The No-Rate-Loss of Wyner-Ziv Coding in the Z-Channel Correlation Case

Nikos Deligiannis, *Member, IEEE,* Andrei Sechelea, Adrian Munteanu, *Member, IEEE,* and Samuel Cheng, *Senior Member, IEEE*

Abstract*—In this letter, we present the exact Wyner-Ziv limit for binary source coding with side information correlated by a Z-channel and with the Hamming-distance as distortion metric. A surprising discovery is that, unlike the general case, Wyner-Ziv coding under this setup does not suffer a rate loss compared to source coding with side information available at both the encoder and the decoder. A detailed proof is presented with further verification through Blahut-Arimoto simulations. Highlighting the significance of our finding, we note that this is only the third case for which Wyner-Ziv coding is shown not to suffer a rate loss after the quadratic Gaussian case, proven by Wyner and Ziv, and its generalization by Pradhan *et al.

***Index Terms*—Source coding with side information, Wyner-Ziv coding, rate-distortion performance, no rate loss.**

I. Introduction

WYNER-ZIV coding [1] refers to the problem of source coding with side information exclusively available at the decoder. The Wyner-Ziv theory states that a rate loss generally occurs compared to source coding in which both the encoder and the decoder are fully informed of the side information. Wyner and Ziv established this rate loss in the case of a binary source with binary symmetric channel correlation and the Hamming-distance distortion metric [1]. They also proved that the rate loss vanishes for memoryless jointly Gaussian sources and the squared-error distortion metric (quadratic Gaussian case) [1]. Pradhan *et al.* [2] generalized this result to sources defined by the sum of arbitrarily distributed side information and independent Gaussian noise. Zamir [3] proved that the Wyner-Ziv rate loss is upper bounded by 0.5 bits/sample for generic sources with squared-error distortion metric and by 0.22 bits/sample for binary sources with Hamming-distance distortion metric.

Although Wyner and Ziv investigated the doubly symmetric binary source coding case, limited light has been shed until now on the case where the correlation is expressed by an asymmetric channel. Recent advances in Wyner-Ziv video coding, however, have demonstrated the benefit of adopting an asymmetric correlation channel model that can lead to performance improvements compared to using a symmetric channel model [4]. In [5], Varodayan *et al.* studied the performance of Low-Density Parity-Check Accumulate codes for practical Slepian-Wolf coding [6] of a binary source with decoder side information under the most representative asymmetric correlation channel model, that is, the Z-channel.

Motivated by these advancements, we study the rate-distortion performance of binary source coding with side information when the correlation is expressed by a Z-channel and the Hamming-distance is used as distortion metric. For coding with side information available to both the encoder and the decoder the rate-distortion function is derived in [7]. However, for the Wyner-Ziv coding setting this function is unknown. In this work, we derive the rate-distortion function for Wyner-Ziv coding under the aforementioned setup and surprisingly, we prove that there is no rate loss compared to source coding with side information available to both the encoder and the decoder. Our theoretical proof, which is further verified by Blahut-Arimoto simulations, contributes a fundamental result as until now the Wyner-Ziv no-rate-loss property is only known and proven for the quadratic Gaussian case [1] and its extension [2].

II. Problem Setting

Consider the problem of source coding with a fidelity criterion and let X, Y and $\hat{X}$ be identically and independently distributed (i.i.d.) random variables, denoting the source, the side information and the reconstructed source, respectively. Moreover, let the corresponding binary alphabets be $\mathcal{X}$, $\mathcal{Y}$ and $\hat{\mathcal{X}}$ and, the distortion measure be the Hamming-distance, with $d(x,\hat{x}) = 0,$ if $x = \hat{x}$, or $d(x,\hat{x}) = 1$, otherwise. The dependence between the source X and the side information Y is expressed by a Z-channel, with crossover probabilities

$$p(y|x) = \begin{cases} 0 \text{ , if } x = 0 \text{ and } y = 1 \\ p_0, \text{if } x = 1 \text{ and } y = 0 \end{cases} \quad (1)$$

