Quantization with Variable Resolution and Coding for Deterministic Broadcast Channels

Eli Haim and Ram Zamir

Abstract

Quantization with variable resolution and channels with varying input alphabet have interesting lines of similarity as coding problems with side information at the encoder. We explore this relationship and propose coding strategies based on erasure correction coding for the discrete case and band-limited interpolation for the continuous case. The varying input alphabet setup is shown to be useful for coding for deterministic broadcast channels.

1 Introduction

Quantization where the resolution changes with time is conceptually similar to channel coding where the input alphabet varies with time. The information about the resolution / alphabet constraints is provided in both these setups to the *encoder*. This is unlike the familiar source-channel coding duality, where the channel encoder plays a role dual to the source decoder [3]. We explore the relationship between the two setups and demonstrate how similar coding concepts apply to both.

In information theoretic terms, quantization with variable resolution is a special case of source coding with a side-information dependent distortion measure. In this problem the distortion measure at the n-th source sample, $d\left(x_n, \hat{x}_n; s_n\right)$, depends on a "state" variable s_n which is available to the encoder [7]. This setup finds applications in context-dependant quantization and in sensor networks. We call "variable resolution" to the case where

$$d(x, \hat{x}; s) = \begin{cases} 0, & \text{if } |\hat{x} - x| < r(s) \\ \infty, & \text{otherwise} \end{cases} \tag{1}$$

with $r(s)$ denoting the required quantization resolution at state s. This is a "hard decision" easy-to-analyze version of the weighted squared distortion measure

$$d(x, \hat{x}; s) = w(s)[\hat{x} - x]^2. \tag{2}$$

Channel coding with variable input alphabet belongs to the family of Gelfand-Pinsker problems, i.e., channels with side information at the encoder [5, 3]. The specific application we consider here is of layered transmission over non-additive broadcast channels, in particular, deterministic and semi-deterministic broadcast channels (DBC) [9, 10], where the transmission to one user puts constraints on the alphabet available to the transmission to the other user. For example, the Blackwell channel is a broadcast channel with three inputs and two binary outputs (users) [2]. See Figure 1. Sending a "1" to user-1

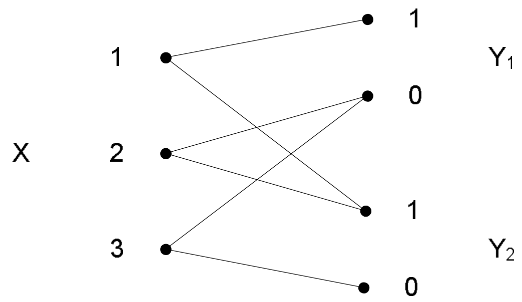

Figure 1: Blackwell Channel

requires to select channel input "1", which constrain user-2 to receive a "1". On the other hand, sending a "0" to user-1 is possible with either channel input "2" or "3", allowing the full binary degree of freedom for user-2 transmission. Hence, the (unconstrained) transmission to user-1 determines whether the effective alphabet for user-2 is unary or binary.

The classical view in information theory is that the Gelfand-Pinsker problem is dual to the Slepian-Wolf / Wyner-Ziv problem, i.e., source coding with side-information at the *decoder* [3]. Similarly, sources with distortion side-information at the encoder have a dual relationship with channels with channel-state information at the decoder. These dualities are not only on a formal level, but have an operational significance. For example, "algebraic binning schemes" prove to be useful for a class of additive / symmetric channels and sources with side-information, such as "writing on dirty paper", MIMO broadcast channels, and "Wyner-Ziv video coding". See [11] and the references therein. Interestingly, the current work finds a common framework for a class of source and channel coding problems which are *not* dual in the classical sense.

The problem of source coding with distortion side-information was proposed recently by Martinian, Wornell and Zamir [7, 8]. They showed that variable resolution coding of a discrete uniform source achieves the *conditional* rate-distortion function, i.e., it does not suffer any loss relative to the case where the side-information is available to both the encoder and the decoder. The same is true for continuous sources in the high resolution limit. Martinian *et al* proposed also efficient schemes which approach this performance based on *erasure correction codes* in the discrete case, and *band-limited interpolation* in the continues case.

