Systematic Lossy Source/Channel Coding

Shlomo Shamai (Shitz)
Electrical Engineering Department
Technion—Israel Institute of Technology
Haifa, 32000, Israel

Sergio Verdú
Electrical Engineering Department
Princeton University
Princeton, NJ 08544
U.S.A.

Ram Zamir
Department of Electrical Engineering - Systems
Tel Aviv University
Tel Aviv 69978, ISRAEL

Abstract

The fundamental limits of "systematic" communication are analyzed. In systematic transmission, the decoder has access to a noisy version of the uncoded raw data (analog or digital). The coded version of the data is used to reduce the average reproduced distortion D below that provided by the uncoded systematic link and/or increase the rate of information transmission.

Unlike the case of arbitrarily reliable error correction ($D \to 0$) for symmetric sources/channels, where systematic codes are known to do as well as non-systematic codes, we demonstrate that the systematic structure may degrade the performance for nonvanishing D. We characterize the achievable average distortion and we find necessary and sufficient conditions under which systematic communication *does not* incur loss of optimality. The Wyner-Ziv rate distortion theorem plays a fundamental role in our setting.

The general result is applied to several scenarios. For a Gaussian bandlimited source and a Gaussian channel, the invariance of the bandwidth-SNR (in dB) product is established, and the optimality of systematic transmission is demonstrated. Bernoulli sources transmitted over binary symmetric channels and over certain Gaussian channels are also analyzed. It is shown that if nonnegligible bit error rate is tolerated, systematic encoding is strictly suboptimal.

Keywords: systematic transmission, uncoded side information, rate-distortion theory, Wyner-Ziv rate distortion, source/channel coding, Gaussian channels and sources. .

June 22, 1997

This work was supported by the US-Israel Binational Science Foundation, Grant: 92–00202, the NSF under Grant NCR95-23805 and the UC Micro program.

1 Introduction

The advent of digital communications in radio and television broadcasting poses the following scenario. Historically, a certain bandwidth has been allocated to the analog transmission of an information source. Then, additional channel bandwidth becomes available over which it is possible to transmit digitally-coded information. This "digital" channel can be used to transmit additional information bandwidth, boost the received fidelity of the original information source, or both. For the sake of back compatibility with existing equipment it may not be possible to convert the "analog" channel into a "digital" one, and the analog uncoded transmission of the original source must be preserved. The principal question addressed in this paper is under what conditions does this restriction incur in no loss of capacity. Analogously, in systems where a direct satellite broadcast signal coexists with a terrestrial analog signal (e.g. DirecTV® and DSS® Digital Satellite System) it is wasteful to simply discard the existing analog transmission in digital broadcast receivers. Through adequate signal processing and coding/decoding, the analog transmission can be used to lower the digital channel bandwidth required to achieve a given degree of fidelity. Such a reduction in bandwidth requirements can be quantified using the results of this paper.

In the context of error-correcting channel coding, *systematic* codes are those where each codeword contains the uncoded information source string plus a string of parity-check symbols; Thus, the multiplexing of coded and uncoded symbols takes place in the time domain. By extension, we call systematic those source/channel codes which transmit the raw uncoded source in addition to the encoded version. The aforementioned compatible analog/digital broadcasting systems are examples of systematic source/channel codes where the multiplexing of coded and uncoded versions takes place in the frequency domain.

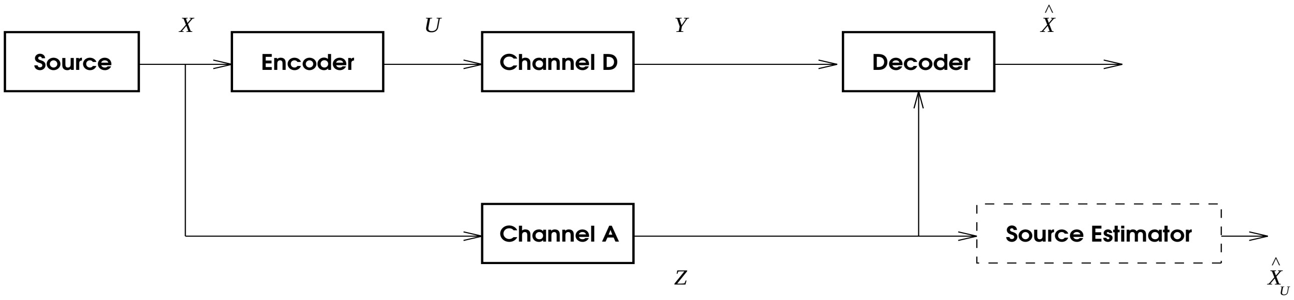

Figure 1: Information Transmission with Uncoded Side Information.

Notwithstanding the above analog+digital motivation, the model of this paper has a wide scope. For example, as a consequence of our results, we characterize the *bit-error-rate* achievable by systematic codes for the *binary symmetric channel* and we show that when the source entropy is higher than the channel capacity, the bit error rate is *strictly higher* than that achievable by non-systematic codes. This is in contrast to the well-known fact [1], [2] that for any rate below capacity, systematic (linear or nonlinear) codes achieve the random coding error-exponent of the binary symmetric channel.

Another motivation for the model of systematic transmission discussed in this paper arises when a source is transmitted over a channel with unknown SNR. In that setting, the systematic part of the transmission enables graceful degradation in reconstruction quality as the SNR decreases, while the digital part allows to boost the fidelity of the reconstruction whenever the SNR is sufficiently high.

The information theoretic problem we consider is depicted in Figure 1. It is desired to reproduce the source with a certain fidelity at the output of the decoder using the outputs of the uncoded channel A ("analog") and the coded channel D ("digital"). A separation theorem for source/channel coding with side information is given in Section 2 characterizing the sources/channels for which communication subject to a prescribed fidelity criterion is possible. For the sake of clarity and conciseness we limit our discussion to memoryless sources and channels as well as additive distortion measures. The achievability part of the separation theorem is a corollary of the Wyner-Ziv direct source coding theorem with side information [4], [5]. The capacity of channels with side information found in [3] corresponds to the zero-distortion special case of the result in Section 2. The Slepian-Wolf theorem [6] on separate coding of correlated sources is used in [3] to show that if the capacity of channel D is larger/smaller than the conditional entropy of the source given the output of channel A, then arbitrarily reliable transmission of the source is feasible/impossible. The separation theorem of Section 2 replaces the conditional entropy by the Wyner-Ziv rate-distortion function evaluated at the prescribed distortion level. From this separation theorem, we derive the necessary and sufficient condition for the optimality of systematic encoding: coding of channel A is superfluous to achieve the optimal distortion/information rate tradeoff, if and only if the following three conditions are satisfied:

1. The source maximizes the input-output mutual information of channel A.
2. Having the output of channel A as side information at the transmitter is superfluous for encoding the source at the desired distortion.
3. The output of channel A is maximally useful[1] at the decoder in order to reduce the rate

[1] In the sense formalized in Section 2.

required to encode the source at the desired distortion.

Note that due to the prohibition of coding in one of the channels (channel A), the setting of this paper is not encompassed by the conventional frameworks of broadcast channels and multiple descriptions [7].

Section 3 applies the results of Section 2 to the special case of Gaussian sources and channels with a mean-square-error fidelity criterion. In the setting where the channel bandwidth is equal to the source bandwidth and all spectra are flat, it is well-known [8] that it is optimal not to encode the Gaussian source in any way and to decode the source with a simple attenuator. As a practical matter, this means that Single-Sideband (SSB) modulation of a Gaussian source is optimal for white Gaussian noise channels [9]–[10]. When the channel bandwidth exceeds the source bandwidth, coding of the source obviously becomes necessary in order to take advantage of the additional channel bandwidth for the sake of improved fidelity. However, systematic encoding turns out to be optimal. This implies that, for Gaussian sources and channels, back compatibility incurs no penalty in the fidelity with which the source can be reproduced by a receiver which observes both the analog and digital channels. For a given bandlimited channel, the higher the bandwidth of the source, the lower the signal-to-noise ratio achievable at the output of the decoder. Under the assumption that the allowed transmission power in any given frequency band is proportional to its bandwidth, we find that for the ideally bandlimited Gaussian channel, the tradeoff information bandwidth/output signal-to-noise ratio is particularly simple: the product of the output signal-to-noise ratio in dB times the transmissible information bandwidth is a constant equal to the channel bandwidth times the signal-to-noise ratio in dB achievable with uncoded transmission.

Sections 4 and 5 deal with the important case of binary information sources and the bit-error-rate as the distortion criterion. Binary symmetric channels are studied in Section 4, while Section 5 considers antipodal modulation in Gaussian channels. Systematic encoding is shown to be strictly suboptimal for both binary symmetric channels and binary-input Gaussian channels. Wyner's [13] interpretation of Slepian-Wolf coding is extended by the explicit construction (proposed in Section 4) of Wyner-Ziv codes for the binary source/channel problem based on channel error-correcting codes.

Section 5 also considers the case where the uncoded and coded channels are mutually interfering. In this alternative scenario, it is assumed that no additional bandwidth or power are available to boost performance. Instead, a percentage of the power originally assigned to the transmission of a binary uncoded source through a Gaussian channel is devoted to the coded transmission superposed to the uncoded transmission. Increasing the power of the coded signal enhances the performance of an optimal decoder while it degrades the performance of a receiver for the uncoded transmission. The tradeoff given in Section 5 generalizes the case of arbitrarily reliable optimum decoding treated in [3].

2 Separation Theorem for Lossy Source/Channel Coding with Side Information

Consider the situation depicted in Figure 1 where an information source X is transmitted to a decoder via two independent channels, only one of which (channel-D) is allowed to be preceded by an encoder. The objective is to reproduce the source by $\hat{X}$ at the output of the decoder within some prescribed distortion. Had we allowed an encoder preceding channel A, then the conventional separation theorem for lossy source/channel coding [7] states that distortion d is achievable/not achievable if the rate-distortion function of the source, $R(d)$, lies below/above the sum, $C_A + C_D$, of the capacities of the respective channels.

In the case of noiseless (arbitrarily reliable) transmission, the separation theorem of [3] states that reliable transmission of the source in the setting of Figure 1, is possible/not possible if its entropy lies below/above $C_D + I(X\,;\,Z)$, where $I(X\,;\,Z)$ is the mutual information between the input and output of channel A. The essential part of the proof of the achievability part of this result from [3] is the Slepian-Wolf theorem [6] (and its ergodic extension [7].), which enables the decoupling of the encoder into a Slepian-Wolf source encoder and a channel encoder (and likewise at the decoder). The converse result of [3] shows that such a decoupling entails no loss of optimality. It follows that in the distortionless case the condition for optimality of systematic encoding is that the source maximize the mutual information of channel A, i.e. $I(X\,;\,Z) = C_A$.

In order to state the separation theorem for lossy source/channel coding with side information we will need a few definitions.

Definition. For a joint distribution P_{VW} define the minimum possible estimation error of V given W by

$$\mathcal{E}(V|W) = \min_f E[D\Big(V, f(W)\Big)] \tag{2.1}$$

where $D(\cdot,\cdot)$ is a distortion measure.

Definition. Let the Wyner-Ziv rate-distortion function of X given that the decoder observes Z

be [11]

$$\overline{R}_{X|Z}(d) = \inf_{\{W:\, Z\to X\to W,\, \mathcal{E}(X|W,Z)\le d\}} I(X\,;W\mid Z) \tag{2.2}$$

where $Z \to X \to W$ denotes that Z and W are conditionally independent given X.

Theorem 2.1 *Let the source and channels A and D be memoryless.*

(a) If

$$\overline{R}_{X|Z}(d) < C_D, \tag{2.3}$$

then blocklength-n encoders/decoders exist for all sufficiently large n such that the source can be reproduced at the output with distortion

$$\frac{1}{n}\sum_{k=1}^{n} E\Big[D(X_k,\, \hat{X}_k)\Big] \le d \tag{2.4}$$

(b) If $\overline{R}_{X|Z}(d) > C_D$, then no such coding scheme can exist.

