State University of New York at Stony Brook
College of Engineering and Applied Sciences

Technical Report No. 766

Joint Source-Channel Codes for Broadcasting and Robust Communications

by

Udar Mittal and Nam Phamdo

Department of Electrical and Computer Engineering
State University of New York
Stony Brook, New York 11794-2350

e-mail: udar@sbee.sunysb.edu, phamdo@sbee.sunysb.edu

Date: November 13, 1998

Joint Source-Channel Codes for Broadcasting and Robust Communications

Udar Mittal and *Nam Phamdo*

Department of Electrical and Computer Engineering
State University of New York at Stony Brook
Stony Brook, NY 11794-2350
udar@sbee.sunysb.edu, phamdo@sbee.sunysb.edu

Abstract

The well-known "threshold effect" dictates that the more powerful a code is, the more sensitive it is to channel noise. A code is said to be *robust* if it is asymptotically optimal for a wide range of channel noise. Thus, robust codes have a "graceful degradation" characteristic and are free of the threshold effect. It is demonstrated that robust codes exist whenever the source and channel bandwidths are equal. In the unequal-bandwidth case, a collection of *nearly* robust joint source-channel codes is constructed using a hybrid of digital and analog coding techniques. The design principle is based on bandwidth/power splitting and matched tandem coding. These nearly robust codes achieve the Shannon limit and have a less severe threshold effect. Finally, for the case of two different noise conditions, the achievable distortion regions of these codes are determined.

1 Introduction

Consider the problem of transmitting a bandlimited analog source on a bandlimited additive noise channel. In many situation, the exact channel signal-to-noise ratio (SNR) may not be known to the transmitter. For example, the *range* of SNR may be known but the true SNR value may be unknown. Given a range of SNR, it is desirable to design a *single* transmitter which performs "well" for all SNRs within this range. There are essentially two ways of designing such a system - (i) using *analog* modulation methods or (ii) using *digital* modulation methods.

The main advantage of analog methods — such as amplitude modulation (AM) and frequency modulation (FM) — is the gradual change (graceful degradation) in the received signal quality with change in SNR [4]. However, these systems are in general suboptimal in the sense that they rarely achieve the theoretically optimal performance. Further, analog systems require expensive circuitry which are generally less reliable.

Digital systems are not only more reliable and cost efficient than analog systems but they can be designed to (asymptotically) achieve the theoretically optimal performance [1, 2, 18]. The inherent problem of digital systems is that they tend to suffer from (a severe form of)[1] the "threshold effect" [3]. This effect can be described as follows. The system achieves a certain performance at a certain designed SNR. The system performance, however, does not improve with increased SNR and it degrades drastically when the true SNR falls below the designed SNR. Further, the threshold effect becomes more pronounced as the system performance approaches the theoretical optimum.

The severeness of the threshold effect in digital systems may be attributed to Shannon's source-channel separation principle [1, 2, 13] which states that no loss of optimality is incurred by separate (independent) design of source and channel codes. Codes designed based on the separation principle are often referred to as tandem source-channel codes. Recent works [5 - 12] [16], however, demonstrated that joint source-channel codes not only outperform tandem codes for a fixed complexity and delay, they are also inherently more robust to channel noise than tandem codes.

A vector quantizer with an optimized index assignment [5] is an example of robust joint source-channel codes. Channel-optimized vector quantizers [7], channel matched tree-structured and multi-stage vector quantizers [9], and channel-optimized trellis coded quantizers [11] are

[1] Analog systems suffer a mild form of the threshold effect [17]. In the mild form, the system performance improves when the SNR is above the designed SNR.

further examples of robust joint source-channel codes.

Recently, Shamai et al. [14] demonstrated that systematic joint source-channel codes are optimal for a wide class of sources and channels. One of the motivations of [14] is that systematic coding allows graceful degradation of performance at low SNR. For Gaussian source and channel, Shamai et al. proposed a type of bandwidth splitting which enables systematic coding on part of the source bandwidth.

In this report, we utilize a combination of bandwidth splitting and power splitting to design a collection of "nearly" robust joint source-channel codes. Furthermore, we introduce the concept of matched tandem code which facilitates linear decoding of (part of) the source signal at low SNR. The use of linear decoder makes the threshold effect less pronounced. The preliminaries, notations and models for source and channel are given in Section 2. In Section 3, we define robustness and proved the existence of a robust code when the channel and source bandwidths are identical. Further, we also conjectured on the non-existence of a robust code when the bandwidths are not identical. In Section 4, we define the modules used in Section 5. In Section 5, we design various joint source-channel coding systems and derive the achievable distortion regions of these systems. In Section 6, we compare the performance of the systems in Section 5. In Section 7, a nearly-robust code is defined. We also showed the existence of a nearly-robust code in this section.

2 Preliminaries and Notation

2.1 The Broadcasting Problem

A broadcast system consists of one transmitter and many fixed (finite) receivers. In Fig. 1. the transmitter broadcasts the same information, say an audio signal, to K different receivers (users). The distance between the transmitter and each user varies. In such a system, one expects that the user closest to the transmitter should receive the lowest distortion (best audio quality). The transmitter sends the same signal to all users, but each user receives the signal with a different SNR. The closer the receiver is to the transmitter, the greater the SNR. It is desired to design an encoder which performs well at all these SNR's. Further, it should give better performance (audio quality) for users which are nearer to the transmitter and the performance should not degrade drastically for those users which are considerably farther from the transmitter.

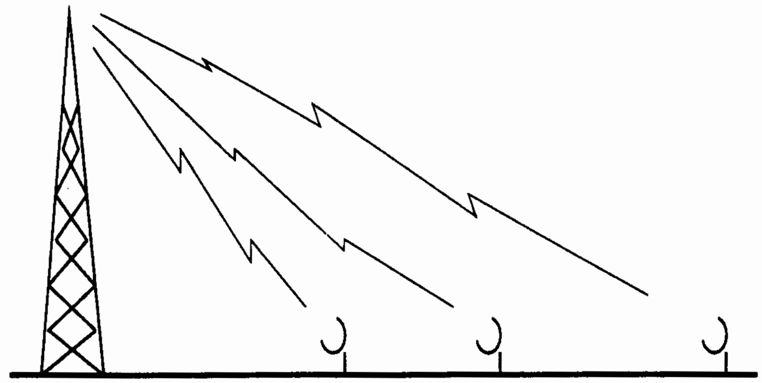

Figure 1: Broadcast System

2.2 The Robustness Problem

In the broadcasting problem, we have many users with different SNR. Now, consider a problem of designing a transmitter for a single mobile user. The channel noise vary from time to time depending on various factors, such as position of the mobile, the speed of the mobile, fading, etc. In such a scenario, we need to design transmitter and receiver whose performance do not degrade drastically because of the mobility of the user. In other words, the transmitter and the receiver should be robust to change in channel noise. One way of doing this is to design transmitter and receiver for the worst case noise. However, this implementation does not utilize the higher capacity of the channel when the channel noise is less than the designed noise.

For many channels of interest, the receiver can accurately estimate the channel noise from the received signal. With this assumption, this problem is now similar to the broadcasting problem. In the broadcasting problem the set of channel noise for which the code should perform well is finite while in the robustness problem this set is uncountable.

2.3 Source and Channel Model

Source: Consider a memoryless Gaussian source, $\{X_i\}_{i=1}^{\infty}$, with zero mean and variance σ^2. Thus, $X_i \sim \mathcal{N}(0, \sigma^2)$ and the sequence $\{X_i\}$ is independent and identically distributed (i.i.d.). We assume the source is obtained from uniform sampling of a continuous-time Gaussian process with bandwidth W_s (Hz). Furthermore, the sampling rate is assumed to be $2W_s$ samples per second.

Channel: The source is to be transmitted over an AWGN channel modeled by $Z_i = Y_i + V_i$, where Y_i, Z_i and V_i are the channel input, output and noise, respectively. We assume $E[Y_i^2] \leq$

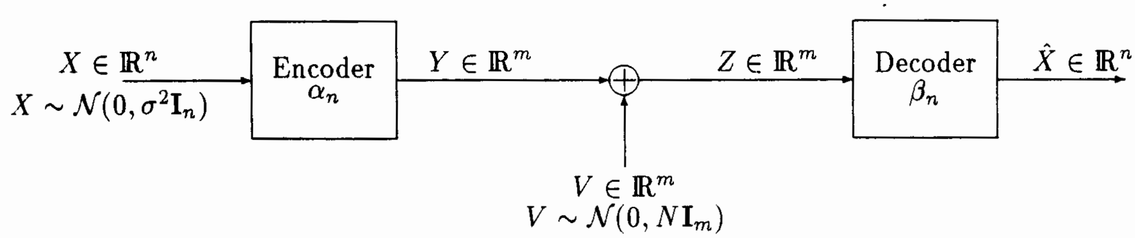

Figure 2: Block Diagram of Joint Source-Channel Coding System.

P and $V_i \sim \mathcal{N}(0, N)$. The channel is derived from a continuous-time AWGN channel with bandwidth W_c (Hz). The discrete-time channel is used at a rate of $2W_c$ channel uses per second.

Coding: The block coding system is depicted in Fig. 2. The source samples are grouped into blocks of size n: $X = (x_1, x_2, \ldots, x_n)$. The encoder is a mapping $\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^m$ which satisfies the power constraint $E[\|\alpha_n(X)\|^2] \leq mP$, where $\rho = m/n = W_c/W_s$. The received signal is given by $Z = Y + V$, where $Y = \alpha_n(X)$, $V \sim \mathcal{N}(0, N\mathbf{I}_m)$ and $\mathbf{I}_m$ is the $m \times m$ identity matrix. The decoder is a mapping $\beta_n : \mathbb{R}^m \mapsto \mathbb{R}^n$. The average squared-error distortion of the coding system is denoted as

$$D_{ave,n}(\alpha_n, \beta_n, N) = \frac{1}{n} E[\|X - \hat{X}\|^2], \tag{1}$$

where $\hat{X} = \beta_n(Z)$. For purpose of analysis, we will consider a sequence of codes (α_n, β_n), where n is increasing but the ratio $\rho = m/n$ is fixed. The asymptotic performance of the code is given by

$$D_{ave}(N) = \lim_{n \to \infty} D_{ave,n}(\alpha_n, \beta_n, N). \tag{2}$$

We assume that $\sigma^2 > 0, P > 0$, and $\rho > 0$ are known and express D_{ave} as a function of N. In subsequent sections distortion means the mean squared distortion.

Theoretical Limit: The rate-distortion function for the memoryless Gaussian source and squared-error distortion measure is given by [2, 18]:

$$R(D) = W_s \log_2 [\text{illegible}] D) \quad \text{bits/sec.}$$

Similarly, the capacity of the AWGN channe[illegible]en by [1, 18]:

$$C(N) = W_c \log_2 [\text{illegible}] N) \quad \text{bits/sec.}$$

If a code has average distortion D, then $R(D) \leq C(N)$. This implies that

$$D \geq \frac{\sigma^2}{(1+P/N)^{\rho}} \triangleq D_{opt}(N). \tag{5}$$

According to classical information theory, $D_{opt}(N)$ provides a lower bound on the average squared-error distortion of any coding system. Furthermore, given σ^2, P, N and ρ, there exists code sequences with $D_{ave}(N)$ equal to $D_{opt}(N)$. However, the optimal code sequence may differ for different values of σ^2, P, N and ρ. In this report, we are interested in code sequences which are robust to channel noise. Thus, we assume that σ^2, P, and ρ are known and N is unknown. Therefore, in (5), we express D_{opt} only as a function of N.

Notations: The significance of symbols X, Y, Z, $D_{opt}(N)$ and $D_{ave}(N)$ is already given in the previous section. We shall be splitting the source X. This splitting may be bandwidth splitting (demultiplexing) or power splitting. In case of bandwidth splitting, $X = [X_1 X_2]$, where X_1 and X_2 represent the two demultiplexed components of X. In case of power splitting, $X = \tilde{X} + e$, where $\tilde{X}$ is a quantized value of X and e is the error due to quantization. For any vector W, let $\tilde{W}$ be a quantized value of W and $\hat{W}$ be an estimate of W. Thus, $\hat{\tilde{W}}$ is an estimate of a quantized value $\tilde{W}$ of W. Let $D_W(N)$ be the limit (as n approaches infinity) of mean squared distortion between two n-dimensional vectors W and $\hat{W}$ when the channel noise is N, i.e.,

$$D_W(N) = \lim_{n\to\infty} \frac{1}{n} E(\|W - \hat{W}\|^2). \tag{6}$$

The symbol $\mathbb{N}$ represents a set of non-negative integers. Further, in the subsequent sections, all the logarithms unless specified are to the base 2.

3 Robustness

Robustness: We define a code to be robust if its average distortion is close to the theoretical limit for a wide range of channel noise conditions. More precisely:

Definition 1 *A code sequence $\{(\alpha_n, \beta_n)\}$ is said to be robust on $\aleph$ if it satisfies the power constraint and*

$$D_{ave}(N) = D_{opt}(N), \quad \forall\, N \in \aleph.$$

We note that the code can not depend on N. For a code to be robust, it must perform close to the optimum for all $N \in \aleph$. Robust codes may or may not exist. In the following sections, we discuss the conditions for the existence of robust codes.

3.1 Robust Codes Exist for $\rho = 1$

In this section, we show that robust codes exist for $\rho = 1$. The proof is based on actual construction of robust codes rather than the traditional technique of random coding.

Theorem 1 *Assume $\rho = 1$. $\forall$ finite set $\aleph$ ($|\aleph| < \infty$), $\exists$ a code sequence $\{(\alpha_n, \beta_n)\}$ which is robust on $\aleph$.*

One can prove this theorem by setting $\alpha_n(X) = \sqrt{P/\sigma^2}X$ and $\beta_n(Z) = \sqrt{\sigma^2 P}Z/(P+N)$. The optimality of this linear code is well-known (see, for example, [14]). However, this decoder depends on N. The proof of this theorem is given in the appendix where we set $P+N = \frac{1}{n}\|Z\|^2$ and applied the law of large numbers.

