Predictive Hybrid Digital-Analog Coding for Correlated Sources

By

Ehsan Montazeri

A Thesis Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements for the Degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg, MB

June 27, 2016

Abstract

Hybrid Digital-Analog (HDA) codes are a class of codes that employ both digital and analog coding for transmission of a source. Such systems can potentially mitigate the drawback associated with digital codes, most importantly the levelling-off effect, where the decoded signal quality does not improve at better channel qualities. This feature makes HDA codes a very good candidate for broadcast systems, where each user has a different received channel signal to noise ratio (CSNR), and hence the code cannot be designed and adapted for a single user. HDA codes are also useful for fading channel, where the transmitter does not have the channel state information.

In this work, we study and evaluate a new class of HDA codes, designed for correlated sources. There is a predictive coder incorporated into the HDA system to use the source correlation, where the predicted source sample are HDA encoded and then transmitted through a communication channel. To do this, the predicted samples are quantized, but rather than merely sending the quantized samples through digital coding, the quantization error is transmitted over the channel as well. The resources, however, must be divided between the digital and analog coders. We found the optimum power allocation factor for this problem, one that would minimize the average end-to-end distortion. Hence, not only this system provides higher average quality than digital codes, it also provides an instantaneous performance improvement with improved channel quality.

Analytical and simulation results are presented to show the performance of the

predictive-HDA system, which is compared to the performance of digital and analog coding, as well as to that of multi-layer (ML) digital codes. ML codes are also meant to provide higher quality to stronger channels by designing a set of codes for a few different received CSNRs, hence providing a staircase instantaneous quality. The superiority of the proposed HDA code is shown in terms of both average and instantaneous signal quality. It is also shown that the power allocation problem in the design of the predictive-HDA system is considerably less complex than that of the ML code.

Acknowledgement

I would like to extend my sincere gratitude to my advisor Dr. Pradeepa Yahampath. This work would simply not be possible without his help, support, guidance, feedback, and patience all along the way.

I am hugely grateful of my parents and brothers for providing me with so much support and keeping me motivated.

I also would like to thank my friends and fellow labmates in the Robert Alan Kennedy Communication Lab, especially Honqia Tian and Kamal Darchini, who were always available to help me either academically or emotionally.

Finally, I would like to express my appreciation to the esteemed committee members for providing me with their invaluable feedback.

Contents

List of Figures

Chapter 1

Introduction

Digital communications has long been the prevalent choice for communication systems. There are numerous examples of applications where digital communications is used, among which TVs, phones, and the Internet are probably the most noteworthy ones. But what is the reason behind this extensive use of digital, compared to the minimal use of analog communications? The answer is the numerous capabilities that digital schemes provide, such as data compression, data encryption, and error correction . Moreover, the advancement of digital signal processing, the prevalence of digital sources of data, such as computers, and the widespread use of computer network also contributed hugely to bring digital communications to the point it is at today [1]. There are, however, some drawbacks associated with digital communications. Many of the data sources that we encounter are in reality analog. This includes voice, speech, video, and the readings from many measurement devices. Hence, the analog samples generated from these sources must be converted into digital data first . This analog to digital conversion at the transmitter side, with a corresponding digital to analog conversion at the receiver side, introduces an unavoidable amount of noise to the transmitted data. This noise is essentially the error that is generated due to quantization. Quantization is the process of converting a signal with continuous amplitude, and thus with infinite resolution, into a

signal with discrete amplitude and finite resolution. This way, the data can be represented using finite number of levels, and consequently, finite number of bits. This provides the digital codes with a noise margin [2], a very strong feature associated with digital schemes, but at the same time, the price paid for it is an irrecoverable distortion caused by quantization [1]. The distortion level can of course be reduced by using a higher bit rate, but since the bit rate is limited by the capacity of the channel over which the data is transmitted, the distortion will be bounded by the Shannon limit [3]. This issue becomes more problematic in the case of a varying channels, such as fading channels, where the channel capacity can vary heavily and randomly. Since the quantizer bit rate is a function of the channel capacity, it must be adapted to the channel state. This, in turn, requires a set of different source and channel codes for each state of the channel, a technique that is currently employed in cellular technologies, which is referred to as link adaptation [4]. The adaptive techniques, however, cannot be used for broadcasting. Broadcasting is a process where the transmitter sends the data to multiple receivers, as opposed to point-to-point communications, where the transmitter sends the data to a particular receiver. For example, TV stations broadcast their signals to many viewers, where as cellular towers send their signals to particular users. In the case of broadcasting, the quality of the channel associated with each receiver varies differently from one another, and hence, unlike the point-to-point communication, the transmitter cannot adapt its coding schemes and its bit rate to the channel condition. A solution to this problem is to allow for an acceptable amount of outage, denoted as outage probability, and then design the system for the worst case scenario. This way, the receivers with channel conditions poorer than the selected threshold will experience outage, a phenomenon named as the threshold effect in the literature. Moreover, all the remaining receivers will receive an identical signal quality, regardless of how good or bad their channel conditions are, a condition referred to as the leveling-off effect. These two effects are considered the main drawbacks of digital communications [10].

The levelling-off effect is the result of the distortion that is caused due to quantization, and the fact that the quantization rate is dictated by the channel coding rate, due to Shannon's source-channel separation theorem [5]. The threshold effect happens due to the failure of channel codes below the required signal to channel noise ratio (CSNR), as explained by the converse of Shannon's channel coding theorem [5], and the sensitivity of source codes to transmission errors. In section 2.1, these issues will be discussed in more details.

Unlike digital coding, however, analog coding does not experience these problems, due to transmission of the unquantized data samples through the channel, as well as omission of source and channel coding. As explained, even though analog coding lacks many of the advantages that digital coding can provide, it does not fail where digital does. This led the research community to the idea of systems where both analog and digital coding are used together, in order to use the advantages of both schemes. Such coding schemes are called Hybrid Digital and Analog (HDA) coding schemes, and even though not currently being used in the industry, they have been studied for a while now in the literature, and are also investigated further in this thesis. Figures 1.1 and 1.2 show examples of an HDA encoder and decoder, respectively, where the analog and digital messages are added together and sent through the channel using a process called superposition coding [3].

The main contributions of this thesis are the study of HDA coding for sources with memory, as well as finding an optimal power allocation between the digital and the analog parts. The performed analysis, as well as the simulation results, confirming the analytical results are included. Moreover, the proposed system is compared to its digital counterpart, and its advantages are demonstrated.

The HDA scheme used in this work includes a predictive coder [6] to take advantage of the source memory, and uses superposition [3] for sharing the bandwidth between the analog and digital messages. These ideas will be explained and expanded further in the following chapters.

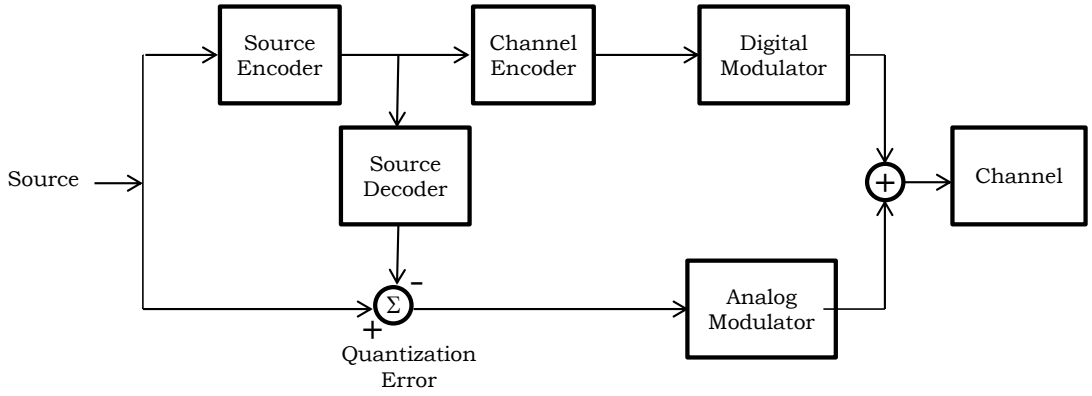

Figure 1.1: An Example of an HDA Encoder

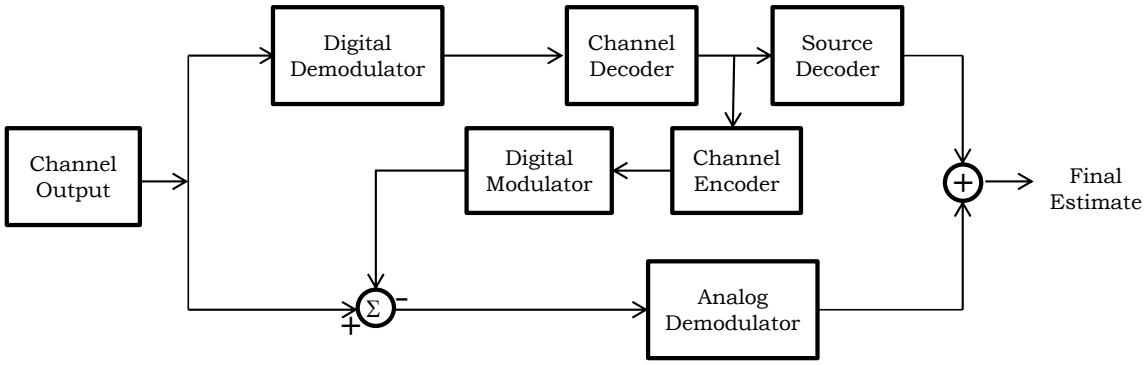

Figure 1.2: An Example of an HDA Decoder

In the rest of this report, first a review of some needed background, as well as an overview of some of the most important work done in the area of HDA coding will be presented, and then, the objective of this thesis will be explained and elaborated. The system model and the problem formulation are presented and discussed in chapter 3, the analytical results are given in chapter 4, and the simulation results and the comparisons are shown in chapter 5. We end the report by our conclusions and the possibilities for the future work, and the list of the reference used for carrying out this study.