In this paper we show that similar concepts apply to the problem of channels with variable input alphabet. This extension allows to find efficient coding schemes for new classes of broadcast channels, such as the DBC, the semi-DBC and the weakly noisy DBC. Similar schemes were proposed for the discrete DBC in a recent paper by Coelman et al [1].

The paper is organized as follows. We start in Section 2 by defining the problem of DBC, and discuss how the GP setup provides a key to its solution. Section 3 describes the Blackwell channel, which is the simplest non-trivial DBC. We then turn in Section 4 to define the problem of sources with distortion side information at the encoder, and describe efficient schemes based on erasure correction codes for the discrete source case. In section 5 we find special cases in which the GP setup does not suffer capacity loss with respect to the case where the side information is available to both the encoder and the decoder. In Section 6 we show how these schemes apply to some interesting examples of DBCs.

Finally in section 7 we describe efficient schemes based on band-limited interpolation for the continuous case.

2 Deterministic Broadcast Channel

Consider a discrete memoryless deterministic broadcast channel (DBC). The channel input is denoted by X, and the channel outputs are deterministic functions of X:

$$\begin{aligned} Y_1 &= f_1(X) \\ Y_2 &= f_2(X) \end{aligned}$$

where the input alphabet $\mathcal{X}$ and the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are all finite. The signals of length n at the channel input are sequences of letters of $\mathcal{X}$, $\mathbf{x} = (x(1), ..., x(n))$; the signals of length n at the channel output of terminal t are sequences of letters of $\mathcal{Y}_t$, $\mathbf{y}_t = (y_t(1), ..., y_t(n))$, where

$$y_t(i) = f_t(x(i)), \quad t = 1, 2.$$

Unlike the general broadcast channel, the capacity region of the DBC is known [9, 10]. It is composed of all rate pairs (R_1, R_2) such that

$$\begin{aligned} &0 \leq R_1 \leq H(Y_1) \\ &0 \leq R_2 \leq H(Y_2) \\ &R_1 + R_2 \leq H(Y_1, Y_2) \end{aligned} \tag{3}$$

for some input distribution P_X. Achievability is shown in general by random binning [9, 10].

The DBC problem can be interpreted as an extension of the Gelfand-Pinsker problem, i.e., channels with channel-state side-information at the encoder. First, information to one terminal is encoded in an unconstrained manner; the channel to the other terminal can be viewed now as a state-varying channel: the encoded information to the former terminal imposes a per-letter constraint on the alphabet of the latter channel. Since the encoder is common, it knows the state of this virtual channel and can use it in order to encode information.

The Gelfand-Pinsker channel capacity is given by:

$$C_{GP} = \max_{U \leftrightarrow (X,S) \leftrightarrow Y} [I(Y;U) - I(U;S)] = \max_{U \leftrightarrow (X,S) \leftrightarrow Y} [H(U|S) - H(U|Y)] \tag{4}$$

where S is the channel-state side-information, X is the channel input, Y is the channel output, U is an auxiliary variable, and the maximization is over (U, X). The GP capacity is upper bounded by the capacity when the side-information is known also at the decoder,

$$C_{both} = \max_{P_{X|S}} I(Y; X|S). \tag{5}$$

In the deterministic channel case, $Y = f(X, S)$, both (4) and (5) are simplified to

$$C_{GP} = C_{both} = \max_{P_{X|S}} H(Y|S). \tag{6}$$

This can be seen by substituting $U = Y$ in the GP capacity formula.

For each given channel input distribution, the achievable region is a pentagon. To see how it can be achieved as a Gelfand-Pinsker capacity, assume that information to the first terminal is sent at maximal rate: $R_1 = H(Y_1)$. By using G-P coding, $R_2 = H(Y_2|Y_1)$ rate can be achieved, when $S = Y_1$ is the state of the constrained channel to terminal 2. Thus the pentagon corner point $(H(Y_1), H(Y_2|Y_1))$ is achieved. The other pentagon vertex point can be achieved in a similar way, and the intermediate points can be achieved by time-sharing between the two corner points.