3.2 Conjecture on Non-Existence of Robust Codes for $\rho \neq 1$

Shannon in [3], argued that topologically it is not possible to map one region to another region in a one-to-one, continuous manner unless both the regions have the same dimensionality. On this basis he explained the threshold effect common to various communication systems. Ziv, by restricting to a class of signals, showed in [15] that for a Gaussian source transmitted on an additive bandlimited Gaussian channel, no single practical modulation scheme can achieve the optimal performance at all noise levels, if the channel bandwidth is greater than the source rate, i.e., $\rho > 2$. In [19, 20], it was conjectured that for $\rho \neq 1$, any system which touches $D_{opt}(N)$ curve at a particular value of N can not touch the curve again. This conjecture is restated below.

Conjecture 1 *Assume $\rho \neq 1$ and $\aleph = \{N_1, N_2\}$, $N_1 \neq N_2$, there does not exist any code sequence $\{(\alpha_n, \beta_n)\}$ which is robust on $\aleph$.*

For the case of AWGN channel, the decoder can always estimate the noise power precisely (see proof of Theorem 1). Hence we may assume that the decoder has full knowledge about the noise power.

Definition 2 *An encoder sequence $\{\alpha_n\}$ is said to be robust on $\aleph$ if it satisfies the power constraint and $\exists$ a collection of decoder sequences $\{\beta_n(N) : N \in \aleph\}$ such that*

$$\lim_{n\to\infty} D_{ave}(\alpha_n, \beta_n(N), N) = D_{opt}(N), \; \forall\, N \in \aleph. \tag{8}$$

Conjecture 2 *Assume $\rho \neq 1$ and $\aleph = \{N_1, N_2\}$, $N_1 \neq N_2$, there does not exist any encoder sequence which is robust on $\aleph$.*

Note that Conjecture 1 follows from Conjecture 2. Even though we have not been able to prove these conjectures, we have derived a collection of codes which are nearly-robust, the exact definition of which is provided in Section 7.

4 Modules Used in the Coding Systems

In the next section, we will be designing many coding systems which are nearly-robust. The modules used in the design of these systems are described here. For each module, we use r to denote the rate (instead of ρ).

4.1 Rate r Tandem Code

A tandem code (α_n, β_n) consists of a tandem encoder $\alpha_n : \mathbb{R}^n \mapsto \mathbb{R}^m$ and a tandem decoder $\beta_n : \mathbb{R}^m \to \mathbb{R}^n$, where $m/n = r$. The encoder α_n consists of a source encoder $\alpha_n^s : \mathbb{R}^m \mapsto \mathbb{N}$ followed by a channel encoder $\alpha_n^c : \mathbb{N} \mapsto \mathbb{R}^n$, i.e., $\alpha_n(X) = \alpha_n^c(\alpha_n^s(X))$. The decoder β_n consists of a channel decoder $\beta_n^c : \mathbb{R}^m \mapsto \mathbb{N}$ followed by a source decoder $\beta_n^s : \mathbb{N} \mapsto \mathbb{R}^n$. Note that $\{(\alpha_n^s, \beta_n^s)\}$ is a source code sequence and $\{(\alpha_n^c, \beta_n^c)\}$ is a channel code sequence. The channel code sequence is such that when it is used for transmitting data on an AWGN channel with noise variance N, input power constraint P_c, and data rate

$$\lim_{n\to\infty} \frac{1}{nr} H(\alpha_n^s(X)) = \frac{1}{2}\log\left(1 + \frac{P_c}{N}\right) \text{ bits/channel use,} \tag{9}$$

it achieves arbitrarily low probability of error, i.e.,

$$\lim_{n\to\infty} \Pr\{\alpha_n^s(X) \neq \beta_n^c(Z)\} = 0. \tag{10}$$

Note that the increase in data rate with n in (9) is governed by the decrease in probability of error in (10).

The source code sequence quantizes an i.i.d. Gaussian source of variance σ_s^2 at rate

$$\lim_{n\to\infty} \frac{1}{n} H(\alpha_n^s(X)) = \frac{r}{2}\log\left(1+\frac{P_c}{N}\right) \text{ bits/source sample,} \tag{11}$$

and distortion

$$D = D(\sigma_s^2, P_c, N, r) = \lim_{n\to\infty} \frac{1}{n}\|X - \beta_n^s(\alpha_n^s(X))\| = \frac{\sigma_s^2}{\left(1+\frac{P_c}{N}\right)^r}. \tag{12}$$

We specify this tandem code sequence by a 5-tuple $(\sigma_s^2, P_c, N, r, D)$ and say that the tandem code is designed for $(\sigma_s^2, P_c, N, r, D)$. Note that D is a function of the other four parameters and is given by (12). Thus, $D_{ave}(\alpha, \beta, N) = D_{opt}(N)$, when this tandem code sequence is used as the code sequence for transmitting an i.i.d. Gaussian source of variance σ_s^2 on an AWGN channel with noise variance N, power constraint P_c, and rate r.

The source code of the tandem code splits X into two orthogonal components, the quantized value $\tilde{X} = \beta_n^s(\alpha_n^s(X))$ of X, and the quantization error $e = X - \tilde{X}$.

Lemma 1 *For a tandem code sequence specified by* $(\sigma_s^2, P_c, N, r, D)$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |E(e_i \tilde{x}_i)| = 0. \tag{13}$$

where $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ *and* $e = (e_1, e_2, \ldots, e_n)$. *Hence*

$$\lim_{n\to\infty} \left|\frac{1}{n}E(\|\tilde{X}\|^2) - (\sigma_s^2 - D)\right| = 0. \tag{14}$$

The proof of this lemma is given in the appendix.

In the rest of the report we will be dealing with tandem code, encoder and decoder sequences. Sometimes, we will specify these sequences as tandem code, encoder and decoder respectively.

4.2 Matched Tandem Code

A tandem code is said to be matched, if the first $l = \min(m, n)$ components of the channel input, $\alpha_n(X)$, are scaled versions of the first l components of the quantizer output, $\beta_n^s(\alpha_n^s(X))$, for all X. In a matched system, if $\beta_n^s(k) = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$ then $\alpha_n^c(k) = (y_1, y_2, \ldots, y_m) = (a\tilde{x}_1, a\tilde{x}_2, \ldots, a\tilde{x}_l, y_{l+1}, \ldots, y_m)$.

Lemma 2 *For any values of* (σ_s^2, P_c, N, r)*, there exists a matched code sequence which is asymptotically optimal.*

The proof of this lemma, based on the random coding argument, is given in the appendix.

The main advantage of the matched tandem code is that it facilitates linear decoding of some components of the transmitted signal when the channel noise is more than the designed noise. Let the matched tandem encoder be designed for $(\sigma_s^2, P_c, N_c, r, D)$. If the channel noise $N > N_c$ then a linear decoder given by

$$\hat{x}_i = \frac{\sqrt{(\sigma_s^2 - D)P_c}}{P_c + N} z_i, \tag{15}$$

obtains a linear estimate of the first $\min(n, rn)$ components of the transmitted signal $X = (x_1, x_2, \ldots, x_n)$. We specify this decoder by a 3-tuple $(\sigma_s^2 - D, P_c, N)$.

In practice, a matched tandem code can be designed using the method of joint trellis coded quantization/modulation (for $r = 1$) [16] and then properly puncturing the code for r less than or greater than one.

4.3 Rate-One Linear Code

A rate-one linear code consists of a linear encoder $\alpha : \mathbb{R}^n \mapsto \mathbb{R}^n$ and a linear decoder $\beta : \mathbb{R}^n \mapsto \mathbb{R}^n$. Let the variance of the input to the linear encoder be σ_s^2, the output power constraint be P_l, and the noise variance be N. Now, the encoder $\alpha(X) = \sqrt{\frac{P_l}{\sigma_s^2}} X$ and the decoder $\beta(Z) = \frac{\sqrt{\sigma_s^2 P_l}}{P_l + N} Z$. Note that the encoder α is independent of the noise variance N. In subsequent sections, the encoder α is specified by an 2-tuple (σ_s^2, P_l) and the decoder by a 3-tuple (σ_s^2, P_l, N).

Even though the rate-one linear code can be implemented by high precision digital circuits it is often referred to as an analog coding system since X is never represented as "bits".

5 Achievable Distortion Region

Consider the broadcast system in Fig. 1. [illegible] source and the channel be as given [illegible] Section 2.3. Let the noise variance of the [illegible] user be N_k $(1 \le k \le K)$. A K-tuple $(d_1, d_2, \ldots, d_K)$ is an *achievable distortion po*[illegible] there exists an encoder sequence $\{\alpha_n\}$ [illegible] decoders sequences $\{\beta_n(N_1), \beta_n(N_2), \ldots, \beta_n(N_K)\}$ such that α_n satisfies the power constraint and $D_{ave}(\alpha, \beta(N_k), N_k) = d_k$ for $1 \leq k \leq K$.

In the previous section, we proved the existence of robust code for $\rho = 1$, and conjectured that there does not exist a robust code for $\rho \neq 1$. Thus, if the conjecture is true, then $(D_{opt}(N_1), D_{opt}(N_2), \ldots, D_{opt}(N_K))$ is achievable iff $\rho = 1$. For simplicity, we assume $K = 2$. Let the noise powers be N_1 and N_2 ($N_2 > N_1$). Note that $(D_{opt}(N_1), \sigma^2)$ is achievable by a code optimized for noise N_1 (for noise N_2, set $\beta_n(N_2)(Z) = 0$). Since $N_1 < N_2$, $(D_{opt}(N_2), D_{opt}(N_2))$ is achievable with a code optimized for N_2. Since $(D_{opt}(N_1), \sigma^2)$ and $(D_{opt}(N_2), D_{opt}(N_2))$ are achievable, any point lying on the line segment joining the two points is also achievable by the principle of time sharing. Fig. 3 shows the achievable and unachievable distortion regions of this system. The points A and C in this figure correspond to $(D_{opt}(N_1), \sigma^2)$ and $(D_{opt}(N_2), D_{opt}(N_2))$, respectively. The region lying above $EACD$ is achievable and the region below EBD is not achievable. To the best of our knowledge, the achievability of the triangular region ABC has so far not been presented for $\rho \neq 1$.

In the subsequent subsections, we propose various hybrids of linear and tandem systems. Using hybrid systems, we show that a significant portion of the triangular region ABC can be achieved. As mentioned before, we shall be splitting the bandwidth and the power of the source and channel. For describing these systems, we use the following nomenclature: A system which employs bandwidth splitting in the source and power splitting in the channel is named as, "B/P". A system with both bandwidth splitting and power splitting in the source and only bandwidth splitting in the channel is named as, "BP/B". A system which does not have any kind of splitting in the source and the channel is named as, "N/N". Further, we shall be using either a purely digital system, or a hybrid of digital and analog (linear) system. Both these system can have matched or unmatched tandem encoders. A purely digital system without a matched tandem encoder is represented by DU, i.e., digital unmatched . A hybrid system with matching is represented by HM. With this naming convention, a time sharing system can be said to be a "B/B DU" system.

5.1 System 1: P/P DU

In [21], it is shown that efficient multi-stage descriptions are possible with i.i.d. Gaussian sources. In this system, an efficient two-stage source encoder and a capacity-achieving broadcast channel encoder are combined. The block diagram of the system is shown in Fig.

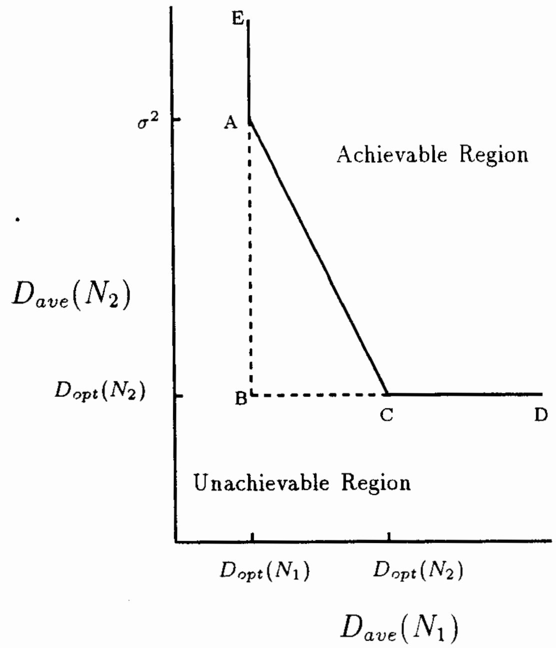

Figure 3: Distortion region for a time sharing digital system.

Tandem Code 1 is designed for $(\sigma^2, aP, (1-a)P + N_2, \rho, \sigma_e^2)$ and Tandem Code 2 is designed for $(\sigma_e^2, (1-a)P, N_1, \rho, D_e(N_1))$. Note that σ_e^2 is the variance of the input of Tandem Encoder 2. ¿From (9), we have

$$R_0 = \lim_{m\to\infty} \frac{1}{m} H(\tilde{X}) = \frac{1}{2}\log\left(1 + \frac{aP}{(1-a)P + N_2}\right), \tag{16}$$

$$R_1 = \lim_{m\to\infty} \frac{1}{m} H(\tilde{e}) = \frac{1}{2}\log\left(1 + \frac{(1-a)P}{N_1}\right), \tag{17}$$

The rate pair (R_0, R_1) lies in the achievable rate region of the above Gaussian broadcast channel [18]. The rate R_0 corresponds to the common information and the rate R_1 corresponds to the information for the higher capacity (less noisy) channel. Thus, there exists channel encoders which can encode the outputs $\tilde{X}$ and $\tilde{e}$ of the source encoders such that both $\tilde{X}$ and $\tilde{e}$ can be decoded correctly (arbitrary low probability of error) when the channel noise is N_1, and $\tilde{X}$ can be decoded correctly when the channel noise is N_2. The channel encoder in Fig. 4 encodes $\tilde{X}$ and $\tilde{e}$ independently. The encoding and decoding procedure for the above channel encoder is described in [18].