Chapter 2

Background and Related Work

2.1 Point-to-Point Communications

In this section, a brief review of some of the most important results in information theory about source and channel coding is presented. These ideas will then be used to explain the drawbacks associated with some of the conventional coding schemes.

2.1.1 Source and Channel Coding

Let $\mathbf{S_D}$ denote an information source, taking values from a discrete alphabet of size M, and let P_i represent the probability of the i^{th} outcome. The entropy of this source is given by [3]:

$$H = -\sum_{i=0}^{M-1} P_i \log P_i \tag{2.1}$$

The entropy of a source gives the lowest number of bits required to represent each source sample.

Let $\mathbf{S_C}$ denote an information source whose values are taken from a continuous alphabet. In order to represent this source using a finite number of bits, it must first be quantized. By quantizing a source with a rate of R_Q, the quantizer output will

take values from an index of size 2^{R_Q}. This introduces an irrecoverable distortion to the signal. The minimum rate required to achieve a particular amount of quantization distortion is given by the rate-distortion function of the source. It can also be viewed as presenting the lowest attainable distortion, given a specific quantization rate. For an iid Gaussian source, with the variance of ${\sigma_s}^2$, the rate distortion function is given by [3]:

$$D(R_Q) = {\sigma_s}^2 \ 2^{-2R_Q} \tag{2.2}$$

A very important result stated by Shannon is the noisy channel coding theorem [5]. It determines the maximum allowable transmission rate for error free communication. Formally, it states that as long $R_C < C$, where R_C is the channel coding rate and C is the channel capacity, there can be a coding scheme with rate R_C that can achieve error-free communication. Conversely, for any rate $R_C > C$, the probability of decoding error will be bounded away from zero, meaning that reliable communication will not be possible. For a continuous-input, continuous-output channel with additive white Gaussian noise (AWGN), the channel capacity is given by

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right) \tag{2.3}$$

where P and N are the signal and noise powers, respectively [3].

Shannon's source-channel separation theorem states that it is possible to send a discrete source with entropy H reliably over a channel with capacity C, as long as $H \leq C$. Similarly, for a source with continuous alphabet, error-free transmission is possible if the quantization rate R_Q satisfies $R_Q \leq C$. The conditions in the theorem can be further modified as follows:

$$\begin{aligned} H &\leq rC \\ R_Q &\leq rC \end{aligned} \tag{2.4}$$

The parameter r is called the bandwidth expansion ratio, and is defined as follows, where W_C and W_S are the channel and the source bandwidths, respectively [10].

$$r = \frac{W_C}{W_S} \tag{2.5}$$

When $r > 1$, the channel bandwidth is greater than the source bandwidth, and we have the case of bandwidth expansion. Conversely, the case of $r < 1$ is called bandwidth compression, meaning that the source bandwidth is less than the channel bandwidth. When $r = 1$, at each channel access, no more than one source sample can be transmitted. This case is referred to as matched bandwidth.

2.1.2 Tandem Codes

Shannon stated that for the case of point-to-point communication, where one transmitter is sending the data over a single channel to one receiver, the source code and the channel codes can be designed separately, without any loss of optimality, given that the codes achieve the optimal bounds. Optimality is defined by the type of the source and the channel that we are dealing with. For example, for an iid Gaussian source that is to be transmitted over an AWGN channel, the optimal distortion is obtained by combining (2.2), (2.3), and (2.4), and is given as follows:

$$D = {\sigma_s}^2 \; 2^{-2rR_Q} \tag{2.6}$$

To achieve optimality, several powerful source and channel coding schemes have ever since been designed; source codes that can achieve the entropy bound [3], quantizers that can get close to the rate distortion limit [6], and channel codes with rates close to channel capacity [7].

Systems where the source and the channel codes are designed separately are called tandem codes. As mentioned, such systems can potentially be optimal, but that does not mean that other types of systems are incapable of optimality. Joint source-channel codes, for example, have been studied extensively, and will be briefly discussed in this chapter. HDA codes, which are the main topic of this thesis, are also another alternative to tandem codes. As previously discussed, for the case of transmission of an iid Gaussian source over an AWGN channel, where optimal distortion is given by (2.6), Goblick [8] proved that a very simple scheme can achieve optimal-

ity: analog or uncoded transmission. Surprisingly, if the source samples are sent uncoded through the channel, and decoded using a linear minimum mean-squared error (MMSE) receiver, (2.6) will be achieved. This scheme is no longer optimal, if the source and channel bandwidths are mismatched, or if the source has memory, while there could still be very good tandem codes in such cases to achieve, or at least, get close to optimality.

2.1.3 The Cliff Effect

Even though tandem codes can achieve optimality for point-to-point communication, they do have some drawbacks associated with them as well. Two of such problems are the levelling-off effect and the threshold effect, which are collectively referred to as the cliff effect [10]. The threshold effect is explained by the converse of Shannon's channel coding theorem [5], as discussed in section 2.1.1. As explained, when $R_C > C$, the decoding error probability will be bounded away from zero, and hence, the transmitted signal cannot be decoded correctly. Since the digital source codes are very prone to errors, even few transmission errors can cause a very large end-to-end distortion in the received signal, resulting in outage. Now, let the channel capacity C in a time varying channel belong to the range $C \in [C_{min}, C_{max}]$. If the digital code is designed for $C_0 > C_{min}$, when the channel is having a capacity in the range $C \in [C_{min}, C_0)$, the system will experience outage, and will not be able to decode the data. A similar problem occurs in broadcast channels, which will be discussed further in section 2.3.

The other problem associated with digital tandem coding, named as the levelling-off effect, is the result of the Rate-Distortion and the source-channel separation theorems, which were also explained in section 2.1.1. Tandem digital codes require the knowledge of the channel capacity to limit the source coding rate. Hence, in a varying channel or in a broadcast channel, where $C \in [C_{min}, C_{max}]$ and the code

is designed for some capacity value of C_0, the resulting distortion will be equal or more than $D_0 = 2^{-2rC_0}$ for all channels whose capacities lie in the range $C \in [C_0, C_{max}]$. This means that the stronger channels cannot take advantage of their better conditions, and will all be limited by the worst case channel. In other words, the output quality saturates in a digital tandem coding system. There have been solutions for this problem, some already implemented in the industry. In fading channels, for example, where the channel conditions vary in time, adaptive rates are used at the transmitter side. The transmitter hence must estimate the channel condition, and given that the channel is slowly varying, the rate and the coding scheme can be adapted to the estimated channel condition [4]. Adaptation, however, is not possible in a broadcast channel, where many different channels, each with a different quality, must be able to decode the data at the same time. A solution that can mitigate this problem to some extend, but not fully, is multi-layered (ML) coding, or progressive coding [11], [12]. In such systems, the signal is coded at a low rate, providing a base layer, or a coarse description of the signal. This coarse layer is meant to be decoded by all the channels. The finer details will be also transmitted, but they will be inaccessible by the weak users. This way, multiple layers of details about the signal can be sent, and depending on their channels strengths, channels can decode a number of these layers. It is shown in [11] that layered coding can achieve optimality for Gaussian sources. Another solution for this problem is the use of analog coding. Analog systems do not experience the cliff effect. By transmitting the data samples unquantized through the channel, and hence not introducing any quantization noise to the signal, analog coding avoids the levelling-off effect. Moreover, due to the absence of data compression in an analog communications system, there won't be an immediate decoding breakdown, once the CSNR falls bellow the minimum required amount. As explained, for Gaussian source, as an example, analog coding does not achieve the optimal performance neither for sources with memory, nor for source-channels pairs with mismatched

bandwidths. It also lacks the many advantage that digital systems are capable of providing. This led the research community to the idea of systems where both analog and digital coding are used together, in order to use the advantages of both schemes. Such coding schemes are called Hybrid Digital and Analog (HDA) coding schemes, and even though not currently being used in the industry, they have been studied for a while now in the literature, and are also investigated further in this thesis. In section 2.4, a review of some of the most important work done in the area of HDA coding will be presented, and then, the objective of this thesis will be explained and elaborated.

Apart form the problems discussed that exist in a digital tandem coding system, there are other scenarios where tandem codes are no longer able to reach the optimal performance level. Broadcast channels, discussed in section 2.3, and multiple user channels are such examples. HDA codes will prove to be a good choice for broadcast channels.

2.2 Joint Source-Channel Coding

As explained in section 2.1, in a tandem coding structure, the source is first compressed through source encoding, and then channel encoded for error correction and detection. Throughout the past several decades, many powerful source and channel codes have been developed; source codes that can achieve near-entropy compression, and channel codes that can achieve near-capacity rates [7]. Such powerful codes, however, can be quite complex, and usually require long blocks of data, causing delay in transmission. Moreover, tandem codes, which are based on Shannon's source-channel separation theorem, are only optimal for memoryless and ergodic sources and channels [3]. In [14], the necessary and sufficient conditions for the separation theorem to hold with no restrictions of memorylessness, stationarity, or ergodicity on either the source or the channel are derived. Applying these condition-

s to a memoryless fading channel shows that better performance can be obtained through a joint source-channel code (JSSC) design [15]. In addition, joint source-channel codes can provide close to optimum performance, without the complexity and delay of tandem codes. JSC codes can outperform tandem codes for a fixed complexity and delay, and are also inherently more robust to change in the channel noise level than tandem codes [10]. In fact, there are cases where the transmission of raw data can result in the optimum performance. In [16], Gastpar et al. state that if the source distribution, the distortion measure, the channel conditional distribution, and the channel input cost function are all matched, there is no need for any source or channel coding. For instance, optimal performance is achieved when a binary uniform source is sent directly over a binary-symmetric channel, provided that the distortion is measured in terms of the Hamming distance [17]. Another well-known example of such behavior, as already explained, occurs when an iid Gaussian source is transmitted across an additive white Gaussian noise (AWGN) channel with matched bandwidth [8],[9]. In all these cases, hence, tandem coding is not the best way to go, even if it does result in optimality.