A. Source Coding with Encoder-Decoder Side Information

In the scenario where the side information is available both at the encoder and the decoder the rate-distortion (R-D) function is given by

$$R_{X|Y}(D) = \inf_{p(\hat{x}|x,y) E[d(X,\hat{X})] \leq D} I(X;\hat{X}|Y), \quad (2)$$

where $E[\cdot]$ denotes the expectation operator and D is the target distortion. A closed form expression of the R-D function for the case of Z-channel correlation was given in [7], that is,

$$R^Z_{X|Y}(D) = (1-q+qp_0)\left[h(\frac{qp_0}{1-q+qp_0}) - h(\frac{D}{1-q+qp_0})\right], \quad (3)$$

N. Deligiannis is with the Electronic and Electrical Engineering Dept, University College London, WC1E 7JE, UK (e-mail: n.deligiannis@ucl.ac.uk).

A. Sechelea and A. Munteanu are with the Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels 1050, Belgium, and also with iMinds, Ghent 9050, Belgium (e-mail: {atsechel, acmuntea}@etro.vub.ac.be).

S. Cheng is with the School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK 74135 USA (e-mail: samuel.cheng@ou.edu)

where $q = \Pr[X = 1]$ parameterizes the probability distribution of the binary source to be encoded and $h(\cdot)$ is the binary entropy function, $h(p) = -p\log_2(p) - (1-p)\log_2(1-p)$, with $p \in [0, 1]$.

B. Source Coding with Decoder Side Information

In the Wyner-Ziv coding case, when the side information is only available at the decoder, the R-D function is given by [1]

$$R_{WZ}(D) = \inf_{p(u|x)p(\hat{x}|u,y)\ E[d(X,\hat{X})]\leq D} I(X;U|Y) \quad (4)$$

where $\hat{X} = f(U, Y)$ and U is an auxiliary random variable, satisfying the Markov inequalities: $U \leftrightarrow X \leftrightarrow Y$ and $X \leftrightarrow (U, Y) \leftrightarrow \hat{X}$. From Carathéodory's theorem [8], the cardinality of U is upper bounded as $|\mathcal{U}| \leq |\mathcal{X}|+1$. Generally, $R_{WZ}(D) \geq R_{X|Y}(D)$, namely, for the same distortion, Wyner-Ziv coding suffers a rate loss compared to coding with encoder and decoder side information [1], [3].

Next, we derive a closed form expression for the R-D function $R^Z_{WZ}(D)$, for the Wyner-Ziv problem, in the case of the Z-channel correlation, and show that, in this particular case, there is no rate loss w.r.t. source coding with encoder and decoder side information, i.e., $R^Z_{WZ}(D) = R^Z_{X|Y}(D)$.

III. Main Result

Theorem 1: Consider binary source coding in the presence of side information with the Hamming-distance as distortion metric. When the correlation between the source and the side information is expressed by a Z-channel, Wyner-Ziv coding does not suffer a rate loss compared to source coding with side information available at both the encoder and the decoder. Specifically,

$$R^Z_{X|Y}(D) = R^Z_{WZ}(D) = (1-q+qp_0)\left[h\left(\frac{qp_0}{1-q+qp_0}\right) - h\left(\frac{D}{1-q+qp_0}\right)\right]. \quad (5)$$

Proof: The quantity to be minimized in equation (4) can be rewritten as

$$I(X;U|Y) = H(U|Y) - H(U|X). \quad (6)$$

For the moment, let U be binary and the transition probabilities given by[1]

$$p(U|X) = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}. \quad (7)$$

Note that we assume $p(U|X)$ to be generally asymmetric, while in the doubly symmetric binary source case, studied by Wyner and Ziv [1], this channel was considered symmetric.

From (1), the inverse Z-channel between Y and X is

$$p(X|Y) = \begin{bmatrix} \frac{1-q}{1-q+qp_0} & \frac{qp_0}{1-q+qp_0} \\ 0 & 1 \end{bmatrix}. \quad (8)$$

[1]For binary channels, we follow the notation of Silverman in [9].

