A Wyner Ziv Codec Based on Equalization at the Decoder*

Azadeh Vosoughi
School of ECE, Cornell University
432 Rhodes Hall, Ithaca, NY 14853
av78@cornell.edu

Anna Scaglione
School of ECE, Cornell University
325 Rhodes Hall, Ithaca, NY 14853
anna@ece.cornell.edu

Abstract

In this paper we consider the problem of lossy coding of correlated vector sources with uncoded side information available at the decoder. In particular, we consider lossy coding of source $\boldsymbol{x} \in \mathbb{R}^N$ which is correlated with source $\boldsymbol{y} \in \mathbb{R}^N$, known at the decoder. The general non-linear mapping between $\boldsymbol{y}$ and $\boldsymbol{x}$ capturing the correlation between the two sources can be approximated through a linear model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}$ in which $\boldsymbol{n}$ is independent of $\boldsymbol{x}$. This model can be viewed as a fictitious communication channel with input $\boldsymbol{x}$ and output $\boldsymbol{y}$. Utilizing a powerful signal processing technique, namely channel equalization, we convert the original vector source coding problem into a set of manageable scalar source coding problems. The scalar source coding problems can be solved using the existing distributed source coding algorithms that are primarily designed for the simple correlation model $y = x + n$ where x and y are scalar jointly Gaussian sources.

1 Introduction

Distributed source coding (DSC) refers to the compression of multiple correlated sources that do not communicate with each other. In [1], Slepian and Wolf showed that two arbitrary correlated memoryless discrete-alphabet scalar sources x and y can be separately compressed at a rate approaching to their joint entropy, provided that the two sources know the joint distribution characterizing their correlation and decoding of the two sources is done jointly. In [2] Wyner and Ziv extended the work in [1] to the case where encoding of source x is with respect to a fidelity criterion rather than lossless, assuming that the decoder has access to the source y, and provided the rate-distortion function for the source encoding. Observing a close connection of DSC to channel coding, Wyner [3] proposed the use of linear binary block codes to achieve the Slepian-Wolf bound for two binary correlated sources when the correlation between the two sources can be modelled by a binary symmetric channel. Nested lattice codes were first introduced in [4] as codes that can achieve the Wyner-Ziv (WZ) limit asymptotically. Based on the partitioning ideas of [4] the authors in [5] considered trellis-based nested codes as a way of realizing nested lattice codes. Practical nested lattice code implementation was first done in [6]. We note that the works in [5] and [6] were restricted to Gaussian scalar sources x and y where $y = x + n$ in which n is Gaussian and is independent of x.

In this paper we propose DSC algorithms which are applicable to a wider range of correlation model. In particular, we consider the scenario where the relationship between

*This research is supported by National Science Foundation under grant CCR-0133635.

the vector source $\boldsymbol{x}$ and the vector side information $\boldsymbol{y}$ can be modelled as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{n}$ is independent of $\boldsymbol{x}$. This model can be viewed as a first order approximation of a general non-linear mapping between $\boldsymbol{y}$ and $\boldsymbol{x}$. Viewing this dependency as a fictitious communication channel with input $\boldsymbol{x}$ and output $\boldsymbol{y}$ and following the ideas in [4], one might think of searching for lattice channel codes suitable for this type of channels [7]. However, the complexity of these constellations are prohibitively high.

As a simple albeit suboptimal architectural alternative the receivers in communication systems incorporate powerful signal processing techniques to combat the effect of the distortion imposed on the signal by $\boldsymbol{H}$ e.g., they apply *equalization*. Using this technique, that has never been explicitly exploited in the DSC arena, we propose to convert the vector source coding problem into a set of parallel scalar source coding problems with the scalar side information at the decoder, where there is a one-to-one correspondence between each source and side information. With this conversion, we are able to use the existing constructive WZ codes to implement the general problem of vector sources.

The paper is organized as follows: Section 2 describes the problem of lossy coding of vector source $\boldsymbol{x}$ with vector side information $\boldsymbol{y}$ when their dependency can be captured through the linear model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$. Section 3 overviews a specific DSC construction, namely DISCUS [5] which we use as the baseline of our work. In Section 4 we clarify the role of equalizer in our proposed algorithm. Section 5 elaborates our proposed coding algorithm and the codec architecture. Section 6 explains the rate allocation procedure which we use to distribute the rate among the entries of the vector source $\boldsymbol{x}$. Section 7 states the rate-distortion bound established in [8] for jointly Gaussian vector sources. Section 8 compares the rate-distortion performance achieved by our codec with the theoretical bounds in [8] and concludes the paper.

Notation: Upper case and lower case boldface letters denote matrices and vectors, respectively. Transpose and pseudoinverse operations are represented by $(.)^T$ and $(.)^\dagger$, respectively. For a random vector $\boldsymbol{a}$, the covariance matrix is denoted as $\boldsymbol{R}_{aa}$. The i-th entry of the vector $\boldsymbol{a}$ is equivalently denoted as a_i and $[\boldsymbol{a}]_i$. The (i, j)-th entry of the matrix $\boldsymbol{A}$ is equivalently denoted as $[\boldsymbol{A}]_{ij}$ and $\boldsymbol{A}_{ij}$.