The block diagram of the decoder at noise N_1 is shown in Fig. 5. Since $\tilde{X}$ and $\tilde{e}$ can be decoded correctly when the channel noise is N_1, the distortion $D_e(N_1)$ is achievable at N_1. When the channel noise is N_2 only $\tilde{X}$ can be decoded correctly. Hence, σ_e^2 is achievable when the channel noise is N_2. From (12), we have

$$D_{ave}(N_1) = \frac{\sigma^2}{\left\{\left(1+\frac{aP}{(1-a)P+N_2}\right)\left(1+\frac{(1-a)P}{N_1}\right)\right\}^{\rho}}, \tag{18}$$

$$D_{ave}(N_2) = \frac{\sigma^2}{\left(1+\frac{aP}{(1-a)P+N_2}\right)^{\rho}}. \tag{19}$$

Note that $D_{ave}(N_1)$ and $D_{ave}(N_2)$ are equal to the distortion values in the distortion-rate function of an i.i.d. Gaussian source at rates $\rho(R_0+R_1)$ and ρR_0, respectively. ¿From the capacity region of a Gaussian broadcast channel [18, 22], (R_0, R_1) is the best achievable rate pair. Thus, any system based on the separation of source and channel coding can not achieve distortion pair better than (18) and (19). The achievable distortion regions for P/P DU systems are shown in Fig. 15, 16 and 17 for $\sigma^2 = 1$ and various values of ρ, P/N_1 and P/N_2. Note that the achievable region of this system is below the time sharing system, i.e., below the line AC in Fig. 3. This system does not achieve any point on the lines AB or BC. In the next two subsections we will show how points on the lines AB and BC can be achieved.

5.2 System 2: P/B HU

Consider the system in Fig. 6. As the name suggest, we first split the channel into two sub-channels. The first channel, referred to as the primary channel, occupies $(\rho-1)W_c/\rho$ Hz of bandwidth. The secondary channel occupies W_c/ρ Hz of bandwidth. We call this "channel bandwidth splitting". Now, a traditional tandem coding system is used in the primary channel and a linear encoder is used in the secondary channel. The encoder for the primary channel is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^{m-n}$. The tandem code is designed for $(\sigma^2, P, N_2, \rho-1, \sigma_e^2)$. Note that this system is valid only for $\rho > 1$.

The source encoder splits X into two components, the encoders output $\tilde{X}$ of variance $\sigma^2-\sigma_e^2$ and the error vector e of variance σ_e^2. We call this "source power splitting". A linear encoder with parameters (σ_e^2, P) transmits the error vector e on the secondary channel. Note that the secondary channel has bandwidth ratio of 1.

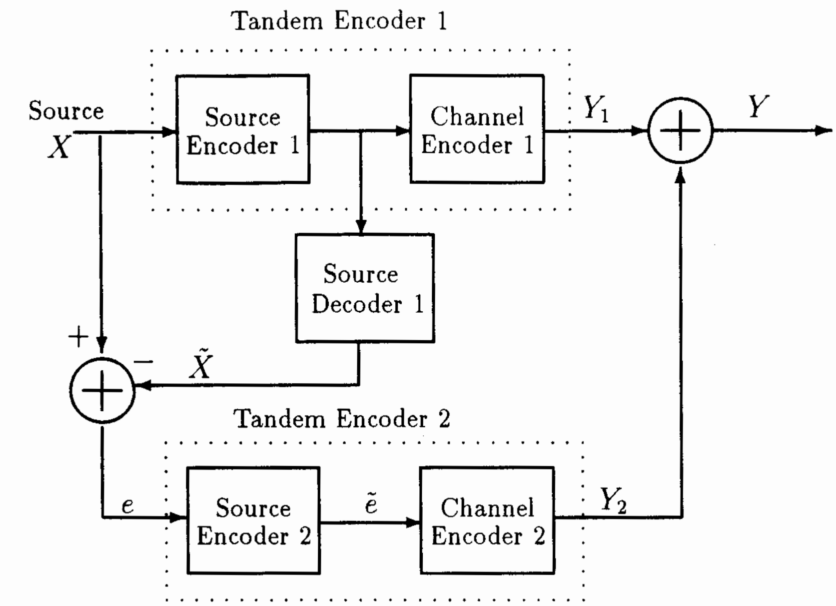

Figure 4: Encoder for System 1: P/P DU.

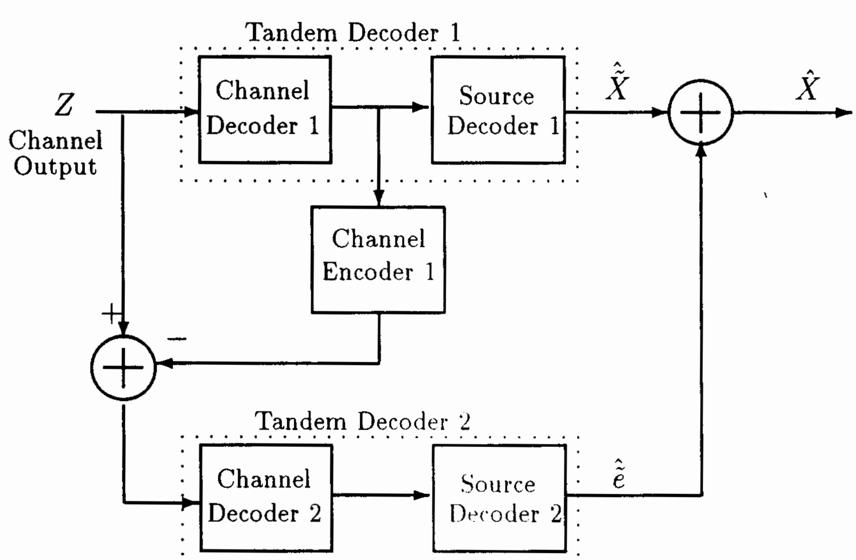

Figure 5: Decoder for System 1: P/P DU for Noise N_1.

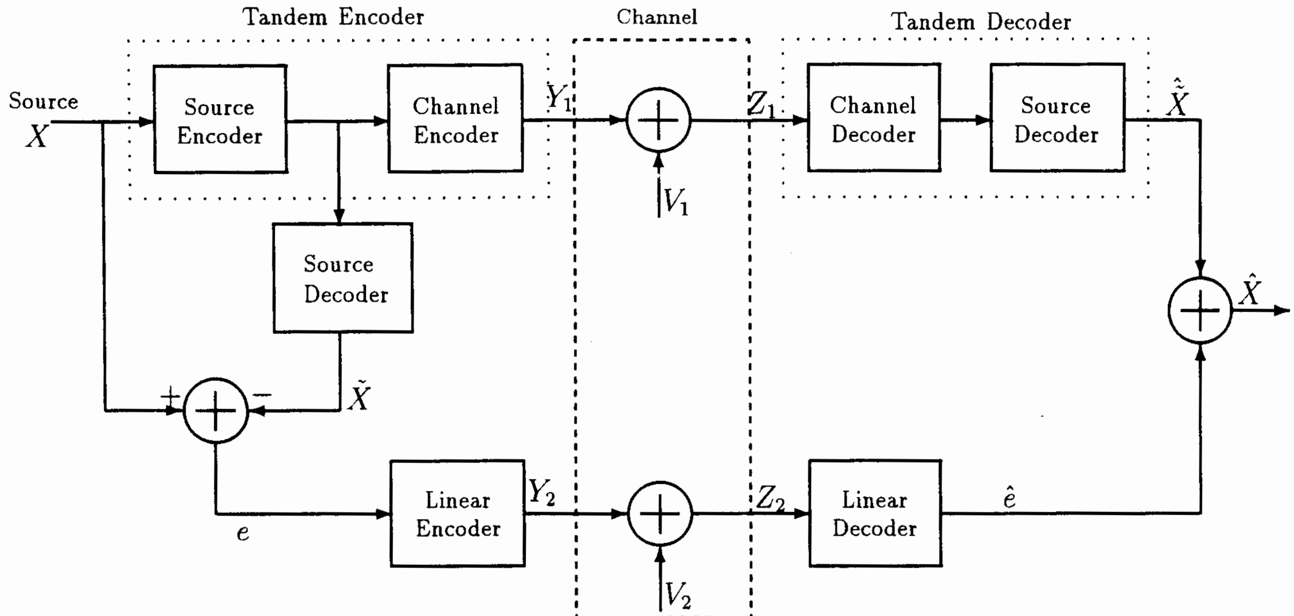

Figure 6: Block Diagram of System 2: P/B HU.

The decoder for the above hybrid encoder when the channel noise $N \leq N_2$ is shown in Fig. 6. The tandem decoder for the primary channel decodes $\tilde{X}$ correctly, i.e.,

$$\lim_{n\to\infty} \Pr(\hat{\tilde{X}} \neq \tilde{X}) = 0. \tag{20}$$

The linear decoder with parameters (σ_e^2, P, N) obtains a linear estimate of e. The estimate $\hat{X}$ of X is the sum of tandem decoder's and linear decoder's outputs, i.e.,

$$\hat{X} = \hat{\tilde{X}} + \hat{e}. \tag{21}$$

The average distortion $D_{ave}(N)$ is given by

$$D_{ave}(N) = \lim_{n\to\infty} \frac{1}{n} E\|X - \hat{X}\|^2 \tag{22}$$

$$= \lim_{n\to\infty} \frac{1}{n} E(\text{[illegible]} - (\hat{\tilde{X}} + \hat{e})\|^2) \tag{23}$$

$$= \lim_{n\to\infty} \frac{1}{n} E(\text{[illegible]} X\|^2) + \frac{1}{n} E(\|e - \hat{e}\|^2) \tag{24}$$

$$= \lim_{n\to\infty} \frac{1}{n} E\text{[illegible]} \tag{25}$$

The equality (24) follows from orthogonality of e and $\tilde{X}$ (Lemma 1), and (25) follows from (20). Note that $D_{ave}(N)$ is equal to the mean square distortion of e, which is equal to the optimal distortion for a Gaussian source of variance σ_e^2 transmitted on a AWGN channel with noise power N, power constraint P and bandwidth ratio 1. Thus

$$D_{ave}(N) = \frac{\sigma_e^2}{1+\frac{P}{N}}. \tag{26}$$

and from (12),

$$D_{ave}(N) = \frac{\sigma^2}{(1+\frac{P}{N})(1+\frac{P}{N_2})^{\rho-1}}. \tag{27}$$

Note that for $N = N_2$ the above coding system achieves $D_{opt}(N_2)$. Further, for $N_1 < N_2$ its performance is better than $D_{opt}(N_2)$ by a factor of $(1 + P/N_1)/(1 + P/N_2)$. Note that this factor is greater than 1. Thus, the encoder achieves a point on the line BC in Fig. 3. The achievable points of System 2 are shown in Fig. 15 and 16. Note that in Fig. 15, $D_{ave}(N_1)$ for System 1 is 17 dB (Gain) less than that of time sharing system provided $D_{ave}(N_2) = D_{opt}(N_2)$ for both the systems. In Fig. 16, this gain is 2.5 dB.

The main advantage of this system is that the tandem encoder of this system can use standard source compression methods as its source encoder and standard channel encoding techniques as its channel encoder. For example, the code-excited linear predictive (CELP) coder can be used for speech compression, and the joint photographic expert group (JPEG) algorithm can be used for image compression. Similarly, standard error control coding techniques such as Reed-Solomon codes, BCH codes, convolution codes, trellis coded modulation and turbo codes can be used as channel encoders.

5.3 System 3: N/N DM

In the previous subsection, we showed that a hybrid of a linear and a tandem system achieves a point on the line BC in Fig. 3. We now show that a matched tandem system can achieve a point on the line AB. Consider a matched tandem encoder with parameters $(\sigma^2, P, N_1, \rho, D_{opt}(N_1))$. As in Section 4.1, let $X = \tilde{X} + e$. Since the matched tandem encoder is optimized for $D_{ave}(N_1) = D_{opt}(N_1)$. When the channel noise is N_2, the use of matched encoder enables the estimation of the first $l = \min(\rho, 1)n$ components of $\tilde{X}$. Let $\tilde{X}_1^l$ be the first l components

$\tilde{X}$, Z_1^l be the first l components of the decoder input Z, and $\hat{\tilde{X}}_1^l$ be an estimate of $\tilde{X}_1^l$. Note that the variance of $\tilde{X}$ is $\sigma^2 - D_{opt}(N_1)$. Thus,

$$\hat{\tilde{X}}_1^l = \frac{\sqrt{(\sigma^2 - D_{opt}(N_1))P}}{P + N_2} Z_1^l. \tag{28}$$

If $n > m$ ($\rho < 1$) then the components $\tilde{X}_{l+1}^n$ are estimated as zero. Since $\tilde{X}$ and e are orthogonal, the distortion in estimating X is the sum of the distortion in estimating $\tilde{X}$ and the variance of e.

The distortion in estimating the first l components of $\tilde{X}$ is $(\sigma^2 - D_{opt}(N_1))/(1 + P/N_2)$ and the distortion in estimating the rest of the components is $\sigma^2 - D_{opt}(N_1)$. Thus, the mean distortion in estimating $\tilde{X}$ is

$$D_{\tilde{X}}(N_2) = \min(\rho, 1)\frac{\sigma^2 - D_{opt}(N_1)}{1 + P/N_2} + \max(1-\rho, 0)(\sigma^2 - D_{opt}(N_1)), \tag{29}$$

and hence

$$D_{ave}(N_2) = \min(\rho, 1)\frac{\sigma^2 - D_{opt}(N_1)}{1 + P/N_2} + \max(1-\rho, 0)(\sigma^2 - D_{opt}(N_1)) + D_{opt}(N_1). \tag{30}$$

Note that $(D_{ave}(N_1), D_{ave}(N_2))$ lies on the line AB in Fig. 3. The performance of this system at various values of ρ, N_1 and N_2 is shown in Fig. 15, 16 and 17.

