

A Universal Wyner-Ziv Scheme for Discrete Sources

Shirin Jalali

Department of Electrical Engineering
Stanford University
Stanford, CA, 94305, USA
shjalali@stanford.edu;

Sergio Verdú

Department of Electrical Engineering
Princeton University
Princeton, NJ 08540, USA
verdu@princeton.edu;

Tsachy Weissman

Department of Electrical Engineering
Stanford University
Stanford, CA, 94305, USA
tsachy@stanford.edu;

Abstract— We consider the Wyner-Ziv (WZ) problem of rate-distortion coding with decoder side information, for the case where the source statistics are unknown or non-existent. A new family of WZ coding algorithms is proposed and its universal optimality is proven. Encoding is based on a sliding window operation followed by LZ compression, while decoding is based on a natural extension of the Discrete Universal DENOISER (DUDE) algorithm to the case where side information is present. The effectiveness of our approach is illustrated with experiments on binary images using a low complexity algorithm motivated by our class of universally optimal WZ codes.

I. INTRODUCTION

Consider the basic setup shown in Fig. 1 which has the following components: a source with unknown statistics, a known discrete memoryless channel (DMC), a noiseless channel with rate constraint R , and a decoder. The goal is to minimize the distortion between the source and the reconstructed signals by optimally designing the encoder and decoder. This is the problem of rate-distortion coding with decoder side information, commonly known as the WZ coding problem since the seminal paper [4]. Like the original rate-distortion problem without side information, the problem of finding practical schemes that get arbitrarily close to a given point on the rate-distortion curve is still largely open. If the source has memory, e.g. a Markov source, the rate-distortion function itself is not explicitly known, let alone a scheme for achieving it.

An alternative view of this problem is as a denoising problem where the denoiser, in addition to the noise-corrupted data, has access to a fidelity-boosting encoded sequence conveyed to it via a noiseless channel of capacity R . Note that in these two viewpoints the role of the main and side-information signals is interchanged. In this paper, we adopt the latter, and suggest a new algorithm for WZ coding of a source with unknown statistics. We show that, for a stationary ergodic source, the algorithm is asymptotically optimal in the sense that its average expected loss per symbol converges to the minimum attainable expected loss.

Some progress towards practical WZ coding schemes has been made in recent years, as seen, e.g. in [9], [10], [11], [12], [13]. The proposed schemes, however, operate under specific assumptions of a known (usually memoryless) source and side information channel. Practical schemes for more general source and/or channel characteristics have yet to be developed

and, a fortiori, no practical universal schemes for this problem are known.

The problem of WZ coding of a source with unknown statistics was recently considered in [7], where existence of universal schemes in a setting similar to ours is established. Thus, our main innovation is not so much in establishing the existence of universal schemes for this setting as it is in the fact that our schemes suggest a paradigm for WZ coding of discrete sources which is not only practical but it is justified through universal optimality results.

The organization of the remainder of this paper is as follows. In Section II, the notation used throughout the paper is introduced. In Section III, the extension of the DUDE (denoising) algorithm [5] to take advantage of a side information sequence is presented, and it is shown how the asymptotic optimality of the original DUDE carries over to this case as well. In Section IV, sliding block WZ coders are introduced and a theorem on their relationship to WZ block codes is presented. In Section V, our new WZ coding algorithm is presented and its optimality is established. In Section VI we present some experimental results.

II. NOTATION

Let \mathcal{X} , $\hat{\mathcal{X}}$, and \mathcal{Z} denote the source, reconstructed signal, and channel output alphabets respectively. In this paper, for simplicity, we restrict attention to the case where

$$\mathcal{X} = \hat{\mathcal{X}} = \mathcal{Z} = \{\alpha_1, \dots, \alpha_N\},$$

though our derivations and results carry over directly to general finite alphabets. Bold low case symbols, e.g. \mathbf{x} , \mathbf{y} , \mathbf{z} , denote individual sequences. The discrete memoryless channel

