Joint Data Hiding and Wyner-Ziv Coding, Theory and Practice

Çagatay Dikici [a,*], Christine Guillemot [b], Caroline Fontaine [b], Khalid Idrissi [a], Atilla Baskurt [a]

[a] *LIRIS, INSA de Lyon, France*

[b] *Project TEMICS, IRISA Rennes, France*

Abstract

This paper considers the joint problem of dirty-paper and Wyner-Ziv coding with partial host information at the receiver. The rate-distortion and the capacity expressions have been derived for the Gaussian case. It is shown that the availability of the partial host information to the receiver has positive effects to the compression performance and the overall capacity. Moreover, a practical coding scheme is proposed which is based on Trellis Coded Quantization (TCQ) codes, and Low Density Parity Check (LDPC) codes. Simulation results show that the Wyner-Ziv and watermarking schemes approach the derived rate-distortion and capacity bounds respectively.

Key words: Source and Channel coding with side informations, Dirty-Paper Coding, Wyner-Ziv Coding, LDPC, TCQ, BCJR, Belief propagation.

1 Introduction

The problem of robust watermarking is to embed a certain amount of message information, under a fixed distortion constraint between the host signal $\boldsymbol{S}$ and the watermarked signal $\boldsymbol{W}$ (i.e., $\mathbb{E}\left[d(\boldsymbol{S},\boldsymbol{W})\right] \leq D_1$), and to allow a reliable recovery of the embedded information from the attacked signal $\hat{\boldsymbol{W}}$. In this paper, we focus on scenarios where watermarked signals are compressed. Compression is, then, considered as one of the attacks that the watermark has

* Corresponding author.
Email address: cdikici@liris.cnrs.fr (Çagatay Dikici).

[1] Manuscript received August 25, 2008; revised February 4, 2008.

to face. Here, we consider one particular form of compression which is based on Wyner-Ziv coding principles. Practical applications of the considered set-up include access to data bases containing Wyner-Ziv layered encoded image and video content, where the low resolution signal is publicly available for preview and the enhancement signal is watermarked. The low resolution version of the content can thus be exploited for Wyner-Ziv decoding and for extracting the mark. The presence of side information at the receiver is also expected, for the same compression rate, to decrease the level of compression attach distortion, hence to increase the embedding rate.

Robust watermarking can be formulated as a problem of channel coding with non-causal channel side information available at the transmitter which has been first considered by Gelfand and Pinsker in [1]. Costa has considered in [2] the particular case of an AWGN channel having white Gaussian interference known to the transmitter only. He has shown that, in this particular set-up, the interference does not incur any loss in capacity. The design of codes for approaching Costas capacity is known as the dirty paper coding problem. The authors in [3] consider a generalized Gel'fand-Pinsker coding problem in the discrete alphabet case. Capacity formulae, as well as random coding exponent are derived for the set-up where partial side information is known to the encoder, attacker and decoder.

The watermarking problem considered here is a dirty paper coding problem with, in addition, partial host information $\hat{\boldsymbol{S}}$ (i.e. a noisy version of the host signal) at the receiver. To be efficiently transmitted, the watermarked signal $\boldsymbol{W}$ is then compressed, under a fixed distortion constraint, i.e., so that the delivered copy $\hat{\boldsymbol{W}}$ satisfies the constraint $\mathbb{E}\left[d(\boldsymbol{W}, \hat{\boldsymbol{W}})\right] \leq D_2$. The compression system is assumed to have no knowledge about the host signal on the transmitting side. This compression problem can be regarded as a form of the Wyner-Ziv coding problem [4], that is of lossy coding with side information at the decoder. Note that the dirty paper and the Wyner-Ziv coding problems are dual: the first one relies on channel coding with side information at the sender whereas the second one relies on source coding with side information at the receiver. The duality between channel coding and source coding, both with side information, is discussed in [5,6].

In the Wyner-Ziv coding set-up, the side information is a noisy version of the signal to be compressed. In the considered setup, the side information $\hat{\boldsymbol{S}}$ is a noisy version of the host signal, whereas the signal to be compressed is the watermarked signal $\boldsymbol{W}$. The side information available can thus be considered as partial information on the signal to be compressed. The rate-distortion function of the Gaussian multiterminal and Wyner-Ziv coding problem with partial information at the decoder has been derived in [7]. Once the watermarked signal has been decompressed, the hidden message $\hat{M}$ must be extracted with a low probability of error P_e.

In this paper, we analyze the theoretical limits of the considered set-up. We first derive the rate-distortion function of the Wyner-Ziv coding component of the chain. In contrast with earlier derivations of rate-distortion functions of Wyner-Ziv coding with partial side information [8], the rate-distortion function must here take into account the targeted embedding power and distortion. The Wyner-Ziv coding and decoding chain is then represented by an equivalent channel model, characterized by the conditional distribution $p(\hat{\boldsymbol{W}}|\boldsymbol{W}, \hat{\boldsymbol{S}})$ of its output $\hat{\boldsymbol{W}}$ given its input $\boldsymbol{W}$ and the partial state information $\hat{\boldsymbol{S}}$. Once this channel is modeled, we are able to derive the capacity of the watermark communication channel (subject to the Wyner-Ziv coding attack channel) and to analyze the rate and capacity gains resulting from the availability of host partial information at the receiver. Costa showed that when the embedder has a perfect knowledge of the *dirt* of the channel, no capacity gain can be expected from the availability of partial host information at the receiver [2]. We show that the availability of a partial host information at the Wyner-Ziv decoder allows to decrease the compression distortion for a given compression rate. Indirectly, this also allows to increase the channel capacity. Moreover, the availability of partial host information at the extraction process has no effect to the channel capacity.

The paper then describes a practical scheme for the entire chain. A lot of effort has already been dedicated in the research community to the design of practical codes to approach the capacity limit established by Costa for the dirty paper coding problem [9–12]. Similarly, in recent years, we have seen intensive research on the design of practical Wyner-Ziv coding solutions [13–17]. The scheme developed here builds upon superposition codes introduced in [18] for side informed channels. The watermarking step makes use of TCQ followed by LDPC codes whereas the Wyner-Ziv step is formed by a scalar quantizer followed by LDPC codes. Simulation results show that the way of combining the capacity approaching superposition code and efficient Wyner-Ziv practical code, with a compression distortion gap of 3.66 dB, a reliable communication of a message M with a coding rate of 1/2 can be achieved.

The remaining of this paper is organized as follows: the problem is stated and the notation is introduced in Section-2. Section-3 focuses on the capacity analysis of the system and rate-distortion function of the modified Wyner-Ziv setup. You can find the proofs of the rate distortion function and the capacity term in this section. We then give the analysis of the rate distortion function and the capacity term if the decoder does not have a noisy side information in Section-4. Finally, in Section-5, a practical scheme is described for the Gaussian case.

2 Problem Statement and Notation

In the following, random variables are denoted by capital letters, their realizations by the corresponding lower case letters, and the calligraphic font is used for sets. Similarly, random vectors, their realizations, and alphabets are denoted respectively, by boldface capital letters, boldface lowercase letters, and calligraphic letters subscripted by the corresponding dimension. $\mathbb{E}[.]$ denotes the expectation operation, $\mathbf{I}$ denotes the $n \times n$ identity matrix. Furthermore, the mean square error distortion metric

$$d(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i)^2 \tag{1}$$

is used throughout in this paper.

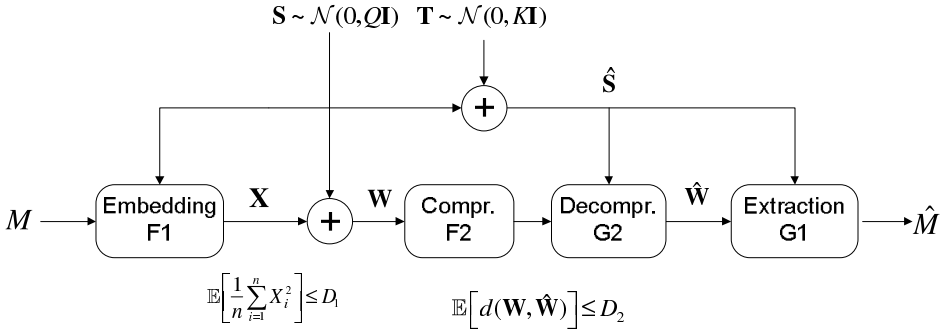


Fig. 1. Dirty paper and Wyner-Ziv coding set-up: overall multivariate Gaussian channel.

The considered communication problem is depicted in Fig. 1. Let $M \in \{1, ..., m\}$ be the discrete-valued hidden message index to be sent to the receiver in n uses of the channel. The quantity m is the greatest integer smaller than or equal to e^{nR_C}, where R_C refers embedding rate or the channel capacity in nats per channel use. Let $\boldsymbol{S} = (S_1, S_2, ..., S_n)$ be the sequence of non-causal channel states for n channel uses. This sequence of states is assumed to be a sequence of i.i.d. random variables $S_i \sim \mathcal{N}(0, Q)$, $i = 1, .., n$, perfectly known to the encoder. We consider the case where this sequence of states is partially known to the decoder. The side information available to the decoder is thus denoted $\hat{\boldsymbol{S}} = (\hat{S}_1, \hat{S}_2, ..., \hat{S}_n)$ and is modeled as $\hat{\boldsymbol{S}} = \boldsymbol{S} + \boldsymbol{T}$, where $\boldsymbol{T}$ is a sequence of i.i.d. random variables $T_i \sim \mathcal{N}(0, K)$. The watermarked signal $\boldsymbol{W}$ is given by $\boldsymbol{W} = \boldsymbol{S} + \boldsymbol{X}$, where $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ represents a sequence of random variables taking their values from the set of real numbers $\mathbb{R}$. $\boldsymbol{X}$ is chosen such that the average embedding power satisfies $\mathbb{E}[d(\boldsymbol{S}, \boldsymbol{W})] \leq D_1$, i.e. $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] \leq D_1$. In the ideal Costa scheme, X_i is defined as $\alpha(\frac{U_i}{\alpha} - S_i)$, where U_i is a codeword taken from a set of $2^{n(I(U;W)-\varepsilon)}$ elements drawn according to the law $\mathcal{N}(0, (D_1 + \alpha^2 Q))$. The term ε is chosen to be very small

as $n \to \infty$. This codebook is partitioned into $2^{n(R_c-\varepsilon)}$ bins. The bin index is associated with the index of the message to be embedded. At the embedding stage, we search for the codevector which is jointly typical with $\boldsymbol{S}$ in the bin associated with the message. The watermarked signal $\boldsymbol{W}$ is then compressed by sending an index $V \in \{1, ..., 2^{nR_S}\}$ under a distortion constraint $\mathbb{E}\left[d(\boldsymbol{W}, \hat{\boldsymbol{W}})\right] \leq D_2$, R_S being the rate per channel use for a given distortion D_2. The entire chain is thus formed by two encoding-decoding pairs: $(F1, G1)$ for the dirty paper coding, and $(F2, G2)$ for the Wyner-Ziv coding. These mappings are defined on the following sets:

$$F1 : \mathcal{M} \times \mathcal{S}^n \to \mathbb{R}^n \ , \ G1 : \hat{\mathcal{W}}^n \times \hat{\mathcal{S}}^n \to \hat{\mathcal{M}}, \tag{2}$$

and

$$F2 : \mathcal{W}^n \to \left\{1, 2, ..., 2^{nR_S}\right\} \ , \ G2 : \left\{1, 2, ..., 2^{nR_S}\right\} \times \hat{\mathcal{S}}^n \to \hat{\mathcal{W}}^n. \tag{3}$$

As both $(F1, G1)$ and $(F2, G2)$ could be referred to as encoders/decoders; in order to avoid any ambiguity, we will refer to $F1$ (resp. $G1$) as *embedding* (resp. *extraction*), and to $F2$ (resp. $G2$) as *compression* (resp. *decompression*). Please note that in this paper, the term *embedding* is not used to refer to the watermarked signal generation $\boldsymbol{W} = \boldsymbol{S} + \boldsymbol{X}$.