Remarks: As usual in separation theorems, the case where capacity and rate-distortion function are equal is not included since, in that case, the feasibility of transmission depends on the source and channel. Extension of Theorem 2.1 to an information-stable channel D is straightforward.

Proof:

a) The proof of achievability follows immediately by considering a Wyner-Ziv code (independent of channel D) followed by a transmission code (independent of P_{XZ}). The achievability parts of the Wyner-Ziv theorem for general memoryless sources [5] and the conventional channel coding theorem [7] imply that a sequence of optimal codes exist as long as (2.3) is satisfied.

b) Consider any encoding/decoding scheme and denote the source word of length n by $\underline{X}^n$; its encoded version (transmitted by channel D) by $\underline{U}^n$; the output due to $\underline{U}^n$ by $\underline{Y}^n$; the output of channel A due to $\underline{X}^n$ by $\underline{Z}^n$; and the output of the decoder by $\underline{\hat{X}}^n$ (see Figure 1).

Furthermore, define

$$\begin{aligned} \underline{Z}_i^c &= (Z_1\,, \ldots,\, Z_{i-1},\, Z_{i+1}\,, \ldots,\, Z_n) \\ \underline{W}_i &= (\underline{Y}^n,\, \underline{X}^{i-1},\, \underline{Z}_i^c)\,. \end{aligned}$$

Note that the following Markov properties hold:

$$\underline{Z}^n \to \underline{X}^n \to \underline{U}^n \to \underline{Y}^n \tag{2.5}$$

$$Z_i \to X_i \to \underline{W}_i \tag{2.6}$$

Consider the following chain of inequalities:

$$
\begin{aligned}
C_D &\geq \frac{1}{n} I(\underline{U}^n ; \underline{Y}^n) && \text{(2.7a)}\\
&\geq \frac{1}{n} I(\underline{X}^n ; \underline{Y}^n) && \text{(2.7b)}\\
&= \frac{1}{n} I(\underline{X}^n, \underline{Z}^n ; \underline{Y}^n) && \text{(2.7c)}\\
&\geq \frac{1}{n} I(\underline{X}^n; \underline{Y}^n \mid \underline{Z}^n) && \text{(2.7d)}\\
&= \frac{1}{n} \sum_{i=1}^{n} I\Big(X_i ; \underline{Y}^n \mid \underline{X}^{i-1}, \underline{Z}^n\Big) && \text{(2.7e)}\\
&= \frac{1}{n} \sum_{i=1}^{n} I\Big(X_i ; \underline{Y}^n, \underline{X}^{i-1}, \underline{Z}_i^c \mid Z_i\Big) - I\Big(X_i ; \underline{X}^{i-1}, \underline{Z}_i^c \mid Z_i\Big) && \text{(2.7f)}\\
&= \frac{1}{n} \sum_{i=1}^{n} I\Big(X_i ; \underline{W}_i \mid Z_i\Big) && \text{(2.7g)}\\
&\geq \frac{1}{n} \sum_{i=1}^{n} \overline{R}_{X|Z}\Big(\mathcal{E}(X_i \mid Z_i , \underline{W}_i)\Big) && \text{(2.7h)}\\
&\geq \frac{1}{n} \sum_{i=1}^{n} \overline{R}_{X|Z}\Big(\mathcal{E}(X_i \mid \underline{Z}^n, \underline{Y}^n)\Big) && \text{(2.7i)}\\
&\geq \frac{1}{n} \sum_{i=1}^{n} \overline{R}_{X|Z}\Big(E[D(X_i, \hat{X}_i)]\Big) && \text{(2.7j)}\\
&\geq \overline{R}_{X|Z}(d) && \text{(2.7k)}
\end{aligned}
$$

which are justified by

(a) the channel coding theorem,

(b) data processing theorem and (2.5),

(c) By (2.5),

(d),(e),(f) chain rule of mutual information,

(g) both the source and channel A are memoryless,

(h) (2.2) and (2.6),

(i) $\overline{R}_{X|Z}$ is decreasing and further conditioning decreases $\mathcal{E}$ (cf. (2.1)),

(j) (2.1) and $\hat{X}_i$ is a function of $\underline{Y}^n$ and $\underline{Z}^n$,

(k) convexity of $\overline{R}_{X|Z}$ [4] and (2.4).

□

We proceed to derive the condition for optimality of systematic coding from Theorem 2.1. Suppose that distortion d^* is the best distortion achievable in a fully coded system (where channel A is preceded by an encoder). From the conventional separation theorem for lossy source/channel coding, d^* is the solution to

$$-R_X(d^*) + C_A + C_D = 0 \; , \tag{2.8}$$

where the rate-distortion function of the source is

$$R_X(d) = \inf_{\{W : \mathcal{E}(X|W) \leq d\}} I(X\,;\,W) \; . \tag{2.9}$$

We will add and subtract the conditional rate distortion function, $R_{X|Z}(d^*)$ (which corresponds to the case where the side information is available also at the encoder [12]), and $I(X\,;\,Z)$ and $\overline{R}_{X|Z}(d^*)$ from the right side of (2.8) to yield:

$$\begin{aligned} \Big(\overline{R}_{X|Z}(d^*) - R_{X|Z}(d^*)\Big) \;&+\; \Big(C_D - \overline{R}_{X|Z}(d^*)\Big) + \Big(C_A - I(X\,;\,Z)\Big) \\ &+\; \Big(-R_X(d^*) + R_{X|Z}(d^*) + I(X\,;\,Z)\Big) \\ &=\; C_A + C_D - R_X(d^*) = 0 \; . \end{aligned} \tag{2.10}$$

We will now show that each of the four terms in parenthesis in the left side of (2.10) is nonnegative. Therefore (2.8) is satisfied if and only if every term is equal to 0. The first term is nonnegative by definition (not having side information at the encoder cannot reduce the required rate); the second term is nonnegative because of Theorem 2.1; the third term is nonnegative because C_A is equal to the maximal input-output mutual information of channel A; the fourth term in (2.10) is nonnegative as can be seen as follows. The conditional rate-distortion function is given by (see

Appendix A and [7])

$$R_{X|Z}(d) = \inf_{\{W:\, \mathcal{E}(X|W,Z)\le d\}} I(X\,;W \mid Z) = \inf_{\substack{P_{XZ\hat{X}}\,: \\ E[D(X,\hat{X})]\le d \\ \sum_c P_{XZ\hat{X}}(a,b,c) = P_{XZ}(a,b)}} I(X\,;\hat{X}\mid Z)\ . \tag{2.11}$$

Using (2.11) and the chain rule for mutual information, we obtain

$$\begin{aligned} R_{X|Z}(d) &= -I(X\,;Z) + \inf_{\substack{P_{XZ\hat{X}}\,: \\ E[D(X,\hat{X})]\le d \\ \sum_c P_{XZ\hat{X}}(a,b,c) = P_{XZ}(a,b)}} I(X\,;\hat{X},Z) && (2.12) \\ &\ge -I(X\,;Z) + \inf_{\substack{P_{XZ\hat{X}}\,: \\ E[D(X,\hat{X})]\le d \\ \sum_{b,c} P_{XZ\hat{X}}(a,b,c) = P_{X}(a)}} I(X\,;\hat{X}) && (2.13) \\ &= -I(X\,;Z) + R_X(d)\ , \end{aligned}$$

where the inequality follows by lower bounding the penalty function and enlarging the optimization set. Thus, we conclude that side information Z at both encoder and decoder can reduce the rate required to encode X by at most $I(X\,;Z)$. We note that inequality (2.13) can be found in [12]. Equality is achieved in (2.13) if and only if $X \to \hat{X} \to Z$ for the optimizing distribution in (2.12).

We have thus established a necessary and sufficient condition on the source and channel for the optimality of systematic lossy source/channel encoding, which consists of the following three simultaneous conditions:

C1. The source maximizes the mutual information of the uncoded channel, i.e.,

$$I(X\,;Z) = C_A\ . \tag{2.14}$$

C2. The output of the channel due to the uncoded source is not needed at the source encoder, i.e.,

$$R_{X|Z}(d) = \overline{R}_{X|Z}(d)\ . \tag{2.15}$$

C3. The output of the channel due to the uncoded source is maximally useful at the source decoder i.e.,

$$R_X(d) = R_{X|Z}(d) + I(X\,;Z)\ . \tag{2.16}$$

It is interesting to note that when condition C2 is not satisfied, $\overline{R}_{X|Z}(d) - R_{X|Z}(d)$ is typically small as shown in [11]. In the special case of zero distortion treated in [3] conditions C2 and C3 are always satisfied.

In the next section we show an important special case where the conditions for optimality of systematic encoding are satisfied.

3 Gaussian Source Through Gaussian Channel

Suppose that a Gaussian continuous-time source, $\{x(t)\}$, with flat power spectral density strictly bandlimited to bandwidth B_A is transmitted (uncoded) through a channel with flat spectral response and additive white Gaussian noise $\{n(t)\}$ with single-sided power spectral density N. Suppose that the transmitted signal single-sided power spectral density is constrained to S. In this uncoded scenario the receiver that minimizes the mean square error (MSE) is an attenuator:

$$\hat{x}(t) = \alpha\Big(x(t) + n(t)\Big) \tag{3.1}$$

with

$$\alpha = \frac{S}{S+N} \, . \tag{3.2}$$

The resulting output signal-to-noise ratio is

$$\mathrm{SNR}_A = \frac{E[x^2(t)]}{E\Big[\Big(x(t) - \hat{x}(t)\Big)^2\Big]} = \frac{SB_A}{\frac{1}{\frac{1}{SB_A} + \frac{1}{NB_A}}} = 1 + \frac{S}{N} \, . \tag{3.3}$$

It is well-known [8]-[10] that if the bandwidth of the channel coincides with the bandwidth of the source B_A, then the signal-to-noise ratio obtained in (3.3) cannot be improved by coding the input signal. This can be immediately checked by noticing that the conventional source-channel separation theorem requires that the rate-distortion function not exceed channel capacity, i.e.,

$$B_A \log\left(\mathrm{SNR}\right) \leq B_A \log\left(1 + \frac{S}{N}\right) ; \tag{3.4}$$

thus SNR_A is the largest SNR for which (3.4) is satisfied.

Suppose now that we are given *additional bandwidth* B_D and we want to explore the enhancement of signal-to-noise ratio achievable thanks to this bandwidth expansion. As before, we constrain the power spectral density transmitted through the channel to S. We will explore two possibilities:

1. Non-systematic: the original uncoded system is scrapped and the full bandwidth of the channel, $B_D + B_A$, is occupied by the encoded signal.

2. Systematic: the uncoded system is retained (source is sent directly through the channel) and only the additional bandwidth B_D is used to transmit the encoded information.

In scenario 1 it is easy to compute the achievable output signal-to-noise ratio, SNR_B, by equating rate-distortion function and channel capacity via the conventional separation theorem: (cf. [8])

$$B_A \log(\mathrm{SNR}_B) = (B_A + B_D) \log\left(1 + \frac{S}{N}\right) \tag{3.5}$$

i.e.,

$$\mathrm{SNR}_B = \mathrm{SNR}_A^{\left(1+\frac{B_D}{B_A}\right)}, \tag{3.6}$$

which admits the simple interpretation that the improvement of the signal-to-noise ratio in dB is a multiplicative factor equal to the bandwidth expansion factor, i.e., if the bandwidth is doubled so is the signal-to-noise ratio figure in dB (recall $\mathrm{SNR}_A \geq 1$).