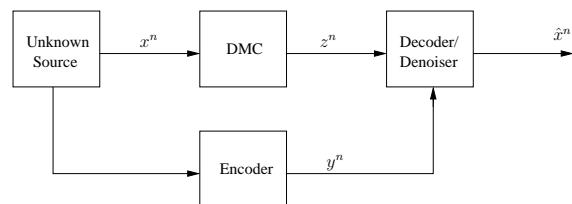


Fig. 1. denoising with side-information

$$\hat{X}^{n,l,m}(z^n, y^n)[i] = \arg \min_{\hat{x} \in \mathcal{A}} \mathbf{r}^T(z^n, y^n, z_{i-l}^{i-1}, z_{i+1}^{i+l}, y_{i-m}^{i+m}) \mathbf{\Pi}^{-1}[\lambda_{\hat{x}} \odot \pi_{z_i}],$$

where, for $\beta \in \mathcal{Z}$, and $t = \max\{l, m\}$

$$\mathbf{r}(z^n, y^n, a^l, b^l, c_{-m}^m)[\beta] = |\{t+1 \leq i \leq n-t : z_{i-l}^{i-1} = a^l, z_{i+1}^{i+l} = b^l, y_{i-m}^{i+m} = c_{-m}^m\}| \quad (1)$$

is described by its transition matrix $\mathbf{\Pi}$, where $\mathbf{\Pi}(i, j)$ denotes the probability of getting α_j at the output of the channel when the input is α_i .

Let $\lambda : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$, be the loss function (fidelity criterion) which measures the loss incurred in denoising (decoding) a symbol α_i to another symbol α_j , which will be represented by a $N \times N$ matrix, $\mathbf{\Lambda} : \{\lambda(\alpha_i, \alpha_j)\}$. Moreover, let $\lambda_m = \max_{i,j} \lambda(\alpha_i, \alpha_j)$, and note that $\lambda_m < \infty$, since the alphabets are finite. The normalized cumulative loss between a source sequence x^n and reconstructed sequence \hat{x}^n , is denoted by

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

Let π_i and λ_j denote the i -th column and the j -th column of $\mathbf{\Pi}$ and $\mathbf{\Lambda}$ matrices respectively, i.e.

$$\mathbf{\Pi} = [\pi_1 | \dots | \pi_N], \quad \mathbf{\Lambda} = [\lambda_1 | \dots | \lambda_N].$$

For two N -dimensional vectors \mathbf{u} and \mathbf{v} , $\mathbf{u} \odot \mathbf{v}$ denotes the N -dimensional vector that results from componentwise multiplication of \mathbf{u} and \mathbf{v} , i.e.

$$\mathbf{u} \odot \mathbf{v}[i] = u_i v_i.$$

III. DUDE WITH SIDE INFORMATION

The problem of noncausal denoising of a sequence corrupted by a DMC whose transition matrix is known is considered in [5]. A practical denoising algorithm, the DUDE algorithm, is presented in [5] along with its asymptotic optimality in both the stochastic setting, where the source is assumed to be stationary, and the semi-stochastic setting, where the source is assumed to be an individual sequence.

Decoding for the WZ problem can also be considered as a denoising problem where the denoiser, in addition to the noisy signal, has access to a fidelity-boosting ‘‘side-information’’ sequence which is designed by the source encoder to be as helpful as possible to the decoder. From this perspective, we are motivated to develop and employ a scheme along the lines of the DUDE that can handle side information, and be asymptotically optimal in senses analogous to those of DUDE.

Consider then the natural extension of the original DUDE algorithm to the problem of denoising with side information presented in (1). The counting process is done simultaneously in both the noisy and the side information sequences. In order to show that the optimality results of [5], carry over to this case, consider a channel $\tilde{\mathbf{\Pi}}$, with input $\tilde{x}_i = (x_i, y_i)$, and output $\tilde{z}_i = (z_i, y_i)$, where z_i is the output of the original channel $\mathbf{\Pi}$, when the input is x_i . Since the newly defined channel inherits its invertibility from that of the original one,

the results of [5] can be applied to this case as well. In other words, provided that $t_n |\mathcal{X}|^{2t_n} = o(n/\log n)$, $\forall \mathbf{x}, \mathbf{y}$

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n - 2t_n} \sum_{i=t_n+1}^{n-t_n} \lambda(x_i, \hat{x}_i) - D_{l_n, m_n}(x^n, y^n, Z^n) \right] = 0 \text{ a.s.},$$

where,

$$D_{l_n, m_n}(x^n, y^n, z^n) = \min_{f: \mathcal{Y} \times \mathcal{Z} \rightarrow \mathcal{X}} \frac{1}{n - 2t_n} \sum_{i=t_n+1}^{n-t_n} \lambda \left(x_i, f(y_{i-l_n}^{i+l_n}, z_{i-m_n}^{i+m_n}) \right),$$

and \hat{x}_i is the output of the denoiser in (1) with parameters l_n, m_n , and $t_n \triangleq \max\{l_n, m_n\}$.