3 Motivation: The Blackwell Channel

One motivation for the coding algorithm that is presented in this paper is a known practical coding algorithm for the Blackwell channel. The Blackwell channel is a deterministic broadcast channel, given by [2]:

$$
\begin{aligned}
Y_1 &= \begin{cases} 1, & x = 1 \\ 0, & x = 2 \text{ or } 3 \end{cases} \\
Y_2 &= \begin{cases} 1, & x = 1 \text{ or } 2 \\ 0, & x = 3 \end{cases}
\end{aligned}
$$

Per transmission, one bit can be sent to receiver Y_1 or one bit can be sent to receiver Y_2 but not simultaneously. However, the channel sum capacity is equal to $\log 3$ bits per transmission.

Gelfand [4] found the Blackwell channel capacity region. He used a "good" binary erasure correction code (near Maximum Distance Separable (MDS) code) in order to encode information. This method is explained in short for a corner point of the capacity region (Figure 2): suppose that information to receiver Y_1 is sent in a rate of $R_1 = H_2(p)$, where $p \triangleq \Pr(Y_1 = 1) = \Pr(X = 1)$. Therefore the information rate receiver Y_2 should be able to reach

$$
R_2 = \max_{P_{Y_2|Y_1}} H(Y_2|Y_1) = 1 - p
$$

In order to achieve this rate, it is required that Y_2 would be conditionally symmetric given $Y_1 = 0$, implying that

$$
\Pr(X = 2) = \Pr(X = 3) = \frac{1-p}{2}.
$$

The problem can now be viewed as *writing to memory with defects* [6]: Y_2 is stuck-at 1 if $x = 1$, therefore a defect occurs with probability p. In order to code information in this problem, an (n, np) near-MDS code is used in the following way: This code has the property that for every set of np stuck-ats, every coset of this code has a word which agrees with the stuck-ats. Therefore the *coset* of the word carries the information. Since there are $2^{n(1-p)}$ cosets, information to Y_2 is coded at a rate of $(1 - p)$ bit per channel use.

4 Quantization with Variable Resolution

A second motivation for the coding algorithm presented here comes from source coding with distortion side information at the encoder [7, 8]. This paper introduces a new

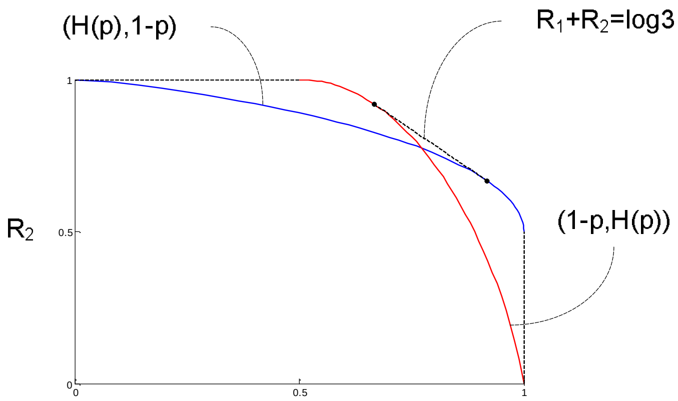

Figure 2: Blackwell Channel Capacity

concept, which lets the distortion measure vary from one sample to another according to the quality of each sample. Suppose the encoder is fed with a random uniform source $X \sim \text{Unif}(\mathcal{X})$, and suppose each source sample comes with a quality side information S, describing how "important" it is. The random variable S is statistically independent of the source X. The way S affects the overall fidelity of the reconstruction is reflected by a three-letter distortion measure $d(x, \hat{x}; s)$.