Over the years, there have been numerous studies on designing JSC codes, [18] – [22] to mention a few. Reviewing these studies is out of the scope of this thesis report, but good papers can be found in [23] and [24] that provide a survey of some of the well known work in this field. Hybrid digital and analog (HDA) codes, however, which can be classified under the umbrella of JSC codes, will be covered in details in section 2.4, after reviewing the concept of broadcast channels in the next section.

2.3 Broadcast Channels

Broadcasting is a form of communication that has been in practice for many decades. TV and radio stations have long been operating, broadcasting their contents to a massive number of customers. In technical terms, broadcasting refers to a process

where a single transmitter sends some data to two or more receivers, each possibly having a different channel. Since each channel may experience different conditions, whether different noise level, or different fading level, what rate should be selected at the transmitter side? And what coding schemes should be used for such channels? These questions have been studies in the literature in the field of network information theory. In this section, a brief overview of some of the most important results of broadcast channels are summarized. These results, which are only applicable to the special case where there are only two receivers, have been taken form [3], [25], and [26].

2.3.1 Capacity Region and Achievable Rates

A broadcast channel consists of an input alphabet X and two output alphabets, Y_1 and Y_2, and a probability transition function, $p(y_1, y_2|x)$. An (M_1, M_2, n) code for such a broadcast channel consists of the following:

1- Two index sets, $\{1, 2, ..., M_1\}$ and $\{1, 2, ..., M_2\}$, representing two message sets, each intended for a receiver.
2- An encoding function that maps the two message sets to a single code of length n:

$$\{1, 2, ..., M_1\} \times \{1, 2, ..., M_2\} \longrightarrow X^n \tag{2.7}$$

3- Two decoding functions that map the received codewords into their corresponding messages:

$$\begin{aligned} g_1 &: {Y_1}^n \longrightarrow \{1, 2, ..., M_1\} \\ g_2 &: {Y_2}^n \longrightarrow \{1, 2, ..., M_2\} \end{aligned} \tag{2.8}$$

The code rates, which define the amount of information transmitted in each channel access, are computed as follows:

$$\begin{aligned} R_1 &= \frac{\log_2(M_1)}{n} \\ R_2 &= \frac{\log_2(M_2)}{n} \end{aligned} \tag{2.9}$$

The probability of decoding error is defined as:

$$P_e{}^{(n)} = Pr\Big(g_1({Y_1}^n) \neq W_1 \ or \ g_2({Y_2}^n) \neq W_2\Big) \tag{2.10}$$

A rate pair (R_1, R_2) is said to be achievable, if there exist a sequence of (M_1, M_2, n) or equivalently $\big((2^{nR_1}, 2^{nR_2}), n\big)$ codes that achieves $P_e \longrightarrow 0$. Notice that from (2.8), we have:

$$\begin{aligned} M_1 &= 2^{nR_1} \\ M_2 &= 2^{nR_2} \end{aligned} \tag{2.11}$$

The capacity region of a broadcast channel is defined as the closure of the set of all the achievable pairs of rates.

The transmitter could also transmit a common message with rate R_0, in addition to messages specific for each receiver, with rates R_1 and R_2.

2.3.2 Degraded Broadcast Channels

A broadcast channel is said to be physically degraded, if the relationship between the transmitter and the two receivers can be modelled as a Markov chain:

$$X \longrightarrow Y_1 \longrightarrow Y_2 \tag{2.12}$$

The channel degrades the signal to some extend between the source and the first receiver, and then degrades it more before the second receiver.

A channel is said to be stochastically degraded, if it is stochastically equivalent to

a physically degraded channel; that is if its marginal distributions are the same as those of a physically degraded broadcast channel. Such two channels have the same capacity regions.

Degraded broadcast channels are a special class of broadcast channels, and are easier to study and analyse. An example of such a channel is the Gaussian broadcast channel, where the transmitter has a total power of P, and the two receivers are both experiencing AWGN channels, with noise variances of N_1 and N_2. We assume that $N_1 < N_2$, meaning that receiver Y_1 has a better channel than Y_2. We can model this channel as

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2 \end{aligned} \tag{2.13}$$

where $Z_1 \sim \mathcal{N}(0, N_1)$ and $Z_2 \sim \mathcal{N}(0, N_2)$, but they can be correlated. The transmitter intends to send independent data to each receiver, at rates R_1 and R_2 to Y_1 and Y_2, respectively. The capacity region of this channel can be found to be as follows:

$$\begin{aligned} C_1 &= \frac{1}{2}\log\left(1 + \frac{\alpha P}{N_1}\right) \\ C_2 &= \frac{1}{2}\log\left(1 + \frac{(1-\alpha)P}{\alpha P + N_2}\right) \end{aligned} \tag{2.14}$$

The parameter α is the power allocation factor between the two messages, hence varying in the range $0 \leq \alpha \leq 1$. The bound can be achieved by a coding scheme called superposition coding. In this scheme, the transmitter generates two codebooks, one with power αP and rate $R_1 \leq C_1$, and the other one with power $(1-\alpha)P$ and rate $R_2 \leq C_2$. The two resulting messages, W_1 and W_2, are then added together to form the channel input $X = W_1 + W_2$ which is then sent through. The weak receiver, Y_2, will experience W_1 as interference, and will be able to only decode W_2. The strong receiver, Y_1, however, will be able to decode both messages, due to the channel being a degraded broadcast channel, and hence can effectively cancel the interference and detect both W_1 and W_2.

If common information is being sent to both receivers, the effective rate of the strong user will be $R_1 + R_2$, while the weak user still having the rate R_2. This is the idea behind successive refinement, where a coarse description of the source is transmitted at a low rate to the weak user, and finer refinements are sent to the strong user, which is able to receive the coarse information as well [13]. Superposition coding can thus reach the capacity region's outer bound for the case of Gaussian broadcast channels, and for all degraded broadcast channels. It is hence a very common coding scheme in HDA system, which will be explored further in the next section. For more information on broadcast channels, [3] and [25]-[28] are recommended.

2.4 Hybrid Digital and Analog Coding

In this section, an overview of some of the most important work in the literature about the HDA codes are presented and studied. We conclude by explaining the problem that is studied in this thesis, and outline the contributions of this work.

One of the earliest work that studies an HDA system is done by Schreiber in [29], where he discusses the requirements for modern television broadcasting systems, and proposes an HDA system for video coding that can meet those requirements to a good extend. The proposed system is based on progressive, or layered coding, with three levels of quality. At the base layer, MPEG is used to source code the video data. The difference between the original signal and the reconstructed version is sent to level 2. At this level, each frame is gone through some processing, including predictive encoding. The prediction error is the signal that is sent by analog transmission, after going through transform coding and adaptive selection. This process is repeated at the third level as well, resulting in three digital and two analog sets of messages. The digital data is channel coded by a concatenated code, comprised of Reed Solomon and Trellis codes, multiplexed, and finally mapped to a 64-PSK constellation, where the constellation points' amplitudes are determined by the analog sample values. Finally, the overall data is transmitted by the use of OFDM. The system in [29] is not claimed by Schreiber to be optimal, but it shows promising results in combating the disadvantages associated with pure digital transmission, by providing improved quality for stronger channels, and by supporting old receiver devices that can decode only the lower levels, resulting in an interoperable design.

Another early study of systems where analog and digital coding are employed together is done by Shamai et al. in [30], also inspired by the challenges in coexistence of digital and analog TV systems. At the time that digital TV receivers were becoming popular, many systems were still using analog signals, and hence, analog transmission had to be retained. Rather than sending a digitally encoded duplicate

signal for the digital receivers, [30] suggests sending the refinement information in the excess bandwidth in digital format, and equipping the new receivers with both digital and analog decoding capabilities. This way, the old analog devices will still be able to decode their analog signals, while the new devices will achieve a higher picture and sound quality. This can also reduce the amount of digital bandwidth required to achieve a certain level of fidelity, compared to the case where analog signal is discarded by the digital receivers. Shamai et al. propose the use of systematic source-channel codes, where a block of the source samples is transmitted in raw format (analog), and the excess bandwidth is used for sending the digitally coded version of that source block to deliver some refinement information. Shamai et al. find the conditions under which systematic codes can achieve optimality. For example, in the specific case of Gaussian IID sources, transmitted over AWGN channels, systematic codes can reach this optimality, while for binary sources sent over binary symmetric channels, they are strictly sub-optimal. For the case of the Gaussian sources, their proposed scheme splits the source in the frequency domain, sending the low pass version of the signal as analog samples, and the high pass content through digital codes. Hence, the receivers capable of decoding both signals will have access to the refinement information about the source as well as the coarse, low frequency data.