IV. WYNER-ZIV CODING PROBLEM

The majority of achievability proofs in the information theory literature are based on the idea of *random block coding*. In rate-distortion coding, besides block codes, another type of codes are *sliding block codes* which were introduced by Gray, Neuhoff, and Ornstein in [3]. In this method a function with a finite number of symbols slides over the original source output sequence, outputting another sequence that has less entropy, but resembles the original sequence as much as the designer desires. In [2], it is shown that the performance of sliding block codes is equivalent to that of the block codes.

In the following, we first briefly review block and sliding block WZ coding, and establish their fundamental limits.

A. Block coding

A WZ block code of length L and rate R consists of encoding and decoding mappings, E_L and D_L respectively, which are defined as follows:

$$E_L : X^L \rightarrow \{1, 2, \dots, [2^{LR}]\},$$

$$D_L : \{1, 2, \dots, [2^{LR}]\} \times \mathcal{Z}^L \rightarrow \hat{\mathcal{X}}^L.$$

The performance of such code is defined as the average distortion per symbol between the source and its reconstruction sequences, i.e.

$$E[\rho_L(X^L, D_L(Z^L, E_L(X^L)))] \triangleq \frac{1}{L} E \left[\sum_{i=1}^L \lambda(X_i, \hat{X}_i) \right],$$

where $\hat{X}^L = D_L(E_L(X^L), Z^L)$.

The rate distortion pair (R, D) is said to be achievable if for any given $\epsilon > 0$, there exists L, E_L , and D_L , such that $E[\lambda(X^L, D_L(Z^L, E_L(X^L)))] \leq D + \epsilon$. For a given source \mathbf{X} , and memoryless channel described by transition matrix $\mathbf{\Pi}$, the

infimum of all achievable distortions at rate R is called $D_{\mathbf{X},\Pi}$, i.e.

$$D_{\mathbf{X},\Pi}(R) = \inf\{D : (R, D) \text{ is achievable}\}.$$

B. Sliding-Block WZ compression

An extension of the idea of sliding block rate distortion codes is sliding block source coding with decoder side information. In this section, using the techniques of [2], we show that similarly in this case, any performance that is achievable by block codes is achievable also by sliding-block codes.

A WZ sliding block code consists of two time-invariant encoding and decoding mappings f and g . The encoding mapping, f with constraint length $2k + 1$, maps every $2k + 1$ source symbols into a symbol of \mathcal{Y} , i.e. $f : \mathcal{X}^{2k+1} \rightarrow \mathcal{Y}$. This encoder moves over the source sequence and generates the fidelity boosting side information sequence $Y_i = f(X_{i-k}^{i+k})$. On the other hand, the decoding mapping, g , with constraint length $\max\{2l + 1, 2m + 1\}$, maps a block of length $2l + 1$ of the noise corrupted signal and a block of length $2m + 1$ of side information to a reconstruction symbol, i.e. $g : \mathcal{Z}^{2l+1} \times \mathcal{Y}^{2m+1} \rightarrow \hat{\mathcal{X}}$. The decoder slides over the noisy and side information sequences and generates the reconstruction sequence as $\hat{X}_i = g(Z_{i-l}^{i+l}, Y_{i-m}^{i+m})$. The following theorem states that sliding block WZ codes perform at least as well as WZ block codes.

Theorem 4.1: Let (R, D) be an interior point in the (block) WZ rate-distortion region of a stationary source \mathbf{X} and memoryless channel Π . For any given $\epsilon_1 > 0$, there exists a sliding block WZ encoder $f : \mathcal{X}^{2k+1} \rightarrow \mathcal{Y}$, where $\log |\mathcal{Y}| \geq R$, and a sliding block decoder g with parameters l and m , such that

- 1) $E[d(X_i, g(Z_{i-l}^{i+l}, Y_{i-m}^{i+m}))] \leq D + \epsilon_1$, where $Y_i = f(X_{i-k}^{i+k})$,
- 2) $H(\mathbf{Y}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(Y_1, \dots, Y_n) \leq R - \epsilon_2$, for some $\epsilon_2 > 0$.