5.4 System 4: P/P DM

In this section, we show the improvement in achievable distortion performance by using a matched tandem encoder instead of a standard tandem encoder in System 1. Let Tandem Encoder 2 in Fig. 4 be matched. The decoder when the channel noise is N_1 is same as in Fig. 5. Thus, the distortion when the channel noise is N_1 is same as in (18). When the channel noise is N_2, a linear decoder with parameters $(\sigma_e^2 - D_{ave}(N_1), (1-a)P, N_2)$ obtains a linear estimate of the first $\min(n\rho, n)$ components of $\tilde{e}$. The distortion for noise N_2 is,

$$D_{ave}(N_2) = D_{ave}(N_1) + D_{\tilde{e}}(N_2). \tag{31}$$

Note that,

$$D_{\tilde{e}}(N_2) = \min(\rho, 1)\frac{(\sigma_e^2 - D_{ave}(N_1))}{1 + \frac{(1-a)P}{N_2}} + \max(1-\rho, 0)(\sigma_e^2 - D_{ave}(N_1)), \tag{32}$$

thus

$$D_{ave}(N_2) = D_{ave}(N_1) + \min(\rho, 1)\frac{(\sigma_e^2 - D_{ave}(N_1))}{1 + \frac{(1-a)P}{N_2}} + \max(1-\rho, 0)(\sigma_e^2 - D_{ave}(N_1)), \tag{33}$$

where

$$\sigma_e^2 = \frac{\sigma^2}{\left(1 + \frac{aP}{(1-a)P+N_2}\right)^{\rho}}. \tag{34}$$

The performance of this system at various values of ρ, N_1 and N_2 is shown in Fig. 15, 16 and 17. Note that, in all the cases this system performs better than its unmatched version, i.e., P/P DU.

5.5 System 5: P/PB HM

In Section 5.2 we showed that System 2 achieves a point on the line AB in Fig. 3 and in the previous subsection we showed that by using matched encoder we can improve the achievable distortion performance. In this section, we combine these two systems. We first split the channel into two sub-channels. The primary channel occupies $(\rho - 1)W_c/\rho$ Hz of bandwidth and the secondary channel occupies W_c/ρ Hz of bandwidth. The block diagram of the encoder is shown in Fig. 7. Tandem Encoder 1 is designed for $(\sigma^2, aP, (1-a)P + N_2, \rho - 1, \sigma_e^2)$. Tandem Encoder 2 is a matched encoder and is designed for $(\sigma_e^2, (1-a)P, N_1, \rho - 1, \sigma_{e_1}^2)$. Note that σ_e^2 is the variance of e and $\sigma_{e_1}^2$ is the variance of e_1. The outputs of Tandem Encoder 1 and Tandem Encoder 2 are added and transmitted on the primary channel. The linear encoder with parameters $(\sigma_{e_1}^2, P)$ transmits e_1 on the secondary channel. Note that System 2 is a special case of this system and can be obtained by fixing $a = 1$. The decoder when the channel noise is N_1 is shown in Fig. 8. The linear decoder has parameters $(\sigma_{e_1}^2, P, N_1)$. The distortion is given by

$$D_{ave}(N_1) = \frac{\sigma^2}{\left\{\left(1 + \frac{aP}{(1-a)P+N_2}\right)\left(1 + \frac{(1-a)P}{N_1}\right)\right\}^{\rho-1}\left(1 + \frac{P}{N_1}\right)}. \tag{35}$$

The decoder when the channel noise is N_2 is shown in Fig. 9. Since $\tilde{X}$, $\tilde{e}$ and e_1 are mutually asymptotically orthogonal,

$$D_{ave}(N_2) = D_{\tilde{X}}(N_2) + D_{\tilde{e}}(N_2) + D_{e_1}(N_2). \tag{36}$$

Since

$$\lim_{n\to\infty} \Pr(\tilde{X} \neq \hat{\tilde{X}}) = 0, \tag{37}$$

$D_{\tilde{X}}(N_2) = 0.$

The first linear decoder in Fig. 9 has parameters $(\sigma_e^2 - \sigma_{e_1}^2, (1-a)P, N_2)$ and estimates the first $l = \min((\rho-1)n, n)$ components of $\tilde{e}$. The other components are estimated as 0. Thus,

$$D_{\tilde{e}}(N_2) = \min(\rho-1, 1)\frac{\sigma_e^2 - \sigma_{e_1}^2}{1 + \frac{(1-a)P}{N_2}} + \max(\rho-2, 0)(\sigma_e^2 - \sigma_{e_1}^2). \tag{38}$$

The second linear decoder with parameter $(\sigma_{e_1}^2, P, N_2)$ estimates e_1. Thus,

$$D_{e_1}(N_2) = \frac{\sigma_{e_1}^2}{1 + \frac{P}{N_2}}, \tag{39}$$

and

$$D_{ave}(N_2) = D_{\tilde{e}}(N_2) + D_{e_1}(N_2). \tag{40}$$

5.6 System 6: PB/PB HM

In the design of Systems 1, 4 and 5, we have only one free parameter: a. The achievable distortion performance was obtained by varying that free parameter. In this, system we generalize most of the above systems by having two free parameters (a and b) instead of one. The block diagram of this system is shown in Fig. 10. We split the channel into two sub channels. The primary channel occupies $(\rho - b)W_c/\rho$ Hz of bandwidth and the secondary channel occupies bW_c/ρ Hz of bandwidth. Tandem Encoder 1, designed for $(\sigma^2, aP, (1-a)P + N_2, \rho - b, \sigma_e^2)$, splits X into $\tilde{X}$ and e. Now, the n-dimensional vector e is split into two subvectors e_1 and e_2 of dimensions $(1-b)n$ and bn (bandwidth splitting), respectively. A matched tandem encoder (Tandem Encoder 2) designed for $(\sigma_e^2, (1-a)P, N_1, (\rho-b)/(1-b), D_{e_1}(N_1))$ encodes e_1. Note that $b \leq \min(\rho, 1)$. The outputs of Tandem Encoder 1 and Tandem Encoder 2 are added and transmitted on the primary channel. A linear encoder with parameters (σ_e^2, P) transmits e_2 on the secondary channel.

The block diagram of the decoder when the channel noise is N_1 is shown in Fig.11. Tandem Decoder 1 and Tandem Decoder 2 decode $\tilde{X}$ and $\tilde{e}_1$, respectively. A linear decoder with

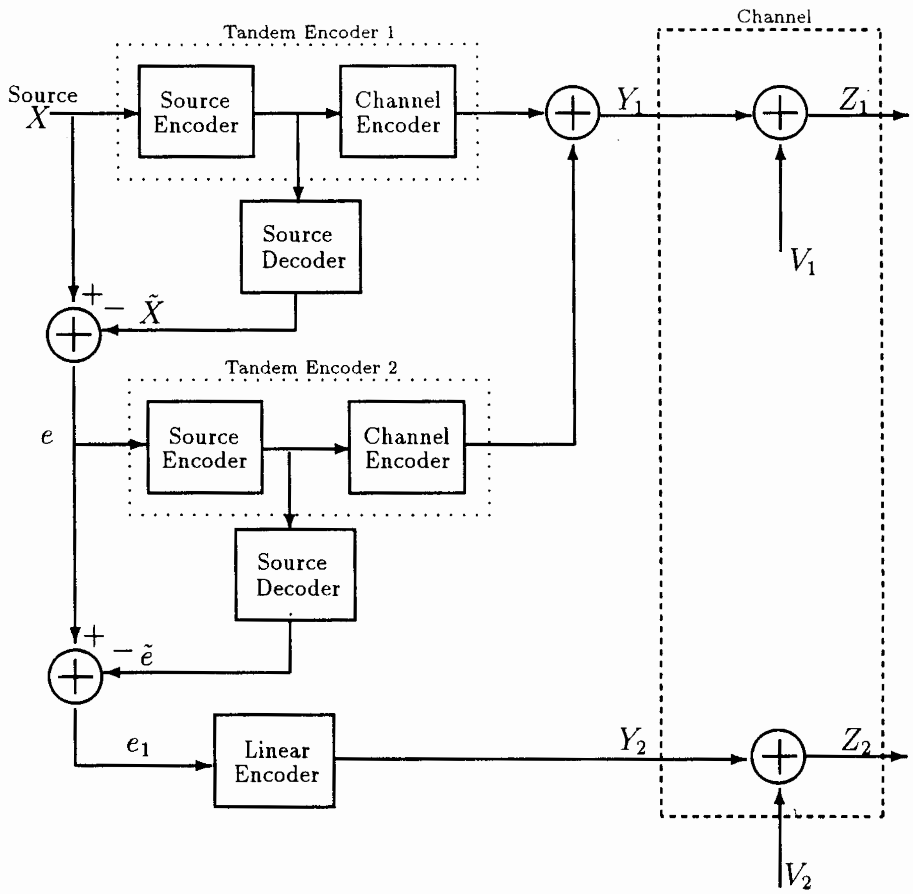

Figure 7: Block Diagram of the Encoder of System 5: P/PB HM.

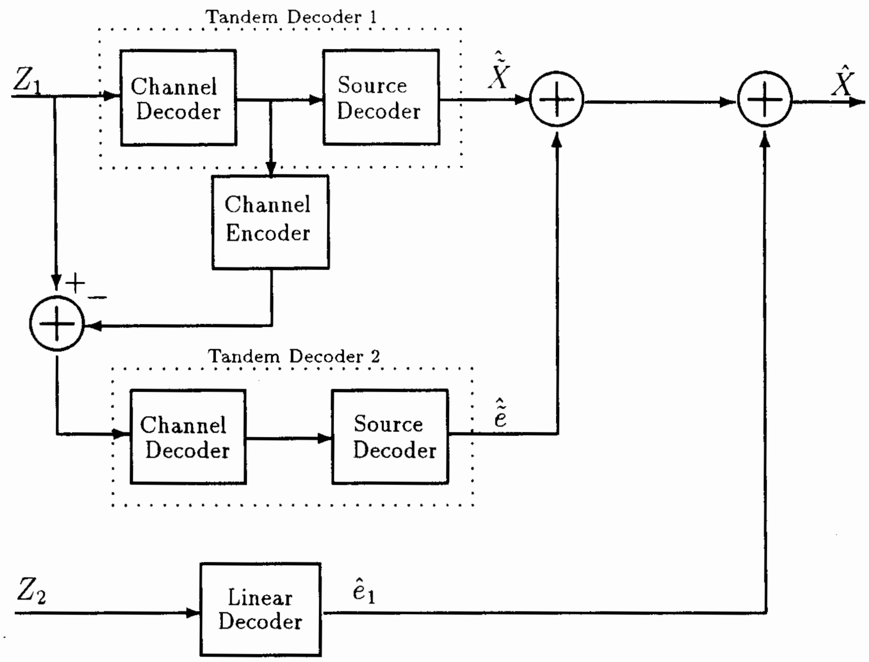

Figure 8: Block Diagram of the Decoder of System 5: P/PB HM when the Channel Noise is N_1.

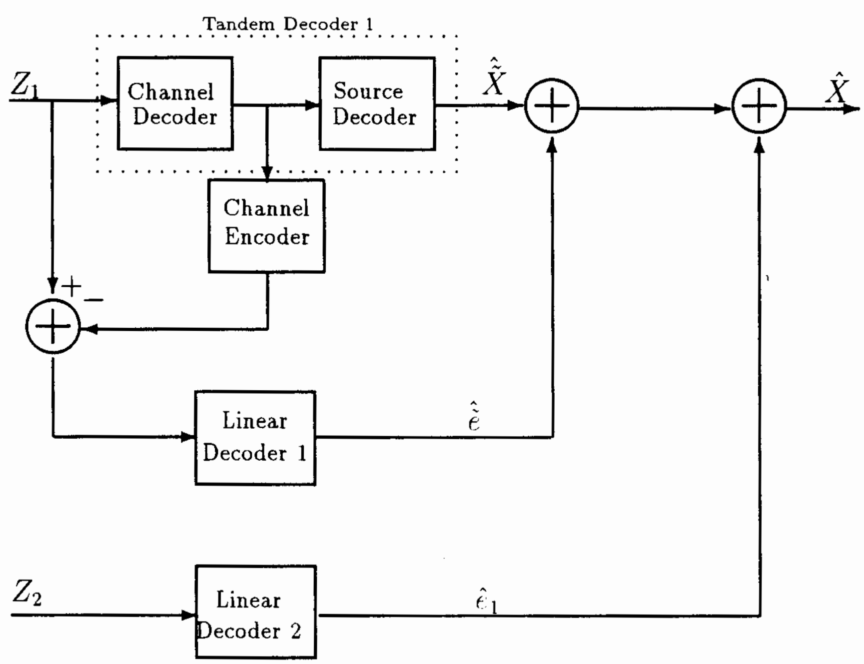

Figure 9: Block Diagram of the Decoder of System 5: P/PB HM when the Channel No N_2

parameters (σ_e^2, P, N_1) obtains a linear estimate of e_2. Now,

$$\begin{aligned} D_{ave}(N_1) &= D_{\tilde{X}}(N_1) + D_e(N_1) & (41) \\ &= (1-b)D_{e_1}(N_1) + bD_{e_2}(N_1), & (42) \end{aligned}$$

where (42) follows from the fact that $P(\hat{\tilde{X}} \neq \tilde{X}) = 0$. ¿From (12),

$$D_{e_1}(N_1) = \frac{\sigma^2}{\left(1 + \frac{aP}{(1-a)P+N_2}\right)^{\rho-b} \left(1 + \frac{(1-a)P}{N_1}\right)^{\frac{\rho-b}{1-b}}}. \tag{43}$$

$D_{e_2}(N)$ is equal to the optimal distortion when an i.i.d. Gaussian source of variance σ_e^2 is transmitted on a rate-one channel with noise variance N ($N = N_1$ or N_2) and power constraint P. Thus,

$$D_{e_2}(N) = \frac{\sigma^2}{\left(1 + \frac{aP}{(1-a)P+N_2}\right)^{\rho-b} \left(1 + \frac{P}{N}\right)}. \tag{44}$$

The block diagram of the decoder when the channel noise is N_2 is shown in Fig. 12. Here, Tandem Decoder 2 is replaced by a linear decoder with parameters $(\sigma_e^2 - D_{e_1}(N_1), (1-a)P, N_2)$ which estimates the first $\min(n, (\rho-b)n/(1-b))$ components of $\tilde{e}_1$. A linear decoder with parameters (σ_e^2, P, N_2) estimates e_2. Now,

$$D_{ave}(N_2) = (1-b)D_{e_1}(N_2) + bD_{e_2}(N_2). \tag{45}$$

Note that,

$$D_{e_1}(N_2) = D_{e_1}(N_1) + D_{\tilde{e}_1}(N_2). \tag{46}$$

It can be easily verified that

$$D_{\tilde{e}_1}(N_2) = \min(1, \frac{\rho-b}{1-b}) \frac{\sigma_e^2 - D_{e_1}(N_1)}{1 + \frac{(1-a)P}{N_2}} + \max(1 - \frac{\rho-b}{1-b}, 0)(\sigma_e^2 - D_{e_1}(N_1)). \tag{47}$$