Equivalent Channel Model

We now describe the equivalent channel model (See Fig. 2) that achieves the rate distortion lower bound and will help us to compute the channel capacity. We follow the steps of Costa [2], Cover and Chiang [19], and Wyner [8]. For any $i = 1, .., n$, X_i, S_i, and T_i be mutually independent i.i.d. random variables, $X_i \sim \mathcal{N}(0, D_1)$, $S_i \sim \mathcal{N}(0, Q)$, $T_i \sim \mathcal{N}(0, K)$. The joint distribution $p(W_i, \hat{S}_i)$ is then a multivariate Gaussian distribution with 0 mean, and covariance matrix

$$\boldsymbol{\Sigma}_{(w_i, \hat{s}_i)} = \begin{bmatrix} Q + D_1 & Q \\ Q & Q + K \end{bmatrix}. \tag{4}$$

$\hat{W}_i$ is modeled as $\hat{W}_i = a(W_i + Z_i) + \beta(1 - a)\hat{S}_i$ where a and β are scalar constants to be determined, and $\boldsymbol{Z}$ is a sequence of i.i.d. random variables $Z_i \sim \mathcal{N}(0, D_2/a)$, and is independent from $\boldsymbol{S}$, $\boldsymbol{T}$ and $\boldsymbol{X}$. This model is a usual way to design $\hat{\boldsymbol{W}}$ that achieves the lower bound of the rate-distortion function [8, Fig. 4]. In our case, the constant β is chosen such that $\mathbb{E}\left[(\boldsymbol{W}|\hat{\boldsymbol{S}})\right] = \beta\hat{\boldsymbol{S}}$. Furthermore, we choose a such that the quantization of $\boldsymbol{W} - \beta\boldsymbol{S}$ achieves the lower bound of the rate-distortion function.

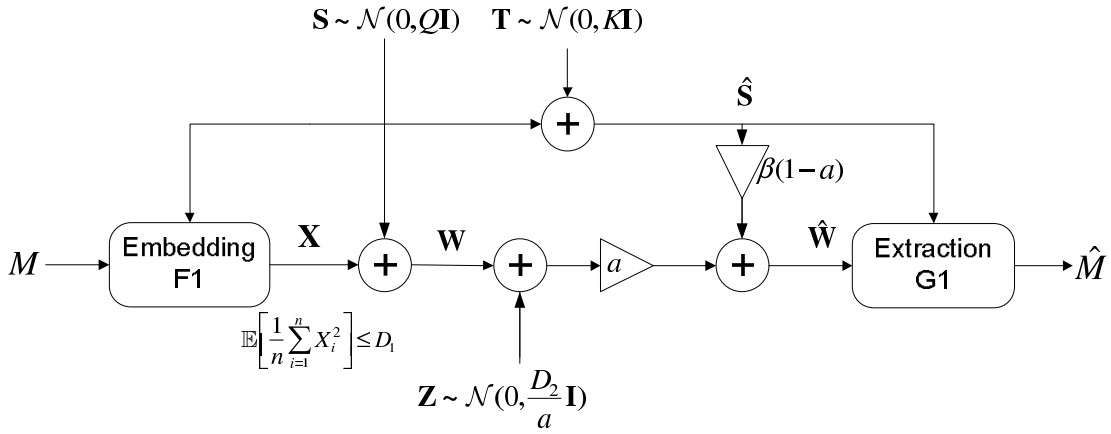


Fig. 2. Equivalent communication channel for noisy side information $\hat{\boldsymbol{S}}$ available to the decoder.

While the constants a and β play a crucial role in the achievability of the rate-distortion bounds of the Wyner-Ziv setup, they will be used in the capacity calculation of the overall system. We employ an auxiliary random vector $\boldsymbol{U}$ for the evaluation of the embedding rate. We propose to use $\boldsymbol{U}$ in a confined class. In particular, consider $\boldsymbol{U} = \boldsymbol{X} + \alpha \boldsymbol{S}$, where $\alpha \in \mathbb{R}$ is a real scalar. We show in Section-3.2 that for an appropriate α value, this coding scheme achieves the capacity.

3 Rate-distortion and capacity limits

In this section, we first derive the rate-distortion function $R_S(D_2)$ of the compression and decompression pair $(F2, G2)$. Let us recall that, in contrast with the classical Wyner-Ziv set-up, here the signal to be coded is the watermarked signal $\boldsymbol{W}$, whereas only host partial side information is available to the decompression process. The capacity R_C of the multivariate Gaussian communication channel (watermark communication channel) depicted in Fig. 1 is then derived.

3.1 Rate-distortion bounds

There has been a lot of effort on the rate distortion function of correlated sources [4,5,8,19]. Fig. 3 shows a general version of the $(F2, G2)$ setup, where the availability of $\hat{\boldsymbol{S}} = \boldsymbol{S} + \boldsymbol{T}$ to the compression and to the decompression processes is controlled by Switch-A and Switch-B. The rate-distortion function is expressed as $R_{AB}(D_2)$, depending on whether the switches are open or closed, corresponding to subindex values 0 or 1 respectively. The rate-distortion function for $(F2, G2)$ pair corresponds to $R_{01}(D_2)$ case. Cover and

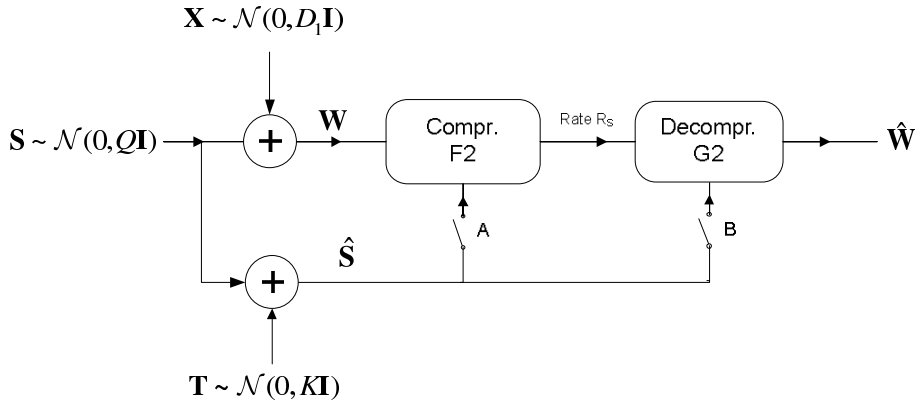


Fig. 3. Modified Wyner-Ziv coding set-up.

Chiang [19] defined the general expressions of the rate-distortion functions $R_{00}(D_2)$, $R_{11}(D_2)$ and $R_{01}(D_2)$. We evaluate the three functions for the appropriate correlation between $\boldsymbol{W}$ and $\hat{\boldsymbol{S}}$ model given in Section-2. Please note that the rate-distortion bounds of our schema can be seen as a modified version of Wyner's [8] setup (See Fig. 3).

Lemma 1 *The three rate distortion functions $R_{00}(D_2)$, $R_{01}(D_2)$, and $R_{11}(D_2)$ are:*

$$R_{00}(D_2) = \frac{1}{2}\ln^{+}\left(\frac{D_1+Q}{D_2}\right), \tag{5}$$

$$R_{01}(D_2) = R_{11}(D_2) = \frac{1}{2}\ln^{+}\left(\frac{D_1}{D_2}+\frac{QK}{(Q+K)D_2}\right), \tag{6}$$

where $\ln^{+}(a) = \max\{0, \ln(a)\}$.

Sketch of Proof: $R_{00}(D_2)$ is obtained in the same way as in [20, Theorem 4.3.2]. We omit the justification of $R_{00}(D_2)$. The proofs of $R_{11}(D_2)$, and $R_{01}(D_2)$ follow the same routines as in [8,21], while the derivations are based on the covariance matrix $\boldsymbol{\Sigma}_{(w,\hat{s})}$ given in (4) and the joint distribution $p(S_i, X_i, T_i)$ is a multivariate Gaussian distribution $\sim \mathcal{N}(0, \boldsymbol{\Sigma}_{S_i,X_i,T_i})$ with a covariance matrix $\boldsymbol{\Sigma}_{S_i,X_i,T_i} = \mathrm{diag}(Q, D_1, K)$. For the sake of clarity, the proofs of $R_{11}(D_2)$, and $R_{01}(D_2)$ are given in Appendix-A. We show that the values a and β that achieve the lower bound of the rate-distortion function are:

$$a = 1 - \frac{D_2(Q+K)}{QD_1 + QK + D_1K}, \tag{7}$$

and

$$\beta = \frac{Q}{Q+K}. \tag{8}$$

3.2 Channel Capacity

The Wyner-Ziv coding pair $(F2, G2)$ can be replaced by the equivalent channel model derived in Section-A.2 (Fig. A.2) where the constants a and β can be found in (7) and (8) respectively. The overall system can thus be sketched as in Fig. 2.