The analysis of the systematic scenario 2 will lead to the same conclusion via the results in Section 2. In particular, we will check that the conditions for optimality of systematic encoding are satisfied in the present scenario:

C1. It is clear that the source maximizes the input-output mutual information of channel A among all sources of power SB_A (consider for example the equivalent discrete-time channel sampled at the Nyquist rate $2B_A$):

$$I(X\,;\,Z) = B_A \log\left(1 + \frac{S}{N}\right). \tag{3.7}$$

C2. According to [5], in the case of a Gaussian memoryless source with side information equal to the source plus Gaussian white noise, the Wyner-Ziv rate-distortion function is given by

$$R_{X|Z}(d) = \overline{R}_{X|Z}(d) = B_A \log \frac{B_A S N}{d(S+N)}. \tag{3.8}$$

C3. The value of side-information is maximally useful at the decoder:

$$\begin{aligned} &R_X(d) - R_{X|Z}(d) - I(X\,;\,Z) = \\ &B_A \log\left(\frac{B_A S}{d}\right) - B_A \log\left(\frac{B_A S N}{d(S+N)}\right) - B_A \log\left(1 + \frac{S}{N}\right) = 0\,. \end{aligned} \tag{3.9}$$

The conclusion is that in order to maximize the output fidelityit is optimal to transmit the uncoded Gaussian source directly through the channel and devote only the excess bandwidth to the transmission of the encoded information. The signal-to-noise ratio enhancement is a function of the bandwidth expansion given by (3.6). We emphasize that the burden to achieve such full bandwidth utilization falls exclusively on the digital encoder and on the decoder for the analog-digital receiver; the existing analog transmitters/receivers are unaffected. In existing analog broadcast systems for radio and television, the analog channel does not quite lend itself to full exploitation of its bandwidth (through suitable coding of the additional channel) because the transmitted signals do not follow the flat spectrum Gaussian model assumed in this work, and, of course, Vestigial Side Band (Television) and Double Side Band (Radio) are inherently wasteful of bandwidth/power. The computation of the Wyner-Ziv rate distortion function for a Double Side Band modulated flat Gaussian source is simple: wasting half of the bandwidth buys a 3 dB enhancement of the analog channel SNR. This follows from the fact that in the Gaussian case parallel independent side information channels are equivalent to a single side information channel whose SNR is equal to the sum of the individual SNRs.

It is interesting to note that the Wyner-Ziv encoding in the Gaussian case can be implemented in a simple way, using a Gaussian source codebook at the encoder, Slepian-Wolf encoding/decoding and a linear transformation at the decoder. This follows from the analysis of this case in [5] and the section on universal quantization with side information in [11]. Specifically, the encoder uses a codebook (implementable with an entropy-coded dithered quantizer [11]) which is optimal for encoding the Gaussian source with a target signal-to-noise ratio of $SNR_{encoder} = SNR_B - S/N$, where SNR_B is the final output signal-to-noise ratio given in (3.6) (Note that $SNR_B \geq SNR_A = 1 + S/N$.) At the decoder, the Slepian-Wolf code is decoded (given the side information), the result is scaled, and then an appropriately scaled version of the side information is subtracted.

The scenario where the digital channel is used for SNR enhancement without bandwidth boosting arises in satellite-enhanced reception, which may be of interest in improving terrestrial broadcast reception quality in remote or mountainous regions where the received SNR of the analog system is low. This requires the consideration of a more general case where the SNRs of the analog and digital channels are not necessarily equal. This is very easy to incorporate into the analysis by, once again, equating the Wyner-Ziv rate-distortion function with the digital channel capacity. Shannon's formula states that the capacity of the ideally bandlimited digital channel is $B_D \log(\mathrm{SNR}_D)$, where SNR_D is equal the SNR at the channel output plus one, (which according to (3.3) is equal to the optimum SNR of an estimator of the channel input from the channel output). Thus, we can write

$$\begin{aligned} B_A \log\left(\frac{B_A S N}{d(S+N)}\right) &= B_A \log\left(\mathrm{SNR}_{A+D}\right) - B_A \log\left(\mathrm{SNR}_A\right) \\ &= B_D \log\left(\mathrm{SNR}_D\right) \end{aligned} \tag{3.10}$$

where SNR_{A+D} is the achievable output SNR of a receiver that processes both channels jointly. Thus, (3.10) results in the conclusion that the improvement of SNR in dB due to the digital channel is proportional to its bandwidth and to its achievable SNR:

$$\frac{\mathrm{SNR}_{A+D}^{dB} - \mathrm{SNR}_A^{dB}}{\mathrm{SNR}_D^{dB}} = \frac{B_D}{B_A},$$

As a point of comparison, note that if the desired output SNR is equal to the SNR of the digital channel, i.e., $\mathrm{SNR}_{A+D} = \mathrm{SNR}_D$, then a stand-alone digital system would require bandwidth B_A, whereas by taking advantage of the analog channel we only need a digital channel bandwidth equal to

$$\frac{B_D}{B_A} = 1 - \frac{\mathrm{SNR}_A}{\mathrm{SNR}_D}.$$

The foregoing results allow us to deal with the more general setting where the additional channel bandwidth, B_D, is used not exclusively to enhance signal-to-noise ratio but to transmit additional information bandwidth, B_E. We are interested in finding the tradeoff between the transmissible source bandwidth $B_E + B_A$ and its decoded signal-to-noise ratio SNR_C, as a function of $\frac{S}{N}$ and the overall bandwidth of the channel $B_A + B_D$. If the target signal-to-noise ratio SNR_C is equal to the original one $\mathrm{SNR_A} = \left(1 + \frac{S}{N}\right)$, then the maximum excess bandwidth, B_E, that can be transmitted is equal to the additional channel bandwidth B_D. As we saw before, this can be accomplished with no coding whatsoever. If $\mathrm{SNR_C} = \mathrm{SNR_B} = \left(1 + \frac{S}{N}\right)^{\left(1+\frac{B_D}{B_A}\right)}$ then $\mathrm{B}_E = 0$. If SNR_C lies strictly between SNR_A and SNR_B then the achievable excess bandwidth is computed as follows.

Let us consider a communication scheme as depicted in Figure 2 where the additional channel bandwidth B_D is divided into

$$B_D = B_{DA} + B_{DE} \tag{3.11}$$

where B_{DA} is devoted to transmitting an encoded version of the original information bandwidth B_A for the sake of its signal-to-noise ratio enhancement; and B_{DE} is devoted to the encoded transmission of the excess information bandwidth B_E. Let the target signal-to-noise ratio be

$\mathrm{SNR}_C \in [\mathrm{SNR}_A, \mathrm{SNR}_B]$. Then, as we saw in (3.5) the necessary bandwidth B_{DA} to support SNR_C is

$$B_A \log(\mathrm{SNR}_C) = (B_A + B_{DA}) \log(\mathrm{SNR}_A) \tag{3.12a}$$

or equivalently,

$$\frac{B_{DA}}{B_A} = \frac{\log(\mathrm{SNR}_C)}{\log(\mathrm{SNR}_A)} - 1 \, . \tag{3.12b}$$

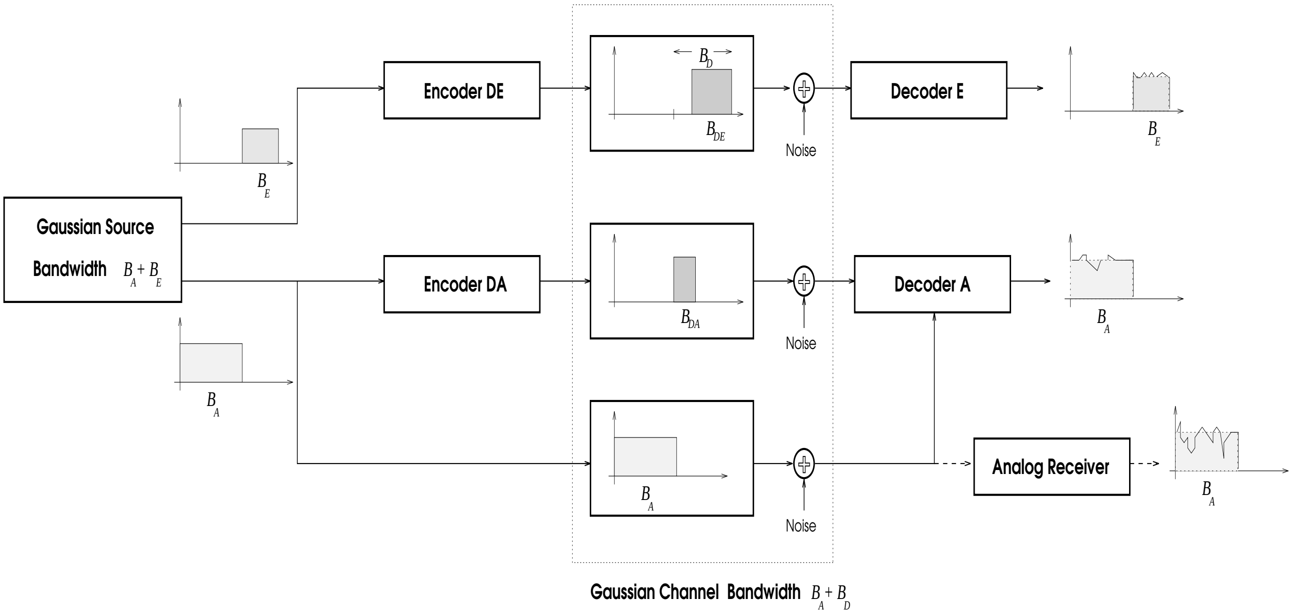

Figure 2: Systematic Transmission of a Gaussian Source through a Gaussian Channel.

The remaining channel bandwidth $B_{DE} = B_D - B_{DA}$ can support the following information bandwidth at signal-to-noise ratio SNR_C:

$$B_E \log(\mathrm{SNR}_C) = (B_D - B_{DA}) \log(\mathrm{SNR}_A) \, . \tag{3.13}$$

Using (3.11) and (3.13) we obtain the sought-after tradeoff:

$$B_E + B_A = (B_D + B_A) \frac{\log \mathrm{SNR}_A}{\log \mathrm{SNR}_C} \, . \tag{3.14}$$

Equation (3.14) implies that the bandwidth reduction factor is equal to the ratio of signal-to-noise ratios in dB. This generalizes the conclusion obtained in (3.6): the product information bandwidth times signal-to-noise ratio in dB remains constant. Again, this can be achieved using systematic encoding of the original information bandwidth B_A.