As an example (from [7]), let the distortion measure be of the form

$$d(x, \hat{x}; s) = s \cdot (x \oplus \hat{x}) = \begin{cases} x \oplus \hat{x}, & \text{if } s = 1 \text{ ("important")} \\ 0, & \text{if } s = 0 \text{ ("non-important").} \end{cases}$$

For a block of n samples, there are approximately $k \triangleq n \cdot \Pr(S = 1)$ positions with "important" information. If the quality side-information S was known at the decoder too, then in order to achieve zero distortion, the minimal rate would be $\frac{k}{n} \log |\mathcal{X}|$ bits per sample. If the quality side-information was not known at the encoder, $\log |\mathcal{X}|$ bits per sample would be required. They showed that if S is known only at the encoder, then there is no rate-loss in comparison to the fully aware system. To see how this can be done, consider the source samples as a codeword of an (n, k) Reed-Solomon (RS) code (or more generally any MDS code) with $s(i) = 0$ indicating an erasure at sample i. The code is used to "correct" the erasures and determine the k corresponding information symbols, which are sent to the receiver. To reconstruct the signal, the receiver encodes the k information symbols using the code, and produces the reconstruction vector. Only symbols with $s(i) = 0$ could have changed, hence the relevant samples are losslessly communicated using only $k \log |\mathcal{X}|$ bits. The RS decoding (at the transmitter) can be viewed as curve-fitting, and RS encoding (at the receiver) can be viewed as interpolation.

This example can be extended to a multi-level quality information. In view of equation (1), we may think of X in the form of a binary word, where $r(s)$ determines how many LSBs of X may be changed without causing notable distortion. The solution in this case would be to encode each bit level in the expansion of X in separate layers, where in layer L we use the erasure correction mechanism above according to important samples probability $\Pr(\log r(S) \leq L)$. See [8]. We next use this multi-resolution framework for coding of channels with variable input constraint.

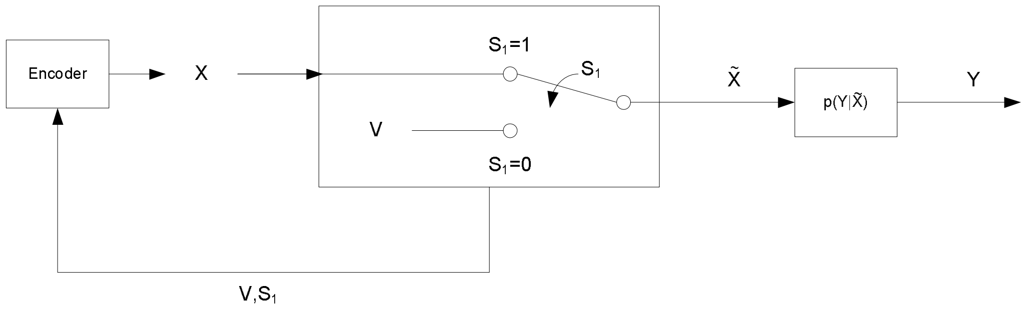

Figure 3: A channel with two-state input alphabet

5 A Channel with Two-State Input Alphabet

Consider a two-state channel, where when $S_1 = 1$ the input X is free to take any value of $\mathcal{X}$, while when $S_1 = 0$ the input X must be equal to an external source V which is independent of the information. An equivalent formulation for this setup is as a channel that randomly switches between "transfer" state and "pure noise" state. Specifically, $p(y|x) = p(\tilde{x}|x) \circ p(y|\tilde{x})$, where $p(y|\tilde{x})$ is a general channel, and

$$\tilde{X} = \begin{cases} X, & \text{if } S_1 = 1 \\ V, & \text{if } S_1 = 0 \end{cases}$$

with

$$S_1 \sim Bernulli\,(q)$$

where X, V, S_1 are independent random variables, $(X, V, S_1) \leftrightarrow \tilde{X} \leftrightarrow Y$ and $q \in [0, 1]$. We define the full channel state S as the pair (V, S_1).

It is easy to verify that the capacity (5) when both the transmitter and the receiver have access to $S = (V, S_1)$ is given by

$$C_{both} = q \cdot \tilde{C}$$

where $\tilde{C}$ is the capacity of the channel $p(y|\tilde{x})$.