Theorem 1 *The capacity R_C for the communication system given in Fig. 2 is*

$$R_C = \frac{1}{2}\ln\left(1 + \frac{D_1}{D_2} - \frac{D_1(Q+K)}{D_1(Q+K)+QK}\right). \tag{9}$$

Proof of Theorem *1*: The joint distribution of $p(X_i, S_i, Z_i, T_i)$ in Fig. 2 is a multivariate Gaussian distribution $\sim \mathcal{N}(0, \mathbf{\Sigma}_{X,S,Z,T})$ with a covariance matrix $\mathbf{\Sigma}_{X,S,Z,T} = \text{diag}(D_1, Q, \tilde{N}, K)$, where $\tilde{N} = \frac{D_2}{a}$, and a is a scalar constant given in (7). The channel output can be expressed as $\hat{W}_i = a \cdot (X_i + S_i + Z_i) + \beta(1-a)(S_i + T_i)$. Assuming $U_i = X_i + \alpha S_i$, where $\alpha \in \mathbb{R}$ is a real constant to be determined, the joint distribution $p(U_i, \hat{W}_i, \hat{S}_i)$ is a multivariate Gaussian distribution with 0 mean and covariance matrix $\mathbf{\Sigma}_{U,\hat{W},\hat{S}} = \mathbf{B}\mathbf{\Sigma}_{X,S,Z,T}\mathbf{B}^{\mathsf{t}}$ where $\mathbf{B}$ is a matrix satisfying the equation

$$\begin{bmatrix} U_i \\ \hat{W}_i \\ \hat{S}_i \end{bmatrix} = \mathbf{B} \begin{bmatrix} X_i \\ S_i \\ Z_i \\ T_i \end{bmatrix}. \tag{10}$$

Resolving (10) yields

$$\mathbf{B} = \begin{pmatrix} 1 & \alpha & 0 & 0 \\ a & (a+\beta-a\beta) & a & \beta(1-a) \\ 0 & 1 & 0 & 1 \end{pmatrix} \tag{11}$$

which in turn leads to the covariance matrix

$$\mathbf{\Sigma}_{U,\hat{W},\hat{S}} = \begin{pmatrix} D_1 + \alpha 2Q & aD_1 + \alpha Q(a+\beta-a\beta) & \alpha Q \\ aD_1 + \alpha Q(a+\beta-a\beta) & a^2 D_1 + (a+\beta-a\beta)^2 Q & (a+\beta-a\beta)Q \\ & \quad +a2\tilde{N} + \beta 2(1-a)^2 K & \quad +\beta(1-a)K \\ \alpha Q & (a+\beta-a\beta)Q + \beta(1-a)K & Q+K \end{pmatrix}. \tag{12}$$

Now, the capacity of the watermark communication channel is given in [19] by

$$C = \max_{p(x,u|s)} [I(U; \hat{W}, \hat{S}) - I(U; S)], \tag{13}$$

where the maximum is over all joint distributions of $p(u)p(s, \hat{s}, x|u)p(\hat{w}|x, s, \hat{s})$, and U is an auxiliary random variable. Selecting U in a confined class such that $U_i = X_i + \alpha S_i$, the terms of mutual information $I(U; \hat{W}, \hat{S})$ and $I(U; S)$ can be expressed as a function of the parameter α as

$$\begin{aligned} I(U; \hat{W}, \hat{S}) &= h(U) + h(\hat{W}, \hat{S}) - h(U, \hat{W}, \hat{S}) \\ &= h(X + \alpha S) + h(a(X + S + Z) + (S + T)\beta(1 - a), S + T) - h(U, \hat{W}, \hat{S}) \\ &= \frac{1}{2} \ln \Big((2\pi e)(D_1 + \alpha^2 Q) \Big) + \frac{1}{2} \ln \Big((2\pi e)^2 (a^2((D_1 + \tilde{N})(Q + K) + QK)) \Big) \\ &- \frac{1}{2} \ln \Big((2\pi e)^3 (a^2(D_1 QK(1 - \alpha)^2 + \tilde{N}K(D_1 + \alpha^2 Q) + D_1 \tilde{N} Q)) \Big) \\ &= \frac{1}{2} \ln \left(\frac{(D_1 + \alpha^2 Q)((D_1 + \tilde{N})(Q + K) + QK)}{(D_1 QK(1 - \alpha)^2 + \tilde{N}K(D_1 + \alpha^2 Q) + D_1 \tilde{N} Q)} \right) \end{aligned} \tag{14}$$

and

$$I(U; S) = h(U) + h(S) - h(U, S) = \frac{1}{2} \ln \left(\frac{D_1 + \alpha^2 Q}{D_1} \right), \tag{15}$$

leading to an achievable rate $R(\alpha) = I(U; \hat{W}, \hat{S}) - I(U; S)$ equal to

$$R(\alpha) = \frac{1}{2} \ln \left(\frac{D_1(K(D_1 + Q + \tilde{N}) + Q(D_1 + \tilde{N}))}{D_1 QK(1 - \alpha)^2 + \tilde{N}K(D_1 + \alpha^2 Q) + D_1 \tilde{N} Q} \right). \tag{16}$$

Maximizing (16) with respect to the parameter α, yields the maximum achievable rate

$$R(\alpha^\nabla) = \frac{1}{2} \ln \left(1 + \frac{D_1}{\tilde{N}} \right) = \frac{1}{2} \ln \left(1 + \frac{D_1}{D_2} - \frac{D_1(Q + K)}{D_1(Q + K) + QK} \right), \tag{17}$$

obtained for

$$\alpha^\nabla = D_1/(D_1 + \tilde{N}). \tag{18}$$

Now, we use an upper bound expression for the capacity where both embedding and extraction processes $(F1, G1)$ know the host sequence $\boldsymbol{S}$, and its noisy version $\boldsymbol{S}$. We first evaluate the upper bound term $C^* = \max_{p(x|s\hat{s})} I(X; \hat{W}|S, \hat{S})$, and show that $R(\alpha) \leq C^*$. We also show that the equality holds for $\alpha^\nabla =$

$D_1/(D_1+\tilde{N})$. Hence

$$\begin{aligned}
C^* &= \max_{p(x|s\hat{s})} I(X;\hat{W}|S,\hat{S}) \\
&= h(\hat{W}|S,\hat{S}) - h(\hat{W}|X,S,\hat{S}) \\
&= h(a(X+Z)|S,\hat{S}) - h(aZ) \\
&\leq h(a(X+Z)) - h(aZ) \qquad (19)\\
&\leq h(\mathcal{N}(0,a^2(D1+\tilde{N})) - h(\mathcal{N}(0,a^2\tilde{N})) \qquad (20)\\
&= \frac{1}{2}\ln((2\pi e)a^2(D_1+\tilde{N})) - \frac{1}{2}\ln((2\pi e)a^2\tilde{N}) \\
&= \frac{1}{2}\ln\left(1+\frac{D_1}{\tilde{N}}\right), \qquad (21)
\end{aligned}$$

where (19) follows from the fact that conditioning reduces the entropy, and (20) from the fact that Gaussian distribution maximizes the entropy for a given variance. Since $C^* = R(\alpha^{\nabla})$, one can conclude that the capacity of this channel is

$$R_C = R(\alpha^{\nabla}) = \frac{1}{2}\ln\left(1+\frac{D_1}{D_2} - \frac{D_1(Q+K)}{D_1(Q+K)+QK}\right). \qquad (22)$$

This completes the proof.

Remark 1 *The equality of* (19)*, and* (20) *hold if* $\boldsymbol{X}$ *is a sequence of i.i.d. random variables* $X_i \sim \mathcal{N}(0,D_1)$*, and mutually independent from* $\boldsymbol{S}$ *and* $\hat{\boldsymbol{S}}$*.*

4 Rate distortion and Capacity analysis

In this section, we analyze the rate-distortion performance and the capacity of the considered set-up for the cases where $\hat{\boldsymbol{S}}$ is not known to $G1$, $G2$, or both.

4.1 Rate distortion analysis

4.1.1 Absence of $\hat{\boldsymbol{S}}$ *to* $G2$

For a fixed distortion D_2 for the Wyner-Ziv coding, the absence of the partial side information $\hat{\boldsymbol{S}}$ to the decompresser process $G2$ results in a rate loss which can be expressed as

$$R_{00}(D_2) - R_{01}(D_2) = \begin{cases} \frac{1}{2}\ln\left(1+\frac{Q^2}{D_1(Q+K)+QK}\right), & 0 < D_2 < D_1 + \frac{QK}{Q+K}, \\ \frac{1}{2}\ln^+\left(\frac{D_1+Q}{D_2}\right), & D_1 + \frac{QK}{Q+K} \leq D_2 \end{cases}. \qquad (23)$$

This is illustrated in Fig. 4 which shows the rate-distortion curves without and with partial side information at the decompression process for different values of the variance K (which reflects the quality of the side information available to the Wyner-Ziv decoder). The embedding power is set to $D_1 = 0.061$ for $Q = 1$, hence is limited with respect to the host signal power ($D_1 << Q$).

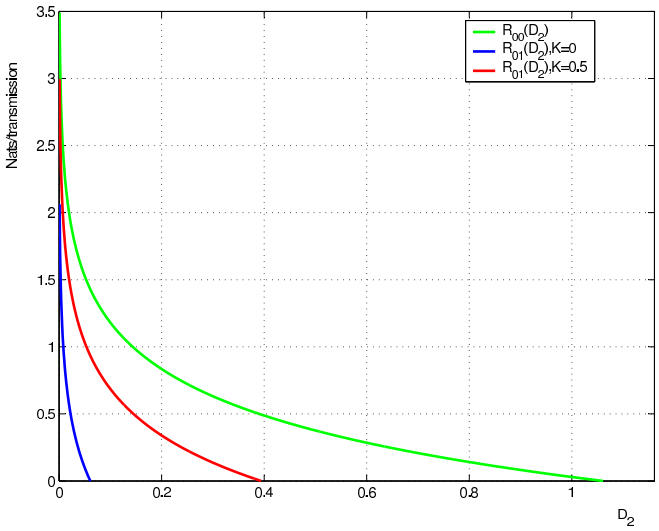


Fig. 4. Rate distortion functions $R_{00}(D_2)$ and $R_{01}(D_2)$ for $Q = 1$, $D_1 = 0.061$ and several K values such as $K = 0$ and $K = 0.5$.

Similarly, in order to obtain an identical compression rate of R, the equivalent distortion D_2^* where $\hat{\boldsymbol{S}}$ is not accessible to $G2$ is

$$D_2^* = \frac{D_2(D_1 + Q)(Q + K)}{D_1(Q + K) + QK}, \tag{24}$$

which is greater than or equal to D_2, where D_2 is the corresponding distortion when $\hat{\boldsymbol{S}}$ is available to the $G2$ for a given rate R, and for a fix variance D_1, Q and K.