What if we are willing to tolerate a signal-to-noise ratio worse than the original SNR_A for the sake of bandwidth expansion beyond that offered by the channel? Then, systematic coding of any

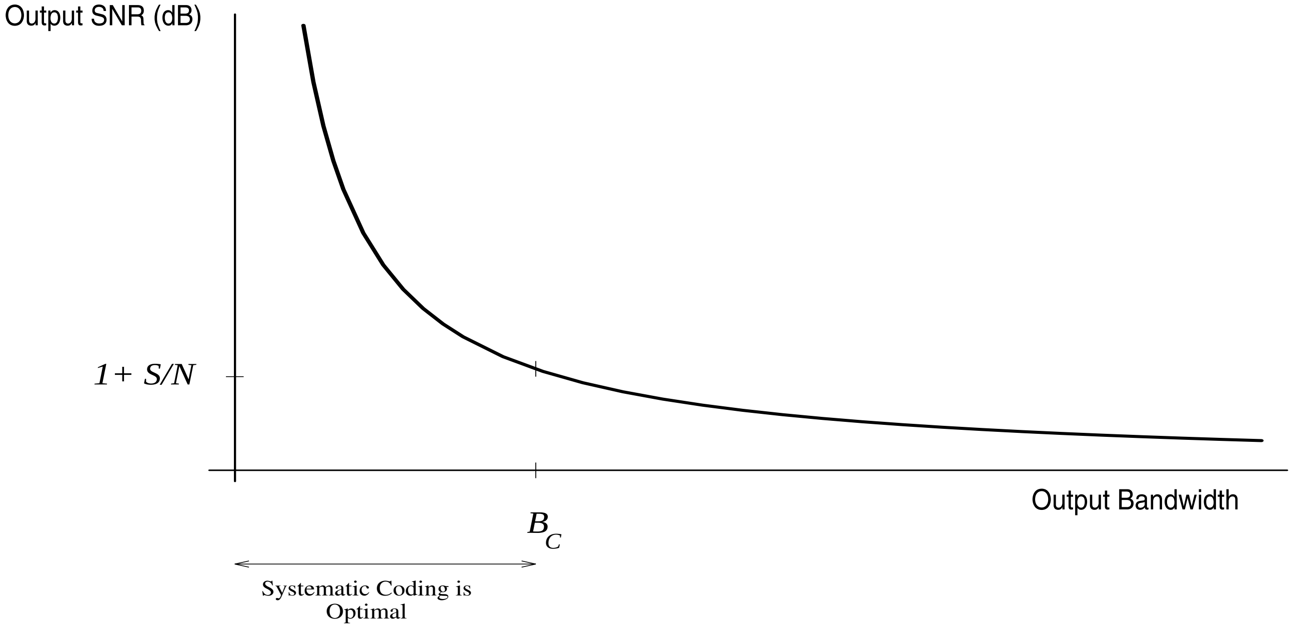

Figure 3: Tradeoff output SNR_C vs. output bandwidth B_I for Gaussian source and channel.

subband of the information source is strictly suboptimal. The conventional theory leads to the conclusion that the product of the transmitted signal bandwidth $B_I = B_A + B_E$ times the decoded signal-to-noise ratio SNR_C (in dB) must be equal to the channel bandwidth $B_C = B_A + B_D$ times the uncoded signal-to-noise ratio, as in (3.14).

$$(B_I)\ \log(\mathrm{SNR}_C) = (B_C)\ \log\left(1 + \frac{S}{N}\right). \tag{3.15}$$

In Figure 3, Equation (3.15) is depicted in terms of the decoded signal-to-noise ratio SNR_C (in dB) versus the transmitted information bandwidth B_I; the region of optimality of systematic transmission is explicitly indicated.

4 Bernoulli Source Through Binary Channel

In this section we examine a special case of practical and theoretical interest of the general framework developed in Section 2 (Figure 1). The source is a Bernoulli symmetric source; the distortion criterion is the bit-error-rate (Hamming distance); channel A [2] is a binary symmetric channel (BSC)

[2]We maintain here the notations channel A and channel D, though in the present case *digital* uncoded information is transmitted through channel A.

with crossover probability p_A; and channel D is a BSC with crossover probability p_D.

The conventional source-channel separation theorem implies that the minimum bit error rate $d \leq \frac{1}{2}$ with full encoding of both channels, is determined by (c.f. (2.8)),

$$R_X^b(d) = \min\{1, C_b(p_A) + C_b(p_D)\} \tag{4.1}$$

where $C_b(p) = 1 - h(p)$ is equal to the capacity of a BSC with crossover probability p, $h(p)$ is the binary entropy function in bits, and $R_x^b(d) = 1 - h(d)$ is the rate-distortion function of the Bernoulli symmetric source with Hamming distortion d.

According to Theorem 2.1, when the systematic uncoded part is transmitted over channel A, the minimum achievable bit-error-rate d is determined by:

$$\overline{R}_{X|Z}^{bb}(d) = C_b(p_D)$$

where $\overline{R}_{X|Z}^{bb}(d)$ is the Wyner-Ziv rate distortion function for a Bernoulli source with a BSC (with crossover probability p_A) side information channel. $\overline{R}_{X|Z}^{bb}$ was obtained by Wyner and Ziv in [4, Section II] (referred to therein as the doubly symmetric binary source):

$$\overline{R}_{X|Z}^{bb}(d) = \left\{ \begin{array}{lll} g(d) & , & 0 \leq d \leq d_c \\ g(d_c)\left(1 - \frac{(d-d_c)}{(p_A-d_c)}\right) & , & d_c < d \leq p_A \end{array} \right\} \tag{4.2a}$$

$$g(d) = \left\{ \begin{array}{ccc} h(p_A * d) - h(d) & , & 0 \leq d < p_A \\ 0 & , & d = p_A \end{array} \right\} \tag{4.2b}$$

where $*$ stands for the binary convolution $a * b = a(1-b) + b(1-a)$ and d_c is the solution of the equation

$$\frac{g(d_c)}{d_c - p_A} = g'(d_c) \ . \tag{4.2c}$$

A simple derivation of $\overline{R}_{X|Z}^{bb}(d)$ is given in [11] in terms of the *additive-noise rate-distortion function*.

The special case where both channels are identical $p_A = p_D = p$ is considered in Figure 4, which shows the achievable bit-error-rate for systematic and nonsystematic rate-$\frac{1}{2}$ coding as a function of the crossover probability of the channels.

It is noted that for reliable communication $d = 0$, no loss is incurred by systematic encoding as is already known [2],[3]. However, we see in Figure 4, that for any bit-error-rate in $0 < d < p$

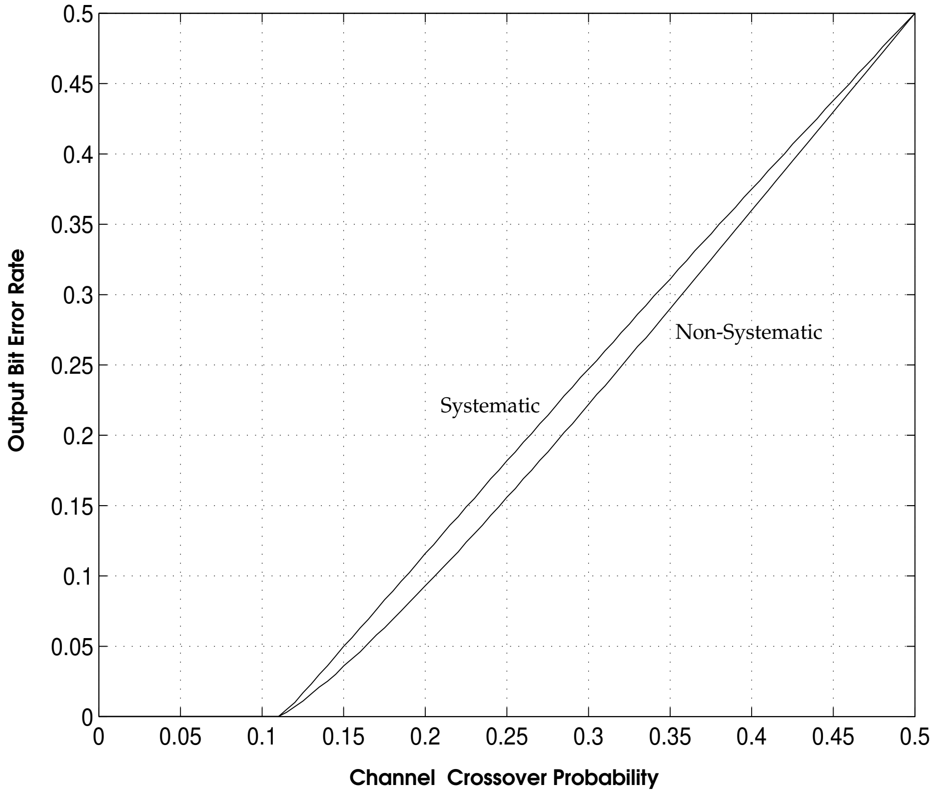

Figure 4: Minimum bit error rate achievable at rate $\frac{1}{2}-$ above the capacity of the binary symmetric channel.

systematic encoding is strictly suboptimal. For any p_A and p_D (not necessarily equal) two out of the three required equalities are satisfied:

$$I(X;Z) = C_A$$

and

$$R^{bb}_{X|Z}(d) = R^b_X(d) - I(X\,;\,Z);$$

only the fact that

$$\overline{R}^{bb}_{X|Z}(d) > R^{bb}_{X|Z}(d)$$

for all $0 < d < p_A$ accounts for the suboptimality of systematic coding.

We now propose a constructive approach to achieve the Wyner-Ziv rate distortion function $\overline{R}^{bb}_{X|Z}$ in (4.2). This approach (for which we do not provide a full proof) is a substantive generalization of Wyner's construction of Slepian-Wolf codes from linear block channel codes [13]. We will assume that channel D is noiseless, i.e., $p_D = 0$, for otherwise, conventional channel codes can be employed to convey the output of the source encoder to the receiver with arbitrary reliability.

We will focus on bit error rates $0 \leq q \leq d_c$; for $d_c < q \leq p_A$ the strategy that follows should be time-shared with no coding. Choose two linear codes defined by parity-check matrices $(m_1 \times n)$ H_1 and $(m_2 \times n)$ H_2, such that

$$\begin{aligned} \frac{m_1}{n} &\simeq h(q) \\ \frac{m_2}{n} &\simeq h(p_A * q), \end{aligned} \tag{4.3}$$

and Code 2 is a subcode of Code 1. Thus, the parity check matrices H_1 and H_2 satisfy

$$H_2 = \left[\begin{array}{c} H_1 \\ --- \\ H_a \end{array} \right] . \tag{4.4}$$

Every codeword of Code 1 $A = (a_1 \cdots a_n)$ satisfies

$$H_1 A^T = 0 . \tag{4.5a}$$

where the superscript T stands for the transpose operation. If, in addition,

$$H_a A^T = 0, \tag{4.5b}$$

then A is also a codeword of Code 2. The decoders for these codes are defined by functions of the corresponding syndromes

$$\begin{aligned} f_1(H_1 Y^T) &= f_1(H_1 U^T) \\ f_2(H_2 Y^T) &= f_2(H_2 U^T) = f_2(H_1 U^T, H_a U^T) \end{aligned} \tag{4.6}$$

where $Y = A \oplus U$, where $\oplus$ denotes modulo-2 addition. According to well-known properties of optimal linear codes and classical random linear coding arguments, it follows that the codebooks and decoding functions of Codes 1 and 2 can be chosen so that

$$\begin{aligned} f_1(H_1 U_1^T) &= U_1 \\ f_2(H_2 U_2^T) &= U_2 \end{aligned} \tag{4.7}$$

for most of the realizations U_1 of a Bernoulli(q) process with $q =$ probability of 1, and for most of the realizations U_2 of a Bernoulli($p_A * q$) process.

The *Wyner-Ziv encoding* of the binary source word consists of two steps.

(a) Among the codewords in Code 1 select X_q which is closest to X in Hamming distance. Let $E_q = X \oplus X_q$.

(b) Output the $m_2 - m_1$ vector $H_a X_q^T$.

Note that the output rate of the Wyner-Ziv encoder is

$$\frac{m_2 - m_1}{n} = h(p_A * q) - h(q) = g(q) = \bar{R}^{bb}_{X|Z}(q)\,. \tag{4.8}$$

The decoder receives $H_a X_q^T$ and the output of the BSC due to X which we denote by

$$\begin{aligned} Z &= X \oplus N \\ &= X_q \oplus E_q \oplus N\,. \end{aligned} \tag{4.9}$$

Since X_q is a codeword of Code 1, $H_1 X_q^T = 0$ and

$$H_2 X_q^T = \begin{bmatrix} 0 \\ ----- \\ H_a X_q^T \end{bmatrix}. \tag{4.10}$$

The decoder outputs the n-codeword

$$Z \oplus f_2(H_2 X_q^T \oplus H_2 Z^T) = Z + f_2(H_2(E_q^T \oplus N^T))\,. \tag{4.11}$$

If $E_q \oplus N$ were a Bernoulli$(p_a * q)$ process, the output of the decoder would yield with high probability

$$Z \oplus f_2(H_2(E_q^T \oplus N^T)) = Z \oplus E_q \oplus N = X_q, \tag{4.12}$$

in which case the decoder obtains a distorted version of the input X within distortion (bit error rate) q, as desired. By assumption, N is Bernoulli(p_A) and E_q is independent of N. Furthermore, for increasingly long blocklength, the distribution of E_q will resemble those of independent binary trials with parameter q. This can be expected by the backward channel interpretation of the rate-distortion function and the asymptotic rate-distortion optimality of linear block codes for the binary/Hamming case [14, Section 7.3] which imply that E_q resembles the noise process of the backward channel (a BSC in this case).