2 Problem Statement

We consider the class of vector sources with corresponding outputs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ whose statistical dependency can be approximately captured through the relationship:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

Regarding the model in (1) we make the following assumptions:
(a1) $\boldsymbol{n}$ and $\boldsymbol{x}$ are independent and $\boldsymbol{H}$ is a constant matrix,
(a2) $\boldsymbol{x}$ is zero mean with a covariance matrix $\boldsymbol{R}_{xx}$ and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{nn})$.
In our setup the decoder has access to the $\boldsymbol{y}$ process (side information). The goal for the encoder is to compress the $\boldsymbol{x}$ process efficiently, considering the fact that the decoder knows $\boldsymbol{y}$. The goal for the decoder is to reconstruct $\boldsymbol{x}$ using the side information $\boldsymbol{y}$ as well as the bitstream transmitted by the encoder. To be able to reconstruct $\boldsymbol{x}$ with a reasonable distortion, the decoder needs to fully exploit the correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$. This requires the knowledge of the correlation model parameters at the decoder, i.e., $\boldsymbol{H}$ and $\boldsymbol{R}_{nn}$[1]. The encoder, however, does not know $\boldsymbol{H}$ and $\boldsymbol{R}_{nn}$. Relying on the

[1]We note that estimating $\boldsymbol{H}$ and $\boldsymbol{R}_{nn}$ at the decoder is itself an interesting problem and deserves a special treatment which is beyond the scope of this paper.

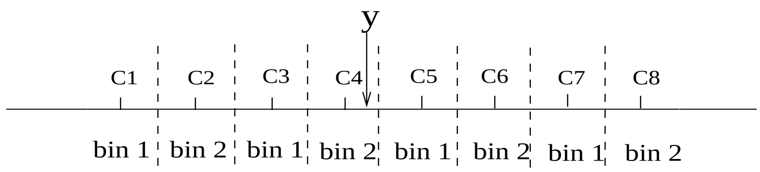

Figure 1: Quantization intervals for a uniform quantizer with $Q = 8$ levels.

knowledge of the correlation model parameters, the decoder calculates the rate at which the encoder should encode the entries of $\boldsymbol{x}$ such that the decoder is able to reconstruct $\boldsymbol{x}$ with a certain distortion. The calculated rate allocation is fed back to the encoder and it is used to encode the entries of $\boldsymbol{x}$ accordingly.

3 Overview of DISCUS

In an attempt to achieve the bounds predicted in [1] and [2] Pradhan and Ramchandran proposed a practical DSC scheme called DISCUS (distributed source coding using syndromes)[5]. Inspired by the random binning idea used in [1] to establish the theoretical results, the authors propose a code construction for the simple correlation model $y = x + n$ where x and y are jointly Gaussian. In their scheme the encoder consists of (i) a source space partition, in which the encoder partitions the real line into disjoint quantization intervals and finds the index of the interval to which the quantized x belongs to[2], and (ii) a coset code partition, in which the encoder partitions the reconstruction points of the quantizer into bins (cosets). Encoding with side information consists of finding the index of the coset containing the quantized x and sending the index at a rate (which we refer to it as *coset rate* and restrict it to be less than or equal to the quantization rate) to the decoder. It is assumed that the index of the coset is available at the decoder error free [5]. The decoder in [5] consists of (i) source code recovery, in which the decoder decides the quantized x as the one which is closest to y (in the sense of Euclidean distance) in the coset whose index is sent by the encoder, and (ii) an estimator, in which the decoder forms the optimal estimator $\hat{x} = E\{x|y, x \in \Gamma_i\}$ where Γ_i is the decided quantization interval. The following example describes the code construction at the encoder and the reconstruction at the decoder:

Example 1 [5]: Consider a uniform quantizer with $Q = 8$ quantization levels. Let $\mathcal{C} = \{c_1, \ldots, c_8\}$ and $\mathcal{G} = \{\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_8\}$ denote the set of reconstruction points and the quantization intervals, respectively. The quantization rate is $\log_2 8 = 3$ bits/sample. To reduce the rate of transmission to 1 bit/sample we partition $\mathcal{C}$ into $2^1 = 2$ cosets as follows: $\mathcal{R}_1 = \{c_1, c_3, c_5, c_7\}$ and $\mathcal{R}_2 = \{c_2, c_4, c_6, c_6\}$. Assuming that $x \in \Gamma_5$, the encoder transmits the coset index 1 at the rate of 1 bit/sample. Knowing that the quantized x belongs to coset 1 and observing y the decoder declares Γ_5 as the interval (i.e., error free detection) and forms $\hat{x} = E\{x|y, x \in \Gamma_5\}$ (See Fig.1).

While applying the minimum distance (MD) rule to decide the quantization interval is justified for the case where the dependency between x and y resembles an additive white Gaussian noise channel, it is not true for the situation where the correlation model is more elaborate, e.g., the dependency model is (1).