Then, the channel between Y and U is given by the concatenation of the two above mentioned channels, namely,

$$p(U|Y) = p(X|Y)\ p(U|X) = \begin{bmatrix} \frac{(1-q)\alpha+qp_0\beta}{1-q+qp_0} & 1-\frac{(1-q)\alpha+qp_0\beta}{1-q+qp_0} \\ \beta & 1-\beta \end{bmatrix}. \quad (9)$$

The entropies in (6) can be therefore expressed as

$$\begin{aligned} H(U|X) &= -\sum_{x,u} p(u,x)\log_2(p(u|x)) \\ &= -\sum_x p(x)\sum_u p(u|x)\log_2(p(u|x)) \\ &= (1-q)\ h(\alpha) + q\ h(\beta), \end{aligned} \quad (10)$$

and

$$\begin{aligned} H(U|Y) &= -\sum_{y,u} p(u,y)\log_2(p(u|y)) \\ &= -\sum_y p(y)\sum_u p(u|y)\log_2(p(u|y)) \\ &= (1-q+qp_0)\ h\left(\frac{(1-q)\alpha+qp_0\beta}{1-q+qp_0}\right) \\ &\quad + q(1-p_0)\ h(\beta). \end{aligned} \quad (11)$$

Replacing (10) and (11) in (6) gives

$$\begin{aligned} I(X;U|Y) &\triangleq R(\alpha,\beta) \\ &= (1-q+qp_0)\ h\left(\frac{(1-q)\alpha+qp_0\beta}{1-q+qp_0}\right) \\ &\quad - (1-q)\ h(\alpha) - qp_0\ h(\beta). \end{aligned} \quad (12)$$

In order to determine the corresponding distortion, we observe that for a fixed pair (u, y), x is governed by the conditional distribution $p(x|u, y)$. Since the decoder can only make a deterministic decision given u and y, the best choice is to output $f(u, y) = \arg\max_x p(x|u, y)$ when it sees (u, y). $1 - p(f(u, y)|u, y)$ is the probability that something else will happen. Therefore, the error rate (distortion) conditioned on (u, y) is $1 - p(f(u, y)|u, y)$ and thus the average distortion is given by

$$D = \sum_{u,y}(1-p(f(u,y)|u,y))p(u,y). \quad (13)$$

Given that

$$p(x|u,y) = \frac{p(x,y,u)}{p(y,u)} = \frac{p(u|x)p(x,y)}{p(y,u)}, \quad (14)$$

we observe the following:

$$p(X,Y) = \begin{bmatrix} (1-q) & 0 \\ qp_0 & q(1-p_0) \end{bmatrix}, \quad (15)$$

$$p(X=0,Y,U) = \begin{bmatrix} \alpha(1-q) & (1-\alpha)(1-q) \\ 0 & 0 \end{bmatrix}, \quad (16)$$

and

$$p(X=1,Y,U) = \begin{bmatrix} \beta qp_0 & (1-\beta)qp_0 \\ \beta q(1-p_0) & (1-\beta)q(1-p_0) \end{bmatrix}. \quad (17)$$

Equations (16) and (17) are straightforward from the multiplication of (15) and (7).

Through direct computation, (13) becomes

$$D = \sum_u \min_x \ p(x, y=0, u) = \min(\alpha(1-q), \beta q p_0) + \min((1-\alpha)(1-q), (1-\beta) q p_0). \quad (18)$$

Letting $0 \le q \le \frac{1}{1+p_0}$ (the complementary case yields the same results), and following the variation of α, the values of the overall distortion will be:

- If $\alpha(1-q) < \beta q p_0$ then $D = \alpha(1-q) + (1-\beta) q p_0$;
- If $\alpha(1-q) > (1-q) - (1-\beta) q p_0$ then $D = \beta q p_0 + (1-\alpha)(1-q)$;
- If $\alpha \in [\beta q p_0, (1-q) - (1-\beta) q p_0]$ then $D = q p_0$.