The first rigorous study of HDA systems is published in [10], by Mittal and Phamdo. Motivated by the ability of analog transmission in mitigating the threshold and the cliff effects, they propose a few HDA systems and present their performance bounds, for transmission of an IID Gaussian source over a Gaussian broadcast channel with two receivers. Their objective is to design codes that are optimal at a target CSNR, but degrade gracefully in performance, should the realized CSNR deviate from the target; a scenario motivated by broadcast channels. They conjecture that no code can be simultaneously optimal at different CSNRs, when the source and channel bandwidths are not equal. They propose two systems for the case of bandwidth

expansion, where the channel bandwidth is greater than the source bandwidth, and one system for the case of bandwidth compression, where the source bandwidth exceeds the channel bandwidth. As explained, when the source and channel bandwidth are equal, a pure analog system reaches the optimal performance [8]. The systems proposed in [10] are all analyzed from an information theoretical point of view, where the theoretical limits of rate distortion and channel capacity are assumed to be possible to reach. They also compute the distortion region achievable by each of their schemes. One of their proposed coding systems, used for bandwidth expansion, is composed of a two-layered digital tandem source and channel coder, with the reconstruction error from the second layer being transmitted as analog samples. The overall bandwidth is split between the analog and digital codes, and the two digital data messages are sent together through superposition. The power and bandwidth division between these layers, however, is arbitrary and not optimized. A variation of this system is also presented, where the data going to the second digital level is first demultiplexed, so that a part of it is digitally encoded, while the other part is sent as analog samples. This way, the analog signal bandwidth will be reduced by an arbitrary factor, set by the demultiplexer. The resulting performance of theses systems demonstrate graceful degradation, while being optimal at the designed C-SNR and beating the purely digital codes designed for the same CSNR, as well as purely analog systems.

In another publication, Mittal and Phamdo [31] use one of their coding schemes in [10] for encoding speech signals. The digital part includes a standard CELP source encoder, in addition to a turbo encoder for channel coding. The analog signal is the residual error signal from noisy source coding, or quantization of the speech signal. For transmission, the bandwidth is split between the analog and digital data. Phamdo and Mittal report that their HDA system shows noticeable improvement in listening tests, compared to the purely digital system operating at the same rate. Furthermore, at low CSNRs, where the digital coding breaks down, the HDA system

generate intelligible outputs. This further demonstrates the ability of HDA coding to mitigate the threshold effect.

The work done by Mittal and Phamdo led the way for many other publications, investigating several other aspects of HDA systems. Reznic and Feder [32] study the case of broadcasting a Gaussian source to two listeners over Gaussian channels with bandwidth expansion. They find the set of all achievable simultaneous distortion pairs, or an outer bound on the distortion region for such a broadcast scenario. They then prove that for a broadcast coding system designed at a target CSNR value $CSNR_{min}$, as the CSNR improves (while the transmitter is held fixed), the distortion cannot decay faster than the rate $\frac{1}{CSNR}$, whereas if a code were to achieve optimal performance at all CSNRs, it would have to decays at the rate of $\frac{1}{CSNR^r}$. Hence, no system can be optimal at all received CSNRs. They also found an inner bound on the distortion region, which is tighter than the inner bounds achieved by the systems proposed in [10] and [30]. This bound is obtained based on an HDA coding scheme, which includes one of the HDA systems of [10] and the systematic coding scheme of [30] as two special cases. Their proposed HDA scheme is based on one of the systems proposed by [10] and is also based on Wyner-Ziv source encoding. It includes a two-layer digital encoder, setting the base layer as a common message to both receivers, and the sencod layer as side information about the source, only decodable by the strong receiver. While the digital messages are sent together by using superposition, the analog signal, carrying the quantization error, is time-multiplexed with the digital data. It should be noted that this system and all its analysis, including the distortion region inner bound, are information theoretical limits that could be very challenging to meet in a practical design.

Skougland et al. [33] propose a simple HDA system for bandwidth expansion, and present an iterative algorithm for designing a vector quantizer for the digital joint source-channel encoder. The algorithm is similar to the well-known LBG algorithm [34]. The encoder is designed to be optimized for a target CSNR, that is, to minimize

the end-to-end distortion for that CSNR value. Hence, the encoder is fixed, while the decoder is variable, adapting to the received noise variance level. The system divides the bandwidth and power between the analog and digital signals, but this division is not optimized. They test their system with an iid Gaussian source, as well as a Gauss-Markov source, and it achieves a close to optimal performance at the designed CSNR, and demonstrates a robust and graceful performance degradation at other CSNRs.

Building on [33], Skuglund et al., in their work in [36], propose a system that can be used for both bandwidth compression and expansion. This system employs a tandem coder for the digital part, using a vector quantizer for source encoding, followed by turbo coder [35] for channel coding. The quantization error, after being adjusted to the proper power limit, is superposed on the digital codes, and sent together over the channel. For bandwidth compression, they use the Karhunen-Loeve Transform (KLT) [6] to decorrelate the source, and compress the signal after. They test their systems for an IID Gaussian source, as well as a Gauss-Markov source. They also propose a non-linear compression system for compressing the analog signal for bandwidth compression, which could be a replacement for the KLT. The system performance demonstrates near-robustness, and outperforms their system in [33].

In a closely related work, in [37], Wang et al. take a closer look at the problem of bandwidth compression in HDA systems. They propose two coding schemes, where one is based on [33] and [36]. They also obtain an information theoretical (mean squared) distortion upper bound for the two systems, in addition to an optimal power allocation formulation between the digital and the analog parts. This power allocation is optimum for a specific target CSNR, but the systems show robust performance over a range of CSNRs. System one uses a vector quantizer to compress the digital signal, and the quantization error is taken for analog transmission. To achieve compression for the analog transmission, a simple method is used, where some of the samples are discarded. If the source has memory, the samples

are first decorrelated using KLT, and the discarded samples are selected from the low-variance samples. System one uses superposition for transmission of the digital and analog data. System two uses source splitting to achieve compression, where a source is split into two vectors, one transmitted digitally, and the other through analog transmission. Superposition is used in this system as well. The analog coding for both systems consists of simple linear encoding and decoding: power adjustment at the encoder, and LMMSE decoding at the receiver. They also test out the systems by concrete implementations, where system one is implemented using a vector quantizer, directly mapped to a binary phase-shift keying (BPSK) modulation system, without the use of a channel coder, with hard decoding at the receiver in the digital part. The employed vector quantizer is similar to the one presented in [33]. The second system, after source splitting, is implemented by a Channel Optimized Vector Quantizer (COVQ) [18], which is a joint source-channel encoder. Two sources are used, an iid Gaussian source, in addition to a Gauss-Markov source, where the latter first goes through KLT for decorrelation. The results show robust performance over a range of different CSNRs, while having some deviation from the theoretical bounds derived in their work, with system two outperforming system one.

In a study done by Wilson et.al, in [38], the problem of transmitting a Gaussian source over a Gaussian channel in two cases is studied:

-where there is interference in the channel that is known only to the transmitter

-where the receiver has some side information about the source

They propose an HDA scheme based on the Costa Dirty Paper coding [39] for case one, and another one based on Wyner Ziv coding [40] for the second problem. They also proposed a system with both Costa and Wyner-Ziv coding, for the case with both interference and side information. They show that their system is optimal for the case of matched bandwidth, and prove that there are infinite coding schemes that could be optimal for this problem, of which uncoded (analog) transmission, and their system, are two examples. They also study their system under bandwidth

mismatch, and for broadcasting over two channels, with bandwidth compression, and derive the corresponding distortion expressions.

In a study presented in [41] by Prabhakaran et.al, the problem of broadcasting parallel Gaussian sources over a Gaussian broadcast channel is studied from an information theoretical point of view, for which three HDA systems are proposed. In the most general case, there are K iid Gaussian sources with different variances, being broadcast to two users. There are two sets of M sub-channels leading to the two receivers, each having the same statistics among themselves, while differing from the second set of sub-channels leading to the other receiver. So, there is a weak and a strong receiver, each having M identical AWGN channels. The case of $M = K$ corresponds to matched bandwidth, while $M > K$ and $M < K$ imply bandwidth expansion and compression, respectively. Prabhakaran et al. consider three cases and propose an HDA system for each that outperforms a similar purely digital scheme.

In case 1, the weak user is able to achieve the optimal point-to-point performance, while in case 2, it is the strong user that can reach the optimal performance. Case 3 is a trade-off between the two other cases. They first study the case of matched bandwidth, with $M = K = 2$, and then extend their systems to other cases. A detailed summary of these three systems is outlined here.

In case 1, the optimal point-to point performance for the weak-user can be achieved by the following two scheme, one purely digital, and one HDA, as presented in [41]: S_1, denoting the first source, which is the one with higher variance, is successively encoded into two layers, while S_2, the second source, is normally encoded. The coarse layer data from S_1, along with the source-coded data from S_2 are sent over the second sub-channel, using superposition coding. The refinement-layer bits from S_1 are sent alone over the first sub-channel. An equivalent distortion performance can be achieved at the weak user while improving the strong user's performance by sending the quantization error from the coarse quantization of S_1 uncoded (scaled)

over the first sub-channel, and by sending S_2 uncoded (scaled) over the second sub-channel and the coarse-layer bits on S_1 channel coded with sufficient power to be decoded (by the weak user), treating the uncoded transmission of S_2 as noise. This way, the strong user benefits from both the uncoded transmissions since it can form better quality estimates than the weak user.