[2]For simplicity we focus on the case of scalar quantization and memoryless coset construction in [5].

4 Equalization in Source Coding

In this section we explain why and how channel equalization techniques can be used as a tool in source coding. Using the equalizer at the decoder we convert the dependency model of (1) into a simple model

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{w} \tag{2}$$

where $\boldsymbol{z}$ and $\boldsymbol{w}$ are the new side information and the new noise term, respectively. By establishing a one-to-one correspondence between the entries of $\boldsymbol{x}$ and $\boldsymbol{z}$ we propose to use a scheme where the decoder performs an entry by entry MD detection as in [5]. Due to the equalization $\boldsymbol{w}$ is in general a colored noise. Although this fact is neglected during the MD detection, by applying a vector estimation (rather than entry by entry estimation) the decoder improves the estimate by taking in to account the noise color. We let $\boldsymbol{y}$ and $\boldsymbol{z}$ be the input and the output of the equalizer, respectively. The equalizer can be either linear (e.g., zero forcing (ZF), linear MMSE (LMMSE)) or non-linear (e.g., decision feedback equalizer (DFE)). In fact $\boldsymbol{z}$ provides us with an *initial estimate* of $\boldsymbol{x}$. Incorporating the data sent by the encoder (i.e., the coset indices) the decoder *refines* this initial estimate and obtains an improved estimate that results in less distortion.

4.1 ZF Equalization

The linear equalizer $\boldsymbol{G}$ is such that $\boldsymbol{G}\boldsymbol{H} = \boldsymbol{I}$, i.e., $\boldsymbol{G} = \boldsymbol{H}^\dagger$. With this choice of $\boldsymbol{G}$ the output of the equalizer $\boldsymbol{z}$ is:

$$\boldsymbol{z} = \boldsymbol{G}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w} \tag{3}$$

where the noise $\boldsymbol{w} = \boldsymbol{G}\boldsymbol{n}$ is zero mean with the covariance $\boldsymbol{R}_{ww} = \boldsymbol{H}^\dagger \boldsymbol{R}_{nn} \boldsymbol{H}^{\dagger^T}$. We note that the mapping between the elements of $\boldsymbol{x}$ and $\boldsymbol{z}$ is one to one. Although $\boldsymbol{w}$ is in general colored, decoding on a entry by entry basis is quite effective.

4.2 LMMSE

The equalizer $\boldsymbol{G}$ is designed such that $E\{||\boldsymbol{G}\boldsymbol{y} - \boldsymbol{x}||^2\}$ is minimized. With this criterion $\boldsymbol{G} = \boldsymbol{R}_{xx}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{R}_{xx}\boldsymbol{H}^T + \boldsymbol{R}_{nn})^{-1}$ [9] and $\boldsymbol{z}$ is:

$$\boldsymbol{z} = \boldsymbol{G}\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w} \tag{4}$$

where $\boldsymbol{w} = (\boldsymbol{G}\boldsymbol{H} - \boldsymbol{I})\boldsymbol{x} + \boldsymbol{G}\boldsymbol{n}$ is zero mean with the covariance $\boldsymbol{R}_{ww} = (\boldsymbol{R}_{xx}^{-1} + \boldsymbol{H}^T\boldsymbol{R}_{nn}^{-1}\boldsymbol{H})^{-1}$ [9]. Although $\boldsymbol{w}$ in general is colored, this choice of $\boldsymbol{G}$ minimizes $tr(\boldsymbol{R}_{ww})$ and thus we expect that the entry by entry detection provides us a better performance compared to that of ZF.

4.3 DFE

In this scheme $\boldsymbol{G}$ is designed such that $\boldsymbol{G}\boldsymbol{R}_{nn}^{-1/2}\boldsymbol{H} = \boldsymbol{R}$ where $\boldsymbol{R}$ is an upper triangular matrix. To accomplish this task we choose $\boldsymbol{G} = \boldsymbol{R}^{-T}\boldsymbol{H}^T\boldsymbol{R}_{nn}^{-1/2}$ where $\boldsymbol{R}$ can be obtained from the cholesky decomposition of $\boldsymbol{H}^T\boldsymbol{R}_{nn}^{-1}\boldsymbol{H} = \boldsymbol{R}^T\boldsymbol{R}$. With this choice of $\boldsymbol{G}$ we have:

$$\boldsymbol{z} = \boldsymbol{G}\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{w} \tag{5}$$

where $\boldsymbol{w} = \boldsymbol{G}\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Equivalently, we may flip the vectors $\boldsymbol{z}, \boldsymbol{x}$ and $\boldsymbol{w}$ to obtain $\boldsymbol{z}^f = \boldsymbol{L}\boldsymbol{x}^f + \boldsymbol{w}^f$ where $\boldsymbol{L}$ is a lower triangular matrix. We observe that z_i^f is related to x_j^f $j = 1, \ldots, i-1$. Hence the decoder performs *successive cancellation* to detect the quantization interval and estimate the elements of $\boldsymbol{x}^f$ in the order of x_i^f $i = 1, \ldots, N$.