4.2 Capacity analysis

4.2.1 Absence of $\hat{\boldsymbol{S}}$ to $G1$

The derivation of the capacity term for the case where $\hat{\boldsymbol{S}}$ is not known to the extraction process $G1$ is similar to the proof of Theorem-1. It is sufficient to evaluate the capacity expression given in [1] as

$$C = \max_{p(x,u|s)} [I(U; \hat{W}) - I(U; S)]. \tag{25}$$

Selecting U in a confined class such that $U_i = X_i + \alpha S_i$, the mutual information term $I(U;\hat{W})$ can be expressed as

$$\begin{aligned} I(U;\hat{W}) &= h(U) + h(\hat{W}) - h(U,\hat{W}) \\ &= h(X+\alpha S) + h(a(X+S+Z) + (S+T)\beta(1-a)) - h(U,\hat{W}) \\ &= \frac{1}{2}\ln\Big((2\pi e)(D_1+\alpha^2 Q)\Big) + \frac{1}{2}\ln\Big((2\pi e)(a^2(D_1+Q+\tilde{N}))\Big) \\ &- \frac{1}{2}\ln\Big((2\pi e)^2(a^2(D_1 Q(1-\alpha)^2 + \tilde{N}(D_1+\alpha^2 Q)))\Big) \\ &= \frac{1}{2}\ln\left(\frac{(D_1+\alpha^2 Q)(D_1+Q+\tilde{N})}{D_1 Q(1-\alpha)^2 + \tilde{N}(D_1+\alpha^2 Q)}\right). \end{aligned} \tag{26}$$

From (26) and (15), the achievable rate $R(\alpha) = I(U;\hat{W}) - I(U;S)$ equals to

$$R(\alpha) = \frac{1}{2}\ln\left(\frac{D_1(D_1+Q+\tilde{N})}{D_1 Q(1-\alpha)^2 + \tilde{N}(D_1+\alpha^2 Q)}\right). \tag{27}$$

Maximizing (27) with respect to the parameter α, yields the maximum achievable rate

$$R(\alpha^\nabla) = \frac{1}{2}\ln\left(1+\frac{D_1}{\tilde{N}}\right) = \frac{1}{2}\ln\left(1+\frac{D_1}{D_2} - \frac{D_1(Q+K)}{D_1(Q+K)+QK}\right), \tag{28}$$

obtained for α^∇ given in (18). Since (2) and (28) are equal, we conclude that the absence of $\hat{\boldsymbol{S}}$ to the extraction process $G1$ has no effect to the overall capacity.

4.2.2 Absence of $\hat{\boldsymbol{S}}$ to $G1$ and $G2$

In order to analyze the effect of the absence of $\hat{\boldsymbol{S}}$ to $G1$ and $G2$ to the overall capacity of the system, we first derive the capacity term of the equivalent schema where there is no $\hat{\boldsymbol{S}}$ available to the $G1$ and $G2$, and compare with the capacity given in 3.2 for an identical compression rate R.

Lemma 2 *The capacity R_C for the communication system given in Fig. 5 is*

$$R_C = \frac{1}{2}\ln\left(1+\frac{D_1}{D_2} - \frac{D_1}{D_1+Q}\right). \tag{29}$$

The proof of Lemma 2 is given in Appendix-B.

Remark 2 *The asymptotic behavior of the capacity term in* (9) *for $K \to \infty$ coincides with the capacity term in* (29).

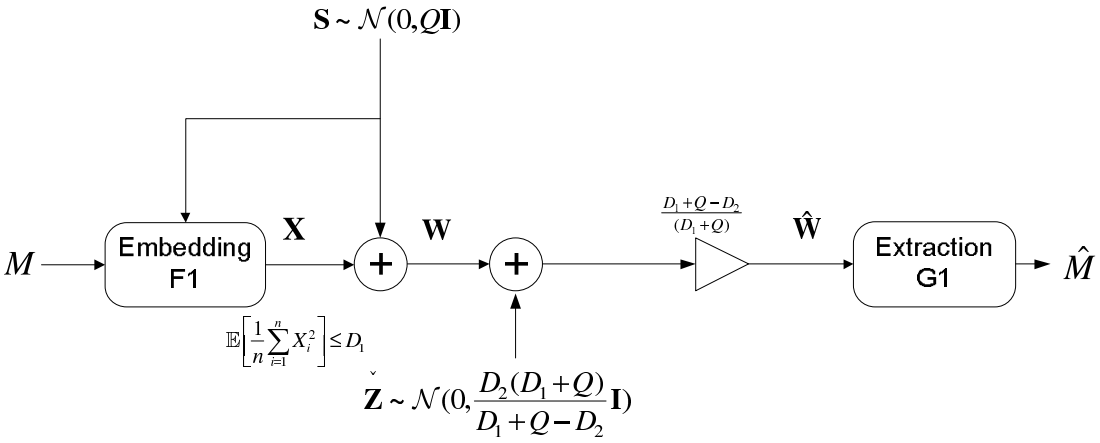


Fig. 5. Equivalent communication channel where noisy side information $\hat{\boldsymbol{S}}$ is not accessible at the decoder.

Now, we can compare the capacity terms that are given in (9) and (29) for an identical compression rate R and under a fixed signal power D_1, Q, and K. The distortion level D_2^* is given in (24) for the absence of $\hat{\boldsymbol{S}}$. If the value D_2 in (29) is replaced by the value of D_2^*, the capacity of the scheme in the absence of $\hat{\boldsymbol{S}}$, but with the same compression rate is then

$$R_C = \frac{1}{2}\ln\left(1 + \frac{D_1[(Q+K)(D_1-D_2)+QK]}{D_2(D_1+Q)(Q+K)}\right), \tag{30}$$

which is less than the capacity term given in (9). Hence the absence of $\hat{\boldsymbol{S}}$ to $G1$ and $G2$ results in a capacity loss.

5 Practical Scheme

It has been shown in the Section-4.2.1 that if the embedder $F1$ uses the optimal value of α at the coding stage, the accessibility of the realization of the noisy side information $\hat{\mathbf{s}}$ to the extraction process $G1$ does not effect the capacity and the rate-distortion function. In the proposed practical coding scheme, the optimal α value is assumed to be known to the embedder $F1$, hence $\hat{\mathbf{s}}$ is not used for the extraction of the watermark in $G1$.

A message M of $n/2$ bits is embedded in a host vector $\mathbf{s}$ of length n, which corresponds to an embedding rate of $R_C = 1/2$ bit/channel use, such that the watermarked signal $\mathbf{w}$ satisfies an average distortion constraint $\mathbb{E}\left[d(\mathbf{s},\mathbf{w})\right] \leq D_1$. The watermarked signal $\mathbf{w}$ is then compressed with a rate R_S bit/channel use. The decompression process $G2$ reconstructs $\hat{\mathbf{w}}$ using the compressed data and noisy version of the host signal $\hat{\mathbf{s}} = \mathbf{s} + \mathbf{t}$. The compression is performed under an MSE constraint of $\mathbb{E}\left[d(\mathbf{w},\hat{\mathbf{w}})\right] \leq D_2$. The hidden message is then extracted as $\hat{M}$ with the help of $\hat{\mathbf{w}}$ and $\hat{\mathbf{s}}$. The decoding error rate can be

calculated as

$$P_e = \frac{\sum_{i=1}^{n/2}(M_i \oplus \hat{M}_i)}{n/2}, \tag{31}$$

where $\sum$ is defined to be a summation over real numbers, and $\oplus$ is the modulo-2 summation operator. In the experiments part, for a fixed R_C, Q, K, and D_1 values, we evaluate the proposed scheme in terms of the message bit error rate P_e, and the rate-distortion performances such as R_S and D_2 of the Wyner-Ziv setup. Moreover, we compare the practical performance of the system with respect to the theoretical bounds which are derived in Sec.-3. For comparison, we also simulate the practical scheme where $\hat{\boldsymbol{S}}$ is absent to both $G1$ and $G2$.

5.1 Dirty Paper Embedding/Extraction Pair $(F1, G1)$

The $(F1, G1)$ pair is based on a superposition of a high performance source code $\mathcal{C}_0$, and a high performance channel code $\mathcal{C}_1$ as proposed in [18] (see Fig. 6). The receiver performs an iterative decoding of the channel and source codes. Due to their high performances, TCQ codes are good candidates for $\mathcal{C}_0$, and LDPC codes are good candidates for $\mathcal{C}_1$. The difference from [18] is mainly in the choice of $\mathcal{C}_0$ and $\mathcal{C}_1$. $\mathcal{C}_1$ is designed to expand the secret message M to a codeword such that 1 bit of the codeword is embedded into one sample of the host signal. Since we want to achieve an embedding rate of 1/2 bit per channel use, we use an irregular LDPC code with rate 1/2 with a variable node edge distribution

$$\lambda(2, 3, 7, 20) = (0.4811, 0.3143, 0.15356, 0.051), \tag{32}$$

and a uniform check node edge distribution as described in [22]. The performance of the LDPC code is 0.11 dB away from BIAWGN limit given in [23]. $\mathcal{C}_0$ is chosen so that the distortion constraint between the host signal $\mathbf{s}$ and the watermarked signal $\mathbf{w}$, $\mathbb{E}\left[d(\mathbf{s}, \mathbf{w})\right] \leq D_1$, is guaranteed. The trellis coded quantization (TCQ) [24] with a 1/2 convolutional code feedback polynomial $(671, 1631)$ in octal digit, and a four-level uniform PAM codebook, is chosen for the code $\mathcal{C}_0$ (See Fig. 7(a) for codebook partitioning of $\mathcal{C}_0$). The source code $\mathcal{C}_0$ can quantize an input $\mathbf{x}$, uniformly distributed in the range [−1,1], to $\mathcal{Q}_{\mathcal{C}_0}(\mathbf{x})$ with a mean distortion of $D_1 = 0.062$ where $\mathcal{Q}_{\mathcal{C}_0}(\mathbf{x})$ is the reconstruction of the quantized vector $\mathbf{x}$. The rate-distortion limit is 0.0585 for $R = 1$ bit/sample, hence $\mathcal{C}_0$ allows to quantize the input source with a gap of 0.19 dB from the theoretical limit.

The embedding algorithm proceeds as follows. Given the $n/2$ bit message M and the host vector $\mathbf{s} = \{s_1, s_2, .., s_n\}$ of dimension n, the embedder starts with the computation of $\mathbf{c_1}$. The 1/2 rate LDPC encoding of the $n/2$ bit message M generates a binary codeword of $\mathbf{k}$ of n bits, such that $\mathbf{k} = \mathbf{G}M$, where

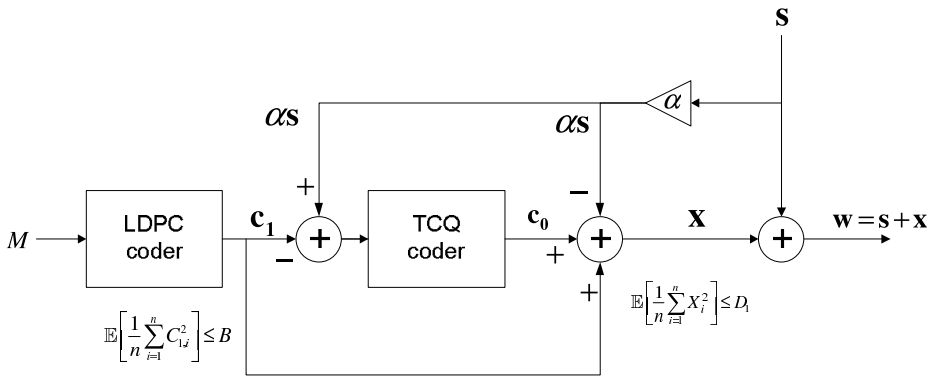
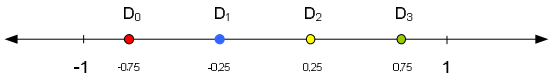
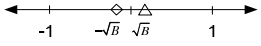


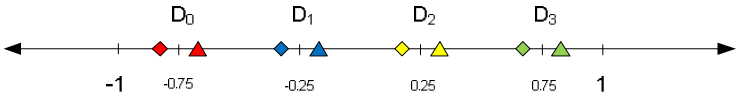
Fig. 6. Block diagram of the embedding process $F1$ in order to embed message M into the host vector $\mathbf{s}$ using superposition coding.