In the second case, where optimal point-to-point separation scheme performance is sought for the strong user, a fully-digital and an HDA coding scheme were presented. In the digital scheme, the coarse-layer bits from a successive refinement source coding of S_1 are sent alone over the first sub-channel, and the source coded bits of the second source component S_2 and the refinement bits of S_1 are sent over the second subchannel, using superposition. These messages are not decodable by the weak user. The counterpart HDA system improves this by providing useful information without compromising the strong user's performance. It does this by sending S_2 uncoded (analog) over the first sub-channel, while sending S_1 uncoded (analog) over the second sub-channel and bits carrying the refinement information about S_1 using Dirty Paper channel coding. The transmission of S_2 acts as Gaussian side-information at the transmitter. Dirty Paper coding ensures that the transmission of S_2 does not affect the rate of transmission of the refinement bits. The bits carrying the refinement information are produced using Wyner-Ziv coding where the noisy observation of S_1 over the first sub-channel acts as side-information at the decoder. While the weak user will be unable to decode the refinement information, it benefits from the two uncoded transmissions.

Finally, there is the case where neither the weak nor the strong user achieves the point-to-point optimum performance, while a performance trade-off can be achieved in between them. An HDA scheme is suggested in [41], which is based on the two previous extreme cases. In this HDA system, the first source component S_1 is sent through three different ways:

- Codewords carrying the coarse layer description which will be decoded by both

users

- An uncoded version of the quantization error which the weak user will estimate from its noisy observation

- Wyner-Ziv bits on S_1 which only the strong user decodes

In decoding the Wyner-Ziv codewords, the strong user uses as side-information a linear estimate of S_1, using the coarse description and the noisy observation it has already decoded. The quantization error is sent uncoded (scaled) over the first sub-channel, using all the power allocated to this sub-channel. Over the second sub-channel, the coarse layer bits are sent using a Gaussian channel code, with part of the power allocated to this sub-channel. This is meant to be decoded by both users, treating the rest of the signals sent over this sub-channel as noise.

Using part of the leftover power, the second source is sent uncoded (scaled). Since the codewords carrying the bits about the coarse quantization of S_1 are assumed to be successfully decoded by both users, they may estimate the second source component, assuming only the rest of the power used in this sub-channel, along with the channel noise, as disturbance.

The leftover power in this sub-channel is used to send the Wyner-Ziv bits. This is done using Dirty-Paper coding, treating the scaled version of the second source component sent over this sub-channel as side-information (at the transmitter).

In short, Prabhakaran et al. show that without compromising the point-to-point optimal performance of either the weak or strong user, the performance of the other user can be improved over what the conventional separation approach offers. For the special case of memoryless sources and channels with bandwidth mismatch, their scheme matches the one presented in [32] by Reznic et al.

In another interesting study presented in [42], Jakubczak et al. design and implement a system named as 'SoftCast', which uses analog communication exclusively for transmission of video over a wireless channel. Their work is mostly practical rather than theoretical, being motivated by the inability of 802.11 (WiFi) to perfor-

m efficiently for video multi-casting, a problem that was previously explained in the context of absence of digital robust codes. SoftCast uses a joint source-channel code for encoding the video pixels, where the distance between the transmitted codewords reflect the difference between the pixel values. It does this by transmitting the analog values of each pixel, and then directly mapping the raw pixel values over OFDM and hence, resulting in a scheme that can be compatible with 802.11. It applies a 2-D Discrete Cosine Transform (DCT) to the raw pixel values, groups the DCT components in chunks, and discards zero and near zero values to achieve compression. For channel error protection, it performs a power allocation based on the variance of the DCT chunks, and by allocating more power to higher-energy components, it in turn makes the more important data more resilient against channel errors. The decoder is also linear, comprised of a linear mean squared error estimator, followed by reversing the other encoding components. Their results show a graceful degradation in the performance, while avoiding the threshold effect. In their demos, for comparison, MPEG fails suddenly as channel conditions deteriorate, while SoftCast does show a graceful performance degradation. The demos can be found at SoftCast web page [1].

In a paper published in [43], Li et al. suggest an HDA system for wireless video multicasting that has similarities to SoftCast. They first apply 2D Discrete Wavelet Transform on each frame to obtain a multi-resolutional description of the frames, and then divide the resulting data into LL, LH, HL, and HH bands. The LL band, acting as the coarse description of the data, is digitally encoded, using standard H.264/AVC video encoder, in conjecture with low rate convolutional codes and Binary Phase Shift Keying. The residual error from the quantization of the LL band, in addition to the other three bands, are sent by analog transmission, in a manner similar to SoftCast[42], as explained previously. The digital and the analog signals are then added together, and mapped to OFDM to be sent through the channel.

[1] http://people.csail.mit.edu/szym/softcast/videos.html

The decoding process is similar to other schemes that employ superposition coding. Their system achieves average performance improvement over SoftCast and shows graceful quality degradation as the channel conditions deteriorate.

In [44], Caire and Narayanan study the use of HDA codes in MIMO channels. They consider an M-input, N-output MIMO block fading channel and find an upper bound for the optimal distortion CSNR component for the channel. The CSNR exponent is the decay rate of the lowest achievable end-to-end distortion as a function of CSNR. They also find the lower bound achievable by tandem coding, in addition to a tighter lower bound, achievable by their proposed HDA system. The HDA scheme achieves the optimal CSNR exponent for the case of $r < 2 * min(M, N)$ for general M and N, as well as for all r for SISO channels, where r is the bandwidth expansion factor, given by (2.5). The HDA scheme is composed of a tandem digital coder with space-time codes, in addition to analog samples obtained from the quantization error. The analog and digital messages are multiplexed for transmission. They also construct practical HDA space-time codes using diversity–multiplexing trade off and scalar quantization.

In [45], a paper published by Rngeler and Vary, the problem of transmitting a correlated source using HDA coding is considered. They incorporate a transform encoder in their system to decorrelate the source vectors. After bit allocation and quantization of the transform coefficients, the digital data is channel encoded and then multiplexed with the analog data,which carry the quantization error. The system outperforms a similar purely digital scheme in a practical simulation, where a Gaussian correlated source first goes through KLT, followed by water filling bit allocation and Turbo coding.

2.5 Contribution of This Work

Though being a quite simple study, [45] shows that HDA coding systems can be quite effective for correlated sources as well. While most of the literature is focused on the transmission of iid sources, in practice, most encountered sources have memory. It is important, hence, to further study such sources in the context of HDA coding. Another problem that is not studied thoroughly in the related work is the issue of resource allocation between the digital and analog parts. This led us to the study of power allocation in an HDA system that is designed for correlated sources. In particular, we incorporated a predictive source encoder into an HDA coding system, where the analog messages carry the quantization error, resulting from quantizing the prediction error samples. The system uses superposition coding, hence the analog and digital messages are sent through the same bandwidth. There is still, however, a need for the distribution of the power between the digital and analog coders. We find an optimal power allocation for our system. The analytical results show that this system can outperform its purely digital counterparts. We also conduct simulations which are in agreement with our analysis.

In the remainder of this thesis, we start by describing the specifics of the problem we studied. We then present the problem formulation, followed by our proposed system. The analysis performed to find the optimal power allocation factor is presented, and the performance results are shown.

Chapter 3

System Model

Let X^D represent a discrete-time source, sampled at the rate $2W_S$ from a continuous-time source X^C, with bandwidth W_S. Let X_n represent the source output at time n. We assume that the source is a first order Gauss-Markov source, with the parameter ρ. Hence, the following relationship holds between the source outcomes:

$$X_n = \rho X_{n-1} + \varepsilon_n \tag{3.1}$$

This is essentially a first-order, auto regressive process [47]. The parameter ε is an iid Gaussian random process.

$$\varepsilon \sim \mathcal{N}(0, {\sigma_\varepsilon}^2)$$

It can be easily shown that X is a stationary, correlated Gaussian random process, with zero mean and the variance of $\dfrac{\sigma_\varepsilon^2}{1-\rho^2}$.

$$X \sim \mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right)$$

In addition, the auto-covariance function[1] of this source is given by the following [47].

$$R_{XX}(k) = E\left\{X_n X_{n-k}\right\} = \sigma_\varepsilon^2 \, \frac{\rho^k}{1-\rho^2} \tag{3.2}$$

[1]This quantity is usually referred to as auto-correlation function in electrical engineering texts

The MMSE linear prediction of this source is given by [48]:

$$\widetilde{X}_n = \rho \; X_{n-1} \tag{3.3}$$

The channel is assumed to be a discrete-time Gaussian channel, derived from a continuous-time AWGN channel with bandwidth W_C, and modelled as,

$$Y = R + Z \tag{3.4}$$

where R and Y are the channel input and output, respectively, and Z is a zero-mean, iid Gaussian random process, with variance ${\sigma_Z}^2$.

$$Z \sim \mathcal{N}\left(0, {\sigma_Z}^2\right)$$

We consider the case of matched bandwidth, where the source and channel bandwidths are equal.

$$W_C = W_S$$

Hence, at every channel access, one source sample is transmitted.

Our aim is to use an HDA scheme to code the source for transmission over this channel. A very commonly used source coder for correlated sources is predictive coder [6], where instead of quantizing the source samples directly, the source samples are predicted, and the prediction error is quantized and transmitted. Predictive encoding is used for speech and video, and is also a good choice for Gauss-Markov sources. This is because a linear predictor of the form (3.3) can achieve a Gaussian uncorrelated prediction error signal, which has many good properties and is easy to analyze. That is why we decided to incorporate a predictive coder in our HDA system to take advantage of the source correlation.

3.1 Predictive Coding

We first start by a brief description of predictive source coding. The material for this part is taken from [6].

A diagram of open-loop and closed-loop predictive source encoders are shown in Fig. 3.1 and Fig. 3.2 respectively.