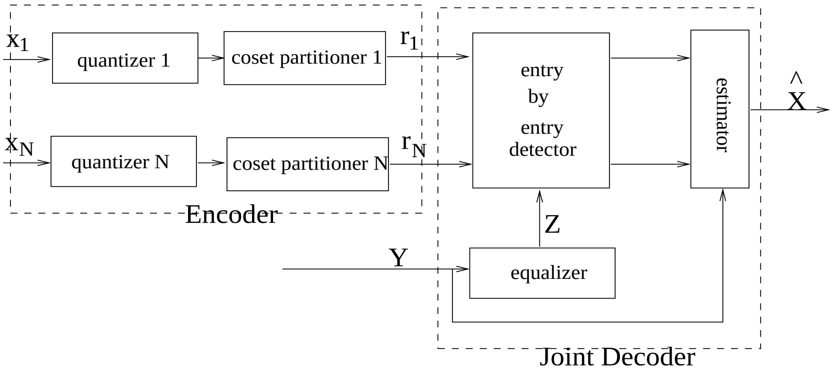

Figure 2: The architecture of the proposed codec.

5 Algorithm Description

The encoder of our proposed system consists of (i) N scalar uniform quantizers and (ii) N coset partitioners that operate in parallel. On an entry by entry basis the encoder quantizes the vector $\boldsymbol{x}$ and finds the indices of the cosets to which the quantized entries belong to. Let r_i bits/sample denote the rate at which the coset index corresponding to encoding x_i is sent to the decoder, i.e., r_i be the coset rate. The encoder sends these indices at the aggregate rate of $R = \sum_{i=1}^{N} r_i$ bits/vector. The decoder of our proposed system has three parts (i) an equalizer, (ii) N MD rule detectors that operate in parallel, and (iii) an estimator.

5.1 Encoder Side: Quantization

We restrict ourselves to scalar fixed rate uniform quantizers. The entry x_i is quantized with the i-th quantizer which has Q_i levels and a step size Δ_i. We define the *quantization rate* as $r_{s_i} = \log_2 Q_i$. The i-th quantizer partitions the real line into Q_i disjoint intervals and $\mathcal{G}_i = \{\Gamma_1^i, \Gamma_2^i, \ldots, \Gamma_{Q_i}^i\}$ denotes the set of these intervals. Each interval $\Gamma_q^i \in \mathcal{G}_i$ corresponds to an input signal amplitude in that interval and is associated with a reconstruction point $c_q^i = c_q^i = (2q - 1 - Q_i)\Delta_i/2 \quad q = 1, \ldots, Q_i$. We call the set $\{c_q^i\}_{q=1}^{Q_i}$ the *source codebook* $\mathcal{C}_i$ associated with the i-th quantizer.

5.2 Encoder Side: Coset Partitioning

Starting with the N source codebooks $\mathcal{C}_i$ we construct the cosets as follows: let r_i bits/sample be the coset rate and define the total number of cosets as $P_i = 2^{r_i}$. We partition $\mathcal{C}_i$ into P_i cosets, namely $\mathcal{R}_p^i$ $p = 1, \ldots, P_i$, such that each coset $\mathcal{R}_p^i$ includes $M_i = Q_i/P_i$ reconstruction points. Clearly we have $r_i = r_{s_i} - r_{c_i}$. Let $\mathcal{R}_p^i = \{r_{p,1}^i, \ldots, r_{p,M_i}^i\}$ for $p = 1, \ldots, P_i$. The coset partitioning is such that the elements of $\mathcal{R}_p^i$ are related to the elements of $\mathcal{C}_i$ through the relationship $r_{p,m}^i = c_{P_i(m-1)+p}^i \quad m = 1, \ldots, M_i \quad p = 1, \ldots, P_i$. In words, considering $\mathcal{C}_i$ we let the first P_i elements of $\mathcal{C}_i$ to be the first elements of the P_i cosets, the second P_i elements of $\mathcal{C}_i$ to be the second elements of the P_i cosets and so on, while we preserve the order during assigning (See Fig. 1). Given the index of the quantization interval q the index of the coset p can be found as $p = q \textit{ mod } P_i$. Let $N = 1$ and consider Example 1. Given $x_1 \in \Gamma_5^1$ the quantizer quantizes x_1 to c_5^1 and finds the index of the coset as $1 = 5 \textit{ mod } 2$.

5.3 Decoder Side: Entry by Entry MD rule Detecting

- **ZF and LMMSE**: The source $\boldsymbol{x}$ and the equalizer output $\boldsymbol{z}$ are related through $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{w}$ where $\boldsymbol{w}$ has the covariance matrix $\boldsymbol{R}_{ww}$. Equivalently, we have $z_i = x_i + w_i \quad i = 1, \ldots, N$ where w_i has the variance $[\boldsymbol{R}_{ww}]_{ii}$. Given the coset index p for each entry x_i, the decoder applies the MD rule to decide the quantized value of x_i in p-th coset that is closet in distance to z_i:

$$r^i_{p,m^*} = \arg\min_{r^i_{p,m} \in \mathcal{R}^i_p} ||z_i - r^i_{p,m}||$$

and declares $P_i(m^* - 1) + p$ as the index of the interval to which x_i belongs.