The following holds

$$D(\alpha,\beta) \triangleq \min(\alpha(1-q) + (1-\beta) q p_0, \beta q p_0 + (1-\alpha)(1-q), q p_0). \quad (19)$$

Equations (12) and (19) enable us to formulate a Lagrangian optimization problem, as follows:

$$J(\alpha,\beta) = R(\alpha,\beta) + \lambda D(\alpha,\beta). \quad (20)$$

We consider $\alpha(1-q) < \beta q p_0$, since the case $D = q p_0$ is not interesting ($\hat{X} = Y$) and $\alpha(1-q) > (1-q) - (1-\beta) q p_0$ yields the same solution. Computing the partial derivatives with respect to α and β and setting them to zero gives

$$\frac{\partial J}{\partial \alpha} = \log_2\left(\frac{(1-\alpha)(1-q) + (1-\beta) q p_0}{\alpha(1-q) + \beta q p_0} \frac{\alpha}{1-\alpha}\right) - \lambda = 0,$$

$$\frac{\partial J}{\partial \beta} = \log_2\left(\frac{(1-\alpha)(1-q) + (1-\beta) q p_0}{\alpha(1-q) + \beta q p_0} \frac{\beta}{1-\beta}\right) + \lambda = 0.$$

By summing up the above we obtain the following:

$$\left(\frac{(1-\alpha)(1-q) + (1-\beta) q p_0}{\alpha(1-q) + \beta q p_0}\right)^2 = \frac{(1-\alpha)(1-\beta)}{\alpha\beta}$$

$$\Leftrightarrow \alpha(1-\alpha)(1-q)^2 = (q p_0)^2 \ \beta(1-\beta). \quad (21)$$

Solving for α and β, the system formed by (19) and (21) gives

$$\begin{cases} \alpha = \frac{D(D - q p_0)}{(1-q)(2D - (1-q+q p_0))} \\ \beta = 1 - \frac{D(D-(1-q))}{(q p_0)(2D - (1-q+q p_0))} \end{cases} \quad (22)$$

For every $D \in [0, q p_0]$, equation (22) defines the parameters of the binary channel $p(U|X)$ that yield the minimum rate.

Replacing the above forms for α and β in (12) gives an expression for the desired R-D function. This expression, however, may only be an upper bound. This is because we have assumed that $|\mathcal{U}| = 2$, while, from Carathéodory's theorem [8], it holds that $|\mathcal{U}| \le |\mathcal{X}| + 1$ and thus we may need to have $|\mathcal{U}| = 3$ to attain the Wyner-Ziv limit. Next, we will show that the expression obtained by replacing α and β in (12) matches the lowest R-D bound, which is $R^Z_{X|Y}(D)$. This proves that the expression corresponds to the desired R-D function $R^Z_{WZ}(D)$ and that Wyner-Ziv coding does not suffer a rate loss in this particular case.

We consider $\alpha(1-q) < \beta q p_0$ (for the other relevant case the proof follows the same steps), and express all the entropy quantities in equations (3) and (12) in terms of α and β. Equations (19) and (21) are the expressions for the distortion and $p(U|X)$ achieving the R-D points and the identity to be proven becomes

$$h\left(\frac{q p_0}{1-q+q p_0}\right) - h\left(\frac{D}{1-q+q p_0}\right) = h\left(\frac{\alpha(1-q) + \beta q p_0}{1-q+q p_0}\right) - \frac{(1-q)}{1-q+q p_0} \ h(\alpha) - \frac{q p_0}{1-q+q p_0} \ h(\beta) \quad (23)$$

or, equivalently

$$h\left(\frac{\alpha(1-q) + \beta q p_0}{1-q+q p_0}\right) + h\left(\frac{\alpha(1-q) + (1-\beta) q p_0}{1-q+q p_0}\right) = h\left(\frac{q p_0}{1-q+q p_0}\right) + \frac{(1-q)}{1-q+q p_0} \ h(\alpha) + \frac{q p_0}{1-q+q p_0} \ h(\beta) \quad (24)$$

Using (21) we can derive the following basic identities

$$\frac{\alpha(1-q) + (1-\beta) q p_0}{1-q+q p_0} = \frac{\alpha(1-q)}{\alpha(1-q) + \beta q p_0}, \quad (25)$$

$$\frac{(1-\alpha)(1-q) + \beta q p_0}{1-q+q p_0} = \frac{\beta q p_0}{\alpha(1-q) + \beta q p_0}, \quad (26)$$

$$\frac{\alpha(1-q) + \beta q p_0}{(1-\alpha)(1-q) + (1-\beta) q p_0} = \frac{\beta q p_0}{(1-\alpha)(1-q)}. \quad (27)$$