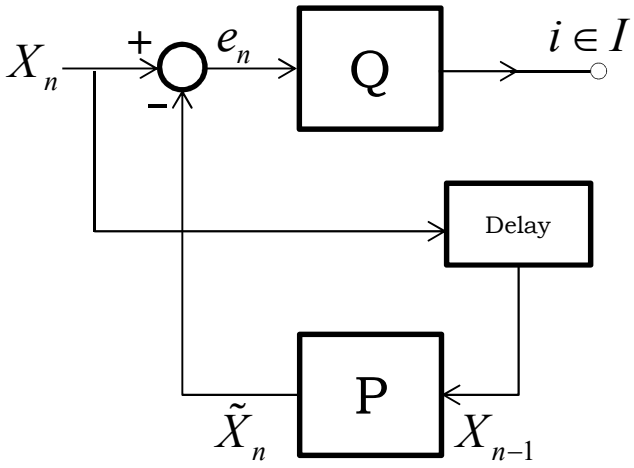

Figure 3.1: Open-Loop Predictive Encoder

- X_n: The value of the source sample at time n
- $\widetilde{X}$: The predicted value of X
- $\widehat{X}$: The reconstructed value of X
- e: The prediction error
- I: The quantizer output index set

The prediction error e has a lower variance than the source does, and hence, the distortion resulting from quantizing the prediction error is less than that of quantizing the source samples directly. For a first-order Gauss Markov source, for

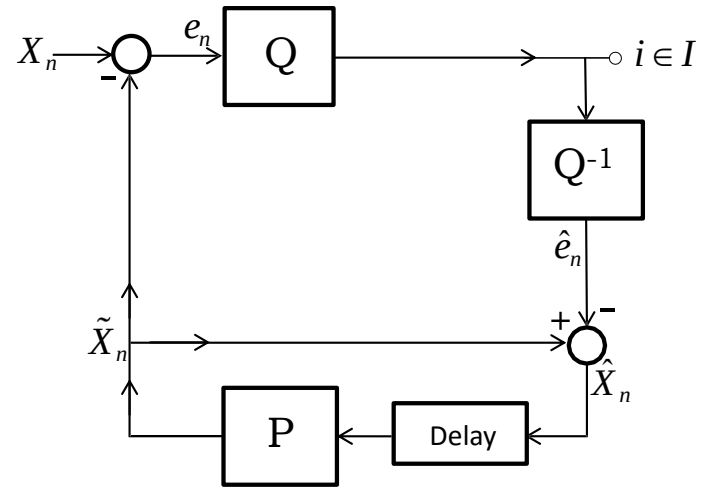

Figure 3.2: Closed-Loop Predictive Encoder

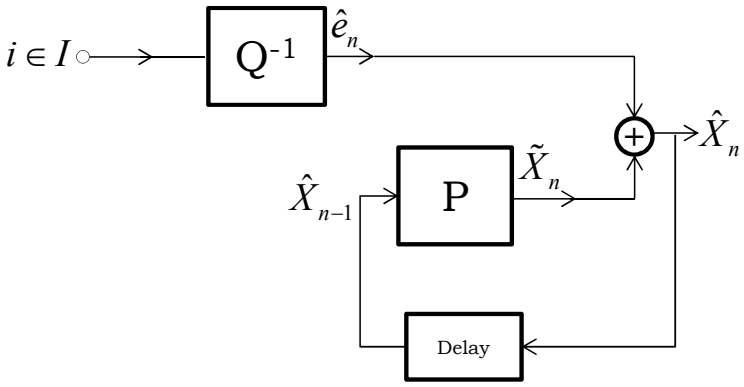

Figure 3.3: Closed-Loop Predictive Decoder

example, with the predictor of form (3.3), the relationship between the two variances is given by the following:

$$\sigma_e^2 = \frac{\sigma_X^2}{1-\rho^2} \tag{3.5}$$

As it can be seen in Fig. 3.1, in open loop encoding, the source sample at time $n-1$ is used to predict the value of X_n, and the prediction error e_n is then quantized. In a communication system, however, the receiver won't have access to the exact source sample values. Hence, the closed-loop configuration is used instead, where

the reconstructed value $\hat{X}_{n-1}$ is used instead of X_{n-1} for predicting X_n.
The closed-loop predictive encoder and decoder are shown in Fig. 3.2 and Fig. 3.3, respectively.

At the decoder at time n, after the lossy reconstruction of the prediction error, e_n is summed up with the predicted value $\tilde{X}_n$ to reconstruct $\hat{X}_n$.

3.1.1 HDA Predictive Coder

We integrate the predictive coder in an HDA framework, with the encoder and the decoder shown in Fig. 3.4 and Fig. 3.6, respectively.

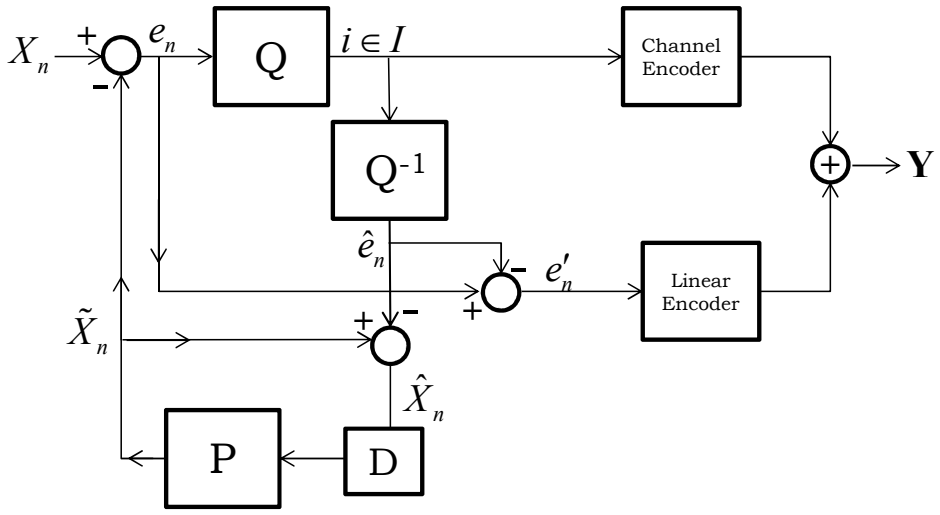

Figure 3.4: Closed-Loop Predictive HDA Encoder

- e_n: The value of the prediction error at time n
- $\hat{e}$: The reconstructed value of e
- e': The error from quantizing e

It can be seen that the prediction error e_n is quantized, and after channel encoding, forms the digital message. The quantization error e' is the content of the

analog message. In our analysis, we assume the quantizer is capable of reaching the rate-distortion bound. Later in section 4.3, where we implement the system, we use a scalar quantizer followed by an entropy block encoder, which gets close to the rate distortion bound.

We also assume that the channel encoder is capable of reaching the channel capacity. Codes such as Turbo codes [35] and LDPC codes [49] are examples of near-capacity reaching channel codes.

The analog message, e', first goes through the linear encoder block, where its power is adjusted to the amount of power that is allocated to the analog part of the system, P_A. Hence, e' is multiplied by $\sqrt{(\frac{P_A}{{\sigma_{e'}}^2})}$, where ${\sigma_{e'}}^2$ is the variance of e'.

The channel input $\mathbf{Y}$ is the superposition, or summation of the channel encoder and the linear encoder outputs. Since channel encoders work in blocks, the output of the linear encoder can be buffered and sent in blocks as well.

The channel is assumed to be an AWGN channel, where each channel input component Y_i is summed by a zero-mean Gaussian random variable Z_i with variance N_0 to form the channel output, or the received component R_i. Fig. 3.5 shows the diagram of such a channel. The noise components are assumed to be independent

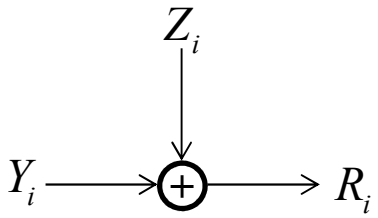

Figure 3.5: AWGN

from each other, as well as from the channel inputs.

In a broadcast setting, there could be multiple channels, where the noise in each one could have different statistics. At the transmitter, hence, the noise variance is

unknown, and is modelled as follows:

$$\begin{aligned} Z_i &\sim \mathcal{N}(0, N_i) \\ N_i &\in [N_{min}, N_{max}] \end{aligned} \tag{3.6}$$

This is the generalization of the two-channel broadcast channel presented in [3], also mentioned in section 2.3. We assume that each receiver can estimate its channel noise variance, while the digital code at the transmitter must be designed for a fixed value of noise variance $N_0 \in [N_{min}, N_{max}]$. This will set the channel coding rate, and subsequently, source coding rate. Channels experiencing a higher level of noise than N_{max} won't be able to decode the digital data.