- **DFE**: The vectors $\boldsymbol{x}$ and $\boldsymbol{z}$ are related through $\boldsymbol{z} = \boldsymbol{L}\boldsymbol{x} + \boldsymbol{G}\boldsymbol{n}$ where $\boldsymbol{L}$ is a lower triangular triangular matrix[3]. In other words:

$$z_i = L_{ii}x_i + \sum_{j=1}^{i-1} L_{ij}x_j + [\boldsymbol{G}\boldsymbol{n}]_i \quad i = 1, \ldots, N \tag{6}$$

where $[\boldsymbol{G}\boldsymbol{n}]_i \sim \mathcal{N}(0, 1)$ and independent of $\boldsymbol{x}_i$. Starting with $i = 1$ given the coset index p corresponding to x_1, the decoder decides the quantization interval to which x_1 belongs and forms the estimate $\hat{x}_1$, which will be described in Section 5.4. Knowing $\hat{x}_1$ the decoder forms a new side information $\overline{z}_2$ in which the interference from x_1 has been suppressed and uses $\overline{z}_2$ to form $\hat{x}_2$ and so on. In general, given the coset index p for each entry x_i, the decoder applies the MD rule to decide the quantized value of x_i in p-th coset that is closest to $\overline{z}_i$:

$$r^i_{p,m^*} = \arg\min_{r^i_{p,m} \in \mathcal{R}^i_p} ||\overline{z}_i - L_{ii} r^i_{p,m}||$$

where:

$$\overline{z}_1 = z_1 \qquad \overline{z}_i = z_i - \sum_{j=1}^{i-1} L_{ij}\hat{x}_j \quad i = 2, \ldots, N \tag{7}$$

and declares $P_i(m^* - 1) + p$ as the index of the interval to which x_i belongs.

5.4 Estimation

The entry by entry MD detector provides the indices of the quantization intervals. Knowing these intervals, for ZF and LMMSE equalizers we form the vector estimate $\hat{\boldsymbol{x}}$. For the DFE the decoder performs an entry by entry detection followed by a symbol by symbol estimation. A vector estimation which uses the decided intervals can be followed in order to improve the performance of the estimate in the sense of MSE. To find closed form expression for $\hat{\boldsymbol{x}}$ in this section we assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are *jointly Gaussian.*

- **Vector Estimation for ZF and LMMSE**: For jointly Gaussian $\boldsymbol{x}$ and $\boldsymbol{y}$ the conditional mean $\boldsymbol{\mu}$ and the conditional covariance $\boldsymbol{R}_{x|y}$ are:

$$\boldsymbol{\mu} = (\boldsymbol{R}^{-1}_{xx} + \boldsymbol{H}^T\boldsymbol{R}^{-1}_{nn}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}_{nn}\boldsymbol{y} \quad \boldsymbol{R}_{x|y} = (\boldsymbol{R}^{-1}_{xx} + \boldsymbol{H}^T\boldsymbol{R}^{-1}_{nn}\boldsymbol{H})^{-1}$$

We introduce the EVD of $\boldsymbol{R}_{x|y} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$.

[3]For simplicity we drop the superscript f which reflects the flipping operation to obtain $\boldsymbol{L}$ from $\boldsymbol{R}$.

lemma 1: *Let $\mathcal{V}_x$ denote the N dimensional volume whose edges are determined by the decided quantization interval. Define $\boldsymbol{u} = \boldsymbol{U}^T(\boldsymbol{x} - \boldsymbol{\mu})$ and let $\mathcal{V}_u$ denote the new volume we wish to integrate over which is obtained from transforming $\mathcal{V}_x$ in the new coordinate system. The optimal vector estimate $\hat{\boldsymbol{x}}$ is (proofs are omitted due to lack of space [10]):*

$$\hat{\boldsymbol{x}} = E\{\boldsymbol{x}|\boldsymbol{x} \in \mathcal{V}_x, \boldsymbol{y}\} = \boldsymbol{\mu} + \sum_{j=1}^{N} \boldsymbol{U}\boldsymbol{e}_j \frac{\int_{\boldsymbol{u}\in\mathcal{V}_u} u_j e^{-\frac{1}{2}\sum_{i=1}^{N} u_i^2/\Lambda_{ii}} d\boldsymbol{u}}{\int_{\boldsymbol{u}\in\mathcal{V}_u} e^{-\frac{1}{2}\sum_{i=1}^{N} u_i^2/\Lambda_{ii}} d\boldsymbol{u}}$$

in which $\boldsymbol{e}_j$ is a zero vector except the j-th entry that is one. The integral can be numerically evaluated using Gaussian quadrature technique in [11].

Remark: The estimator $\hat{\boldsymbol{x}}$ is decomposed into two terms: $\boldsymbol{\mu}$ is the optimal estimate of $\boldsymbol{x}$ when the encoder does not send any data to the decoder and the decoder forms an estimate using the side information $\boldsymbol{y}$. The second term is the adjustment in the estimation resulting from the extra bits (coset indices) that the encoder sends.