Substituting the values of σ_e^2, $D_{e_1}(N_2)$, and $D_{e_2}(N_2)$ in (42) we obtain $D_{ave}(N_2)$. Thus, $D_{ave}(N_1)$ and $D_{ave}(N_2)$ are functions of a and b for given values of σ^2, P, N_1, N_2, and ρ. The best achievable region of System 6 can be obtained by solving the constrained optimization problem

$$\min_{a,b} D_{ave}(N_2). \tag{48}$$

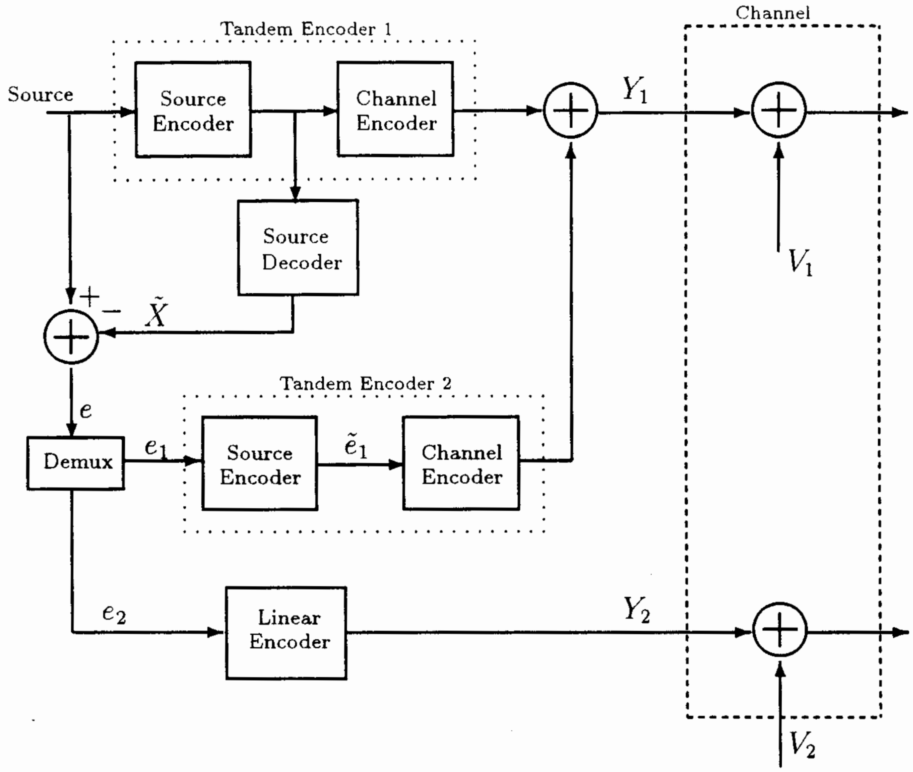

Figure 10: Block Diagram of the Encoder of System 6: PB/PB HM.

subject to:

$$D_{ave}(N_1) \leq D, \tag{49}$$

for various values of D.

Note that this system is a generalization of System 2, System 3 and System 4. These systems are obtained by fixing $(a = 1, b = 1)$, $(a = 1, b = 0)$ and $b = 0$, respectively.

5.7 System 7: BP/BP HM

The systems discussed thus far perform well for $\rho > 1$. A system is said to be a dual of another system if its encoder is "similar" to the decoder at noise N_1 of the other system [24]. The duals of System 1 to 6 perform well for $\rho < 1$. Here, we consider the dual of System 6. The name of this system is BP/BP instead of PB/PB because the bandwidth splitting of the source and the channel are done before the power splitting. The block diagram of the encoder is shown in Fig. 13. Note that the encoder here is similar to the decoder in Fig. 11. The source X is split (bandwidth splitting) into two constituents X_1 and X_2 of

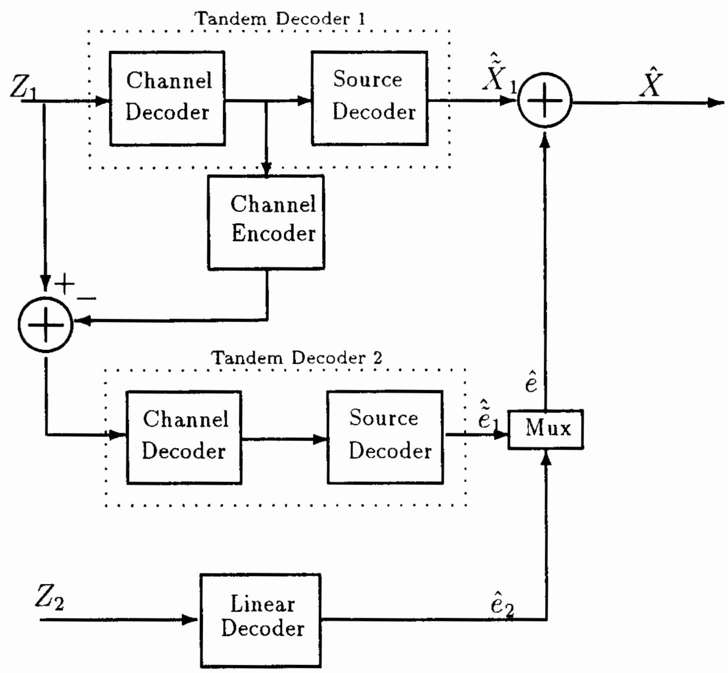

Figure 11: Block Diagram of the Decoder of System 6: PB/PB HM when the Channel Noise is N_1.

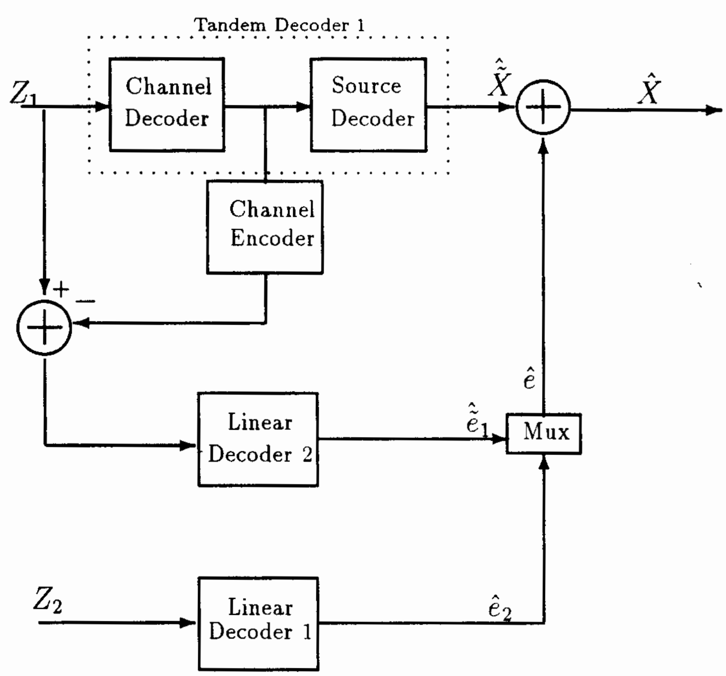

Figure 12: Block Diagram of the Decoder of System 6: PB/PB HM when the Channel Noise is N_2

bandwidths $(1-b\rho)W_s$ and $b\rho W_s$ ($b \leq \min(1, 1/\rho)$), respectively. Tandem Encoder 1 designed for $(\sigma^2, aP, N_2 + (1-a)P, \rho/(1-b\rho), \sigma_e^2)$ splits X_1 into $\tilde{X}_1$ and e (power splitting). A matched tandem encoder (Tandem Encoder 2) designed for $(\sigma_e^2, (1-a)P, N_1, (\rho - b\rho)/(1-b\rho), D_e(N_1))$ encodes e. A linear encoder with parameters $(\sigma^2, (1-a)P)$ encodes X_2. The outputs of the linear encoder and Tandem Encoder 2 are multiplexed and added to the output of Tandem Encoder 1. Note that the bandwidth of the multiplexed output is the same as the bandwidth of the output of Tandem Encoder 1. The block diagram of the decoder when the channel noise is N_1 is shown in Fig. 14. Tandem Decoder 1 and Tandem Decoder 2 decodes $\check{X}_1$ and $\tilde{e}$, respectively. A linear decoder with parameters $(\sigma^2, (1-a)P, N_1)$ finds a linear estimate of X_2. The distortion (for general N) is given by

$$D_{ave}(N) = (1-b\rho)D_{X_1}(N) + b\rho D_{X_2}(N). \tag{50}$$

Since $P(\hat{\tilde{X}}_1 \neq \tilde{X}_1) = 0$, $D_{X_1}(N) = D_e(N)$. From (12),

$$D_e(N_1) = \frac{\sigma^2}{\left(1 + \frac{aP}{(1-a)P+N_2}\right)^{\frac{\rho}{1-b\rho}} \left(1 + \frac{(1-a)P}{N_1}\right)^{\frac{\rho - b\rho}{1-b\rho}}}. \tag{51}$$

Since the input to the linear decoder is equal to the sum of linear encoders output and the channel noise (with large probability),

$$D_{X_2}(N_1) = \frac{\sigma^2}{1 + \frac{(1-a)P}{N_1}}. \tag{52}$$

When the channel noise is N_2, a linear decoder with parameter $(\sigma_e^2, (1-a)P, N_2)$ replaces Tandem Decoder 2. This decoder finds a linear estimate of $\tilde{e}$. Using the same approach as in previous sections, we get

$$D_{X_1}(N_2) = D_e(N_1) + \min\left(1, \frac{\rho - b\rho}{1-b\rho}\right)\frac{\sigma_e^2 - D_e(N_1)}{1 + \frac{(1-a)P}{N_2}} + \max\left(1 - \frac{\rho - b\rho}{1-b\rho}, 0\right)(\sigma_e^2 - D_e(N_1)). \tag{53}$$

and

$$D_{X_2}(N_2) = \frac{\sigma^2}{1 + \frac{(1-a)P}{N_2}}. \tag{54}$$

The achievable region now can be obtained by solving a constrained optimization problem similar to the one in Section 5.6.

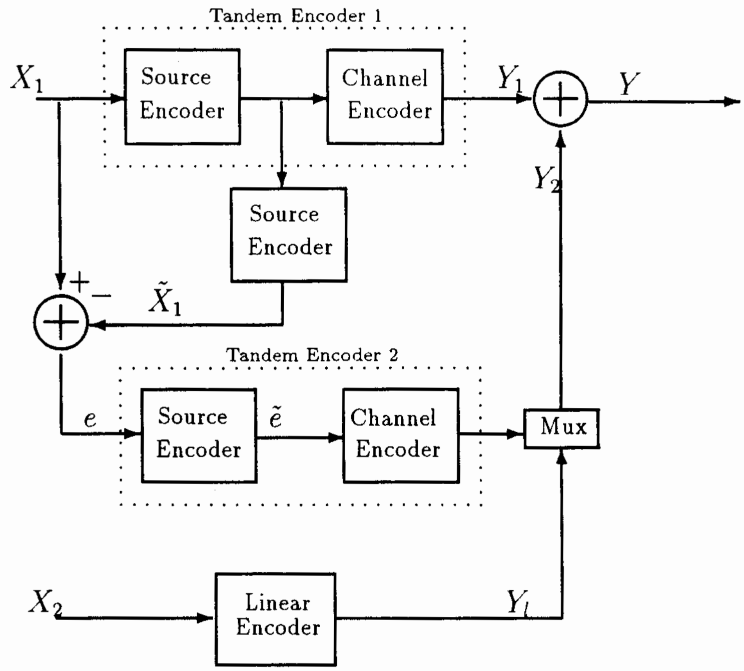

Figure 13: Block Diagram of the Encoder of System 7: BP/BP HM.

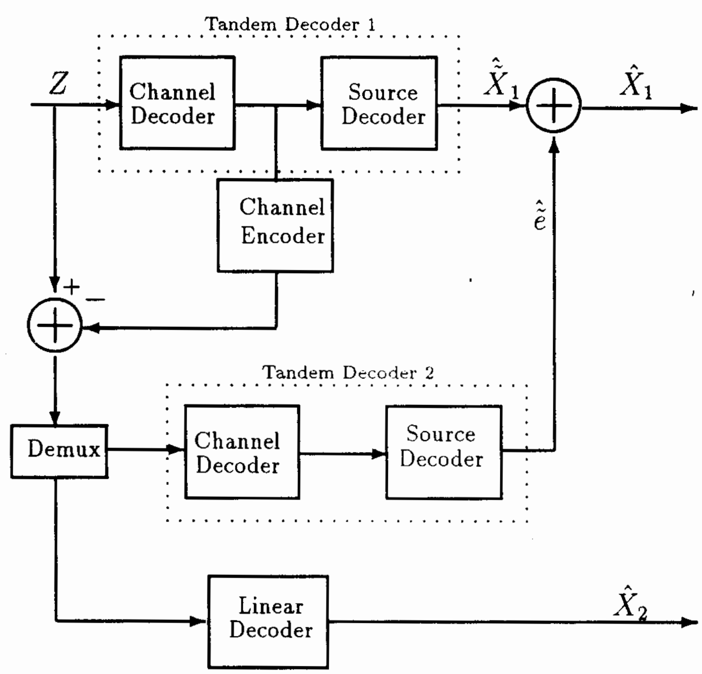

Figure 14: Block Diagram of the Decoder of System 7: BP/BP HM when the Channel N is N_1.