5 Bernoulli Source through Gaussian Channel

In this section, Channel A is an additive Gaussian noise discrete-time channel (variance σ_A^2) with binary antipodal $\left(\pm\sqrt{P_A}\right)$ inputs. Channel D is an additive Gaussian noise discrete-time channel (variance σ_D^2) with P_D-power-constrained (continuous) input, operating at a rate of ρ_D channel uses per source information bit. The distortion measure remains the bit-error-rate.

The minimum bit-error-rate d achievable when channel A is coded is given by the conventional source/channel separation theorem

$$R_X^b(d) = C_b(d) \leq C_{bg}(\mathrm{SNR}_A) + \rho_D C_g(\mathrm{SNR}_D) \ , \tag{5.1}$$

where $\mathrm{SNR}_A = P_A/\sigma_A^2$, $\mathrm{SNR}_D = P_D/\sigma_D^2$ are respectively the signal-to-noise ratios of channels A and D, and $C_{bg}(s)$ and $C_g(s)$ are the capacities of the Gaussian channel with binary ($\pm\sqrt{s}$) inputs and average power-constrained s inputs, respectively:

$$C_{bg}(s) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log_2\left(1 + e^{-2\sqrt{s}\,\tau - 2s}\right) d\tau \tag{5.2}$$

$$C_g(s) = \frac{1}{2} \log(1+s) \ . \tag{5.3}$$

If channel A is connected directly to the source, then the achievable bit-error-rate is given by (Theorem 2.1)

$$\overline{R}_{X|Z}^{bg}(d) \leq \rho_D C_g(\mathrm{SNR}_D) = \frac{\rho_D}{2} \log(1+\mathrm{SNR}_D) \tag{5.4}$$

where the right-hand-side of the inequality is the capacity per source information bit of channel D. The function $\overline{R}_{X|Z}^{bg}(d)$ stands for the Wyner-Ziv rate distortion of a Bernoulli source where the side-information is given by the outputs of a binary-input Gaussian channel driven by the uncoded source with signal-to-noise ratio SNR_A. In Appendix B, it is shown that $\overline{R}_{X|Z}^{bg}(d)$ is equal to the lower convex envelope[3] (conv)

$$\overline{R}_{X|Z}^{bg}(d) = \mathrm{conv}\Big(F(d)\Big) \tag{5.5}$$

of the function $F(d)$, $d \geq 0$ which is defined in a parametric form with the parameter $0 \leq q \leq 1/2$ via:

$$\begin{aligned} d = d(q) \quad &= \quad qQ\left(\sqrt{\mathrm{SNR}_A} - \frac{1}{2\sqrt{\mathrm{SNR}_A}} \log \frac{1-q}{q}\right) \\ &\qquad +(1-q)\,Q\left(\sqrt{\mathrm{SNR}_A} + \frac{1}{2\sqrt{\mathrm{SNR}_A}} \log \frac{1-q}{q}\right), \quad 0 \leq q \leq \frac{1}{2} \ , \end{aligned} \tag{5.6a}$$

[3]An alternative description is given in terms of an auxiliary time-sharing random variable. See [11] and Appendix B, eq. (B.7).

$$F(d) = F\Big(d(q)\Big) \quad = \quad \int_{-\infty}^{\infty} \frac{du}{\sqrt{2\pi}}\, e^{\frac{-u^2}{2}}\, h\left(\frac{(1-q) + q\, e^{-2\left(u\sqrt{\mathrm{SNR}_A}+\mathrm{SNR}_A\right)}}{1 + e^{-2\left(u\sqrt{\mathrm{SNR}_A}+\mathrm{SNR}_A\right)}} \right) - h(q)\ , \quad (5.6\mathrm{b})$$

$$0 \le q \le \tfrac{1}{2}$$

Note that the distortion achievable at zero rate, d_m, is the bit-error-rate of the uncoded channel,

$$p_u = Q\Big(\sqrt{\mathrm{SNR_A}}\Big) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\mathrm{SNR_A}}}^{\infty} e^{-x^2/2}\, dx,$$

Thus,

$$F(d) \equiv 0\ , \qquad d \ge d_m = p_u.$$

In Figure 5, we illustrate (5.1) and (5.4) in the case SNR=SNR_A =SNR_D and $\rho_D = 1$.

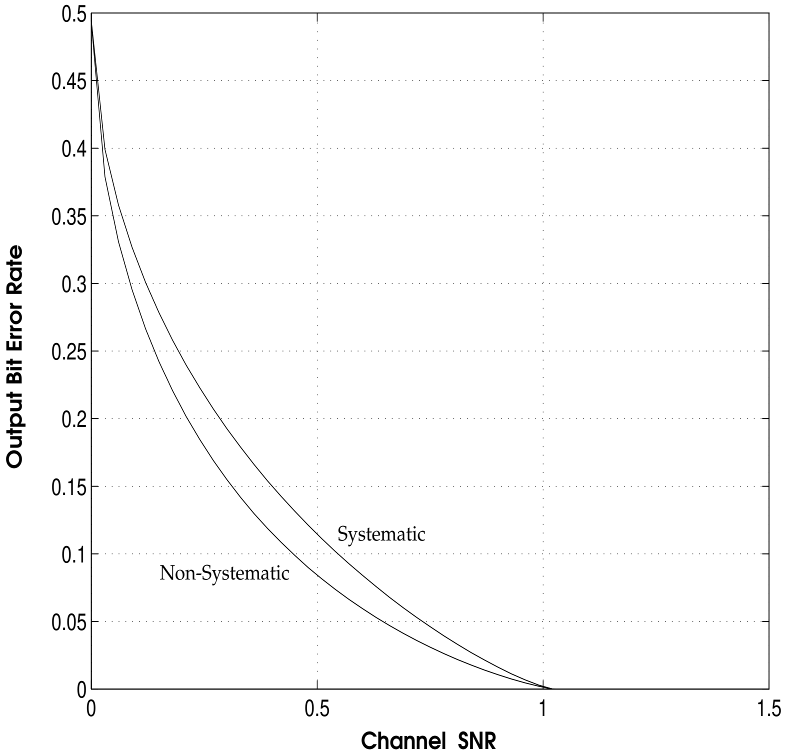

Figure 5: Minimum Bit Error Rate for transmission of a Bernoulli Source via a Gaussian Channel.

As we would expect, systematic coding is strictly suboptimal for $0 < d < d_m = Q(\sqrt{\mathrm{SNR}_A})$. In this case, only one of the three required equalities in (2.10) is satisfied, i.e. $I(X\,;\,Z) = C_A$ (condition C1), as equiprobable inputs maximize the capacity of the binary-input Gaussian channel. Thus, even in the hypothetical case in which the uncoded channel output were available at the encoder, the performance would still be suboptimal since condition C3 is not satisfied.

In the second part of this section we examine the case of overlaid communication over a single Gaussian channel, where neither additional bandwidth nor additional power are allocated to boost performance, see [3]. Degradation of communication performance over the existing channel A is traded off for enhanced performance of a receiver which accesses the outputs of both channels A and D. In this scenario channel A and D are in fact induced by the same physical channel over which the overlaid communication is operated.

The block diagram of the system is shown in Figure 6, where the original power P_I is divided into P_D (assigned to the overlaid coded system) and $P_I - P_D$ (left for the use of the uncoded system). The coding on channel D consists of Wyner-Ziv source encoding followed by optimal channel coding.

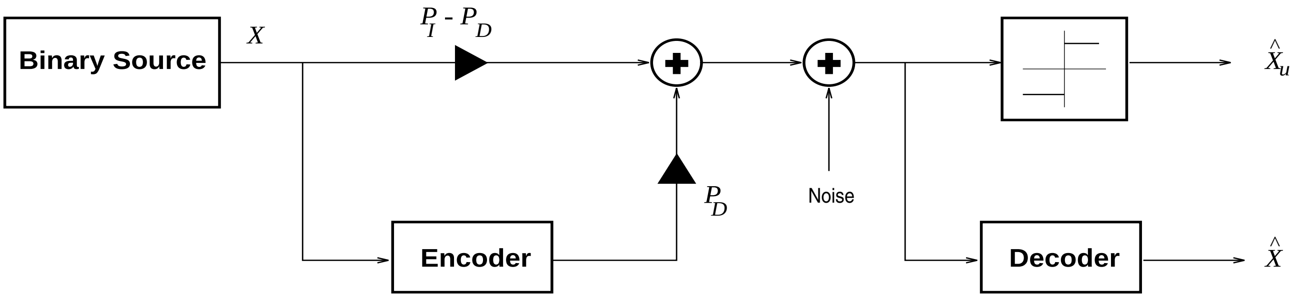

Figure 6: Overlaid encoded communication of a binary source on a Gaussian channel.

The overlaid communication link has been analyzed in [3] where the goal was to provide arbitrarily reliable communication for the enhanced system at the expense of degrading performance of the uncoded link. It was shown in [3] that minimizing the effect on the uncoded channel while preferring arbitrarily reliable communication at the highest possible rate requires that P_D be chosen as the smallest value satisfying

$$1 - C_{bg}\Big((P_I - P_D)/(\sigma^2 + P_D)\Big) \leq \frac{1}{2}\ \log(1 + P_D/\sigma^2)\ , \qquad 0 < P_D \leq P_I \tag{5.7}$$

and the corresponding performance of the uncoded link is degraded from $p_u = Q\left(\sqrt{\frac{P_I}{\sigma^2}}\right)$, to:

$$p_{ud} = Q\left(\sqrt{\frac{P_I - P_D}{\sigma^2 + P_D}}\right)\ . \tag{5.8}$$

We will extend the result of [3] on overlaid communication by considering an ad-hoc scheme which trades off degradation of the error probability of the uncoded link for improvement of the error probability of the coded link. The optimal trade off between these error probabilities with the overlayed communication scheme where no additional power or bandwidth is available and therefore the encoder must share resources with the uncoded raw information is unknown. This is why we adhere to a specific systems which offers a reasonable tradeoff.

In the extreme case where the error probability of the coded system d vanishes, our result particularizes to [3], whereas in the other extreme where no power is allocated to the coded system ($P_D = 0$), we get

$$d = p_{ud} = p_u = Q\left(\sqrt{\frac{P_I}{\sigma^2}}\right) . \tag{5.9}$$

The performance in [3] was achieved by an overlaid look-ahead communication scheme which guarantees no interference from the uncoded part to the coded part. This is no longer feasible when distortion is allowed since the receiver cannot replicate the source noiselessly. In this case, we devise a more involved version of the look-ahead encoding, as described in the following.

We now examine two different strategies. The first, which is somewhat easier to describe, is usually useful only when a minuscule degradation in the performance of the uncoded link is allowed. In this case the channel D coded part (channel coding only) gets first decoded reliably and is canceled out. Since reliable decoding of channel D codes is possible as long as the capacity of channel D is not exceeded, the interference from the coded part to the uncoded systematic transmission can be eliminated when we consider the uncoded channel as the side information channel. The coded transmission however suffers from the interference due to the uncoded communication. Thus the signal-to-noise ratios over the uncoded and coded channels are given, respectively, by

$$\mathrm{SNR}_A^{(1)} = (P_I - P_D)/\sigma^2 \ , \ \mathrm{SNR}_D^{(1)} = P_D/\Big(\sigma^2 + (P_I - P_D)\Big) . \tag{5.10}$$

The capacity of channel D is lower bounded by

$$C_D \geq \frac{1}{2} \log\left(1 + \mathrm{SNR}_D^{(1)}\right) = C_g(\mathrm{SNR}_D^{(1)}) \tag{5.11}$$

where the lower bound is obtained by replacing the binary interference by the worst case Gaussian interference with the same power $P_I - P_D$.