- **Symbol by Symbol Estimation for DFE** Let $\tilde{x}_i = x_i - \hat{x}_i$ be the estimation error. Combining (6) and (7) we find:

$$\overline{z}_i = L_{ii}x_i + \sum_{j=1}^{i-1} L_{ij}\tilde{x}_j + [\boldsymbol{G}\boldsymbol{n}]_i \quad i = 1, \ldots, N \tag{8}$$

Let $v_i = \sum_{j=1}^{i-1} L_{ij}\tilde{x}_j + [\boldsymbol{G}\boldsymbol{n}]_i$ and $\sigma_{v_i}^2$ be the corresponding variance. Assuming that $\tilde{x}_j$ $j = 1, \ldots, i-1$ are zero mean and uncorrelated we obtain $\sigma_{v_i}^2 = \sum_{j=1}^{i-1} L_{ij}^2 E\{\tilde{x}_j^2\} + 1$. We rewrite (8) as $\overline{z}_i = L_{ii}x_i + v_i$ and establish a one-to-one correspondence between $\overline{z}_i$ and x_i. To find a closed form expression for $\hat{x}_i$ we make two more assumptions: (i) $\overline{z}_i$ and x_i are jointly Gaussian, (ii) v_i is Gaussian and is independent of x_i.

lemma 2: *Let $\Gamma_j^i = (a_j^i, b_j^i)$ be the decided quantization interval and $\sigma_{x_i}^2 = [\boldsymbol{R}_{xx}]_{ii}$. Under all the above assumptions the optimal scalar estimate $\hat{x}_i$ is:*

$$\hat{x}_i = E\{x_i|x_i \in \Gamma_j^i, \overline{z}_i\} = \mu_{x_i|\overline{z}_i} + \sigma_{x_i|\overline{z}_i}\sqrt{\frac{2}{\pi}} \frac{e^{-(\breve{a}_j^i)^2} - e^{-(\breve{b}_j^i)^2}}{erfc(\breve{a}_j^i) - erfc(\breve{b}_j^i)}$$

$$\mu_{x_i|\overline{z}_i} = \frac{L_{ii}\sigma_{x_i}^2}{L_{ii}^2\sigma_{x_i}^2 + \sigma_{v_i}^2}\overline{z}_i \quad \sigma_{x_i|\overline{z}_i}^2 = \frac{\sigma_{x_i}^2\sigma_{v_i}^2}{L_{ii}^2\sigma_{x_i}^2 + \sigma_{v_i}^2} \quad \breve{a}_j^i = \frac{a_j^i - \mu_{x_i|\overline{z}_i}}{\sqrt{2}\sigma_{x_i|\overline{z}_i}} \quad \breve{b}_j^i = \frac{b_j^i - \mu_{x_i|\overline{z}_i}}{\sqrt{2}\sigma_{x_i|\overline{z}_i}}$$

6 Rate Allocation

Given $\boldsymbol{R}_{xx}$ at the encoder we choose the number of quantization levels $Q_i = 2^{r_{s_i}}$ of the i-th quantizer proportional to σ_{x_i} [12], i.e., we set $r_{s_i} = \max\left\{0, \left\lceil \frac{1}{2}\log_2 \alpha\sigma_{x_i}^2 \right\rceil\right\}$ where α is a parameter which is identical for $\forall i$. For simplicity we assume $\sigma_{x_i}^2 = \sigma_x^2$ and hence $r_{s_i} = r_s$ $i = 1, \ldots, N$. In words, the entries of $\boldsymbol{x}$ are quantized with identical quantizers.

Suppose that the encoder is equipped with a finite set of quantizers so that it can quantize the entries of $\boldsymbol{x}$ at different resolutions[4], i.e., $r_s \in \mathcal{R}_s$ where $\mathcal{R}_s$ is a finite set of

[4]Each quantizer can be associated with a choice of α.

quantization rates. Given the constraints $\sum_{i=1}^{N} r_i = R$ and $r_i \leq r_s$ ($r_s \in \mathcal{R}_s$ $i = 1, \ldots, N$) the decoder finds the *optimal coset rates* r_i^* $i = 1, \ldots, N$ and the *optimal quantization rate* r_s^* such that the total distortion due to the construction of the entries of $\boldsymbol{x}$ is minimized. Let $d_i(r_s, r_i)$ denote the distortion occurred during x_i reconstruction and is a function of r_s and r_i $i = 1, \ldots, N$ as well as the correlation model parameters $\boldsymbol{R}_{nn}$ and $\boldsymbol{H}$. In the following we provide an approximate expression for d_i.