The following then hold

$$h\left(\frac{\alpha(1-q) + \beta q p_0}{1-q+q p_0}\right) = \log_2(1-q+q p_0) - \frac{\alpha(1-q) + \beta q p_0}{1-q+q p_0} \log_2(\alpha(1-q) + \beta q p_0) - \frac{(1-\alpha)(1-q) + (1-\beta) q p_0}{1-q+q p_0} \log_2((1-\alpha)(1-q) + (1-\beta) q p_0) \quad (28)$$

and, using (25),

$$h\left(\frac{\alpha(1-q) + (1-\beta) q p_0}{1-q+q p_0}\right) = h\left(\frac{\alpha(1-q)}{\alpha(1-q) + \beta q p_0}\right) = \log_2(\alpha(1-q) + \beta q p_0) - \frac{\alpha(1-q)\log_2(\alpha(1-q))}{\alpha(1-q) + \beta q p_0} - \frac{\beta q p_0 \log_2(\beta q p_0)}{\alpha(1-q) + \beta q p_0}. \quad (29)$$

Summing up (28) and (29), and using (27), gives

$$h\left(\frac{\alpha(1-q) + \beta q p_0}{1-q+q p_0}\right) + h\left(\frac{\alpha(1-q) + (1-\beta) q p_0}{1-q+q p_0}\right) = \log_2(1-q+q p_0) + \frac{(1-\alpha)(1-q) + (1-\beta) q p_0}{1-q+q p_0} \times \log_2\left(\frac{\beta q p_0}{(1-\alpha)(1-q)}\right) - \frac{\alpha(1-q)\log_2(\alpha(1-q))}{\alpha(1-q) + \beta q p_0} - \frac{\beta q p_0 \log_2(\beta q p_0)}{\alpha(1-q) + \beta q p_0}. \quad (30)$$

By using (25) and (26) for the last two terms, and following basic arithmetic operations, and convenient regrouping of the

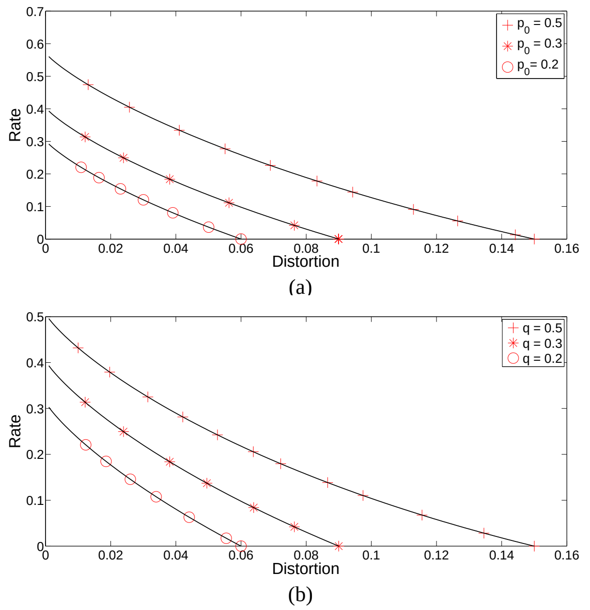

Fig. 1. Derived $R^Z_{WZ}(D) = R^Z_{X|Y}(D)$ and corresponding R-D points from the Blahut-Arimoto algorithm: (a) $p_0 = 0.5, p_0 = 0.3, p_0 = 0.2$ and $q = 0.3$; (b) $p_0 = 0.3$, and $q = 0.5, q = 0.3, q = 0.2$