Proof: The proof is given in the full version of the paper. ■

V. WYNER-ZIV DUDE

In this section, we propose a new WZ coding scheme and prove its asymptotical optimality based on the results established so far. For any given block length n , let E_n^* and D_n^* denote the encoder and the decoder of the scheme respectively. The scheme has a number of parameters, namely l , k , m and δ , that their specific functionality will be clarified as the scheme is described in the sequel.

- 1) **Encoder:** For a given source sequence x^n define $S(x^n, k, R)$ to be the set of all sliding-block mappings of window length $2k + 1$ with the property that their output is a sequence of Lempel-Ziv description length no larger than nR , i.e.

$$S(x^n, k, R) \triangleq \left\{ f : \mathcal{X}^{2k+1} \rightarrow \mathcal{Y} : \frac{1}{n} LZ(f(x^n)) \leq R \right\}.$$

Note that $f(x^n)$ is assumed to be equal to y^n , where $y_i = f(x_{i-k}^{i+k})$ for $k + 1 \leq i \leq n - k$, and $y_i = 0$

otherwise. For each $f \in S$, and integers l and m define

$$V(f, l, m) = \min E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, g(Z_{i-l}^{i+l}, y_{i-m}^{i+m})) \right], \quad (2)$$

where the minimization is over all decoding mappings $g : \mathcal{Z}^{2l+1} \times \{1, \dots, \lceil 2^R \rceil\}^{2m+1} \rightarrow \hat{\mathcal{X}}$. Let $f^*(l, m)$ be the mapping in S that minimizes $V(f)$, i.e.

$$f^*(l, m) = \arg \min_{f \in S} V(f, l, m). \quad (3)$$

Then, the fidelity-boosting encoded sequence is the LZ description of $f_n^*(x^n)$ which is sent to the decoder.

- 2) **Decoder:** Upon obtaining $f_n^*(x^n)$ with an LZ decompressor the decoder employs the DUDE with side information described in Section III, i.e. let the reconstruction signal be $\hat{X}^{n, k_n, m_n}(Z^n, y^n)$, where $y^n = f^*(x^n)$.

The main result of this paper is the following theorem, which shows that the described WZ coding algorithm is asymptotically optimal. Let k_n , l_n , and m_n increase without bound with n , but sufficiently slowly that $t_n |\mathcal{X}|^{t_n} = o(n / \log(n))$. Then,

Theorem 5.1: For any $R \geq 0$, and any stationary ergodic source \mathbf{X} ,

$$\limsup_{n \rightarrow \infty} \rho_n(X^n, D_n^*(Z^n, E_n^*(X^n))) \leq D_{\mathbf{X},\Pi}(R) \quad a.s. \quad (4)$$

Proof: See Appendix A. ■

VI. EXPERIMENTAL RESULTS

In this section we present some experimental results. As mentioned earlier, the demanding aspect of the WZ DUDE algorithm is finding the optimal mapping f^* . Here, instead of looking for the optimal mapping, we use a lossy JPEG encoder. Since except for the encoding of the DC component, JPEG works on the 8×8 blocks separately, it can be considered as a sliding-block encoder of window length 1 working on the super-alphabets formed by 8×8 binary blocks. Fig. 2 and Fig. 3 show the original binary image and its noise-corrupted version under a binary symmetric channel with transition probability $P_e = 0.15$. Fig. 4 shows the JPEG encoded image which requires 0.22 bit per pixel (b.p.p.) after JPEG lossless compression, compared to 0.6 b.p.p. required by the original image. The average distortion between the original image and the encoded one is 0.0556. Fig. 5 shows the result of denoising the noise corrupted image with DUDE algorithm ignoring the fidelity-boosting side information sequence. In this case the resulting average distortion would be 0.0635. On the other hand, Fig. 6 shows the result of denoising the noisy signal when the side information is also taken into account. The decoder/denoiser in this case is WZ DUDE with parameters $l = 1$ and $m = 1$. The final average distortion between the reconstructed image and the original image would be 0.0407.

VII. CONCLUSION

In this paper, the problem of WZ coding of a source with unknown statistics was investigated, and a new WZ coding algorithm, WZ DUDE, was presented and its asymptotical



Fig. 2. Original binary image, 0.6 b.p.p.