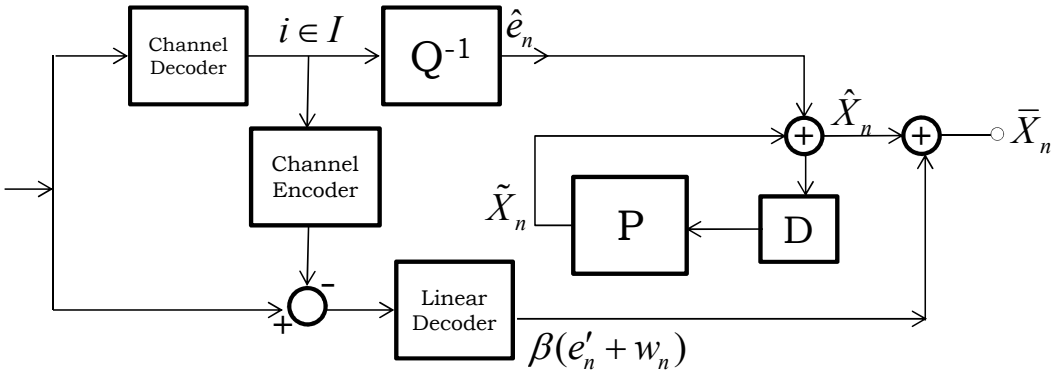

Figure 3.6: HDA Predictive Coding Decoder

The diagram of the decoder which is placed at the receiver is shown in Fig. 3.6. The received signal contains a block of digitally encoded messages, superposed with the block of analog messages, and disturbed by noise. The digital channel coding rate has already been adjusted to the extra interference introduced due to superposition, and hence, the channel decoder can cancel the noise and the interference to detect the digital data message. By subtracting this signal from the received message, the remaining signal will be the analog message, containing the quantization error e', obtained from quantizing the prediction error e, plus the added noise w.

The linear decoder multiplies this signal by parameter β, selected to minimize the mean squared-error (MSE) between the analog signal and its estimate. The value of β is a function of the channel noise variance, and hence, is different at each receiver. This value will be given in the next section.

The digital message, after channel decoding, will be an index set $i \in I$, wchich after being fed to the quantization decoder, will correspond to the reconstructed prediction error, denoted by $\hat{e}_n$ at time n. There is also a predictor block at the receiver, which is identical to the one at the transmitter. The summation of the reconstructed prediction error and the predicted value results in the reconstructed source value of $\hat{X}_n$ at time n. Finally, by adding the output of the linear decoder, which is the estimate of the quantization error, the final reconstructed source value $\overline{X}_n$ will be obtained. The following equations can show the logic behind this decoding procedure:

$$\begin{aligned} X_n &= \tilde{X}_n + e_n \\ e_n &= \hat{e}_n + e'_n \\ X_n &= \tilde{X}_n + \hat{e}_n + e'_n \\ \overline{X}_n &= \tilde{X}_n + \hat{e}_n + \beta(\alpha e'_n + w_n) \end{aligned} \tag{3.7}$$

3.2 End-to-End Distortion

Our aim is to minimize the end-to-end distortion. By taking a second look at (3.7), we can see that the end-to-end distortion can be obtained as the following:

$$
\begin{aligned}
D &= E\left\{\left|X-\overline{X}\right|^2\right\} \\
&= E\left\{\left|X-\left(\tilde{X}+\hat{e}+\beta(\alpha e'+w)\right)\right|^2\right\} \\
&= E\left\{\left|\left(X-\tilde{X}-\hat{e}\right)-\alpha\beta e'-\beta w)\right|^2\right\} \\
&= E\left\{\left(e'-\alpha\beta e'-\beta w\right)^2\right\} \\
&= E\left\{\left(e'\left(1-\alpha\beta\right)-\beta w\right)^2\right\} \\
&= E\left\{e'^2\left(1-\alpha\beta\right)^2+\beta^2 w^2-2\beta\left(1-\alpha\beta\right)e'w\right\}
\end{aligned}
\tag{3.8}
$$

In order to further simplify 3.8, we notice that

$$
\begin{aligned}
E\left\{e'^2\right\} &= \sigma_{e'}^2 \\
E\left\{w^2\right\} &= \sigma_w^2 \\
E\left\{e'\ w\right\} &= 0
\end{aligned}
$$

and thus, the distortion can be reduced to the following.

$$
D = \left(1-\alpha\beta\right)^2\sigma_{e'}^2+\beta^2\sigma_w^2 \tag{3.9}
$$

Notice that $\sigma_{e'}^2$ is the error generated due to the quantization of the prediction error. To find the decoder parameter β that would minimize this distortion, given α, the scaling parameter at the transmitter, we can set the partial derivative of D to zero and solve for β, while verifying the second partial derivative is positive at the reflection point.

$$
\begin{aligned}
&\frac{\partial D}{\partial \beta} = 0 \\
&-2\alpha\left(1-\alpha\beta\right)\sigma_{e'}^2+2\beta\sigma_w^2 = 0 \\
&\beta\left(2\alpha^2\sigma_{e'}^2+2\sigma_w^2\right)-2\alpha\sigma_{e'}^2 = 0 \\
&\beta = \frac{\alpha\sigma_{e'}^2}{\alpha^2\sigma_{e'}^2+\sigma_w^2} \\
&\beta = \frac{\alpha}{\alpha^2+\frac{\sigma_{e'}^2}{\sigma_w^2}}
\end{aligned}
\tag{3.10}
$$

$$\frac{\partial^2 D}{\partial \beta^2} = \left(2\alpha^2 \sigma_{e'}^2 + 2\sigma_w^2\right) \geq 0 \tag{3.11}$$

Having the total power at the transmitter limited to P, this power must be divided between digital and analog powers, denoted by P_D and P_A respectively.

$$\begin{aligned} P_D &= \lambda P \\ P_A &= (1-\lambda) P \end{aligned} \tag{3.12}$$

The parameter λ is the power allocation factor.

$$0 \leqslant \lambda \leqslant 1 \tag{3.13}$$

The case where $\lambda = 1$ corresponds to purely digital transmission, and similarly, the case where $\lambda = 0$ to fully analog transmission. The parameter α is the scaling required to adjust the analog signal power to the allowed level.

$$\alpha = \sqrt{\frac{P_A}{\sigma_{e'}^2}} = \sqrt{\frac{(1-\lambda)\, P}{\sigma_{e'}^2}} \tag{3.14}$$

Substituting (3.14) into (3.10), we can write the expression for β as:

$$\beta = \frac{\sqrt{(1-\lambda)\, P \sigma_{e'}^2}}{(1-\lambda)\, P + \sigma_w^2} \tag{3.15}$$

Now, we can use the expressions for α and β and plug them into (3.9) to obtain the following distortion expression.

$$D = \frac{\sigma_{e'}^2}{1 + \frac{(1-\lambda)P}{\sigma_w^2}} \tag{3.16}$$

The term in the numerator is the variance of the analog message, whereas the denominator is 1+CSNR. It is noteworthy that in a purely digital system, we would have the end-to-end distortion as:

$$D = \sigma_{e'}^2 \tag{3.17}$$

In other words, in a digital system, with close to zero probability of channel errors due to channel coding, the end-to-end distortion is caused by the quantization error,

$\sigma^2_{e'}$, where as in this HDA system, this error can be reduced. Of course, if all the power is allocated to digital, the amount of the quantization error will decrease. This is why the distribution of power between the analog and digital parts must be optimized in order to minimize the distortion. We start by taking a closer look at the term $\sigma^2_{e'}$ through the following steps:

$$\begin{aligned} X_n &= \rho X_{n-1} + \varepsilon_{n-1} \\ \tilde{X}_n &= \rho \hat{X}_{n-1} \qquad \Longrightarrow \qquad (closed-loop\ predictor) \\ e_n &= \rho \left(X_{n-1} - \rho \hat{X}_{n-1} \right) + \varepsilon_{n-1} \\ e_n &= \rho e'_{n-1} + \varepsilon_n \end{aligned} \tag{3.18}$$

Assuming the quantization error is wide-sense stationary, we will have:

$$E\left\{e'^2_{n-1}\right\} = \sigma^2_{e'} \tag{3.19}$$

Also, the random sequence ε at the current instant is uncorrelated with its past values, as well as the past values of e'. Hence,

$$E\left\{e'_{n-1}\varepsilon_n\right\} = 0 \tag{3.20}$$

We can therefore arrive at the following:

$$\sigma^2_e = \rho^2 \sigma^2_{e'} + \sigma^2_\varepsilon \tag{3.21}$$

Now, assuming e is a Gaussian process, we can also get the quantizer rate distortion expression. This assumption is valid for the case of open-loop prediction, where $\tilde{X}_n = \rho X_{n-1}$, and hence $e_n = \varepsilon_n$.

$$\sigma^2_{e'} = \sigma^2_e 2^{-2R_Q} \tag{3.22}$$

Combining (3.21) and (3.22), we can obtain the following.

$$\begin{aligned} \sigma^2_e &= \frac{\sigma^2_\varepsilon}{1 - \rho^2\ 2^{-2R_Q}} \\ \sigma^2_{e'} &= \frac{\sigma^2_\varepsilon}{1 - \rho^2\ 2^{-2R_Q}}\ 2^{-2R_Q} \end{aligned} \tag{3.23}$$

Substituting σ_e^2 into the distortion expression given by (3.16), we can rewrite it as follows.

$$\begin{aligned} D &= \frac{\sigma_\varepsilon^2 \; 2^{-2R_Q}}{1-\rho^2 \; 2^{-2R_Q}} \cdot \frac{1}{1+\frac{(1-\lambda)P}{\sigma_w^2}} \\ D &= \frac{\sigma_\varepsilon^2}{2^{\;2R_Q}-\rho^2} \cdot \frac{1}{1+\frac{(1-\lambda)P}{N}} \end{aligned} \tag{3.24}$$

We replaced the term σ_ε^2 with N, representing the noise variance each receiver experiences. Recalling the source-channel separation, stated in (2.4), we need to determine the capacity of the digital channel, since it is the upper bound of the quantization rate R_Q. With an input power of λP, while being designed for the noise variance level of N_0, we get the following capacity for the digital channel.

$$C = \frac{1}{2}\log\left(1+\frac{\lambda P}{N_0+(1-\lambda)\,P}\right) \tag{3.25}$$