Consider x_i reconstruction at the decoder given a set of coset rates r_i $i = 1, \ldots, N$ and a quantization rate r_s. Let p_{ei} represent the average probability of error occurred during the detection of the quantization interval to which x_i belongs, $E\{\varepsilon^2\}$ represent the MSE of the quantization error corresponding to quantizing x_i, and δ_i indicate the distance between two adjacent points in a coset. Given the simple coset partitioning strategy we adopted in Section 5.2 it can be verified that $\delta_i = 2^{r_i}\Delta$ where Δ is the quantization step size[5]. A rough approximate of d_i is:

$$d_i(r_s, r_i) \approx (1 - p_{ei})\theta_i^2 E\{\varepsilon^2\} + p_{ei}(\theta_i^2 E\{\varepsilon^2\} + \delta_i^2) = \theta_i^2 E\{\varepsilon^2\} + p_{ei}\delta_i^2 \tag{9}$$

where $\theta_i = 1$ for ZF and LMMSE and $\theta_i = L_{ii}$ for DFE. To obtain (9) we implicitly assume that within a coset the detection error occurs only between the adjacent points. This assumption is justified for the case when δ_i is relatively large compared to $\sigma_{x_i|z_i}$.

Recall that we established a one-to-one correspondence between $\boldsymbol{x}$ and the side information for the equalizers we utilize, namely, $z_i = x_i + w_i$ for ZF and LMMSE and $\overline{z_i} = L_{ii}x_i + v_i$ for DFE where $\sigma_{w_i}^2$ and $\sigma_{v_i}^2$ are given. First consider ZF and LMMSE and let $\nu_i = \varepsilon + w_i$ where ε is the quantization error. Assuming that ε and w_i are uncorrelated we obtain $\sigma_{\nu_i}^2 = E\{\varepsilon^2\} + \sigma_{w_i}^2$. Similarly for DFE let $\nu_i = L_{ii}\varepsilon + v_i$. Assuming that ε and v_i are uncorrelated we obtain $\sigma_{\nu_i}^2 = L_{ii}^2 E\{\varepsilon^2\} + \sigma_{v_i}^2$[6].

Lemma 3: *An upper bound on* p_{ei} *is:*

$$p_{ei} < \frac{1}{2} e^{-\frac{\delta_i^2}{8\sigma_{\nu_i}^2}} \tag{10}$$

where σ_{ν_i} *depends on the type of equalizer and is given above.*

Using the upper bound as the actual p_{ei} and substituting (10) in (9) we define:

$$(r_s^*, r_1^*, \ldots, r_N^*) = \arg\min \sum_{i=1}^{N} \left(\theta_i^2 E\{\varepsilon^2\} + \frac{\delta_i^2}{2} e^{-\frac{\delta_i^2}{8\sigma_{\nu_i}^2}} \right) \tag{11}$$

where the minimum is over all $r_s \in \mathcal{R}_s$ and all integer valued sets $\{r_i\}_{i=1}^{N} \leq r_s$ satisfying $\sum_{i=1}^{N} r_i = R$. The minimization is performed numerically. The decoder computes $\sum_{i=1}^{N} d_i(r_s^*, r_i^*)$ and compares it with $trace(\boldsymbol{R}_{x|y})$[7]. Only if $\sum_{i=1}^{N} d_i(r_s^*, r_i^*) < trace(\boldsymbol{R}_{x|y})$ the decoder feeds back $(r_s^*, \{r_i^*\}_{i=1}^{N})$ to the encoder and requires the encoder to send the coset indices. This is the scenario where the decoder can improve the $\boldsymbol{x}$ estimation by incorporating the data received from the encoder. if $\sum_{i=1}^{N} d_i(r_s^*, r_i^*) \geq trace(\boldsymbol{R}_{x|y})$ the decoder asks the encoder not to send any data.

[5]We note that $E\{\varepsilon^2\}$ and Δ vary for different quantizers. Each $r_s \in \mathcal{R}_s$ is associated with a pair of ($E\{\varepsilon^2\}$,Δ). To run the simulation in Section 8 we use the optimal uniform quantizer for normal random variables in [12].

[6]In the expression for $\sigma_{v_i}^2$ we assume that $E\{\tilde{x}_j^2\} = E\{\varepsilon^2\}$. This implies that we assume during the successive cancellation no detection error happens and the distortion is due to the quantization error.

[7]$trace(\boldsymbol{R}_{x|y})$ is the MSE of the optimal $\boldsymbol{x}$ estimate based on the side information $\boldsymbol{y}$, i.e., it is the MSE of $\hat{\boldsymbol{x}} = E\{\boldsymbol{x}|\boldsymbol{y}\}$.

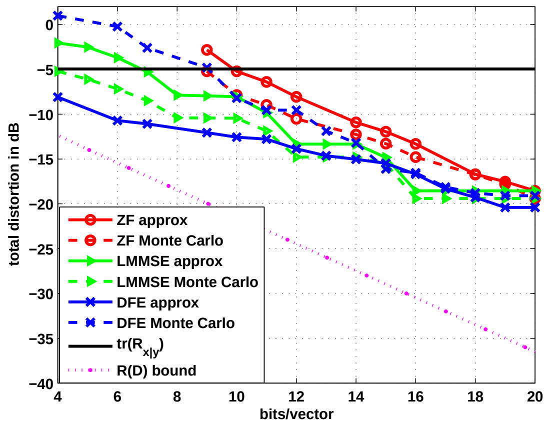

Figure 3: The rate-distortion of the proposed codec is compared to the theoretical bound.