The Gelfand-Pinsker (GP) capacity is given in general by (4). It is not hard to verify that for any pair (U, X) such that $U \leftrightarrow (X, S) \leftrightarrow Y$ form a Markov chain, and X is a deterministic function of (U, S), we have

$$I(X;Y|S) = [I(Y;U) - I(U;S)] + I(U;S|Y).$$

It follows that the capacity loss in the GP problem is *zero* whenever $I(U;S|Y) = 0$, i.e., $U \leftrightarrow Y \leftrightarrow S$ form a Markov chain, for some admissible pair (U, X) such that X achieves C_{both} (5). A simple interpretation for this condition is that given Y, the channel state information S does not give any information about the information bearing variable U. In any case, the for the discrete state we have that the rate loss is bounded by $H(S)$.

The zero loss situation occurs when the channel $p(y|\tilde{x})$ is deterministic, i.e., $Y = f(\tilde{X})$. In this case Y is a deterministic function of $(X, S) = (X, V, S_1)$, implying by (6) that the choice $U = Y$ is optimal, and we obtain

$$C_{SI@enc} = C_{both} = q \cdot \tilde{C}$$

where $\tilde{C} = \max_{P_X} H(Y)$.

We next extend the discussion above to a class of channels with *multiple states*, corresponding to communicating to the second terminal in the DBC corner point.

6 Coding under a Variable Resolution DBC Model

As noted above, the capacity region of the deterministic broadcast channel can be derived by using Gelfand and Pinsker's channel capacity. Therefore we can consider only one vertex point. Assume that a distribution P_{Y_1} of the output of one terminal is set, and the information rate for this terminal is $R_1 = H(Y_1)$. For any $y_1 \in \mathcal{Y}_1$, the mappings f_1, f_2 determine the achievable values at $\mathcal{Y}_2$:

$$\mathcal{Y}_2(y_1) \triangleq f_2\left(f_1^{-1}(y_1)\right)$$

This implies that the maximal value of R_2 is $E\{\log|\mathcal{Y}_2(Y_1)|\}$. Equivalently, we may say that an optimum pair (Y_1, Y_2) in (3) satisfies that Y_2 is *conditionally uniform* over the constrained alphabet $\mathcal{Y}_2(y_1)$ given $Y_1 = y_1$. To see why, note that the event $Y_1 = y_1$ imposes a constraint on the transmission to the second terminal: only $x \in \mathcal{X}$ such that $f_1(x) = y_1$ can be used as the channel input. Therefore $f_2\left(f_1^{-1}(y_1)\right)$ is the set of the (instantaneous) achievable symbols at terminal 2. This implies that the conditional entropy is bounded by $H(Y_2|y_1) \leq \log|\mathcal{Y}_2(y_1)|$, and the inequality is achievable for a conditionally uniform Y_2. Here B and W correspond to the V and S_1 from section 5.

In order to consider one terminal as a "channel" and the other terminal as known "state", it would be convenient to denote the former channel output by Y, and the known channel-state by S. This channels capacity is denoted by C.

We consider a special case of a deterministic broadcast channel, which is an extended case of the Blackwell channel. The state in this case has a "metrical" meaning, and information can be coded using erasure (near- MDS) codes. This "metrical" meaning is related to resolution in source coding. In this case, each input alphabet $x \in \mathcal{X}$ is mapped to a different binary number of length $\log|\mathcal{X}|$, such that the side-information S is characterized by two components: W which is the number of LSBs that are free in the set, and B which are the MSBs imposed by S. The metrical meaning is that for a given B, the allowable channel inputs are those in the interval which starts at B, with length of $2^W - 1$. Notice that since the number of degrees of freedom of Y in the current state is 2^W, W bounds the current possible transmission rate. It follows that

$$C = \sum_{s \in S} \Pr(S = s) \log|w_s| = E\{W\}$$

where w_s is the value of W at state s. Therefore B doesn't influence the capacity (but it does influence the information coding).