(a) Codebook of $\mathcal{C}_0$.



(b) PAM codebook of $\mathcal{C}_1$.



(c) Codebook of $\mathcal{C}_0 + \mathcal{C}_1$.

Fig. 7. Codebook partitioning of $\mathcal{C}_0$, $\mathcal{C}_1$, and $\mathcal{C}_0 + \mathcal{C}_1$.

$\mathbf{G}$ is the generator matrix of the LDPC coder. Then, n-length vector $\mathbf{c_1}$ is constructed from a two-level PAM as

$$c_{1,i} = -1^{(k_i+1)}\sqrt{B}, \tag{33}$$

where $k_i \in 0, 1$, B is a scalar constant $B = \alpha D_1$, and α is the optimal scalar constant which is found in (18).

The second step mainly searches a n-length $\mathbf{c_0}$ vector such that $\mathbf{c_0} + \mathbf{c_1}$ is jointly typical with the scaled host vector $\alpha\mathbf{s}$. The vector $\alpha\mathbf{s} - \mathbf{c_1}$ is thus fed into the TCQ quantizer. The Viterbi algorithm searches the path on the trellis that corresponds to the minimum quantization error. The quantized vector is assigned to $\mathbf{c_0}$ as

$$\mathbf{c_0} = \mathcal{Q}_{\mathcal{C}_0}(\alpha\mathbf{s} - \mathbf{c_1}). \tag{34}$$

The superposition code $\mathbf{c}$ is then given by

$$\mathbf{c} = \mathbf{c_0} + \mathbf{c_1} = \mathcal{Q}_{\mathcal{C}_0}(\alpha\mathbf{s} - \mathbf{c_1}) + \mathbf{c_1}. \tag{35}$$

The codebook partitioning of $\mathcal{C}_0$, $\mathcal{C}_1$, and $\mathcal{C}_0 + \mathcal{C}_1$ is given in Fig. 7. The em-

bedding noise x is computed as

$$\mathbf{x} = \mathbf{c_0} + \mathbf{c_1} - \alpha \mathbf{s}, \tag{36}$$

leading to the watermarked signal

$$\mathbf{w} = \mathbf{s} + \mathbf{x}. \tag{37}$$

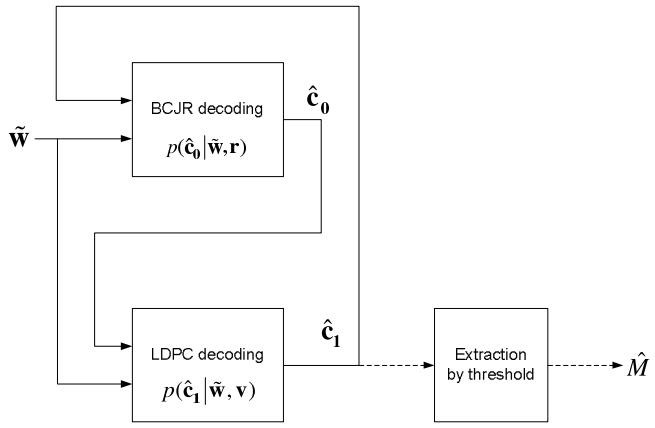


Fig. 8. Block diagram of the extraction process $G1$.

The Wyner-Ziv coding pair $(F2, G2)$ performs a compression of the watermark signal $\mathbf{w}$ with an MSE constraint $\mathbb{E}\left[d(\mathbf{w}, \hat{\mathbf{w}})\right] \leq D_2$. The extraction process $G1$ uses a scaled version of the decompressed signal with the form $\tilde{\mathbf{w}} = \alpha\hat{\mathbf{w}}/a$. $G1$, searches a $\hat{\mathbf{c}}_\mathbf{0}$ and $\hat{\mathbf{c}}_\mathbf{1}$ pair such that the conditional probability

$$P(\tilde{\mathbf{w}}|(\hat{\mathbf{c}}_\mathbf{0} + \hat{\mathbf{c}}_\mathbf{1})) \tag{38}$$

is maximized. The reason behind the use of scaled version $\tilde{\mathbf{w}}$ is that it can be expressed as the sum $\mathbf{c_0} + \mathbf{c_1} + \hat{\mathbf{z}}$, and an estimation of $\hat{\mathbf{c}}_\mathbf{0}$ and $\hat{\mathbf{c}}_\mathbf{1}$ can be performed iteratively. Hence, $\tilde{\mathbf{w}}$ yields

$$\begin{aligned}
\tilde{\mathbf{w}} &= \alpha\hat{\mathbf{w}}/a \\
&= \alpha\mathbf{s} + \alpha\mathbf{x} + \alpha\mathbf{z} \\
&= \mathbf{c_0} + \mathbf{c_1} - (1-\alpha)\mathbf{x} + \alpha\mathbf{z} && (39) \\
&= \mathbf{c_0} + \mathbf{c_1} + \hat{\mathbf{z}}, && (40)
\end{aligned}$$

where (39) follows from (36). The effective noise $\hat{\mathbf{z}}$ is defined as $\hat{\mathbf{z}} = -(1-\alpha)\mathbf{x} + \alpha\mathbf{z}$, hence is Gaussian distributed with mean 0, and a variance $\sigma^2_{\hat{z}}$ given by

$$\sigma^2_{\hat{z}} = (1-\alpha)^2 D_1 + \alpha^2 \tilde{N} = \alpha\tilde{N}. \tag{41}$$

The main steps of the decoding process can be seen in Fig. 8. The decoding process iterates between a BCJR decoder and a LDPC belief propagation decoder which produces soft outputs on $\hat{\mathbf{c}}_\mathbf{0}$ and $\hat{\mathbf{c}}_\mathbf{1}$ respectively. The iterative

decoding process comprises three main steps which are described in detail in [10]. For sake of completeness, these steps are being recalled here:

- **Updating likelihood values:** Two likelihood matrixes $\mathbf{v}$ and $\mathbf{r}$ are calculated, which contain a priori information that initiate the cost functions on the trellis path of TCQ and the bipartite graph of the LDPC decoder respectively. From the scaled decompressed signal $\tilde{\mathbf{w}}$, the extraction process $G1$ first computes an $n \times 4$ matrix $\mathbf{v}$ to be used by the BCJR decoder. The second likelihood computation concerns a $n \times 2$ matrix $\mathbf{r}$ which initializes the cost functions on the bipartite graph of the LDPC decoder. At the beginning of the decoding, the LDPC decoder initializes $r_{i0} = r_{i1} = 0.5$. The (i,t)-th entry of $\mathbf{v}$, v_{it}, contains the likelihood of $\hat{c}_{0,i} = \delta(t)$, given $\tilde{w}_i$ and $\mathbf{r}$. The term $\delta(t)$ denotes the t-th quantization level of the TCQ coder which is the closest to $\tilde{w}_i$. Each element of $\mathbf{v}$ is calculated as

$$v_{it} = \frac{\sum_{b=0}^{1} r_{ib} \cdot f_{\sigma_{\hat{z}}}(\tilde{w}_i - \delta(t) + (-1)^{(b+1)}\sqrt{B}))}{\sum_{k=1}^{4}\sum_{b=0}^{1} r_{ib} \cdot f_{\sigma_{\hat{z}}}(\tilde{w}_i - \delta(k) + (-1)^{(b+1)}\sqrt{B}))} \tag{42}$$

  for $i = 1, 2, .., n$ and $t = 1, 2, 3, 4$, where $f_{\sigma_{\hat{z}}}$ is the probability density function of a Gaussian r.v. $\mathcal{N}(0, \alpha\tilde{N})$, and r_{i0}, r_{i1} are the messages coming from the LDPC node which indicate whether the likelihood of the i-th element of $\hat{\mathbf{c}}_\mathbf{1}$, $p(\hat{c}_{1i} = (-1)^{(b+1)}\sqrt{B})$ for $b \in 0, 1$.

  The (i,b)-th entry of $\mathbf{r}$, r_{ib} corresponds to the likelihood of $p(\hat{c}_{1i} = (-1)^{(b+1)}\sqrt{B})$ given $\tilde{w}_i$ and $\mathbf{v}$ for $b \in 0, 1$. Each element of $\mathbf{r}$ is calculated as

$$r_{ib} = \frac{\sum_{t=1}^{4} v_{it} \cdot f_{\sigma_{\hat{z}}}(\tilde{w}_i - \delta(t) + (-1)^{(b+1)}\sqrt{B})}{\sum_{t=1}^{4} v_{it} \cdot f_{\sigma_{\hat{z}}}(\tilde{w}_i - \delta(t) - \sqrt{B}) + \sum_{t=1}^{4} v_{it} \cdot f_{\sigma_{\hat{z}}}(\tilde{w}_i - \delta(t) + \sqrt{B})} \tag{43}$$

  for $i = 1, 2, .., n$ and $b = 0, 1$. Similarly $f_{\sigma_{\hat{z}}}$ is the probability density function of a Gaussian r.v. $\mathcal{N}(0, \alpha\tilde{N})$.
- **Iteration BCJR:** The branch metrics of the trellis are initialized with the received vectors $\mathbf{r}$ for each sample. The BJCR algorithm [25] computes the probability $P(\mathbf{c_0}|\tilde{\mathbf{w}}, \mathbf{r})$, which is then mapped to the message matrix $\mathbf{v}$.
- **Iteration LDPC:** The variable node likelihoods $\mathbf{v}$ are calculated as explained above. Then, 10 LDPC iterations are executed between the variable and the check nodes as in [26]. The LDPC decoder generates the likelihood probability $P(\mathbf{c_1}|\tilde{\mathbf{w}}, \mathbf{v})$, which is then mapped to the message vector $\mathbf{r}$.

For a unit host variance $Q = 1$, an embedding noise variance $D_1 = 0.062$, and an embedding rate of $R_C = 1/2$ bit per channel use, the message can be perfectly decoded with an AWGN variance that corresponds to 1.5 dB away from the theoretical limit [2].

[2] The theoretical AWGN noise limit is given as $N = \frac{D_1}{e^{2R_c}-1}$ in [2].