Fig. 3. Noise corrupted image, generated by passing the original signal through a BSC(0.15)



Fig. 4. The fidelity-boosting side information image, y , generated by lossy JPEG coding of the original image, 0.22 b.p.p., $\rho(x^n, y^n) = 0.0556$

optimality was established. In order to optimize the scheme one would list all possible mappings that have a certain property and look for the one that gives minimum expected loss. However, we saw that even by a simple encoding mapping, namely an off-the-shelf lossy compressor, it is possible to get considerable improvement compared to the absence of fidelity-boosting compressed information, or to the absence of the noisy signal at the decoder.

APPENDIX A: PROOF OF THEOREM 5.1

First, we prove that for any given $\epsilon > 0$, there exists $N_\epsilon > 0$, such that for $n > N_\epsilon$,

$$E[\rho_n(X^n, D_n^*(Z^n, E_n^*(X^n)))] < D_{\mathbf{X}, \Pi}(R) + \epsilon. \quad (\text{A-1})$$



Fig. 5. Output of DUDE for $l = 1$. $\rho_n(x^n, \hat{x}^n) = 0.0635$



Fig. 6. Output of the WZ-DUDE decoder for $l = m = 1$, and $R = 0.22$ b.p.p., $\rho_n(x^n, \hat{x}^n) = 0.0407$

By definition, $D_{\mathbf{X}, \Pi}(R)$ denotes the infimum of all distortions achievable by Wyner-Ziv coding of source \mathbf{X} at rate R when the DMC is described by Π . Therefore, for any $\epsilon > 0$, $(D + \frac{\epsilon}{4}, R)$ would be an interior point of the rate-distortion region. Hence, by theorem 4.1 for $\epsilon_1 = \frac{\epsilon}{4} > 0$, there exist some $\epsilon_2 > 0$, and a sliding block Wyner-Ziv code with mappings f and g , each one having a finite window length, such that

- 1) $E[d(X_i, g(Z_{i-l}^{i+l}, Y_{i-m}^{i+m}))] \leq D + \frac{\epsilon}{2}$, where $Y_i = f(X_{i-k}^{i+k})$,
- 2) $H(\mathbf{Y}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(Y_1, \dots, Y_n) \leq R - \epsilon_2$, for some $\epsilon_2 > 0$.

On the other hand, the side information process $\{Y_i\}$ generated by sliding-windowing a stationary ergodic process $\{X_i\}$ with a time invariant mapping f , is also a stationary ergodic process. Consequently, since for any stationary ergodic process Lempel-Ziv coding algorithm is an asymptotically optimal lossless compression scheme [6], for any given $\sigma > 0$, there exists $N_\sigma > 0$, such that for $n > N_\sigma$,

$$\frac{1}{n} LZ(Y_1, \dots, Y_n) \leq H(\mathbf{Y}) + \sigma. \quad (\text{A-2})$$

Letting $\sigma = \frac{\epsilon_2}{2}$, and choosing n greater than the corresponding N_σ , yields

$$\frac{1}{n} LZ(Y_1, \dots, Y_n) < R. \quad (\text{A-3})$$

Therefore, for any given $\epsilon > 0$, and any source output sequence, by choosing the block length n sufficiently large, the

mapping f would belong to $S(x^n, k, R)$. On the other hand, since for any individual source sequence x^n , f^* is the mapping in S that defines the side information sequence minimizing the expected distortion, it follows that

$$V(f^*, l, m) < V(f, l, m). \quad (\text{A-4})$$

Moreover, since $V(f, l, m)$ is the minimum accumulated loss attainable by the mappings in $S(n, l, m)$, when the decoder is constrained to be a sliding window decoder with parameters l and m , it is in turns less than the expected distortion obtained by the specific mapping g given by Theorem 4.1, i.e.

$$E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, g^*(Z_{i-l}^{i+l}, \tilde{y}_{i-m}^{i+m})) \right] \leq E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, g(Z_{i-l}^{i+l}, y_{i-m}^{i+m})) \right], \quad (\text{A-5})$$

where $y_i = f(x_{i-k}^{i+k})$ and $\tilde{y}_i = f^*(x_{i-k}^{i+k})$.