In order to encode information block coding is used, with block length=N. For a given channel state vector $\mathbf{s} = (s_1, ..., s_N)$, encoding is done in layers, each layer is encoded independently of the others. The number of layers is:

$$L \triangleq \log|\mathcal{W}|$$

Denote the channel input vector $\mathbf{x} = (x_1, ..., x_N)$, and $x_i^{(m)}$ is the mth bit of x_i in it's binary representation (starting from the LSB): $x_i = \sum_{m=0}^{L-1} 2^m x_i^{(m)}$. At each layer $m \in \{0, ..L-1\}$ we look at the vector of bits: $\mathbf{x}^{(m)} = \left(x_1^{(m)}, ..., x_N^{(m)}\right)$. At layer m, if $w_i > m$, then $x_i^{(m)}$can be either $\{0, 1\}$. Otherwise, $x_i^{(m)}$ is stuck-at. It is useful to denote

$$q_w \triangleq \Pr(W = w) = \sum_{s \in \mathcal{S}: w_s = w} \Pr(S = s)$$

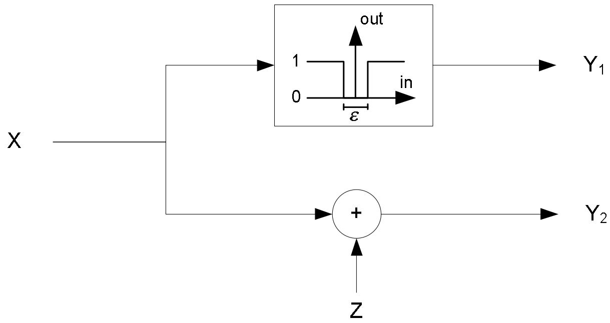

Figure 4: Semi-discrete semi-continuous channel

Hence a stuck-at occurs with probability

$$p_m \triangleq \Pr(W_i \leq m) = \sum_{w=0}^{m} q_w$$

Therefore each layer m can be used to encode information by using a near-MDS $(N, p_m N)$ code, with an erasure probability of $1 - p_m$ (every coset of the code has a word which agrees with the stuck-at constraints; see [1] for construction of such codes). So at layer m information is encoded at a rate of $1 - p_m$. It can be easily shown that the total rate that is achieved is equal to C:

$$R = \sum_{m=0}^{L-1} (1 - p_m) = C$$

7 Band-Limited Interpolation for Variable Resolution Coding

As mentioned above, the workings of the erasure-correction mechanism can be explained in terms of the notion of band-limited (BL) interpolation. The codebook corresponds to a collection of bandlimited signals, that can meet a certain number of constraints in time domain no matter in which order they come.

This idea was used in [7] to extend the discrete variable resolution quantization example of Section 4 to the quadratic-Gaussian source case. Specifically, the scheme of [7] uses a low-pass filter (LPF) codebook of DFT bandwidth k to quantize a block of n samples from an i.i.d. Gaussian source where only k samples are "important". In terms of the weighted squared-error distortion measure (2), the state is binary $s \in \{0, 1\}$, where $w(s) = s$ and $\Pr(S = 1) = k/n$. Furthermore, only the encoder needs to know the exact location of the important samples in order to do "correct" interpolation.

We can use BL interpolation also to extend the discrete GP and DBC coding schemes above to Gaussian signaling under various types of constraints in time domain. Consider first the semi-discrete / semi-continuous BC problem described in Figure 4:

$$\begin{aligned} Y_1 &= \begin{cases} 0 & if \quad X \in (-\varepsilon, \varepsilon) \\ 1 & if \quad X \notin (-\varepsilon, \varepsilon) \end{cases} \\ Y_2 &= X + Z \end{aligned}$$

where $\varepsilon > 0$, and $Z \sim N(0, \sigma_Z^2)$, and there is a channel input power constraint $E(X^2) \leq P$. The first channel output is discrete and a deterministic function of the channel input. The second channel output is continuous and noisy version of the channel input. Marton (&G-P) found the capacity region of a broadcast channel with one deterministic component. This region consolidates with the Marton inner-bound, i.e. it is the convex-closure of the vertex points, where each vertex point is a point where information to one terminal is sent in a certain rate, while information to the other terminal is send while the message to the first terminal is viewed as a channel-state side information.