$(F2, G2)$ pair can be summarized in 4 consecutive steps: quantification, syndrome coding, decoding, and final estimation (See Fig. 9). The watermarked signal $\mathbf{w}$ is d level quantized using a TCQ code with a $1/2$ convolutional code feedback polynomial $(561, 753)$ in octet digit, and a $2 \times d$ level Max-Lloyd partitioning of the codewords (See Fig. 10 for code partitioning example for Gaussian distribution $\sim \mathcal{N}(0, 1.0613)$ and $d = 4$). The quantization output can be grouped in $l = \log_2(d)$ bitplanes which can be stated as $[\mathbf{b_0}, \mathbf{b_1}, .., \mathbf{b_{l-1}}]$, where $\mathbf{b_0}$ governs the trellis path, and $[\mathbf{b_1}, .., \mathbf{b_{l-1}}]$ are the sub-codebook indices of the TCQ from the most significant bit to the least significant one.

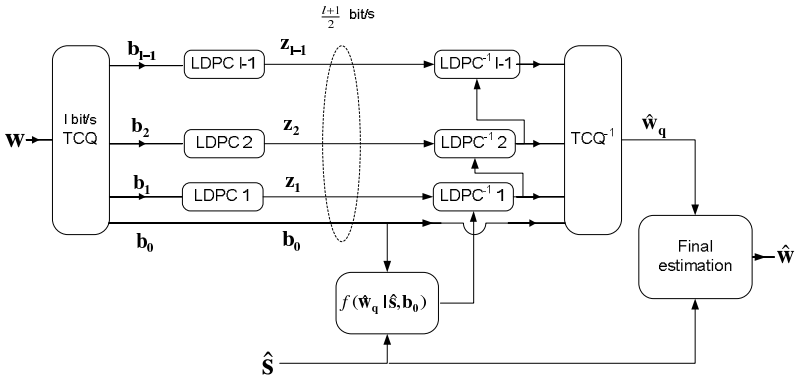


Fig. 9. Block diagram of the Wyner-Ziv coding $(F2, G2)$.

In syndrome coding step, each bitplane of the sub-codebook indices $\mathbf{b_i}$ is Slepian-Wolf encoded with irregular LDPC code with rate $1/2$ (as in Sec-5.1), resulting $\mathbf{z_i} = \mathbf{H}\mathbf{b_i}$ for $i \in 1, .., l-1$, where $\mathbf{H}$ is the parity check matrix of the LDPC code and $\mathbf{z_i}$ is the syndrome of the i-th bitplane. The proposed coding scheme does not compress the trellis path $\mathbf{b_0}$ because of its low correlation nature with the noisy side information $\hat{\mathbf{s}}$. Hence, the sequence $\mathbf{b_0}, \mathbf{z_1}, .., \mathbf{z_{l-1}}$ is sent to the decompression process $G2$ that yields a total rate of $R_S = 1 + \frac{l-1}{2} = \frac{l+1}{2}$ bits per sample.

Fig. 10. Eight level Max-Lloyd codebook partitioning of $\mathcal{N}(0, 1.0613)$ input distribution.

The decompression process $G2$ first calculates the a priori probability $f(\mathbf{b_1}|\hat{\mathbf{s}}, \mathbf{b_0})$, then the bitplane $\mathbf{b_1}$ is estimated from LDPC syndrome decoding of $\mathbf{z_1}$. A modified version of the belief propagation algorithm in [27] is applied for the LDPC syndrome decoding. Consecutively, for the i-th bitplane, $f(\mathbf{b_i}|\hat{\mathbf{s}}, \mathbf{b_{i-1}}, .., \mathbf{b_0})$ is calculated, and LDPC decoding is applied for the i-th bitplane. The inverse TCQ reconstructs $\hat{\mathbf{w}}_\mathbf{q}$ from the decoded bitplanes. The final step is the esti-

mation of $\hat{\mathbf{w}}$, which can be stated as the conditional mean

$$\hat{w}_i = \mathbb{E}\left[W_i = w|\hat{w}_{qi}, \hat{s}_i\right] = \int_{-\infty}^{+\infty} \frac{w f(\hat{w}_{qi}|w) f(w|\hat{s}_i)}{f(\hat{w}_{qi}|\hat{s}_i)} dw, \tag{44}$$

where $f(.|.)$ is the corresponding conditional probability density function. Please note that $f(\hat{w}_{qi}|w)$ is a zero except in a region where w_i has been been quantified as $\hat{w}_{qi}$, and the denominator in (44) is a normalization constant. The intuition of the final estimation is to exploit the availability of $\hat{\boldsymbol{S}}$ to $G2$ for low Wyner-Ziv rate R_S. Hence our scheme is close to the [27], the difference is that the trellis path $\mathbf{b_0}$ is not compressed in the proposed scheme, and the a priori probabilities are calculated using the corresponding distribution between W and $\hat{S}$.

For a unit host variance $Q = 1$, an embedding noise variance $D_1 = 0.062$, a Wyner-Ziv compression rate of 1.5 bit per sample, a correlation noise of $K = 0.3$, and a block length of $n = 20000$, the watermarked signal $\mathbf{w}$ is compressed with a MSE of $\mathbb{E}\left[d(\mathbf{w}, \hat{\mathbf{w}})\right] \leq 0.075$ which is 3.11 dB away from the theoretical limit given in (6).

5.3 Performance Analysis of the Proposed System

In order to evaluate the performance of the proposed coding scheme, we analyze the gap between the theoretical and practical scheme for a fix operating point where the embedding rate R_C is 1/2 bit/channel use, the embedding power D_1 is 0.062, the variance of the host signal Q is 1, and the correlation noise K is 0.1 (The data set and the parameters of the practical scheme can be found in Table-1, and Table-2 respectively). The theoretical maximum compression distortion $D_{2\text{-theo}}$ which allows to achieve the targeted capacity is given by (9) as

$$\frac{1}{2} = \frac{1}{2}\log_2\left(1 + \frac{0.062}{D_{2\text{-theo}}} - \frac{0.062(1+0.1)}{0.062(1+0.1)+0.1}\right). \tag{45}$$

which yields a maximum compression distortion $D_{2\text{-theo}} = 0.0441$.

For a block length of $n = 20000$ samples, we evaluate the performance of the proposed practical code designs $(F1, G1)$ and $(F2, G2)$. The simulation results are calculated from an average of 10^7 samples. We assumed that a message error rate of $P_e \leq 10^{-5}$ is enough for a reliable communication, hence we search operating points where P_e vanishes. Fig. 11 shows the P_e performance of the proposed system for different compression rates R_S. The leftmost lineplot in Fig. 11 corresponds to Wyner-Ziv setup where a noisy side information $\hat{\boldsymbol{S}}$ is available to the decompression process $G2$ where $K = 0.1$. We achieve

Table 1
Data Set.

Data Set	
n	20000
M	Bernoulli(0.5) with length 10000
Q	1
K	0.1
R_C	$10000/20000 = 0.5$
Target P_e	$\leq 10^{-5}$
Number of simulation	10^7 samples for each compression rate

Table 2
Parameter list.

Parameters of $F1$	
LDPC code	1/2 rate with edge distributions $\lambda(2,3,7,20) = (0.4811, 0.3143, 0.15356, 0.051)$, $\rho(8) = 1$.
TCQ code	1/2 rate convolutional code feedback polynomial $(671, 1631)$ in octal digit.
TCQ codebook	four-level uniform PAM with reconstruction levels $[-0.75, -0.25, 0.25, 0.75]$.
D_1	0.062
Parameters of $G1$	
iterations	For 1 iteration of BCJR, 10 iteration of LDPC.
max # of iteration of BCJR	15
Parameters of $F2$	
LDPC code	1/2 rate with variable node edge distribution $\lambda(2,3,7,20) = (0.4811, 0.3143, 0.15356, 0.051)$, and check node edge distribution $\rho(8) = 1$.
d	Tested for values 4,8 and 9.
TCQ code	1/2 rate convolutional code feedback polynomial $(561, 753)$ in octal digit.
TCQ codebook	$2 \times d$ Max-Lloyd partitioning.
R_S	$\frac{\log_2(d)+1}{2}$
Parameters of $G2$	
max # of iteration of LDPC	150

a decoding error rate $P_e = 4.3 \times 10^{-6}$ for a $d = 9$ level compression scheme corresponding to a rate $R_S = 2.085$ and distortion $D_2 = 0.019$. Hence we have a $10\log_{10}(\frac{0.0441}{0.019}) = 3.66$ dB gap with respect to the theoretical $D_{2\text{-theo}}$ value found in (45). Furthermore, if we analyze the Wyner-Ziv performance for this operational D_2 value, then the theoretical rate $R_{S\text{-theo}}$ can be evaluated by resolving (6) as

$$R_{S\text{-theo}} = \frac{1}{2}\log_2\left(\frac{0.062}{0.019} + \frac{0.1}{(1+0.1)0.019}\right), \tag{46}$$

which yields $R_{S\text{-theo}} = 1.5043$ bits/sample. Hence the Wyner-Ziv part of our proposed system has a rate gap of $2.085 - 1.5043 = 0.5807$ bit/sample.

While the stand alone performance of $(F1, G1)$, and $(F2, G2)$ are given in Sec.-5.1 and Sec.-5.2 as within the magnitude of 1.5 dB and 3 dB respectively; the gaps of the joint-coding scheme from the theoretical bounds are in the order of 3.66 dB and 0.5807 bit/sample in D_2 and R_S respectively.

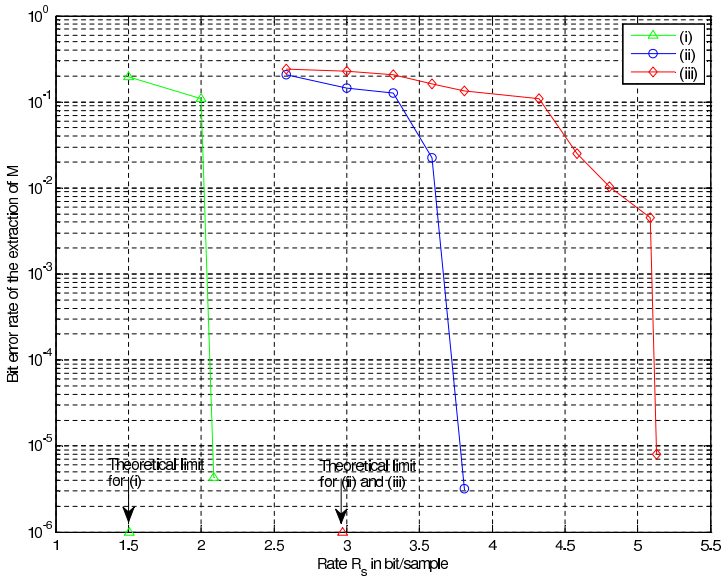


Fig. 11. Message error rate P_e versus compression rate R_S for three cases: *i)* $\hat{\boldsymbol{S}}$ is available to $G2$ where $K = 0.1$, the proposed Wyner-Ziv compression scheme is applied to $\boldsymbol{W}$. *ii)* $\hat{\boldsymbol{S}}$ is absent to $G2$, TCQ quantification is applied to $\boldsymbol{W}$. *iii)* $\hat{\boldsymbol{S}}$ is absent to $G2$, scalar quantification is applied to $\boldsymbol{W}$.

