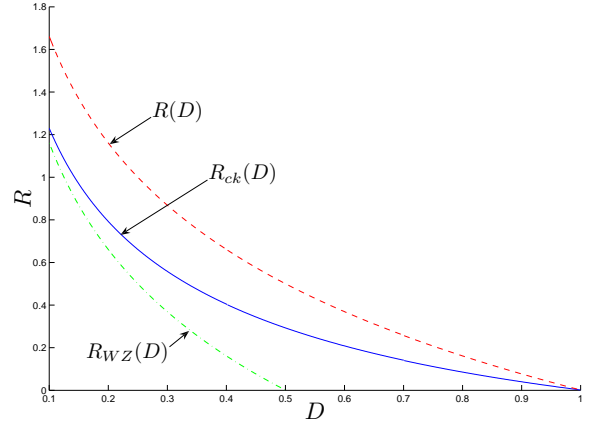


Fig. 3. Rate distortion curves for the Gaussian source  $X \sim \mathcal{N}(0, 1)$ , with side information  $Y = X + Z$ , where  $Z \sim \mathcal{N}(0, 1)$ , independent of  $X$ .

refinement rate region of the source with distortions  $(D_1, D_2)$ , intersects the capacity region of the DBC. That is, under a common knowledge requirement, a separation theorem holds. While this result can sound intuitive at first, note that in general a common knowledge constraint does not imply the optimality of separation. For example, common knowledge is automatically satisfied in lossless transmission. However, as seen from the transmissibility result of Cover, El Gamal, and Salehi [1], separation strategy is not optimal in lossless transmission of a joint source over the MAC.

Lossy transmission of a source  $X$  over the general broadcast channel (BC) under common knowledge constraint is defined as follows. Let  $P_{V_1, V_2|U}$  be a BC with input  $U$  and outputs  $V_1, V_2$ . Let the finite sets  $\hat{\mathcal{X}}_1$  and  $\hat{\mathcal{X}}_2$  stand for the two reconstruction alphabets at user 1 and user 2, respectively. Let  $d_j : \mathcal{X} \times \hat{\mathcal{X}}_j \rightarrow [0, \infty)$ , be a single letter distortion measure at user  $j$ ,  $j = 1, 2$ . As in Section II, both distortion measures are finite. Distortion between sequences is measured in an additive-normalized manner, as in (1).

*Definition 2:* An  $(n, m, D_1, D_2, \epsilon)$  CK code for transmission of the source  $X$  over the BC  $P_{V_1, V_2|U}$  consists of an encoder map

$$f : \mathcal{X}^n \rightarrow \mathcal{U}^m$$

a pair of decoding maps

$$g_j : \mathcal{V}_j^m \rightarrow \hat{\mathcal{X}}_j^n, \quad j = 1, 2$$

and a pair of sender reconstruction maps

$$\psi_j : \mathcal{X}^n \rightarrow \hat{\mathcal{X}}_j^n$$

such that the distortions are bounded by  $D_1$  and  $D_2$

$$\mathbb{E}d_j(X^n, g_j(V_j^m)) \leq D_j, \quad j = 1, 2$$

and with high probability, the maps  $\psi_j$  can produce, at the sender side, the reconstructions available at the decoders

$$P(\psi_j(X^n) \neq g_j(V_j^m)) \leq \epsilon, \quad j = 1, 2.$$

A distortion pair  $(D_1, D_2)$  is said to be *CK-achievable with bandwidth expansion factor*  $\rho$  ( $\rho > 0$ ) if for every  $\epsilon > 0$  and sufficiently large  $n$  there exists an  $(n, \rho n, D_1 + \epsilon, D_2 + \epsilon, \epsilon)$  CK code for transmission of  $X$  over  $P_{V_1, V_2|U}$ . The *CK-distortion region*  $\mathcal{D}_{ck}(\rho)$ , is the set of all distortion pairs  $(D_1, D_2)$  that are CK-achievable with bandwidth expansion ratio  $\rho$ .