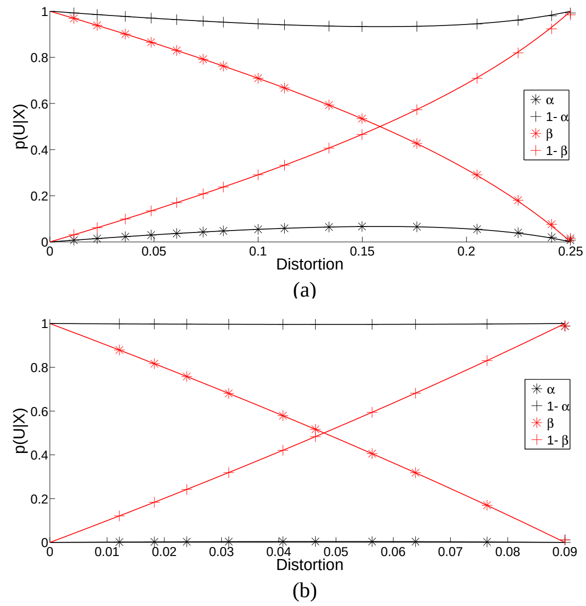

Fig. 2. $p(U|X)$ obtained from the Blahut-Arimoto algorithm: (a) $p_0 = 0.5$ and $q = 0.5$; (b) $p_0 = 0.3$ and $q = 0.3$

terms, this becomes

$$h\left(\frac{\alpha(1-q)+\beta q p_0}{1-q+qp_0}\right) + h\left(\frac{\alpha(1-q)+(1-\beta)qp_0}{1-q+qp_0}\right) = \log_2(1-q+qp_0) - \frac{1-q}{1-q+qp_0}\log_2(1-q) - \frac{qp_0}{1-q+qp_0}\log_2(qp_0) - \frac{1-q}{1-q+qp_0}\times [-\alpha\log_2(\alpha)-(1-\alpha)\log_2(1-\alpha)] - \frac{qp_0}{1-q+qp_0}[-\beta\log_2(\beta)-(1-\beta)\log_2(1-\beta)]. \quad (31)$$

It is straightforward to show that this is equivalent to (24). ■

IV. Blahut-Arimoto Simulations

In order to verify our theory, we used an implementation of the Blahut-Arimoto algorithm for the R-D problem with two-sided state information [10], adapted such that it generates the R-D points for binary source coding with decoder side information, under Z-channel correlation. We vary the distribution of the source X, by modifying $q = \Pr[X = 1]$, as well as the crossover probability of the Z-channel, by modifying p_0.

Fig. 1(a) presents the Wyner-Ziv R-D performance for a uniform source with varying crossover probability for the Z-channel, while Fig. 1(b) depicts the R-D when the correlation channel is kept constant, and the source distribution is varying. The figures corroborate the perfect match of our theoretical R-D function [see (5)] with the experimental R-D points obtained with the Blahut-Arimoto algorithm, full lines for the former and, respectively, discrete values for the latter.

Moreover, the variation of $p(U|X)$ with the distortion, as obtained by the Blahut-Arimoto algorithm for two different pairs of p_0 and q, is depicted in Figs. 2(a) and (b). It can be observed that $p(U|X)$ exhibits an asymmetric behavior. In particular, it is noteworthy that, for $(p_0, q) = (0.3, 0.3)$, $p(\bar{U}|X)$—i.e., the channel between X and the inverse of U—approximates a Z-channel ($\alpha \approx 1$) for any $D \in [0, 0.09]$.

V. Conclusion

Motivated by practical applications, we have derived the R-D function for Wyner-Ziv coding for a binary source with side information correlated by a Z-channel and with the Hamming-distance as distortion metric. A surprising discovery is that, under this setup, Wyner-Ziv coding does not suffer a rate loss compared to source coding with encoder and decoder side information. This is only the third case for which the Wyner-Ziv no-rate-loss property is shown and proven after the quadratic Gaussian case [1] and its extension [2].

References

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[2] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, 2003.

[3] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 2073–2084, 1996.

[4] N. Deligiannis, J. Barbarien, M. Jacobs, A. Munteanu, A. Skodras, and P. Schelkens, "Side-information-dependent correlation channel estimation in hash-based distributed video coding," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1934–1949, Apr. 2012.

[5] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Process., Special Issue on Distributed Source Coding*, vol. 86, no. 11, pp. 3123–3130, Nov. 2006.

[6] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[7] Y. Steinberg, "Coding and common reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4995–5010, Nov. 2009.

[8] H. G. Eggleston, *Convexity*. Cambridge University Press, 1966, vol. 47.

[9] R. Silverman, "On binary channels and their cascades," *IRE Trans. Inf. Theory*, vol. 1, no. 3, pp. 19–27, 1955.

[10] S. Cheng, V. Stankovic, and Z. Xiong, "Computing the channel capacity and rate-distortion function with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4418–4425, Dec. 2005.