7 Information Theoretic Rate-Distortion Bound

As a benchmark to evaluate the distortion-rate performance of our proposed algorithm, we use the theoretical bounds established in [8] for jointly Gaussian vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.
Lemma 4 [11,Thm.9]: *The rate-distortion function is given by:*

$$R(D) = \min_{D_i} \sum_{i=1}^{N} max \left\{ \frac{1}{2} \log_2 \frac{\lambda_i^2}{D_i}, 0 \right\} \tag{12}$$

where λ_i^2 *are the* N *eigenvalues of the matrix* $\boldsymbol{R}_{x|y}$ *and the minimum is over all sets* $\{D_i\}_{i=1}^N$ *satisfying* $\sum_{i=1}^N D_i \leq D$.

8 Simulation Results and Conclusion

In Section 6 we proposed a rate allocation procedure based on minimizing an approximate expression of the total distortion occurred during $\boldsymbol{x}$ reconstruction. Conducting a numerical optimization [c.f. (11)] the decoder determines the coset rates and the quantization rate at which the encoder should operate. In this section we evaluate the performance of our proposed algorithms, assuming that the encoder operates at the optimal quantization/coset rates provided by the decoder. The decoder reconstructs $\boldsymbol{x}$ and calculates the actual distortion averaged over the length of vector source sequence. We compare the performance against the theoretical bound in Section 7 and $trace(\boldsymbol{R}_{x|y})$. Our choice of parameters are: $N = 4$, $\boldsymbol{H}$ and $\boldsymbol{R}_{xx}$ are two Toeplitz matrices with first row $[1\ \rho\ \rho^2\ \rho^3]$ where $\rho = 0.6$ for $\boldsymbol{H}$ and $\rho = 0.8$ for R_{xx}, $\boldsymbol{R}_{nn} = \sigma_n^2 \boldsymbol{I}$ where $\sigma_n^2 = 0.0631$ (-12 dB). The input $\boldsymbol{x}$ is a vector sequence of length 10^5 where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{xx})$ and the quantization rate set is $\mathcal{R}_s = \{2, 3, 4, 5\}$ bits/sample. As suggested by Fig. 3 for each equalizer there is a critical sum rate $R = \sum_{i=1}^N r_i$ bits/vector below which the total attainable distortion is higher than $trace(\boldsymbol{R}_{x|y})$ and hence the encoder should not send any data. In [8] it is shown that the $R(D)$ bound is achievable if the encoder applies conditional Karhünen-Loève transform (KLT) to $\boldsymbol{x}$ and encode the transformed coefficients. We note that attaining the bound requires a vector quantization over an infinite length sequence of transformed coefficients. Our preliminary results show that the gap is reduced if we encode (using the scalar uniform qunatizers) the conditional KLT transformed coefficients as opposed to the entries of $\boldsymbol{x}$, assuming that $R_{x|y}$ is available at the encoder.

In summary, in this paper we proposed a WZ codec which is suitable for a wider class of vector sources $\boldsymbol{x}$ and $\boldsymbol{y}$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are related as $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$. Viewing the dependency between $\boldsymbol{x}$ and $\boldsymbol{y}$ as a virtual fading channel with input $\boldsymbol{x}$ and output $\boldsymbol{y}$, we utilize equalizer at the decoder, a technique used in communication systems to combat the effect of channel distortion, to convert the problem of vector source coding into a set of scalar source coding problems. Using this technique, we are able to use the WZ codes designed for the simple model $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ for the more general model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$.

References

[1] D. Slepian and J. K. Wolf,"Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, Vol. IT-19, pp. 471-480, July 1973.

[2] A. D. 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, January 1976.

[3] A. D. Wyner,"Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, Vol. IT-20, pp. 2-10, January 1974.

[4] R. Zamir, S. Shamai and U. Erez, "Nested linear lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, Vol. 48, No.6, pp. 1250-1276, 2002.

[5] S.Sandeep and K. Ramchandran, "Distributed source coding using syndroms (DISCUS): design and construction", *IEEE Trans. Inform. Theory*, Vol. 49, Issue 3, pp. 626-643, March 2003.

[6] S. Servetto,"Lattice quantization with side information," in Proc. DCC 2000, UT 2000, pp.510-519.

[7] X. Giraud and J. C. Belfiore, "Constellations matched to the Rayleigh fading channel" *IEEE Trans. Inform. Theory*, Vol. 42, No. 1, pp. 106-115, January 1996.

[8] M. Gastpar, P. L. Dragotti and M. Vetterli, "The distributed Karhunen-Loeve transform, " *EPFL technical report IC/2003/12.*

[9] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, vol. 1, Princton Hall 1993.

[10] A. Vosoughi and A. Scaglione, "Precoding and Decoding Paradigms for Distributed Data Compression, " under preparation, available at *http://crisp.ece.cornell.edu/.*

[11] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press 1984.

[12] A. Gersho and R. M. Gray, *Vector qunatization and signal compression*, Kluwer Academic Publishers 1991.