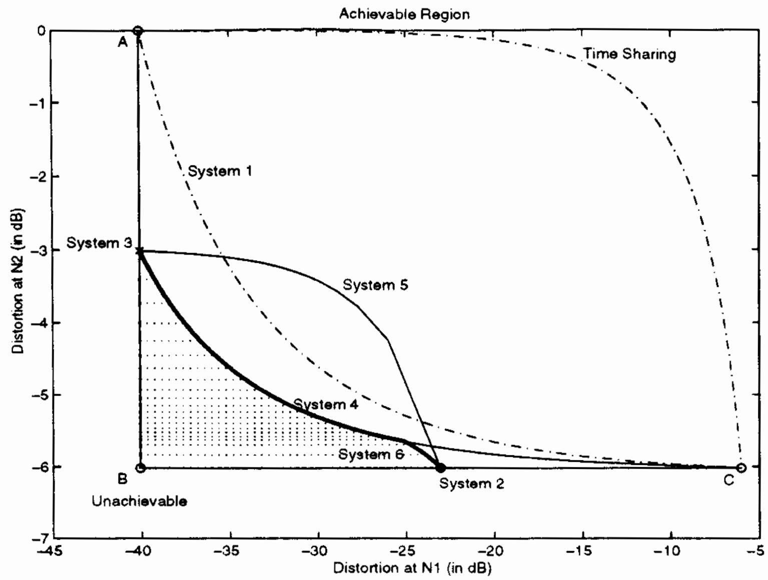

Figure 15: Achievable Region of Various Systems at $\rho = 2$, $SNR_1 = 20$ dB and $SNR_2 = 0$ dB.

6 Comparison of Performance of Various Systems

In Fig. 15, 16 and 17, we compare the performance of the systems in Section 5. Fig. 15 corresponds to $\rho = 2$, $N_1 = P/100$ and $N_2 = P$, Fig. 16 corresponds to $\rho = 1.5$, $N_1 = P/2.5$ and $N_2 = P$, and Fig. 17 corresponds to $\rho = 0.5$, $N_1 = P/100$ and $N_2 = P$. In these figures, the distortions $(D_{ave}(N_1), D_{ave}(N_2))$ are plotted in logarithmic scale. Thus the time-sharing curve is no longer a straight line. Note that System 1 (P/P DU) outperforms the time sharing system (B/B DU). Further, the matched version of System 1, i.e., System 4 (P/P DM) outperforms System 1. Since System 4 is a special case of System 6, its performance is worse than System 6. In Fig. 15, the performance of System 6 is better than the performance of System 5, while in Fig. 16, System 5 outperforms System 6 for smaller values of $D_{ave}(N_2)$. For $\rho < 1$, the dual of System 6, i.e., System 7 outperforms all the other systems in Fig. 17.

7 Nearly-Robust Code

A code is said to be nearly robust if it does not suffer from the (severe form of the) threshold effect.

Definition 3 *A code sequence $\{(\alpha_n, \beta_n)\}$ is said to be nearly-robust on $\aleph$ at N, if it satisfies the power constraint and*

$$D_{ave}(N) = D_{opt}(N), \tag{55}$$

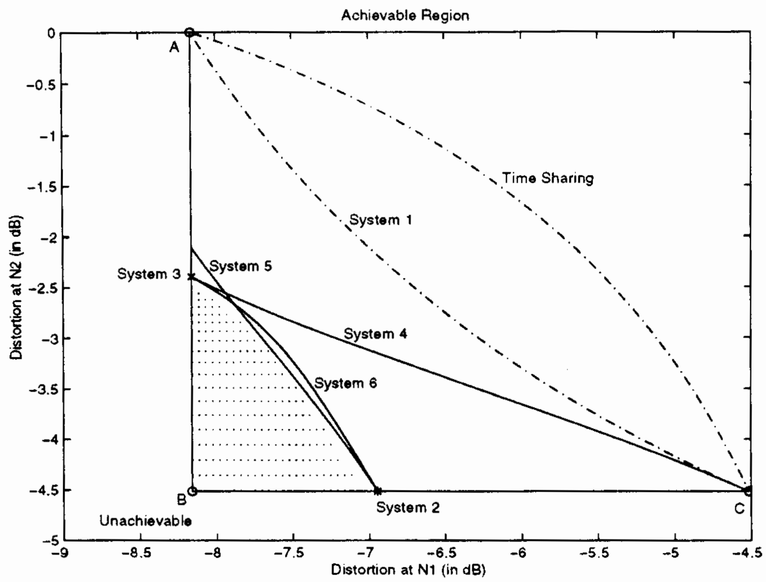

Figure 16: Achievable Region of Various Systems at $\rho = 1.5$, $SNR_1 = 4$ dB and $SNR_2 = 0$ dB.

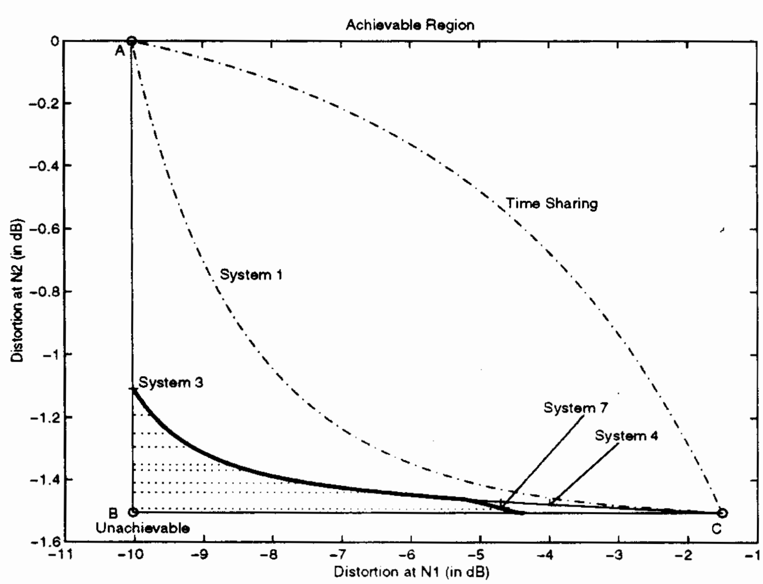

Figure 17: Achievable Region of Various Systems at $\rho = 0.5$, $SNR_1 = 20$ dB and $SNR_2 = 0$ dB.

$$D_{ave}(N_i) < D_{opt}(N) \tag{56}$$

$\forall\ N_i \in \aleph$ which are less than N and

$$D_{ave}(N_i) < \sigma^2 \tag{57}$$

$\forall\ N_i \in \aleph$ which are greater than N. Further, if $N_i < N_j$ then

$$D_{ave}(N_i) < D_{ave}(N_j), \forall N_i, N_j \in \aleph. \tag{58}$$

Note that if a code is robust on $\aleph$ then it is nearly-robust on $\aleph$ at all $N_i \in \aleph$. A code sequence is said to be nearly-robust at N, if it is nearly robust at N on all finite $\aleph$.

Consider an encoder of System 2. Let the tandem encoder be matched (P/B HM). We will show that the above code is nearly-robust at N_2. Note that, for $N < N_2$, from (27)

$$D_{ave}(N) = \frac{\sigma^2}{\left(1 + \frac{P}{N}\right)\left(1 + \frac{P}{N_2}\right)^{\rho-1}}. \tag{59}$$

Whenever $N > N_2$, the use of matched encoder facilitates linear estimation of $\tilde{X}_1$. Proceeding as in Section 5, we have

$$D_{\tilde{X}_1}(N) = \min(\rho - 1, 1)\frac{\sigma^2 - \sigma_e^2}{1 + \frac{P}{N}} + \max(\rho - 2, 0)(\sigma^2 - \sigma_e^2), \tag{60}$$

and

$$D_e(N) = \frac{\sigma_e^2}{1 + \frac{P}{N}}, \tag{61}$$

where

$$\sigma_e^2 = \frac{\sigma^2}{\left(1 + \frac{P}{N_2}\right)^{\rho-1}}. \tag{62}$$

The net distortion

$$D_{ave}(N) = D_{\tilde{X}_1}(N) + D_e(N). \tag{63}$$

It can be easily verified that this code is nearly-robust at N_2. The above result holds provided the decoder has a prior knowledge about the noise power N. Since the set $\aleph$ is finite, noise power can be estimated (decoded) with probability one. Note that System 3 and System 4 are not nearly-robust at N_2. Fig. 18 shows the near robust performance of the above mentioned system for $\rho = 2$. Note that the system is nearly-robust at $N = P/10$.

Now consider System 7 and $\rho < 1$. Fix $b = 1$ and

$$a = \frac{N_2}{P}\left(1 + \frac{P}{N_2}\right)^{\rho}\left[\left(1 + \frac{P}{N_2}\right)^{1-\rho} - 1\right]. \tag{64}$$

Since $\rho < 1$, $0 < a < 1$. Let Tandem Encoder 1 be a matched encoder designed for $(\sigma^2, aP, (1-a)P + N_2, \rho/(1-\rho), \sigma_e^2)$. Since $b = 1$, the rate is 0 for Tandem Encoder 2. Thus, Tandem Encoder 2 can be removed from System 7 whenever $b = 1$. Note that this system is a dual of System 2. From (50), (53), (54) and (64), we have

$$D_{X_2}(N_2) = D_{X_1}(N_2) = \frac{\sigma^2}{\left(1 + \frac{P}{N_2}\right)^{\rho}}, \tag{65}$$

and

$$D_{ave}(N_2) = D_{opt}(N_2) = \frac{\sigma^2}{\left(1 + \frac{P}{N_2}\right)^{\rho}}. \tag{66}$$

Further, if the noise power $N \leq N_2$,

$$D_{ave}(N) = (1-\rho)D_{opt}(N_2) + \rho\frac{\sigma^2}{1 + \frac{(1-a)P}{N}}. \tag{67}$$

Since $\sigma^2/(1+(1-a)P/N_2)$ is equal to $D_{opt}(N_2)$, $\sigma^2/(1+(1-a)P/N) < D_{opt}(N_2)$. This implies that $D_{ave}(N) < D_{opt}(N_2)$, for $N < N_2$.

For $N > N_2$, Tandem Decoder 1 is replaced by a linear decoder with parameters $(\sigma^2 - D_{opt}(N_2), aP, (1-a)P + N)$. Thus,

$$D_{X_1}(N) = \min(1, \frac{\rho}{1-\rho})\left[D_{opt}(N_2) + \frac{\sigma^2 - D_{opt}(N_2)}{1 + \frac{aP}{(1-a)P+N}}\right] + \max(1 - \frac{\rho}{1-\rho}, 0)\sigma^2. \tag{68}$$

Since $N > N_2$, X_1 can not be decoded correctly. Now the linear decoder which estimates X_2 has parameters $(\sigma^2, (1-a)P, aP+N)$. Note that in this case the output of Channel Encoder act as noise for the linear decoder. Thus,

$$D_{X_2}(N) = \frac{\sigma^2}{1 + \frac{(1-a)P}{aP+N}}. \tag{69}$$

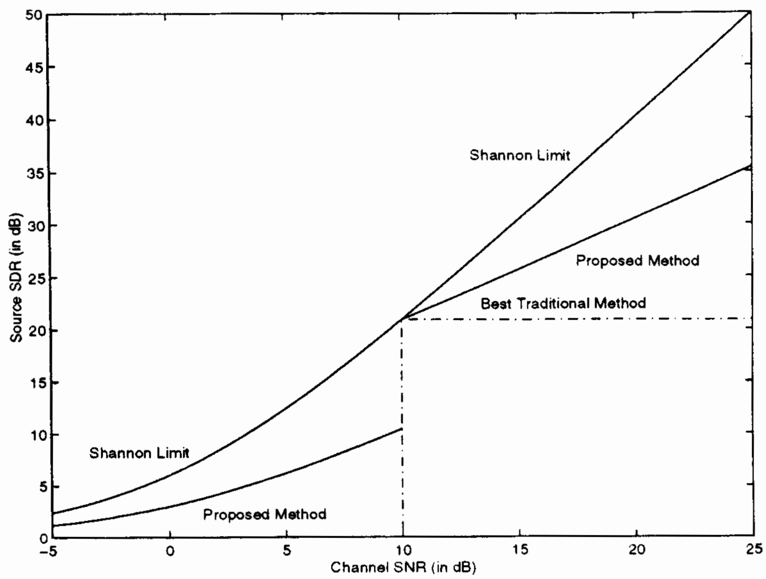

Figure 18: Near Robust Performance of System: P/B HM for $\rho = 2$.

Since

$$D_{ave}(N) = (1-\rho)D_{X_1}(N) + \rho D_{X_2}(N), \tag{70}$$

and $D_{X_1}(N), D_{X_2}(N) < \sigma^2$, $D_{ave}(N) < \sigma^2$. All the properties of nearly-robust code at N_2 are satisfied by this code. Fig. 19 shows the near robust performance of the above mentioned system for $\rho = 1/2$. Note that the system is nearly-robust at $N = P/4$. Furthermore, the best traditional digital coding method suffers from the severe form of the threshold effect, while the threshold effect of the proposed system is mild.

8 Conclusion

We proved the existence of a robust code when the source bandwidth and the channel bandwidth are equal. Hybrids of analog and digital systems with achievable distortion regions better than the achievable distortion regions of time sharing and tandem systems are designed. A nearly-robust code is defined and it is shown that a hybrid system is nearly-robust when the source bandwidth is not equal to the channel bandwidth.

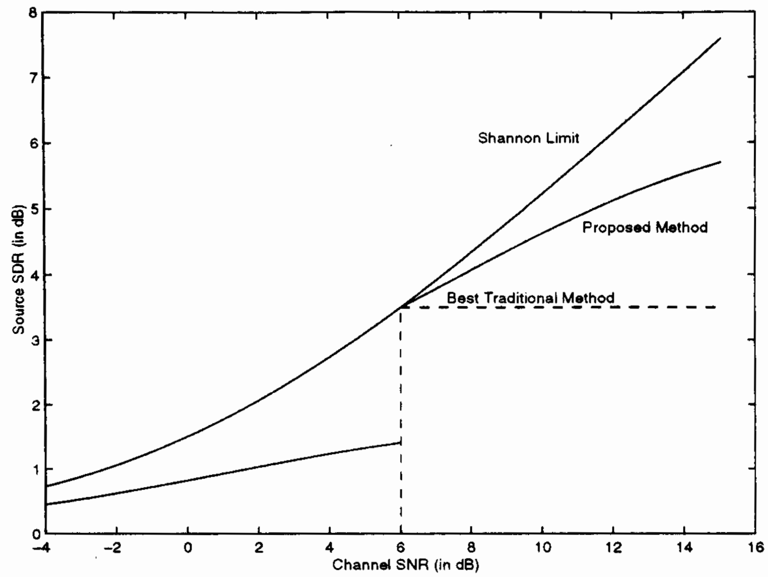

Figure 19: Near Robust Performance of a Special case of System 7: B/P HM for $\rho = 1/2$.