The final step, is using the asymptotic optimality of Wyner-Ziv DUDE algorithm in the semi-stochastic setting, which was discussed in Section III. From that result, by choosing the parameters $l < m = o(\log(n))$, the difference between the performance of the Wyner-Ziv DUDE decoding algorithm and the optimal sliding window decoder of the same order goes to zero as the block length goes to infinity. In other words, for any given $\epsilon > 0$, there exists $N'_\epsilon > 0$, such that for $n > N'_\epsilon$,

$$\frac{1}{n-2k} E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, \hat{x}_i) \right] \leq \frac{1}{n-2k} E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, g^*(Z_{i-l}^{i+l}, \tilde{y}_{i-m}^{i+m})) \right] + \frac{\epsilon}{4}, \quad (\text{A-6})$$

where $\hat{x}^n = D_n^*(Z^n, E_n^*(x^n))$. Note that the only uncertainty in (A-6) is due to the channel noise, and the source and side information are assumed to be individual sequences. Combining (A-5) and (A-6), it follows that with probability one

$$\frac{n}{n-2k} E [\rho_n(x^n, D_n^*(Z^n, E_n^*(x^n)))] \leq \frac{1}{n-2k} E \left[\sum_{i=k+1}^{n-k} \lambda(x_i, g(Z_{i-l}^{i+l}, y_{i-m}^{i+m})) \right] + \frac{\epsilon}{4}. \quad (\text{A-7})$$

On the other hand, since $\{(X_i, Y_i)\}_{i=0}^\infty$ is also a stationary ergodic process with super-alphabet $\mathcal{X} \times \mathcal{Y}$, by the ergodic theory, with probability one,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n-2k} \sum_{i=k+1}^{n-k} E [\lambda(x_i, g(Z_{i-l}^{i+l}, y_{i-m}^{i+m}))], \\ = E [\lambda(X_0, g(Z_{-l}^l, Y_{-m}^m))], \\ \leq D_{\mathbf{X}, \Pi} + \frac{\epsilon}{2}. \end{aligned}$$

This means that with probability one, there exists $N''_\epsilon > 0$, such that for $n > N''_\epsilon$,

$$\begin{aligned} \frac{1}{n-2k} \sum_{i=k+1}^{n-k} E [\lambda(x_i, g(Z_{i-l}^{i+l}, y_{i-m}^{i+m}))], \\ \leq D_{\mathbf{X}, \Pi} + \frac{\epsilon}{2} + \frac{\epsilon}{4}. \end{aligned} \quad (\text{A-8})$$

Finally, combining (A-7) and (A-8), and taking $n > N_\epsilon$, where $N_\epsilon = \max\{N_\sigma, N'_\epsilon, N''_\epsilon\}$, yields the desired result as follows

$$\begin{aligned} \frac{n}{n-2k} E [\rho_n(X^n, D_n^*(Z^n, E_n^*(X^n)))] \\ \leq D_{\mathbf{X}, \Pi}(R) + \epsilon. \end{aligned} \quad (\text{A-9})$$

ACKNOWLEDGMENT

The authors would like to thank Erik Ordentlich and Pascal Vontobel for the helpful discussions. They also acknowledge Stanford Graduate Fellowship supporting the first author, and NSF Grant CCR-0312839, and NSF CAREER grant supporting the second and third authors respectively.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.
- [2] R. M. Gray, "Block, Sliding-Block, and Trellis Codes," *János Bolyai Colloquium on Info. Theory*, Keszthely, Hungary, August 1975.
- [3] R. M. Gray, D. L. Neuhoff, and D. S. Ornstein, "Nonblock Source Coding with a Fidelity Criterion," *The Annals of Probability*, vol. 3, pp. 478–491, 1975.
- [4] A. Wyner, and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.
- [5] T. Weissman, and E. Ordentlich, G. Seroussi, S. Verdu and M. Weinberger, "Universal Discrete Denoising: Known Channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 5–28, January 2005.
- [6] T. M. Cover, and J. A. Thomas, *Elements of information theory*, New York: Wiley, 1991.
- [7] N. Merhav and J. Ziv, "On the Wyner-Ziv problem for individual sequences," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 867–873, March 2006.
- [8] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, 93(1):71–83, Jan. 2005.
- [9] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," In *Proc. IEEE Data Compression Conference*, pp. 158–167, Snowbird, UT, March 1999.
- [10] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," In *Proc. IEEE Data Compression Conference (DCC)*, page 1322, Snowbird, UT, March 2003.
- [11] S. D. Servetto, "Lattice quantization with side information," In *Proc. IEEE Data Compression Conference (DCC)*, pp. 510–519, Snowbird, UT, March 2000.
- [12] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes", In *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003.
- [13] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. on Inf. Theory*, 48(6):1250–1276, June 2002.