If we assume a Gaussian-distributed random codebook for channel D-codes and if the Euclidean decoding metric is adopted, then as shown in [15], the right-hand-side of (5.11) constitutes indeed the highest possible reliably transmitted rate, under these mismatched metric conditions and random coding. Since the uncoded transmission for the standard existing receiver is contaminated by the coded transmission, the resulting bit error probability increases to

$$p_{ud} = Q\left(\sqrt{\frac{P_I - P_D}{\sigma^2 + P_D}}\right) . \tag{5.12}$$

Note that the coded transmission resembles Gaussian noise as a 'Gaussian' codebook is used. The degradation in performance as compared to $p_u = Q\left(\sqrt{\mathrm{SNR}}\right)$, where $\mathrm{SNR} = P_I/\sigma^2$, follows by the signal power reduction on one hand and noise enhancement due to the coded part on the other. The achievable distortion d of the upgraded system which decodes the information on the basis of both the coded and systematic parts of the transmission is characterized by Theorem 2.1:

$$\overline{R}^{bg}_{X|Z}(d) \leq C_g\left(\mathrm{SNR}^{(1)}_D\right) , \tag{5.13}$$

where channel A - the side information channel enjoys $\mathrm{SNR} = \mathrm{SNR}^{(1)}_{\mathrm{A}}$. The tradeoff parameter is $0 \leq P_D \leq P_I$, the power allocated for the coded transmission. The curve of d versus p_{ud} for this strategy is shown in Figure 7 where it is produced in a parametric form, with P_D being the parameter.

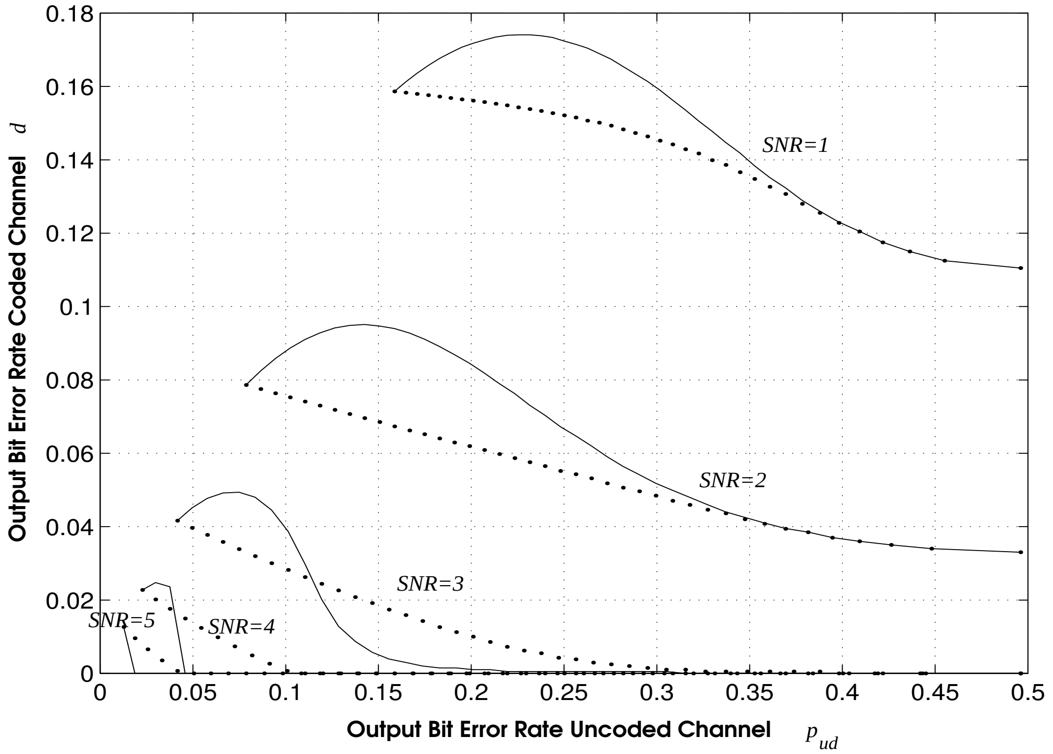

Figure 7: Bernoulli source through Gaussian channel. Dotted/Solid line is strategy 1/2.

We turn now to the second approach in which the interference caused by the uncoded transmission to the coded part is to be reduced. Towards this end we shall employ a version of look-ahead encoding. The input information bits of the source are grouped into super-blocks, where each super-block consists of N blocks of raw information bits and where each block comprises N bits of information. The blocks are encoded separately by the Wyner-Ziv source encoder which produces about $NR^{bg}_{X|Z}(d)$ bits and then re-encoded by an appropriate channel code of length N. The consecutive N-blocks of the coded information are grouped together and form a coded super-block. The i-th super block of uncoded raw information is interleaved by an $N \times N$ interleaver and transmitted. Superposed on this transmission is the $(i+1)$ coded super-block, which constitutes the look-ahead part of the encoding.

The i-th super-block of the uncoded information is received by the decoder when its coded counterpart is already available as it has been superimposed on the $(i-1)$-th unencoded super-block. The uncoded transmission suffers the interference from the coded part which again can be assumed Gaussian, as we specialize here to a "Gaussian" codebook [7, 15]. Therefore, the SNR in the uncoded channel is: $\mathrm{SNR}^{(2)}_{\mathrm{A}} = (P_1 - P_D)/(\sigma^2 + P_D)$ which accounts for both the signal power reduction and the noise enhancement due to the coded transmission. In contrast to the case examined in [3], no ideal cancellation of the interference due to the systematic transmission is possible, as in general the achievable bit-error-rate d is not negligible. Nevertheless, the channel code can partially reduce the interference, by using the Wyner-Ziv source code that is associated with the interfering systematic transmission. Let $\underline{U}_i, \sqrt{P_I - P_D}\underline{X}_i$ and $\underline{N}_i$ denote the coded, the systematic and the Gaussian noise components at the transmission of the i-th block, having powers $P_D, P_I - P_D$ and σ^2, respectively. According to the look-ahead strategy, at the time block $\underline{U}_i$ is transmitted, block $\underline{U}_{i-1}$ is already available for the digital link. Thus, the discrete-time channel viewed by the coded transmission is given by

$$\underline{Y}_i = \underline{U}_i + \sqrt{P_I - P_D}\underline{X}_i + \underline{N}_i;\ \underline{U}_{i-1} \text{ is available}. \tag{5.14}$$

The block $\underline{U}_{i-1}$ carries the Wyner-Ziv code for the source block $\underline{X}_i$. For an optimal Wyner-Ziv coder, the statistical relation between $\underline{U}_{i-1}$ and $\underline{X}_i$ corresponds to the Wyner-Ziv "test" channel between X and W; see definition (2.2) and in [11]. In fact, using standard random coding arguments and due to the interleaving, the channel in (5.14) can be re-written in terms of single-letter quantities, as

$$Y = U + \sqrt{P_1 - P_D}X + N;\ W \text{ is available }, \tag{5.15}$$

where W is the random variable (jointly distributed with X) which achieves the Wyner-Ziv function (2.2), and where (U, X, N) are independent. Clearly the channel encoder can hope to achieve the conditional capacity of this channel given W. We proceed by finding a lower bound for this capacity. For that, we assume that the code $\underline{U}_i$ is Gaussian (i.e. a Gaussian-like codebook is employed), and white (due to the interleaver), so the analog (side-information) channel (channel A) is effectively an Additive White Gaussian Noise (AWGN). For this case, Appendix B shows that the Wyner-Ziv function is achieved by a binary-symmetric channel $W = X \oplus V$ (or by a convex combination of such channels), where V is independent of X, and the crossover probability is determined by the overall bit-error-rate d via the parameter $0 \leq q \leq 1/2$ (5.6). Now, for convenience replace the modulo-2 sum above by the equivalent real multiplication, assuming that X and V are bipolar variables. Substituting $W = XV$ in (2), and manipulating the condition by multiplying the received signal by $W = XV$, while noting that $X^2 = 1$, we obtain the equivalent channel $Y = XVU + \sqrt{P_I - P_D}V + XVN$; XV is available. But since X, U and N are symmetric and independent, the random variables XVU and XVN are statistically independent of both V and XV, and are distributed as U and N, respectively. We thus conclude that the channel viewed by the coded transmission is effectively

$$\tilde{Y} = U + \sqrt{P_I - P_D}V + N \, . \tag{5.16}$$

Employing, as mentioned above, a Gaussian codebook for U, we can achieve at least the Gaussian capacity (i.e., the capacity calculated as if the bipolar variable V was Gaussian with the same variance) [15],

$$C_g\left(\frac{P_D}{\sigma^2 + 4q(1-q)(P_I - P_D)}\right) , \tag{5.17}$$

where $q = \Pr(V = -1)$ is determined by the overall bit-error-rate d; see Appendix B and (5.6). This is the desired lower bound for the capacity C_D of the digital link.

The distortion d is now determined by the least q for which the equation

$$\overline{R}^{bg}_{X|Z}(d) = C_g\left(\frac{P_D}{\sigma^2 + 4q(1-q)(P_I - P_D)}\right) \tag{5.18}$$

holds (note that $\overline{R}^{bg}_{X|Z}(d) = I(X; X \oplus V|Z)$ depends on q and on the Gaussian side-information channel (channel A) with $\mathrm{SNR} = \mathrm{SNR}^{(2)}_{\mathrm{A}}$). The degraded performance of the systematic uncoded transmission equals in both strategies (5.8), i.e. $p_{ud} = Q(\sqrt{\mathrm{SNR}^{(2)}_{\mathrm{A}}})$.

The plot of d versus p_{ud} is shown in Figure 7 for several $\mathrm{SNR} = P_I/\sigma^2$ values. It is noted that when p_{ud} must be selected close to p_u, the first approach is usually preferable while in cases where small d is the target, the second approach is mostly advantageous. In fact, when $\mathrm{SNR} = P_I/\sigma^2$ is less than 2 (3 dB), the first approach dominates the second one for all values of P_D, while for higher SNR values there is a cross over point and for larger values of P_D, the second approach is advantageous. For SNR $= p_I/\sigma^2$ larger than 5 (7 dB) the second approach dominates. The second approach reveals an interesting anomaly in performance for small values of P_D where d and p_{ud} increase together. This occurs because the degradation suffered by the uncoded channel (as reflected by $\mathrm{SNR}_{\mathrm{A}}^{(2)}$) is not counter-balanced by the capacity of the digital channel (channel D) as given in (5.17). In contrast, the performance of the first approach is monotonic. Note that when enough P_D is allocated so that (5.7) is satisfied, $d = 0$ and fully reliable performance of the enhanced system becomes possible. By time-sharing both approaches the lower convex envelope of the two curves in Figure 7 reflects the tradeoff between boosting the performance of the upgraded system and degrading the uncoded link.

We emphasize that the specific overlaid communication scheme considered here is by no means optimal[4]; it constitutes a particularly interesting example with reasonably good performance where a non-trivial performance tradeoff is demonstrated. Other interesting options can be treated: for example, when the Wyner-Ziv encoder is replaced by the Kaspi-Heegard-Berger [16],[17] encoder which operates either with or without the outputs of uncoded channel A (the side information channel). In this case, the partial cancellation of the interference caused by the uncoded transmission can be done without the assistance of outputs of the uncoded channel.