Furthermore, we simulate the effect of the absence of $\hat{\boldsymbol{S}}$ to the decompression process $G2$ [3]. In this case, $(F2, G2)$ pair is replaced by the conventional single source quantization set-up. The conventional quantization is tested with both

[3] See Sec.-4 and Fig. 5 for the theoretical background.

scalar and TCQ quantization with Max-Lloyd partitioning. The scalar quantification achieves a compression rate of $R_S = 5.1293$ bit/sample for $P_e \leq 10^{-5}$. The TCQ compression has 1.3219 bit/sample gain with respect to the scalar quantization for a compression distortion of $D_2 = 0.0172309$ and a compression rate of 3.8074 bit/sample. The compression rate bound for the absence of the noisy side information to the decompression process $G2$ case can be evaluated using (5) as $R_{00}(0.0172309) = 0.5 \times \log_2((0.062+1)/0.0172309) = 2.9728$. Hence there exists a rate gap of $3.8074 - 2.9728 = 0.8346$ bit/sample. As seen in Fig. 11, the performance of our joint dirty paper and Wyner-Ziv scheme outperforms the case where $\hat{\boldsymbol{S}}$ is absent to $G2$. Therefore the availability of a noisy side information at the decompression process $G2$ is beneficial for both the compression performance and the dirty paper coding capacity.

5.4 Computational Complexity

The time complexity of the joint data hiding and Wyner-Ziv coding scheme can be summarized in component based as in below:

- Embedding process **F1**: is constituted of LDPC encoding (one matrix multiplication of length $n \times n/2$ by $n/2 \times 1$), TCQ compression (viterbi algorithm on trellis with an output of one-bit/sample), TCQ reconstruction (n times lookup table), and final embedding (n summation operations).
- Compression process **F2**: is constituted of TCQ compression (viterbi algorithm on trellis with an output of l-bit/sample), $l-1$ LDPC syndrome calculation ($l-1$ matrix multiplication of length $n/2 \times n$ by $n \times 1$). Please note that there is no TCQ reconstruction stage in $F2$.
- Decompression process **G2**: is constituted of a likelihood calculation $f(W_q|\hat{S}, b_0)$, $l-1$ independent LDPC decoding given likelihood calculations based on previously decoded bitplanes $b_0, .., b_{i-1}$, and final estimation of $\hat{\boldsymbol{W}}$ given $\boldsymbol{W_q}$ and $\hat{\boldsymbol{W}}$.
- Extraction **G1**: is constituted of joint LDPC and BCJR decoding (one iteration of BCJR for ten iterations of LDPC, with a maximum of 15 BCJR iterations).

Table-3 gives the time consumption for two encoding and two decoding processes for $R_S = 2.085$ bit/sample (on a Intel Core 2 Duo processor at 2.0 GHz). As seen, the time consumption for $F1$ and $F2$ is in the order of 1 second. The decompression process $G2$ consumes 76 seconds for decoding of 3 bitplanes, and the extraction process $G1$ consumes 3 seconds which corresponds 25 iterations of LDPC and 2 iterations of BCJR. Please note that, in our implementation of $G2$, the likelihood calculations are not optimized. Due to the use of high precision integral operations, likelihood calculations consume 66 seconds of 76 seconds. LDPC decoding and final estimation of $\hat{\boldsymbol{W}}$

are relatively less complex (9 and 1.2 seconds).

Table 3
Computation time of each block in seconds.

$F1$	$F2$	$G2$	$G1$
1.312	0.312	76.532	2.953

6 Conclusion

This paper derives the theoretical limits of a joint dirty paper coding and Wyner-Ziv coding for the Gaussian case. Moreover a practical coding scheme is proposed which is based on TCQ and LDPC codes. We have promising simulation results that approach the derived embedding rate and the rate-distortion bounds respectively. A variation of dirty paper coding and Wyner-Ziv compression is independently proposed by [28,29] where the decoder has access to a noisy version of the embedded message M. Our proposed scheme has several potential practical applications such as multimedia watermarking and efficient compression of the watermarked signal where a noisy version of the cover signal is available to the decoder. The availability of this noisy version yields to a better compression rate and a higher embedding rate. For future work, the theoretical setup can be extended to an arbitrary distribution of host signal $\boldsymbol{S}$, using the duality between source and channel coding [30], [5, Sec.III.C.3]. Similarly, the rate and capacity loss analysis can be stated for the cases where the correlation noise $\boldsymbol{T}$, and the compression noise $\boldsymbol{Z}$ have arbitrary distributions, using the duality between source and channel coding [31,32].

A Proof of Lemma-1

A.1 Part $R_{11}(D_2)$

We evaluate the general rate distortion formula given in [19]

$$R_{11}(D_2) = R_{W|\hat{S}}(D_2) = \min_{p(\hat{w}|w,\hat{s}):\mathbb{E}[d(w,\hat{w})]\leq D_2} I(W;\hat{W}|\hat{S}), \qquad \text{(A.1)}$$

for our case. We first find a lower bound for the rate distortion function. We then prove that this lower bound is achievable. Since $\mathbb{E}\left[d(w,\hat{w})\right] \leq D_2$, we

observe

$$
\begin{aligned}
I(W;\hat{W}|\hat{S}) &= h(W|\hat{S}) - h(W|\hat{S},\hat{W}) \\
&= h(W,\hat{S}) - h(\hat{S}) - h(W-\hat{W}|\hat{S},\hat{W}) \\
&\geq h(W,\hat{S}) - h(\hat{S}) - h(W-\hat{W}) && \text{(A.2)} \\
&\geq h(W,\hat{S}) - h(\hat{S}) - h(\mathcal{N}(0,\mathbb{E}\left[d(W,\hat{W})\right])) && \text{(A.3)} \\
&= h(W,\hat{S}) - \frac{1}{2}\ln((2\pi e)(Q+K)) - \frac{1}{2}\ln((2\pi e)D_2) \\
&= \frac{1}{2}\ln((2\pi e)^2((Q+D_1)(Q+K)-Q^2)) - \frac{1}{2}\ln((2\pi e)^2(Q+K)D_2) && \text{(A.4)} \\
&= \frac{1}{2}\ln\left(\frac{D_1(Q+K)+QK}{(Q+K)D_2}\right) \\
&= \frac{1}{2}\ln\left(\frac{D_1}{D_2}+\frac{QK}{(Q+K)D_2}\right). && \text{(A.5)}
\end{aligned}
$$

where h is the differential entropy. Please note that (A.2) stems from the fact that conditioning reduces entropy, (A.3) follows from the fact that the Gaussian distribution maximizes the entropy for a given variance, and (A.4) follows from the fact that the joint probability of $p(w,\hat{s})$ is multivariate Gaussian distribution with mean 0 and covariance matrix (4). Hence

$$
R_{11}(D_2) \geq \frac{1}{2}\ln\left(\frac{D_1}{D_2}+\frac{QK}{(Q+K)D_2}\right) \tag{A.6}
$$

in nats.

Remark 3 *Equation* (A.2) *is achieved with equality if* $W-\hat{W}$ *is independent from* $\hat{S}$*, and* $\hat{W}$*. Equation* (A.3) *is achieved with equality if* $W-\hat{W}$ *is Gaussian with variance* D_2*.*

Now, we will show that this lower bound is achievable. The random vector $\hat{\boldsymbol{W}}$ that achieve can be realized by the equivalent system given in Fig. A.1, with $Z_i \sim \mathcal{N}(0,\frac{D_2}{a})$; and a, β are fixed to (7) and (8) respectively.

Fig. A.1 is the same as in [33, Fig. 9.7.3] with conditional mean, $\beta\hat{\boldsymbol{S}}$, subtracted out at the input and added in at the output. For the nonzero rate case (with distortion level $0 \leq D_2 < D_1 + \frac{QK}{Q+K}$), a straightforward calculation of $I(W;\hat{W}|\hat{S})$ for the equivalent channel in Fig. A.1 achieves the lower bound in (A.6). For the zero rate case, if $D_1 + \frac{QK}{Q+K} \leq D_2 < Q + D_1$, we choose $\hat{\boldsymbol{W}} = \hat{\boldsymbol{S}}$; and if $Q + D_1 \leq D_2$, we choose $\hat{\boldsymbol{W}} = \boldsymbol{0}$. This completes the proof.

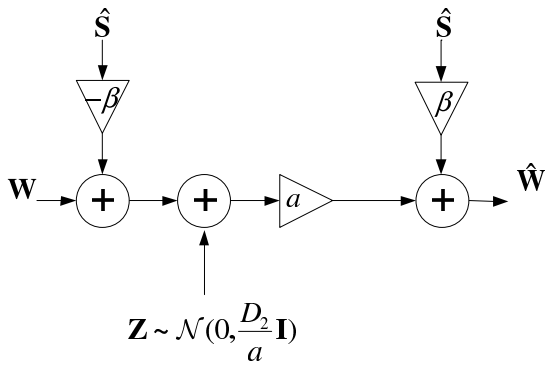


Fig. A.1. Equivalent channel model for the case $R_{11}(D_2)$.

A.2 Part $R_{01}(D_2)$

We evaluate the general rate distortion formula given in [4]

$$R_{01}(D_2) = R^*(D_2) = \inf_{p(\hat{w}|w,\hat{s}):\mathbb{E}[d(w,\hat{w})]\leq D_2} \Big[I(W;E) - I(\hat{S};E) \Big], \tag{A.7}$$

where E is an auxiliary variable and inf is the infimum operator. The infimum is over all E such that $\hat{S} \leftrightarrow W \leftrightarrow E$ form a Markov Chain. Let $\hat{W}$, E are conditionally independent given X, then the term $I(W;E) - I(\hat{S};E)$ in (A.7) is

$$\begin{aligned} I(W;E) - I(\hat{S};E) &= h(E|\hat{S}) - h(E|W) \\ &= h(E|\hat{S}) - h(E|W,\hat{S}) \\ &= I(W;E|\hat{S}) \qquad \text{(A.8)} \\ &\geq I(W;\hat{W}|\hat{S}), \qquad \text{(A.9)} \end{aligned}$$

where (A.8) follows from the assumption that $\hat{W}$, E are conditionally independent given X, and (A.9) follows from the data processing inequality. The equality in (A.9) holds if and only if

$$I(W;E|\hat{W},\hat{S}) = 0. \tag{A.10}$$

Random vector $\hat{\boldsymbol{W}}$ that achieves the lower bound can be chosen as in Fig. A.2, where a and β are the scalar constants given in (7) and (8). Please note that the random vectors $\hat{\boldsymbol{W}}$ in Fig. A.1 and Fig. A.2 are same. Moreover, given the values of $(\hat{\boldsymbol{W}}, \hat{\boldsymbol{S}})$, the random vector $\boldsymbol{E}$ is a constant, which implies (A.10). This completes the proof.