Appendix

Proof of Theorem 1: We rely on the following lemma for proving Theorem 1.

Lemma 3 *Let A_n be a sequence of random variables which converges with probability one (w.p.1) to a. Let B_n converges w.p.1 to b. Let $F : \mathbb{R}^2 \mapsto \mathbb{R}$, $C_n = F(A_n, B_n)$ and $c = F(a, b)$. If F is continuous at (a, b) then C_n converges w.p.1 to c.*

The proof of this lemma follows directly from the continuity of F at (a, b). We will only be interested in the case where $F(A_n, B_n) = A_n/B_n$ and $b \neq 0$. Suppose $\alpha(X) = \sqrt{P/\sigma^2}\,X$ and $\beta(Z) = \sqrt{\sigma^2 P} Z/(P+N)$, then it can be easily shown that the optimal distortion (5) can be achieved. This code, however, is not robust since β depends on N. We will replace $P + N$ by $(1/n)\|Z\|^2$ in the proof, hence making β independent of N.

Let $\alpha(X) = \sqrt{P/\sigma^2}X$. Note that $E[\|\alpha(X)\|^2] = (P/\sigma^2)E[\|X\|^2] = nP$. Thus this encoder satisfies the power constraint. Also let

$$\hat{X} = \beta(Z) = \begin{cases} \frac{n\sqrt{\sigma^2 P}}{\|Z\|^2} Z & \text{if } \frac{1}{n}\|Z\|^2 \geq \frac{P}{2} \\ 0 & \text{if } \frac{1}{n}\|Z\|^2 < \frac{P}{2} \end{cases} . \qquad (71)$$

Observe that neither the encoder nor the decoder depends on N. The squared-error distortion can be expressed as:

$$D_n \triangleq \frac{1}{n}\|X - \hat{X}\|^2 = \frac{1}{n}\left\{\|X\|^2 + \|\hat{X}\|^2 - 2\langle X, \hat{X}\rangle\right\}. \tag{72}$$

Note that

$$\frac{1}{n}\|\hat{X}\|^2 = \begin{cases} \frac{\sigma^2 P}{\frac{1}{n}\|Z\|^2} & \text{if } \frac{1}{n}\|Z\|^2 \geq \frac{P}{2} \\ 0 & \text{if } \frac{1}{n}\|Z\|^2 < \frac{P}{2} \end{cases}. \tag{73}$$

We apply Lemma 3 by setting $A_n = \sigma^2 P = a$, $B_n = \|Z\|^2/n$ and $C_n = \|\hat{X}\|^2/n$. Note that $B_n \to P + N = b$ w.p.1 by the strong law of large numbers (SLLN). Obviously, the function F given in (73) is continuous at $(a, b) = (\sigma^2 P, P + N)$. Thus $\|\hat{X}\|^2/n \to \sigma^2 P/(P + N)$ w.p.1.

Also note that

$$\frac{2}{n}\langle X, \hat{X}\rangle = \begin{cases} \frac{\frac{2\sqrt{\sigma^2 P}}{n}\langle X, Z\rangle}{\frac{1}{n}\|Z\|^2} & \text{if } \frac{1}{n}\|Z\|^2 \geq \frac{P}{2} \\ 0 & \text{if } \frac{1}{n}\|Z\|^2 < \frac{P}{2} \end{cases}. \tag{74}$$

Again we apply Lemma 3, setting $A_n = 2\sqrt{\sigma^2 P}\langle X, Z\rangle/n$, $B_n = \|Z\|^2/n$ and $C_n = 2\langle X, \hat{X}\rangle/n$. As before, $B_n \to P + N = b$ w.p.1. Also, by the SLLN, $A_n \to 2r\sqrt{\sigma^2 P}\sqrt{\sigma^2(P + N)}$ w.p.1, where r is the correlation coefficient between X_i and Z_i. It can be easily shown that $r = \sqrt{\sigma^2 P}/\sqrt{\sigma^2(P + N)}$. Thus $A_n \to 2\sigma^2 P = a$ w.p.1. By Lemma 3, we have $2\langle X, \hat{X}\rangle/n \to 2\sigma^2 P/(P + N)$ w.p.1.

Lastly, we have $\|X\|^2/n \to \sigma^2$ w.p.1 by the SLLN. Combining these results to (72), we get

$$D_n = \frac{1}{n}\|X - \hat{X}\|^2 \xrightarrow{w.p.1} \frac{\sigma^2}{1 + P/N}. \tag{75}$$

The above limit is D_{opt} when $\rho = 1$. The remainder of this proof is to show that the expectation of D_n converges to D_{opt}. To do this, we need to show that D_n is uniformly integrable (see [23], pages 219-220, 347-348). By the triangular inequality,

$$D_n \leq \frac{1}{n}\|X\|^2 + \frac{1}{n}\|\hat{X}\|^2 + \frac{2}{n}\|X\|\|\hat{X}\|. \tag{76}$$

Note that $\|\hat{X}\|^2/n$ is bounded by $2\sigma^2$. Thus

$$D_n \leq \gamma_n + 2\sigma^2 \tag{77}$$

where $\gamma_n \triangleq \|X\|^2/n + 2\|X\|\|\hat{X}\|/n$. We now show that γ_n is uniformly integrable. This is equivalent to showing that (i) $E[\gamma_n]$ is bounded and (ii) $\forall\ \epsilon > 0, \exists\ \delta$ s.t. $P(A) < \delta \implies E[\gamma_n \mathcal{I}_A] < \epsilon$, where A is a measurable set and $\mathcal{I}_A$ is its indicator function (see [23], problem 16.19). By Schwartz Inequality $E[\gamma_n] \leq \sigma^2(1 + 2^{\frac{3}{2}})$ so (i) is proven. To prove (ii), Again by Schwartz Inequality

$$E[\gamma_n \mathcal{I}_A] \leq \frac{1}{n} E[\|X\|^2 \mathcal{I}_A] + 2^{\frac{3}{2}} \sigma \left(E[\frac{1}{n} \|X\|^2 \mathcal{I}_A] \right)^{\frac{1}{2}}. \tag{78}$$

Now,

$$\frac{1}{n} E[\|X\|^2 \mathcal{I}_A] = \frac{1}{n} \sum_{i=1}^{n} E[X_i^2 \mathcal{I}_A]. \tag{79}$$

Define $\eta_i \triangleq \mathcal{I}_{[X_i^2 \leq \lambda]}$, where λ is a positive number to be chosen later. Note that $X_i^2 \eta_i \leq \lambda$. Thus

$$\begin{aligned} E[X_i^2 \mathcal{I}_A] &= E[X_i^2 \mathcal{I}_A \eta_i] + E[X_i^2 \mathcal{I}_A (1 - \eta_i)] \\ &\leq \lambda E[\mathcal{I}_A] + E[X_i^2 (1 - \eta_i)] \\ &= \lambda P(A) + E[X_i^2 (1 - \eta_i)]. \end{aligned} \tag{80}$$

Note that

$$E[X_i^2 (1 - \eta_i)] = \int_{x^2 > \lambda} x^2 f_X(x) dx = 2 \int_{\sqrt{\lambda}}^{\infty} x^2 f_X(x) dx, \tag{81}$$

where $f_X(\cdot) \sim \mathcal{N}(0, \sigma^2)$. Given $\epsilon > 0$, $\exists$ a sufficiently large λ s.t. $E[X_i^2(1 - \eta_i)] < \epsilon/2$. For this λ, we choose $\delta = \epsilon/(2\lambda)$. If $P(A) < \delta$, we have $E[X_i^2 \mathcal{I}_A] < \epsilon$, which implies that $E[\gamma_n \mathcal{I}_A] < \epsilon$. This result indicates that γ_n is uniformly integrable, which in turn implies that D_n is uniformly integrable. This along with the fact that $D_n \to \sigma^2/(1 + P/N)$ w.p.1 implies that $E[D_n] \to \sigma^2/(1 + P/N)$.

Let $\aleph = \{N_1, N_2, \ldots, N_K\}$. Thus, we have shown

$$\lim_{n \to \infty} D_{ave}(N_k) = D_{opt}(N_k). \tag{82}$$

$\forall\ N_k \in \aleph$. □

Proof of Lemma 1: Let source rate be R_s.

$$R_s = \frac{1}{n} H(\tilde{X}) \tag{83}$$

$$= \frac{1}{n} I(X;\tilde{X}) \tag{84}$$

$$= \frac{1}{n}\left(h(X) - h(X|\tilde{X})\right) \tag{85}$$

$$\geq \frac{1}{2}\log(2\pi e\sigma_s^2) - \frac{1}{n}\sum_{i=1}^{n} h(e_i|\tilde{x}_i) \tag{86}$$

$$= \frac{1}{2}\log(2\pi e\sigma_s^2) - \frac{1}{n}\sum_{i=1}^{n} E_x[h(e_i|\tilde{x}_i = x)] \tag{87}$$

$$\geq \frac{1}{2}\log(2\pi e\sigma_s^2) - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2} E_x[\log(2\pi e \text{VAR}(e_i|\tilde{x}_i = x))] \tag{88}$$

$$\geq \frac{1}{2}\log(2\pi e\sigma_s^2) - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log\left(2\pi e E_x(\text{VAR}(e_i|\tilde{x}_i = x))\right), \tag{89}$$

where VAR denotes the variance of a random variable. (83), (84) and (85) follows from the definition of R_s, the fact that $\tilde{X}$ is a function of X and the definition of mutual information. (86) follows from the chain rule of differential entropy and the fact that conditioning reduces differential entropy. (87) follows from the definition of conditional differential entropy. (88) follows from the fact that a Gaussian distribution has the highest entropy for a given variance, (89) follows from Jensen's inequality.

Without loss of generality, we assume that e_i and $\tilde{x}_i$ have zero mean. We further assume that $E(e_i^2)$ and $E(\tilde{x}_i^2)$ are bounded. Let c_i be the coefficient of correlation between e_i and $\tilde{x}_i$. Note that $E_x(\text{VAR}(e_i|\tilde{x}_i = x))$ is the minimum mean square distortion in estimating e_i given $\tilde{x}_i$. If we estimate e_i from $\tilde{x}_i$ as $\hat{e}_i = c_i\sqrt{\frac{VAR(e_i)}{VAR(\tilde{x}_i)}}\tilde{x}_i$ then the distortion in estimating e_i from $\tilde{x}_i$ is less than $(1 - c_i^2)D_i$, where D_i is $E(e_i^2)$. Thus,

$$E_x(\text{VAR}(e_i|\tilde{x}_i = x)) \leq (1 - c_i^2)D_i. \tag{90}$$

From (11),

$$R_s = \frac{r}{2}\log\left(1 + \frac{P_c}{N_c}\right), \tag{91}$$

and from (12),

$$D = [illegible]. \tag{92}$$

Thus,

$$R_s = \frac{1}{2}\log\left(\frac{\sigma_s^2}{D}\right). \tag{93}$$

¿From (89), (90) and (93), we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Big(\log(1-c_i^2)+\log(D_i)\Big) \geq \log(D). \tag{94}$$

Thus,

$$\begin{aligned} -\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log(1-c_i^2) &\leq \lim_{n\to\infty}\log\left(\frac{(\prod_{i=1}^{n}D_i)^{\frac{1}{n}}}{D}\right), & (95)\\ &\leq \lim_{n\to\infty}\log\left(\frac{\frac{1}{n}\sum_{i=1}^{n}D_i}{D}\right), & (96)\\ &= 0. & (97)\end{aligned}$$

Since $-\log(1-c_i^2) > c_i^2$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}c_i^2 = 0. \tag{98}$$

Now, the result follows from the fact that $|E(e_i\tilde{x}_i)| = |c_i|[E(e_i^2)E(\tilde{x}_i^2)]^{\frac{1}{2}}$. □.

Proof of Lemma 2: Let

$$\mathcal{N}_x^{(n)}(\mu,\sigma^2) \triangleq (2\pi\sigma^2)^{-n/2}\exp\left(\frac{-(x-\mu)^T(x-\mu)}{2\sigma^2}\right). \tag{99}$$

That is, $\mathcal{N}_x^{(n)}(\mu,\sigma^2)$ is the joint probability distribution function (pdf) of an n-dimensional Gaussian vector with mean vector μ and covariance matrix $\sigma^2\mathbf{I}_n$. Let $\text{Proj}(W,k)$ denotes a k-dimensional projection of a l-dimensional $(l>k)$ vector W, i.e., if $W=(w_1,w_2,\ldots,w_k,\ldots,w_l)$ then $\text{Proj}(W,k)=(w_1,w_2,\ldots,w_k)$. We use the random coding argument to prove this lemma. Let $x,\tilde{x},y,z\in\mathbb{R}$ and define

$$\begin{aligned} f_s(x,\tilde{x}) &\triangleq \mathcal{N}_{\tilde{x}}^{(1)}(0,\sigma_s^2-D-D\delta_2)\mathcal{N}_x^{(1)}(\tilde{x},D+D\delta_2) & (100)\\ f_c(y,z) &\triangleq \mathcal{N}_y^{(1)}(0,P_c-N\delta_1)\mathcal{N}_z^{(1)}(y,N), & (101)\end{aligned}$$

where D is given in (12). We define the two sets $A_{s,\epsilon}^{(n)}$, and $A_{c,\epsilon}^{(m)}$. The set $A_{s,\epsilon}^{(n)}$ is a collection of all n-tuple pairs $(X,\tilde{X}) = ((x_1^1,x_2^1,\ldots,x_n^1),(\tilde{x}_1^1,\tilde{x}_2^1,\ldots,\tilde{x}_n^1))$ which satisfy

1. $|-\frac{1}{n}\log f_s(X) - \frac{1}{2}\log(2\pi e\sigma_s^2)| < \epsilon$.

2. $|-\frac{1}{n}\log f_s(\tilde{X}) - \frac{1}{2}\log(2\pi e(\sigma_s^2 - D - D\delta_2))| < \epsilon$.

3. $|-\frac{1}{n}\log f_s(X,\tilde{X}) - \frac{1}{2}\log((2\pi e)^2(D + D\delta_2)(\sigma_s^2 - D - D\delta_2))| < \epsilon$.