6 Conclusions

It is wasteful to simply discard the existing analog channel in digital broadcast receivers. Through adequate signal processing and coding/decoding, this paper has established that if the source and noise are Gaussian and the analog channel is transmitted in Single Side Band, its bandwidth/power can be used as efficiently as if we were able to design a completely digital system from scratch. We emphasize that the burden to achieve such full bandwidth utilization falls exclusively on the digital encoder and on the decoder for the analog-digital receiver; the existing analog transmitters/receivers are unaffected.

The designer can choose how to use the "digital" channel bandwidth. It can improve the fidelity of the output signal relative to that demodulated by the analog receiver; it can add new source information bandwidth (for example, additional screen lines for high definition television system,); or it can do both. Clearly, the more ambitious the bandwidth boosting is, the lower the capabilities for increased output signal-to-noise ratio. At one extreme, the designer may want to transmit an

[4]For example, a Gaussian-like codebook was assumed which is not necessarily optimal in the presence of the non-Gaussian residual interference.

information bandwidth which exceeds the channel bandwidth; in this case, the analog transmission is strictly suboptimum. Assuming Gaussian sources and channels, if the information bandwidth to be transmitted is equal to the total channel bandwidth (analog plus digital), then we have shown that not only the existing analog channel incurs no loss of capacity provided it is Single Side Band modulated, but there is no modulation method for the digital channel –analog or digital– which gives better SNR than Single Side Band. If the desired information bandwidth is less than the transmitted bandwidth (analog plus digital), we have shown that sophisticated encoding for the digital channel can render SSB modulation optimal for the analog channel. We have established the fundamental tradeoff between the output SNR and the transmitted information bandwidth: the product of the information bandwidth and its SNR (in dB) is constant and equal to its value when the reproduced signal bandwidth is equal to the channel bandwidth, i.e.

$$B_c \log\left(1 + \frac{S}{N}\right),$$

where S and N are the spectral levels of signal and noise, respectively. In particular, this formula enables the computation of the SNR achievable when the digital channel is used for SNR enhancement and no bandwidth boosting is required.

The analysis of binary sources has yielded the conclusion that systematic transmission through either binary symmetric channels or Gaussian channels is suboptimal in terms of resulting bit error rate. This result is somewhat surprising in view of the almost-noiseless case where it is well known that not only systematic codes (either linear or not) do not incur loss in capacity, but they also enjoy the best error exponent at rates between the critical rate and capacity.

We have also introduced a general structure of the optimal (in terms of Theorem 2.1) systematic code for the BSC channel, via the implementation of the Wyner-Ziv rate-distortion function in terms of good structured binary codes.

This paper and its predecessor [3] have shown that the fundamental information theoretic results of Slepian-Wolf [6] and Wyner-Ziv [4] lead to important conclusions in channel coding with uncoded sided information.

Appendix A: Conditional Rate-Distortion Function

Fact: The conditional rate-distortion function (2.11) admits the following expression

$$R_{X|Z}(d) = \inf_{\{W:\, \mathcal{E}(X|W,Z)\leq d\}} I(X\,;\,W \mid Z)\,. \tag{A.1}$$

Proof: We can expand the optimization set in (A.1) as the set of joint distributions $P_{XWZ\tilde{X}}$ such that if we sum out W and $\tilde{X}$ the resulting distribution is P_{XZ},

$$\sum_{(X,\tilde{X},W,Z)} D(X,\tilde{X}) P_{X,W,Z,\tilde{X}}(X,W,Z,\tilde{X}) \leq d \tag{A.2}$$

and

$$X \to (W,Z) \to \tilde{X} \ . \tag{A.3}$$

Under condition (A.3), the last term in the following equation is zero

$$\begin{aligned} I(X\,;W \mid Z) &= I(X\,;\tilde{X} \mid Z) + I(X\,;W \mid Z,\tilde{X}) - I(X\,;\tilde{X} \mid W,Z) \\ &\geq I(X\,;\tilde{X} \mid Z) \ . \end{aligned} \tag{A.4}$$

Thus, the right-side of (A.1) is greater or equal than inf $I(X\,;\tilde{X} \mid Z)$ where the infimum is over the set of joint distributions $P_{XWZ\tilde{X}}$ that satisfy the above conditions. Dropping (A.3) can only further lower bound the resulting expression, at which point W can be eliminated from consideration since it does not affect either the penalty function or the optimization set. Thus, we have established

$$\inf_{\{W:\,\mathcal{E}(X|W,Z)\leq d\}} I(X\,;W \mid Z) \geq \inf_{\substack{P_{XZ\hat{X}}\ : \\ E[D(X,\hat{X})]\leq d \\ \sum_c P_{XZ\hat{X}}(a,b,c) = P_{XZ}(a,b)}} I(X\,;\hat{X} \mid Z) \ . \tag{A.5}$$

To show the reverse inequality, we add the following additional condition to the feasible set of joint distributions $P_{XWZ\tilde{X}}$ in (A.1):

$$\sum_{a,b,c} P_{XWZ\tilde{X}}(a,\, b,\, c,\, b) = 1 \tag{A.6}$$

which implies that $\tilde{X} = W$, (A.3) is then automatically satisfied, and both the penalty function and the feasible set become those in the right side of (A.5).

Appendix B: Wyner-Ziv Rate Distortion of a Binary Source with a Gaussian Side Information Channel

In this appendix we calculate the Wyner-Ziv rate distortion function (2.2)

$$\overline{R}^{bg}_{X|Z}(d) = \min_{\{W:\,Z\to X\to W,\,\mathcal{E}_h(X|Z,W)\leq d\}} I(X;W \mid Z) \tag{B.1}$$

of a binary symmetric source $X \sim \mathcal{U}(-1,+1)$, with respect to the Hamming distortion measure, when the decoder has access to the noisy side information $Z = \sqrt{P}X + N$, where N is Gaussian $\sim \mathcal{N}(0,\sigma^2)$ and where P is here a scaling factor which stands for power. Our technique is inspired by the general approach for evaluating and bounding the Wyner-Ziv-rate distortion function recently introduced in [11].

Let $\mathcal{E}_h(\cdot \mid \cdot)$ be the minimum estimation error defined in (2.1) with the Hamming distortion. Let $U \in \{-1,+1\}$ be a binary random variable, independent of (X,Z), such that $\Pr(U=-1)=q$. Define the function $F(d)$ in parametric form:

$$F = I(X\,;\,X \oplus U \mid Z) \tag{B.2}$$

$$d = \mathcal{E}_h(X \mid X \oplus U, Z)\ , \tag{B.3}$$

where $\oplus$ denotes modulo-2 addition. Since we are dealing with the bipolar alphabet $\{-1,+1\}$ rather than $\{0,1\}$, it is convenient to adopt the convention that $\oplus$ is equivalent to real multiplication, i.e., $-1 \oplus -1 = +1 \oplus +1 = +1$ ("no error"), while $-1 \oplus +1 = +1 \oplus -1 = -1$ ("error"). Note that as q varies from 0 to 0.5, the function F ranges from $H(X|Z)$ to 0, while d ranges from 0 to $\mathcal{E}_h(X|Z) = d_{max} = Q\Big(\sqrt{P/\sigma^2}\Big)$. By symmetry it is easy to verify that the maximum a posteriori (MAP) estimator of X from Z is $\mathrm{sign}(Z)$. Thus, $Q = X \oplus \mathrm{sign}(Z)$ is the error variable and

$$\alpha(z) \triangleq \Pr(Q=-1 \mid Z=z) = \frac{1}{1+\exp(2\sqrt{P}|z|/\sigma^2)} \tag{B.4}$$

is the conditional probability of error given $Z = z$ in MAP estimation of X from Z. To get an explicit form of $F(d)$ using $\alpha(z)$ (though still in a parametric form), we decompose the mutual information of the additive channel in (B.2) into a difference of binary entropies, noting that $Z \to X \to X \oplus U$ implies $H(X \oplus U|XZ) = H(X \oplus U|X) = H(U)$, to obtain

$$F = H(X \oplus U|Z) - H(U) = E_Z\Big\{h\Big(\alpha(Z) * q\Big)\Big\} - h(q)\ , \tag{B.5}$$

where $h(\cdot)$ is the binary entropy, $*$ denotes the binary convolution operator (Section 4) and the expectation is taken with respect to $Z \sim \frac{1}{2\sqrt{2\pi\sigma^2}}\{\exp(-[z-\sqrt{P}]^2/2\sigma^2) + \exp(-[z+\sqrt{P}]^2/2\sigma^2)\}$.

The MAP estimation rule for X based on the two measurements $(X \oplus U, Z)$ in (B.3) boils down to

$$\left\{\begin{array}{ccc} \mathrm{sign}\ (Z) & , & \alpha(Z) > q \\ X \oplus U & , & \alpha(Z) < q \end{array}\right\} = \mathrm{sign}\ \left(Z + (X \oplus U)\sigma^2/2P\ \log\frac{1-q}{q}\right) \tag{B.6a}$$

and the corresponding distortion, i.e. error probability, is:

$$\begin{aligned} d &= E_Z\Big\{\min\Big(\alpha(Z), q\Big)\Big\} \\ &= qQ\left[\sqrt{\frac{P}{\sigma^2}} - \frac{1}{2}\sqrt{\frac{\sigma^2}{P}}\log\left(\frac{1-q}{q}\right)\right] + (1-q)Q\left[\sqrt{\frac{P}{\sigma^2}} + \frac{1}{2}\sqrt{\frac{\sigma^2}{P}}\log\left(\frac{1-q}{q}\right)\right]. \quad \text{(B.6b)} \end{aligned}$$

Next define the function,

$$\begin{aligned} F^*(d) &= \text{conv}\left\{F(d)\right\} \\ &= \min_{\{0\le\theta\le1, d_1, d_2 :\, \theta d_1 + (1-\theta)d_2 = d\}} \theta F(d_1) + (1-\theta)F(d_2)\ . \end{aligned} \tag{B.7}$$

The main result of this appendix is summarized as follows.

Proposition:

$$\overline{R}^{bg}_{X|Z}(d) = F^*(d)\ . \tag{B.8}$$

Proof: We first prove the "direct" part of (B.8), i.e.,

$$\overline{R}^{bg}_{X|Z}(d) \le F^*(d)\ . \tag{B.9}$$

¿From Definition (B.7), for each d the value $F^*(d)$ is a convex combination of two points on $F(d)$. Thus, for each d we can find a binary random variable ("time sharing variable") Θ, such that the pair (U, Θ) is independent of (X, Z), and

$$F^*(d) = I(X; X \oplus U \mid Z, \Theta) = I(X\,;\, X \oplus U, \Theta \mid Z)\ , \tag{B.10}$$

where

$$d = \mathcal{E}_h(X \mid X \oplus U, Z, \Theta)\ , \tag{B.11}$$

and where the second equality in (B.10) follows from the chain rule noticing that: $I(X; \Theta \mid Z) = 0$. Observe that since (U, Θ) is independent of (X, Z), we have the following Markov chain

$$Z \to X \to (X \oplus U, \Theta) \tag{B.12}$$

By (B.11) and (B.12) the pair $(X \oplus U, \Theta)$ satisfies the conditions of the minimization (B.1), so by (B.10) $F^*(d)$ must upper bound $\overline{R}^{bg}_{X|Z}(d)$, and (B.9) follows.

We turn to show the converse part of (B.8), i.e.,

$$\overline{R}^{bg}_{X|Z}(d) \geq F^*(d) \tag{B.13}$$

by means of four lemmas.