We will use this equivalent channel in the derivation of the overall communication channel capacity.

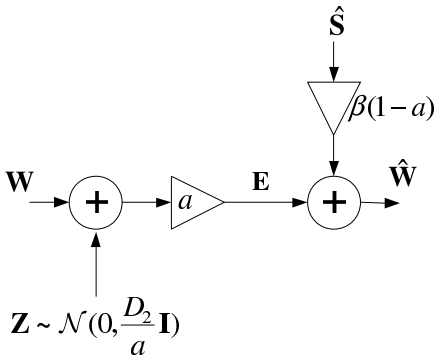


Fig. A.2. The equivalent forward channel of Wyner-Ziv compression of Gaussian source W under MSE criterion.

B Proof of Lemma-2

Let X_i, S_i, and $\check{Z}_i$ be i.i.d. random variables with respective Gaussian distributions $\mathcal{N}(0, D_1)$, $\mathcal{N}(0, Q)$, and $\mathcal{N}(0, \frac{D_2(D_1+Q)}{D_1+Q-D_2})$. Let $\check{N}$ denote the variance of the r.v. $\check{Z}_i$, i.e., $\check{N} = \frac{D_2(D_1+Q)}{D_1+Q-D_2}$, and let $a = \frac{D_1+Q-D_2}{D_1+Q}$. Then, the joint distribution $p(X_i, S_i, \check{Z}_i)$ is a multivariate Gaussian distribution $\sim \mathcal{N}(0, \mathbf{\Sigma}_{X,S,\check{Z}})$ with a covariance matrix $\mathbf{\Sigma}_{X,S,\check{Z}} = \mathrm{diag}(D_1, Q, \check{N})$.

The channel output can be expressed as $\hat{W}_i = a \cdot (X_i + S_i + Z_i)$. Assuming $U_i = X_i + \alpha S_i$, where α is a constant to be determined, the joint distribution $p(U_i, S_i, \hat{W}_i)$ is a multivariate Gaussian distribution with mean 0 and covariance matrix $\mathbf{\Sigma}_{U_i,S_i,\hat{W}_i} = \mathbf{B}\mathbf{\Sigma}_{X,S,Z}\mathbf{B}^{\mathrm{t}}$ where $\mathbf{B}$ is a matrix satisfying the equation

$$\begin{bmatrix} U \\ S \\ \hat{W} \end{bmatrix} = \mathbf{B} \begin{bmatrix} X \\ S \\ Z \end{bmatrix}. \tag{B.1}$$

Resolving (B.1) yields

$$\mathbf{B} = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ a & a & a \end{pmatrix} \tag{B.2}$$

which in turn leads to the covariance matrix

$$\boldsymbol{\Sigma}_{U_i,S_i,\hat{W}_i} = \begin{pmatrix} D_1 + \alpha^2 Q & \alpha Q & a(D_1 + \alpha Q) \\ \alpha Q & Q & aQ \\ a(D_1 + \alpha Q) & aQ & a^2(D_1 + Q + \check{N}) \end{pmatrix}. \tag{B.3}$$

Now, the capacity expression of the communication channel given in (25) can be evaluated. If we employ the same confined class $U_i = X_i + \alpha S_i$, the terms of mutual information $I(U;\hat{W})$ and $I(U;S)$ can be expressed as a function of the parameter α as

$$\begin{aligned} I(U;\hat{W}) &= h\,(U) + h(\hat{W}) - h(U,\hat{W}) \\ &= h(X + \alpha S) + h(a(X + S + \check{Z})) - h(U,\hat{W}) \\ &= \frac{1}{2}\ln\Big((2\pi e)(D_1 + \alpha^2 Q)\Big) + \frac{1}{2}\ln\Big((2\pi e)(a^2(D_1 + Q + \check{N}))\Big) \\ &- \frac{1}{2}\ln\Big((2\pi e)^2(a^2(D_1 Q(1-\alpha)^2 + \check{N}(D_1 + \alpha^2 Q)))\Big) \\ &= \frac{1}{2}\ln\left(\frac{(D_1 + \alpha^2 Q)(D_1 + Q + \check{N})}{D_1 Q(1-\alpha)^2 + \check{N}(D_1 + \alpha^2 Q)}\right) \end{aligned} \tag{B.4}$$

and

$$I(U;S) = h(U) + h(S) - h(U,S) = \frac{1}{2}\ln\left(\frac{D_1 + \alpha^2 Q}{D_1}\right), \tag{B.5}$$

leading to an achievable rate $R(\alpha) = I(U;\hat{W}) - I(U;S)$ equals to

$$R(\alpha) = \frac{1}{2}\ln\left(\frac{D_1(D_1 + Q + \check{N})}{D_1 Q(1-\alpha)^2 + \check{N}(D_1 + \alpha^2 Q)}\right). \tag{B.6}$$

Maximizing (B.6) with respect to the parameter α, yields the maximum achievable rate

$$R(\alpha^\diamond) = \frac{1}{2}\ln\left(1 + \frac{D_1}{\check{N}}\right) = \frac{1}{2}\ln\left(1 + \frac{D_1}{D_2} - \frac{D_1}{D_1 + Q}\right), \tag{B.7}$$

obtained for

$$\alpha^\diamond = D_1/(D_1 + \check{N}). \tag{B.8}$$

From [2], we know that the capacity of the channel can not exceed $C^\dagger = \max_{p(x|s)} I(X;\hat{W}|S)$, the case where both embedding and extraction processes have perfect knowledge of the state information $\boldsymbol{S}$. We first calculate the term $C^\dagger$, and show that $R(\alpha) \leq C^\dagger$. We also show that the equality holds for

$\alpha^\diamond = D_1/(D_1 + \check{N})$. Hence

$$\begin{aligned}
C^\dagger &= \max_{p(x|s)} I(X; \hat{W}|S) \\
&= h(\hat{W}|S) - h(\hat{W}|X, S) \\
&= h(a(X + S + \check{Z})|S) - h(a(X + S + \check{Z})|X, S) \\
&= h(a(X + \check{Z})|S) - h(a\check{Z}|X, S) \\
&\leq h(a(X + \check{Z})) - h(a\check{Z}) \qquad \text{(B.9)} \\
&\leq h(\mathcal{N}(0, a^2(D_1 + \check{N})) - h(\mathcal{N}(0, a^2\check{N})) \qquad \text{(B.10)} \\
&= \frac{1}{2}\ln((2\pi e)a^2(D_1 + \check{N})) - \frac{1}{2}\ln((2\pi e)a^2\check{N}) \\
&= \frac{1}{2}\ln\left(1 + \frac{D_1}{\check{N}}\right), \qquad \text{(B.11)}
\end{aligned}$$

where (B.9) follows from the fact that conditioning reduces the entropy, and (B.10) from the fact that Gaussian distribution maximizes the entropy for a given variance. Since $C^\dagger = R(\alpha^\diamond)$, one can conclude that the capacity of this channel is

$$R_C = R(\alpha^\diamond) = \frac{1}{2}\ln\left(1 + \frac{D_1}{D_2} - \frac{D_1}{D_1 + Q}\right) \qquad \text{(B.12)}$$

This completes the proof.

References

[1] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[2] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[3] P. Moulin and Y. Wang, "Capacity and random-coding exponents for channel coding with side information," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1326–1347, 2007.

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

[5] 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. Inform. Theory*, vol. 49, no. 5, pp. 1181–1203, 2003.

[6] J. K. Su, J. J. Eggers, and B. Girod, "Illustration of the duality between channel coding and rate distortion with side information," in *34th Asilomar Conf Signals, Systems and Computers, Pacific Grove, CA, USA*, vol. 2, 2000, pp. 1841–1845.

[7] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912–1923, 1997.

[8] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and Control*, vol. 38, no. 1, pp. 60–80, 1978.

[9] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[10] 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.

[11] Y. Sun, M. Uppal, A. D. Liveris, S. Cheng, V. Stankovic, and Z. Xiong, "Nested turbo codes for the Costa problem," *IEEE Transactions on Communications*, vol. 56, no. 3, pp. 388–399, 2008.

[12] U. Erez and S. Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.

[13] A. M. Aaron, E. Setton, and B. Girod, "Towards practical Wyner-Ziv coding of video," in *Proceedings of the IEEE Image Processing, ICIP*, vol. 2,3, 2003, pp. 869–872.

[14] R. Zamir and S. Shamai, "Nested linear/ lattice codes for Wyner-Ziv encoding," in *IEEE Information Theory Workshop, Killarney, Ireland*, 1998, pp. 92–93.

[15] D. Rebollo-Monedero and B. Girod, "Design of optimal quantizers for distributed coding of noisy sources," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Philadelphia*, vol. 5, 2005, pp. v/1097–v/1100.

[16] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4358–4379, 2006.

[17] K. Lajnef, C. Guillemot, and P. Siohan, "Distributed coding of three binary and Gaussian correlated sources using punctured turbo codes," *EURASIP Journal on Applied Signal Processing*, vol. 86, no. 11, pp. 3131–3149, 2006.

[18] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1872–1889, 2006.

[19] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, 2002.

[20] T. Berger, *Rate-Distortion Theory: A mathematical basis for data compression.* Prentice-Hall, 1971.

[21] G. Kramer, *Topics in Multi-User Information Theory.* Foundations and Trends in Communications and Information Theory, 2007, vol. 4, no. 4-5.

[22] A. Amraoui, S. Y. Chung, and R. L. Urbanke, "Lthc: Ldpcopt." http://lthcwww.epfl.ch/research/ldpcopt/, 2003.

[23] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, 2001.

[24] M. W. Marcellin and T. R. Ficher, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, 1990.

[25] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[26] D. J. C. Mackay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 33, no. 6, pp. 457–458, 1997.

[27] W. Z. Yang Yang Cheng, S. Zixiang Xiong, "Wyner-Ziv coding based on TCQ and LDPC codes," in *International Symposium on Turbo Codes and Related Topics*, vol. 1, 2003, pp. 825–829.

[28] N. Merhav and S. Shamai, "On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2844–2855, 2003.

[29] Y. Kochman and R. Zamir, "Joint Wyner-Ziv/Dirty Paper coding by modulo-lattice modulation," *ArXiv e-prints*, vol. 801, Jan. 2008.

[30] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, 2002.

[31] A. S. Cohen and R. Zamir, "Entropy amplification property and the loss for writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 54, no. 4, pp. 1477–1487, 2008.

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

[33] R. G. Gallager, *Information Theory and Reliable Communication.* New York, NY, USA: John Wiley & Sons, Inc., 1968.