4. $|\frac{1}{n}\sum_{i=1}^{n}||x_i - \tilde{x}||^2 - (D + D\delta_2)| < \epsilon$.

The distribution $f_s(X,\tilde{X})$ is a *product* distribution, i.e., $f_s(X,\tilde{X}) = \prod_{i=1}^{n} f_s(x_i,\tilde{x}_i)$. $A_{s,\epsilon}^{(n)}$ is called the *distortion typical set* w.r.t. f_s [18] (with distortion $D + D\delta_2$). The set $A_{c,\epsilon}^{(m)}$ is a collection of all m-tuple pairs (y,z) which satisfies the first three conditions in the above enumeration with f_s replaced by f_c. $A_{c,\epsilon}^{(m)}$ is called the *jointly typical set* w.r.t. f_c [18].

Generation of Codebook. Generate a codebook $\mathcal{C}$ comprising of 2^{nR}, $l = \max(m,n)$-dimensional i.i.d. Gaussian vectors of zero mean and variance $\sigma_s^2 - D - D\delta_2$. The rate R will be decided later. Enumerate the codebook from 1 to 2^{nR}. A special codeword consisting of all-zero elements is added to the codebooks and is referred to as the 0^{th}codeword. The codebook $\mathcal{C}$ is revealed to the encoder and the decoder.

Encoding. For an incoming n-dimensional source vector X, choose a codeword $W \in \mathcal{C}$, s.t. $(X, \text{Proj}(W,n)) \in A_{(s,\epsilon)}^{(n)}$. If more than one of such codeword exist then choose one of them randomly. If no such codeword exists then choose the 0^{th} codeword. If the i^{th} codeword W_i is chosen then transmit $\sqrt{(P_c - N\delta_1)/(\sigma_s^2 - D - D\delta_2)}\text{Proj}(W_i,m)$ on the channel provided the code to be transmitted satisfies the power constraint. If the power constraint is not satisfied send an all zero codeword on the channel and also say that the 0^{th} codeword is chosen.

Decoding. For a channel output Z, choose a codeword $W \in \mathcal{C}$, s.t. $(\sqrt{(P_c - N\delta_1)/(\sigma_s^2 - D - D\delta_2)}\text{Proj}(W,m), Z) \in A_{(c,\epsilon)}^{(m)}$. If more than one or no such codeword exists then choose and decode an all-zero codeword. If only one such codeword exists then choose and decode $\hat{X} = \text{Proj}(W,n)$ provided $||\text{Proj}(W,n)||^2/n < \sigma_s^2 - D - D\delta_1 + \epsilon$. Otherwise. choose and decode an all-zero codeword.

Calculation of Distortion. The expected distortion is calculated over the random codebook $\mathcal{C}$. Let X denote a source vector and $\hat{X}$ be the decoded value. Define

$$\bar{D}_{\mathcal{C}} = E_{\mathcal{C}}[E(d(X,\hat{X}))|\mathcal{C}], \quad (102)$$

where $d(X, \hat{X})$ is the mean squared distortion. Denote $d(X, \hat{X})$ by d. Let $\mathcal{E}_{enc}$ be the event that the 0^{th} codeword is chosen by the encoder (encoding error event). Let $\mathcal{E}_{dec}$ be the event that the chosen codeword of the decoder is *not* the same as the chosen codeword of the encoder (decoding error event).

$$\begin{aligned}\bar{D}_{\mathcal{C}} = E_{\mathcal{C}}[E(d|\mathcal{C})] &= E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}^c \cap \mathcal{C}\right)\Pr\left(\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}^c|\mathcal{C}\right)\right] \\ &+ E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec} \cap \mathcal{C}\right)\Pr\left(\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec}|\mathcal{C}\right)\right] \\ &+ E_{\mathcal{C}}\left[E\left(d|\mathcal{E}_{enc} \cap \mathcal{C}\right)\Pr\left(\mathcal{E}_{enc}|\mathcal{C}\right)\right] \qquad (103)\end{aligned}$$

Note that, the first term in the above summation is less than $D + D\delta_2 + \epsilon$. Now consider the second term in the above summation. Since the source vector belongs to $A_{s,\epsilon}^{(n)}$ and since the mean squared value of the decoder output is always bounded by $\sigma^2 - D - D\delta_2 + \epsilon$,

$$E(d|\mathcal{E}_{enc}^c \cap \mathcal{E}_{dec} \cap \mathcal{C})) \leq 4(\sigma^2 + \epsilon). \qquad (104)$$

Let I denotes the codeword *chosen* by the encoder:

$$\Pr(\mathcal{E}_{dec} \cap \mathcal{E}_{enc}^c|\mathcal{C}) = \sum_{i=1}^{2^{nR}} \Pr(I = i|\mathcal{C})\Pr(\mathcal{E}_{dec}|(I = i) \cap \mathcal{C}). \qquad (105)$$

Let $\mathcal{B}_j$ be the event that $(Z, \text{Proj}(W_j, m)) \in A_{c,\epsilon}^{(m)}$, where Z is the decoder input and W_j is the j^{th} codeword of $\mathcal{C}$ and let $\mathcal{E}_{dec}^i = \mathcal{E}_{dec} \cap (I = i)$. Note that, the event

$$\mathcal{E}_{dec}^i \subset \mathcal{B}_i^c \cup \left(\bigcup_{j \neq i} \mathcal{B}_j\right). \qquad (106)$$

Applying the union bound, we get

$$\Pr(\mathcal{E}_{dec}^i|(I = i) \cap \mathcal{C}) \leq \Pr(\mathcal{B}_i^c|(I = i) \cap \mathcal{C}) + \sum_{j \neq i} \Pr(\mathcal{B}_j|(I = i) \cap \mathcal{C}). \qquad (107)$$

Thus,

$$\begin{aligned}\Pr(\mathcal{E}_{dec} \cap \mathcal{E}_{enc}^c|\mathcal{C}) &< \sum_{i=1}^{2^{nR}} \Pr(I = i|\mathcal{C})\Pr(\mathcal{B}_i^c|(I = i) \cap \mathcal{C}) \\ &+ \sum_{i=1}^{2^{nR}} \sum_{j \neq i} \Pr(I = i|\mathcal{C})\Pr(\mathcal{B}_j|(I = i) \cap \mathcal{C}) \qquad (108)\end{aligned}$$

Taking the expectation over the random codebook $\mathcal{C}$, we have

$$E_{\mathcal{C}}[\Pr(\mathcal{E}_{dec} \cap \mathcal{E}_{enc}^c|\mathcal{C})] < \sum_{i=1}^{2^{nR}} E_{\mathcal{C}}[\Pr(I = i|\mathcal{C})\Pr(\mathcal{B}_i^c|(I = i) \cap \mathcal{C})] + \sum_{i=1}^{2^{nR}} \sum_{j \neq i} E_{\mathcal{C}}[\Pr(I = i|\mathcal{C})\Pr(\mathcal{B}_j|(I = i) \cap \mathcal{C})]. \quad (109)$$

Since $\Pr(B_i^c|\mathcal{C} \cap (I = i))$, depends only on the channel noise, $\Pr(B_i^c|\mathcal{C} \cap (I = i)) < \epsilon$ for sufficiently large m. Thus, the first term on the RHS of (109) is less than ϵ. We need to show that the second term on the RHS of (109) is also less than ϵ. In order to prove this we rely on the following lemma.

Lemma 4 *In the above encoding procedure,*

$$\Pr(I = i|\mathcal{C}) \leq 2^{-\frac{n}{2}\left(\log \frac{\sigma_s^2}{D+D\delta_2} - 6\epsilon\right)}, \quad (110)$$

for $i \neq 0$ *and for any codebook* $\mathcal{C}$.

Proof: Let W_i be the chosen codeword. Now,

$$\Pr(I = i|\mathcal{C}) = \int_{(X,\mathrm{Proj}(W_i,n)) \in A_{s,\epsilon}^{(n)}} f_s(X) \mathrm{d}x. \quad (111)$$

Note that, $f_s(X) \leq 2^{\frac{n}{2}(-\log(2\pi e \sigma_s^2)+2\epsilon)}$ for $(X, \mathrm{Proj}(W_i, n)) \in A_{s,\epsilon}^{(n)}$. Thus,

$$\Pr(I = i|\mathcal{C}) \leq 2^{\frac{n}{2}(-\log(2\pi e \sigma_s^2)+2\epsilon)} \mathrm{Vol}(\{X : (X, \mathrm{Proj}(W_i, n)) \in A_{s,\epsilon}^{(n)}\}) \quad (112)$$

From Theorem 9.2.2 in [18]

$$\mathrm{Vol}(\{X : (X, \mathrm{Proj}(W_i, n)) \in A_{s,\epsilon}^{(n)}\}) \leq 2^{\frac{n}{2}(\log(2\pi e(D+D\delta_2))+4\epsilon)}. \quad (113)$$

From (112) and (113), we get the required result. □

From the above lemma, we have

$$\sum_{i=1}^{2^{nR}} \sum_{j \neq i} E_{\mathcal{C}}[\Pr(I = i|\mathcal{C})\Pr(\mathcal{B}_j|(I = i) \cap \mathcal{C})] \leq 2^{-\frac{n}{2}\left(\log(\frac{\sigma_s^2}{D+D\delta_2}) - 6\epsilon\right)} \sum_{i=1}^{2^{nR}} \sum_{j \neq i} E_{\mathcal{C}}[\Pr(\mathcal{B}_j|(I = i) \cap \mathcal{C}$$

Following similarly as in the proof of Theorem 10.1.1 [18], we get

$$E_{\mathcal{C}}[\Pr(\mathcal{B}_j|(I=i)\cap\mathcal{C})] \leq 2^{-\frac{m}{2}\left(\log(1+\frac{P_c-N\delta_1}{N})-6\epsilon\right)}. \tag{115}$$

From (114) and (115), we get

$$\sum_{i=1}^{2^{nR}}\sum_{j\neq i} E_{\mathcal{C}}[\Pr(I=i|\mathcal{C})\Pr(\mathcal{B}_j|(I=i)\cap\mathcal{C})] \leq 2^{n\left(2R-\frac{\rho}{2}\log(1+\frac{P_c-N\delta_1}{N})-\frac{1}{2}\log\frac{\sigma_s^2}{D-D\delta_2}+6\epsilon\right)}. \tag{116}$$

The above term can be made arbitrarily small if

$$R < \frac{1}{2}\left(\frac{\rho}{2}\log(1+\frac{P_c-N\delta_1}{N})+\frac{1}{2}\log\frac{\sigma_s^2}{D+D\delta_2}\right)-3\epsilon, \tag{117}$$

and hence the second term in (103) can be made arbitrarily small.

Now consider the third term of (103). Proceeding as in the proof of Theorem 3 in [24], we can show that this term can be made arbitrary small for large n provided

$$R > \frac{1}{2}\log\frac{\sigma_s^2}{D+D\delta_2}+3\epsilon \tag{118}$$

It can be easily verified that for $\delta_1 << 1$ and $\delta_2 = \rho\delta_1$, there exists R which satisfies (117) and (118).

Thus, we have shown that averaged over all such codebooks the distortion is less than $D+\delta$. Thus, there exists a codebook for which the distortion is less than $D+\delta$. □

Acknowledgement

The authors wish to thank Mitch Trott for valuable comments and for bringing to our attention references [4], [8], [15], and [19].

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, Vol. 27, pp. 379–423 and 623–656, 1948.

[2] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pp. 142–163, Mar. 1959.

[3] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, Vol. 37, pp 10–21, Jan. 1949.

[4] T. J. Goblick Jr., "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, pp. 558–567, Oct. 1965.

[5] N. Farvardin, "A study of vector quantization for noisy channels,"*IEEE Transactions on Information Theory*, pp. 799–809, July 1990.

[6] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Transactions on Information Theory*, pp. 827–838, Nov. 1987.

[7] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Transactions on Information Theory*, pp. 155–160, Jan. 1991.

[8] J. M. Lervik, A. Grϕvlen and T. A. Ramstad, "Robust digital signal compression and modulation exploiting the advantages of analog communications," *Proc. IEEE GLOBE-COM*, pp. 1044–1048, Singapore, Nov. 1995.

[9] N. Phamdo, N. Farvardin and T. Moriya, "A unified approach to tree-structured and multistage vector quantization for noisy channels," *IEEE Transactions on Information Theory*, pp. 835–850, May 1993.

[10] J. M. Kroll and N. C. Phamdo, "Analysis and design of trellis codes optimized for a binary symmetric Markov source with MAP detection," *IEEE Transactions on Information Theory*, to appear, Nov. 1998.

[11] M. Wang and T. R. Fischer, "Trellis-coded quantization designed for noisy channels," *IEEE Transaction on Information Theory*, pp. 1792–1802, Nov. 1994.

[12] E. Ayanoğlu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Transactions on Information Theory*, pp. 855–865, Nov. 1987.

[13] S. Vembu, S. Verdú and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Transactions on Information Theory*, pp. 44–54, Jan. 1995.

[14] S. Shamai, S. Verdú and R. Zamir, "Systematic lossy source/channel coding,", *IEEE Transactions on Information Theory*, pp. 564–579, March 1998.

[15] J. Ziv, "The behavior of analog communication systems," *IEEE Transactions on Information Theory*, pp. 587–594, Sept. 1970.

[16] T. R. Fischer and M. W. Marcellin, "Joint trellis-coded quantization/modulation," *IEEE Transactions on Communications*, Feb. 1991.

[17] R. J. McAulay and D. J. Sakrison, "A PPM/PM hybrid modulation system," *IEEE Transactions on Communication Technology*, pp. 458–469, Aug. 1969.

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

[19] M. Trott, "Unequal error protection: Theory and practice," Talk presented on March 1996.

[20] U. Mittal and N. Phamdo, "Robust joint source-channel coding theory," *IEEE International Symposium on Information Theory*, pp. 445, Germany, June 29–July 4, 1997.

[21] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, pp. 269–275, March 1991.

[22] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory*, pp. 279–280, March 1974.

[23] P. Billingsley, *Probability and Measure*, 2nd Edition, New York: John Wiley and Sons, 1986.

[24] U. Mittal and N. Phamdo, "Duality theorems for joint source-channel coding," *Submitted to IEEE Transactions on Information Theory*, Feb. 1998.