Since the two decoders in Definition 2 do not cooperate, the region  $\mathcal{D}_{ck}(\rho)$  depends on the channel  $P_{V_1, V_2|U}$  only via its conditional marginal distributions  $P_{V_1|U}$  and  $P_{V_2|U}$ . This resembles the situation in channel coding for the general BC. The capacity region  $\mathcal{C}$  of the DBC  $P_{V_1, V_2|U}$  is given by the convex hull of the set of all rate pairs  $(R_1, R_2)$  satisfying

$$\begin{aligned} R_1 &\leq I(Z; V_1) \\ R_2 &\leq I(U; V_2|Z) \end{aligned} \quad (27)$$

for some  $P_{Z,U,V_1,V_2} = P_Z P_{U|Z} P_{V_1,V_2|U}$ , where  $Z$  is an external random variable, taking values in a set  $\mathcal{Z}$ , with  $|\mathcal{Z}| \leq \min\{|\mathcal{U}|, |\mathcal{V}_1|, |\mathcal{V}_2|\}$ . The stronger user of a DBC can detect all the messages that are intended for the weaker user, which is user 1 in our case. Denote by  $R'_2$  the total rate that can be detected by the stronger user, and denote by  $\mathcal{C}'$  the corresponding capacity region. The region  $\mathcal{C}'$  is the convex hull of the set of all rate pairs  $(R_1, R'_2)$  satisfying

$$\begin{aligned} R_1 &\leq I(Z; V_1) \\ R'_2 = R_1 + R_2 &\leq I(Z; V_1) + I(U; V_2|Z) \end{aligned} \quad (28)$$

for some  $P_{Z,U,V_1,V_2} = P_Z P_{U|Z} P_{V_1,V_2|U}$  as in (27). The inclusion  $\mathcal{C} \subseteq \mathcal{C}'$  follows immediately by summing the two coordinates in (27). The other direction is a result of the channel being degraded.

Successive refinement (SR), or hierarchical coding of sources, was first defined and studied by Koshelev [5], and later independently by Equitz and Cover [3] and Rimoldi [6]. In particular, it is shown in these works that a rate pair  $(R_1, R'_2)$  is in  $\mathcal{R}_X(D_1, D_2)$  if and only if there exists a pair of random variables  $\hat{X}_1, \hat{X}_2$ , with alphabets  $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2$ , such that

$$\begin{aligned} R_1 &\geq I(X; \hat{X}_1) \\ R'_2 &\geq I(X; \hat{X}_1 \hat{X}_2), \end{aligned} \quad (29)$$

and

$$\begin{aligned} \mathbb{E}d_1(X, \hat{X}_1) &\leq D_1 \\ \mathbb{E}d_2(X, \hat{X}_2) &\leq D_2. \end{aligned} \quad (30)$$

A formal definition of a successive refinement code, an operational definition of the successive refinement region

$\mathcal{R}_X(D_1, D_2)$ , and the proof that it indeed satisfies a single letter characterization as in (29) and (30), can be found in [5] or [6]. For a given bandwidth expansion factor  $\rho > 0$ , denote by  $\rho\mathcal{C}'$  the set of rate pairs  $(R_1, R'_2)$  such that  $(R_1/\rho, R'_2/\rho) \in \mathcal{C}'$ . The main result on  $\mathcal{D}_{ck}(\rho)$  is stated next.

*Theorem 2:* For any discrete memoryless degraded broadcast channel and discrete memoryless source with bounded distortion measures,  $\mathcal{D}_{ck}(\rho)$  consists of all distortion pairs  $(D_1, D_2)$  for which

$$\mathcal{R}_X(D_1, D_2) \cap \rho\mathcal{C}' \neq \emptyset.$$

Thus, for the transmission of a source over the DBC under the CK constraint, a separation strategy yields optimal distortion pairs. However, unlike the situation in single user systems, here the separation does not imply that the design of the channel code and of the source code are completely decoupled of each other. Unlike the situation in single user systems, the code designer has degrees of freedom in deciding where on the boundary of the capacity/rate distortion region the rate pair resides. Some coordination between the designers of the source code and the channel code is necessary, to guarantee that the designed rate pairs match each other.

#### ACKNOWLEDGEMENT

This research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 280/07).

#### REFERENCES

- [1] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648-657, November 1980.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, pp. 269-274, Mar. 1991.
- [4] M. Gastpar, *To code or not to code*, Ph.D. Dissertation, EPFL Lausanne, 2003.
- [5] V. N. Koshelev, "Hierarchical coding of discrete sources," *Probl. Peredachi Inform.*, vol. 16, no. 3, pp. 31-49, 1980. English translation: vol. 16, pp. 186-203, 1980.
- [6] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, pp. 253-259, Jan. 1994.
- [7] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, No. 12, pp. 5218-5239, December 2006.
- [8] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: general sources," *Inf. Contr.*, vol. 38, pp. 60-80, 1978.
- [9] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1-10, Jan. 1976.