Lemma B.1 *X is independent of $(Q, |Z|)$. Furthermore, sign(Z) is independent of $(Q, |Z|)$.*

Proof: The first part of the lemma is equivalent to $\Pr(X = +1 \mid Q, |Z|) = \Pr(X = -1 \mid Q, |Z|) = 1/2$, or, by the Bayes rule, to

$$\Pr(X = +1, Q \mid |Z|) = \Pr(X = -1, Q \mid |Z|) \tag{B.14}$$

for every value of Q and $|Z|$. To show (B.14), note that if $X = +1$ then $\{Q = -1\} \Leftrightarrow \{N < -1\}$ ("error"), and $\{Q = +1\} \Leftrightarrow \{N > -1\}$ ("no error"). Similarly, if $X = -1$ then $\{Q = -1\} \Leftrightarrow \{N > +1\}$ ("error"), and $\{Q = +1\} \Leftrightarrow \{N < +1\}$ ("no error"). Thus,

$$\begin{aligned} \Pr(Q = -1 \mid X = +1, |Z|) &\overset{(a)}{=} \Pr(N < -1 \mid |N + 1|) \\ &\overset{(b)}{=} \Pr(N > 1 \mid |N - 1|) \\ &\overset{(c)}{=} \Pr(Q = -1 \mid X = -1, |Z|) \end{aligned} \tag{B.15}$$

where (a) follows from the independence of X and N; (b) follows from the symmetry of N; and (c) follows like (a). Similarly, we have $\Pr(Q = +1 \mid X = +1, |Z|) = \Pr(Q = +1 \mid X = -1, |Z|)$. Thus we can write

$$\Pr(Q \mid X = +1, |Z|) = \Pr(Q \mid X = -1, |Z|) \ . \tag{B.16}$$

Now, from the symmetry of N it follows that the conditional density of Z given $X = +1$ is equal to the conditional density of $-Z$ given $X = -1$, so $|Z|$ is independent of X. This fact, together with the symmetry of X implies by the Bayes rule that

$$\Pr(X = +1 \mid |Z|) = \Pr(X = -1 \mid |Z|) = 1/2 \ . \tag{B.17}$$

Multiplying the left / right sides of (B.16) and (B.17), we obtain (B.14), and the first part of the lemma is proved.

The second part of the lemma follows straightforwardly from the first part and the fact that $X = Q \oplus \text{sign}(Z)$. $\square$

Lemma B.2 *If W is such that $Z \to X \to W$ form a Markov chain, then the pair (X, W) is independent of the pair $(Q, |Z|)$.*

Proof: Note that $Z = (\text{sign}(Z), |Z|) = (X \oplus Q, |Z|)$. By standard manipulations of mutual information, we obtain

$$I(Z; W \mid X) = I(X \oplus Q, |Z|; W \mid X) = I(Q, |Z|; W \mid X) = I(Q, |Z|; W, X) \ , \tag{B.18}$$

where the last equality follows since by Lemma B.1, $I(|Z|, Q; X) = 0$. Now, if Z and W are conditionally independent given X, then the first term in (B.18) is 0, implying that the last term is also zero, which implies that $(Q, |Z|)$ are independent of (W, X). □

Lemma B.3 *For any* W

$$\mathcal{E}_h(Q \mid Z, W) = \mathcal{E}_h(X \mid Z, W) \ . \tag{B.19}$$

Proof: Suppose that the measurements $(Z, W) = (X \oplus Q, |Z|, W)$ are given. Then, any estimator $\widehat{Q}$ of Q induces an estimator of X: $\widehat{X} = (X \oplus Q) \oplus \widehat{Q}$, whose error is $X \oplus \widehat{X} = Q \oplus \widehat{Q}$. Thus, $P\left[X \neq \hat{X}\right] = P\left[Q \neq \hat{Q}\right]$, and (B.19) follows. □

Lemma B.4 *If* V *is independent of* (Q, Z)*, then*

$$I(Q; Q \oplus V \mid |Z|) = F[\mathcal{E}_h(Q \mid Q \oplus V, |Z|)] \tag{B.20}$$

Proof: We have

$$\begin{aligned}
I(Q; Q \oplus V \mid |Z|) &= I(Q; Q \oplus V \mid \text{sign}(Z), |Z|) && \text{(B.21a)}\\
&= I(X; X \oplus V \mid Z) && \text{(B.21b)}\\
&= F[\mathcal{E}_h(X \mid X \oplus V, Z)] && \text{(B.21c)}\\
&= F[\mathcal{E}_h(Q \mid X \oplus V, Z)] && \text{(B.21d)}\\
&= F[\mathcal{E}_h(Q \mid Q \oplus V, |Z|)] \ , && \text{(B.21e)}
\end{aligned}$$

where (a) follows by Lemma B.1 ($\text{sign}(Z)$ is independent of $(Q, |Z|)$); (b) follows by manipulating the condition; (c) follows by the definition of $F(d)$ in (B.2)-(B.3); (d) follows from Lemma B.3; and (e) follows from Lemma B.1. □

We are now in a position to show (B.13). Let W be any random variable such that $Z \to X \to W$ form a Markov chain, and $\mathcal{E}_h(X \mid Z, W) \leq d$. Let $E_W\{\cdot\}$ denote expectation over W. Consider the following chain of equalities / inequalities:

$$
\begin{aligned}
I(X;W \mid Z) &= I(X;W \mid \mathrm{sign}(Z), |Z|) && \text{(B.22a)}\\
&= I(Q;W \mid X \oplus Q, |Z|) && \text{(B.22b)}\\
&= I(Q;W \mid |Z|) + I(Q;Q \oplus X \mid W, |Z|) - I(Q;Q \oplus X \mid |Z|) && \text{(B.22c)}\\
&= I(Q;Q \oplus X \mid W, |Z|) && \text{(B.22d)}\\
&= E_W\{I(Q;Q \oplus X \mid W = w, |Z|)\} && \text{(B.22e)}\\
&= E_W\{F[\mathcal{E}_h(Q \mid Q \oplus X, |Z|, W = w)]\} && \text{(B.22f)}\\
&= E_W\{F[\mathcal{E}_h(X \mid Z, W = w)]\} && \text{(B.22g)}\\
&\geq F^*[\mathcal{E}_h(X \mid Z, W)] && \text{(B.22h)}\\
&\geq F^*[d] && \text{(B.22i)}
\end{aligned}
$$

which are justified by

(a)-(c) standard manipulations and the chain rule,

(d) first term in (c) is zero by Lemma B.2, and last term in (c) is zero by Lemma B.1 (note that $Q \oplus X = \mathrm{sign}(Z)$),

(e) the definition of conditional mutual information,

(f) Lemma B.4, noting that by Lemma B.1 X is independent $(Q, |Z|)$ for any value of w (X here plays the role of V in Lemma B.4),

(g) Lemma B.3, and the fact that $(Q \oplus X, |Z|) \Leftrightarrow Z$,

(h) Jensen inequality, using the fact that $F^*(d)$ lower bounds $F(d)$ and is convex $\cup$,

(i) $F(d)$ is monotonically decreasing and $d \geq \mathcal{E}_h(X \mid Z, W)$.

The desired inequality (B.13) now follows from (B.22a) and the definition of $\overline{R}^{bg}_{X|Z}(d)$ in (B.1). Direct evaluation of F in (B.5) along with some standard properties of the binary entropy function (such as $h(p) = h(1-p)$, $0 \leq p \leq 1$) yield the expression in (5.6). $\square$

We call attention to the fact that the function $\overline{R}^{bb}_{X|Z}(d)$ in case the side information is supplied by a BSC (4.3) can also be expressed in the form of (B.8) [11].

Acknowledgements

Discussions with Simon Litsyn are gratefully acknowledged.

References

[1] P. Elias, "Coding for Noisy Channels," IRE Convention Record, Vol. 4, pp. 37–46, March 1955.

[2] E. M. Gabidulin, "Limits for the Decoding Error Probability When Linear Codes are Used in Memoryless Channels," Problemy Peredachi Informatsii, Vol. 3, pp. 55–62, 1967.

[3] S. Shamai (Shitz) and S. Verdú, "Capacity of Channels with Uncoded Side Information," *European Trans. Telecommunications*, Vol. 6, No. 5, pp. 587–600, Sept.-Oct. 1995.

[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. IT–22, No. 1, pp. 1–10, January 1976.

[5] A.D. Wyner "The Rate-Distortion Function for Source Coding with Side Information at the Decoder—II: General Sources," *Information and Control*, Vol. 38, pp. 60–80, 1978.

[6] J.D. Slepian and J.K. Wolf, "Noiseless Coding of Correlated Information Sources, *IEEE Trans. Inform. Theory*, Vol. IT–19, No. 4, pp. 471–480, July 1973.

[7] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[8] T.J. Goblick, "Theoretical Limitations on the Transmission of Data from Analog Sources," *IEEE Trans. Inform. Theory*, Vol. IT–11, No. 4, pp. 558–567, October 1965.

[9] J. Ziv, "The Behavior of Analog Communication Systems," *IEEE Trans. Inform. Theory*, Vol. IT–16, No. 5, pp. 587–594, September 1970.

[10] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Pt.–II: Nonlinear Modulation Theory*, John Wiley & Sons, NY 1971.

[11] R. Zamir, "The Rate Loss in the Wyner-Ziv Problem," *IEEE Trans. Inform. Theory*, Vol 42, No 6, pp. 2073-2084,, Nov. 1996.

[12] R.M. Gray, "A New Class of Lower Bounds to Information Rates of Stationary Sources via conditional Rate-Distortion Functions," *IEEE Trans. Inform. Theory*, Vol. IT–19, No. 4, pp. 480–489, July 1973.

[13] A.D. Wyner, "Recent Results in the Shannon Theory", *IEEE Trans. Inform. Theory*, Vol. IT–20, No. 1, pp. 2–10, January 1974.

[14] A.J. Viterbi and J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, New York, 1979.

[15] A. Lapidoth, "On Information Rates for Mismatched Decoders," *Proc. 2nd Inter. Winter Meeting on Coding and Information Theory*, pp. 26, Dec. 12–15, 1993. Institut für Experimentelle Mathematik, Universität GH Essen, Ellernstrasse 29, 45326 Essen, Germany.

[16] A. H. Kaspi, "Rate-Distortion Function When Side-Information May Be Present At The Decoder," *IEEE Trans. Inform. Theory*, Vol. 40, No. 6, pp. 2031–2034, November 1994.

[17] C. Heegard and T. Berger, "Rate Distortion When Side Information May Be Absent," *IEEE Trans. Inform. Theory*, Vol. IT–31, No. 6, pp. 727–734, November 1985.

[18] S. Shamai (Shitz), S. Verdú, R. Zamir, "Digital Broadcasting Back-Compatible with Analog Broadcasting: Information Theoretic limits," *Fifth European Space Agency International Workshop on Digital Signal Processing Techniques Applied to Space Communications*, Sitges, Barcelona, Spain, Sep 25-27, 1996, pp. 8.25-8.39

[19] S. Shamai (Shitz), S. Verdu and R. Zamir, "Information Theoretic Aspects of Systematic Coding," *Proceedings of the International Symposium on Turbo Codes and Related Topics*, Sep. 1997, ENST de Bretagne, Brest, France.

Figure Captions

Figure 1: Information Transmission with Uncoded Side Information.

Figure 2: Systematic Transmission of a Gaussian Source through a Gaussian Channel.

Figure 3: Tradeoff output SNR_C vs. output bandwidth B_I for Gaussian source and channel.

Figure 4: Minimum bit error rate achievable at rate $\frac{1}{2}$– above the capacity of the binary symmetric channel.

Figure 5: Minimum Bit Error Rate for transmission of a Bernoulli Source via a Gaussian Channel.

Figure 6: Overlaid encoded communication of a binary source on a Gaussian channel.

Figure 7: Bernoulli source through Gaussian channel. Solid/Dotted is strategy 1/2.

Footnotes

1. In the sense formalized in Section 2.

2. We maintain here the notations channel A and channel D, though in the present case digital uncoded information is transmitted through channel A.

3. An alternative description is given in terms of an auxiliary time-sharing random variable. See [11] and Appendix B, eq. (B.7).

4. For example, a Gaussian-like codebook was assumed which is not necessarily optimal in the presence of the non-Gaussian residual interference.