Specifically, suppose that information to the first terminal is encoded in a rate of $R_1 = H(q)$. This rate is achieved by setting $\Pr(Y_1 = 1) = q$. We use the following auxiliary random variable U as the channel input:

$$X = U \triangleq \begin{cases} \tilde{U} \sim N(0, P/\alpha) & if \quad Y_1 = 1 \\ 0 & if \quad Y_1 = 0 \end{cases}$$

The channel input power constraint holds: $E(X^2) = P$. It can be easily shown that at high SNR this setting achieves the capacity, moreover, there is no rate loss due to not knowing Y_1 at the second terminal. At high SNR:

$$R_2 \simeq \alpha \frac{1}{2} \log \left(\frac{P/\alpha}{N} \right)$$

This GP performance can be achieved by a superposition of two n-dimensional codebooks: a high-pass filter (HPF)-codebook of DFT bandwidth qn which carries the information, and a LPF auxiliary codebook of bandwidth $(1-q)n$ which enforces the time domain constraint of $(1-q)n$ zeros. The receiver detects the information from the HPF band alone, just as in the discrete case the information was extracted from the "coset" (or syndrome).

A second variant of the problem is *writing on dirty paper (WDP) with variable resolution* ("information embedding with distortion side-information"). Here the signal V corresponds to the clean source, in which information can be embedded with variable resolution, and it is observed through an AWGN channel with noise variance σ_z^2. Assume the distortion constraint of (2), with a binary side information S_1, so embedding can distort "sensitive" samples of V at a level D_l and the rest of the samples at level D_h. Both S_1 and V are known to the encoder. Assuming a strong source and high signal-to-noise ratio in the channel ($\sigma_v^2 \gg D_h > D_l \gg \sigma_z^2$) we can show that

$$C_{GP} = C_{both} \approx q \log \left(\frac{D_l}{\sigma_z^2} \right) + (1-q) \log \left(\frac{D_h}{\sigma_z^2} \right)$$

where $q = \Pr(S = \text{``}sensitive'')$. This performance can be achieved by concatenation of a BL-interpolator and a standard WDP encoder. The BL interpolator uses a superposition of HPF and LPF codebooks to enforce zero (or small) modulating signal values at the more sensitive positions. This is done in a way that the uninformed decoder can still extract the embedded information from the HPF band of the demodulated signal. Efficient methods for BL-interpolation using discrete codebooks are currently under investigation.

References

[1] T. P. Coleman, E. Martinian, M. Effros, and M. Medard. Interference management via capacity-achieving codes for the deterministic broadcast channel. In *Proc. Int. Symp. Info. Theory*, 2005.

[2] T. M. Cover. Comments on broadcast channels. *IEEE Trans. Information Theory*, IT-44:2524–2530, October 1998.

[3] T. M. Cover and M. Chiang. Duality between channel capacity and rate distortion with two-sided state information. *IEEE Trans. Information Theory*, IT-48:1629–1638, June 2002.

[4] S. I. Gelfand. Capacity of one broadcast channel. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 13, No. 3:106–108, July-Sept. 1977.

[5] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Probl. Contr. Inform. Theory*, 9, No. 1:19–31, Jan.-Mar. 1980.

[6] A. V. Kuznetsov and B. S. Tsybakov. Coding in memory with defective cells. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 10, No. 2:52–60, 1974.

[7] E. Martinian, G. W. Wornell, and R. Zamir. Encoder side information is useful in source coding. In *Proc. Int. Symp. Info. Theory*, page 211, June 2004.

[8] Emin Martinian. *Dynamic information and constraints in source and channel coding.* PhD thesis, Massachusetts Institute of Technology, September 2004.

[9] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Information Theory*, IT-25:306–311, May 1979.

[10] M. S. Pinsker. Capacity of noiseless broadcast channels. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 14, No. 2:28–34, Apr.-June 1978.

[11] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Information Theory*, IT-48:1250–1276, June 2002.