The denominator also counts for the interference from the analog signal, with power $(1-\lambda)\,P$. Setting $R_Q = C$, the distortion expression can be expressed as:

$$D = \frac{\sigma_\varepsilon^2}{1+\frac{\lambda P}{N_0+(1-\lambda)P}-\rho^2} \cdot \frac{1}{1+\frac{(1-\lambda)P}{N}} \tag{3.26}$$

Notice that at the transmitter, the only parameters that can be varied to change D are λ, the power allocation factor, and N_0, the target noise level at which the digital data is aimed to be decodable. The other parameters are either the source characteristics, such as σ_ε^2 and ρ, or the broadcast channel characteristics, N. The distortion D is a random variable, since it is a function of the random variable N. In order to find the average distortion over different values of N, we need to take the following expectation.

$$E\{D\} = \frac{\sigma_\varepsilon^2}{1+\frac{\lambda P}{N_0+(1-\lambda)P}-\rho^2} \cdot E\left\{\frac{1}{1+\frac{(1-\lambda)P}{N}}\right\} \tag{3.27}$$

Evaluatoin of this expectation requires the pdf for the variable N. In order to have a reasonable assumption, we consider a slightly different problem, where instead of having a broadcast channel with different noise variance levels, we are dealing with

a fading channel without the channel state information present at the transmitter. In a fading channel, the amplitude of the transmitted signal can randomly vary, and hence, the receiver CSNR will be random as well [4]. To the transmitter, this is essentially equivalent to transmitting to a broadcast channel where the received CSNR is also random and unknown.

There are several probability distributions that can be used to model fading channels, examples of such include Rayleigh and Rycian distributions. In general, a fading channel can be modelled as shown below in Fig. 3.7.

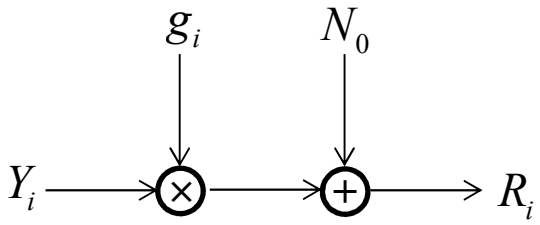

Figure 3.7: Fading Channel

The channel noise is assumed to be white Gaussian. The channel gain g is also a random variable, and it could have either a discrete or continuous sample space, following some distribution such as Rayleigh. Such a channel is hence called the Rayleigh channel [4].

Since the channel state information is not available at the transmitter, the digital code must be designed for a fixed rate, and hence, for a fixed CSNR. This corresponds to the most severe fading level the digital code will work at, below which the digital data will be undecodable. We designate this fading level by g_0. The probability that the channel gain goes lower than this level is denoted as outage probability P_{out}.

$$P_{out} = Pr\left(g < g_0\right) \tag{3.28}$$

Since severe fading is inevitable, systems must tolerate a certain level of outage probability [4], where higher tolerance to outage probability results in a higher

transmission rate. In this model, the noise variance is considered to be constant, and it is represented by N.

The steps in (3.8)-(3.26) can be repeated in a very similar manner to arrive at the following distortion expression.

$$D = \begin{cases} D_1 & g \geq g_0 \\ \sigma_X^2 & g < g_0 \end{cases} \tag{3.29}$$

$$D_1 = \frac{\sigma_\varepsilon^2}{1 + \frac{\lambda\ g_0^2\ P}{N+(1-\lambda)g_0^2 P} - \rho^2} \cdot \frac{1}{1 + \frac{(1-\lambda)g^2 P}{N}}$$

$$E\{D\} = E\{D_1 | g \geq g_0\}\ .\ Pr(g \geq g_0) + \sigma_X^2 . Pr(g < g_0) \tag{3.30}$$

$$E\{D_1 | g \geq g_0\} = \frac{\sigma_\varepsilon^2}{1 + \frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P} - \rho^2}\ .\ E\left\{ \frac{1}{1 + \frac{(1-\lambda)g^2 P}{N}} \mid g \geq g_0 \right\} \tag{3.31}$$

Evaluation of the conditional expectation requires a pdf for fading gain g. We consider two types of fading: discrete fading and Rayleigh fading. For the discrete case, (3.31) can be rewritten in the following way, where the fading sample space is of size n, taking values in the range $\{g_1, g_2, ...g_n\}$, with the corresponding probabilities $\{q_1, q_2, ...q_n\}$.

$$E\{D_1 | g \geq g_0\} = \frac{\sigma_\varepsilon^2}{1 + \frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P} - \rho^2} . \frac{\sum_{i=j}^{n} \frac{q_i}{1+\frac{(1-\lambda)g_i^2 P}{N}}}{\sum_{i=j}^{n} q_i} \tag{3.32}$$

$$g_0 < g_j$$

The outage probability can be written as follows.

$$P_{out} = Pr(g < g_0) = \sum_{i=1}^{j-1} q_i \tag{3.33}$$

We can then rewrite the average distortion expression (3.30) as follows.

$$\begin{aligned} E\{D\} &= \frac{\sigma_\varepsilon^2}{1+\frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P}-\rho^2}\cdot\frac{\sum_{i=j}^{n}\frac{q_i}{1+\frac{(1-\lambda)g_i^2 P}{N}}}{\sum_{i=j}^{n} q_i}\ \sum_{i=j}^{n} q_i\ +\ \sigma_X^2\sum_{i=1}^{j-1} q_i \\ &= \frac{\sigma_\varepsilon^2}{1+\frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P}-\rho^2}\cdot\sum_{i=j}^{n}\frac{q_i}{1+\frac{(1-\lambda)g_i^2 P}{N}}\ +\ \sigma_X^2\sum_{i=1}^{j-1} q_i \end{aligned} \tag{3.34}$$

For the case of Rayleigh fading, the channel gain g is assumed to follow a Rayleigh distribution with parameter θ_g, which results in the following pdf, mean and variance [46].

$$f_g(g) = \frac{g}{\theta_g^2} e^{\left(-\frac{g^2}{2\theta_g^2}\right)} \tag{3.35}$$

$$\mu_g = \theta_g\sqrt{\frac{\pi}{2}} \tag{3.36}$$

$$\sigma_g^2 = \frac{4-\pi}{2}\theta_g^2 \tag{3.37}$$

Using the Rayleigh pdf, we can expand the conditional and average distortions.

$$E\{D_1|g\geq g_0\} = \frac{\sigma_\varepsilon^2}{1+\frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P}-\rho^2}\cdot\frac{\int_{g_0}^{\infty}\frac{1}{1+\frac{(1-\lambda)g^2 P}{N}}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g}{\int_{g_0}^{\infty}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g} \tag{3.38}$$

$$P_{out} = \int_0^{g_0}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g = 1-e^{\frac{-g_0^2}{2\theta_g^2}} \tag{3.39}$$

$$\tag{3.40}$$

$$E\{D\} =$$

$$\frac{\sigma_\varepsilon^2}{1+\frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P}-\rho^2}.\frac{\int_{g_0}^{\infty}\frac{1}{1+\frac{(1-\lambda)g^2P}{N}}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g}{\int_{g_0}^{\infty}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g}\int_{g_0}^{\infty}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g\ +\ \sigma_X^2\int_0^{g_0}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g$$

$$=\frac{\sigma_\varepsilon^2}{1+\frac{\lambda\ g_0^2 P}{N+(1-\lambda)g_0^2 P}-\rho^2}.\int_{g_0}^{\infty}\frac{1}{1+\frac{(1-\lambda)g^2P}{N}}\frac{g}{\theta_g^2}e^{\frac{-g^2}{2\theta_g^2}}d_g+\ \sigma_X^2\left(1-e^{\frac{-g_0^2}{2\theta_g^2}}\right) \tag{3.41}$$

Having obtained the average distortion expressions, we can proceed with formulating our problems.

3.3 Problem Formulation

By looking more closely at the expressions (3.41) and (3.34), it can be seen that for a given source and a channel, both average distortion expressions are functions of the deterministic parameters λ and $g0$, which are in fact the only design parameters. We hence formulate two problems of interest. First, for the case where there is a ceiling for the tolerable outage probability (corresponding to a maximum g_0), we find the optimum power allocation factor λ^* which minimizes the average distortion. The second considered problem is the case where outage probability is flexible, and thus we find the optimum parameters λ^* and g_0^* that lead to the minimum average distortion. The first problem can mathematically be represented as follows.

$$\begin{aligned}&\min_{\lambda}\ E\{D|\,P_{out}\}\\ &s.t.\ 0\leq\lambda\leq 1\end{aligned} \tag{3.42}$$

The constraint arises from the fact that λ is an allocation factor, where the case $\lambda = 0$ corresponds to all the power being allocated to analog, and $\lambda = 1$ to digital. Notice that P_{out} and g_0 are fixed in this case. For the case where there is no

restriction placed on the outage probability, however, P_{out} is allowed to vary, and hence, the optimum value of g_0 must be determined as well. We arrive at the following second optimization problem.

$$\begin{aligned} &\min_{\lambda, g_0} \; E\{D\} \\ &s.t.\; 0 \leq \lambda \leq 1 \\ &and\; 0 \leq g_0 \leq 3\theta_g \end{aligned} \tag{3.43}$$

In this formulation, the additional constraint placed on g_0 is due to the fact that in a Rayleigh distribution, the interval $[0, 3\theta_g]$ covers 99% of all the possible occurrences, where θ_g is the Rayleigh distribution parameter.

We essentially have four optimization problems, which can be summarized as follows:

- Specified Outage - Discrete Fading
- Specified Outage - Rayleigh Fading
- Unspecified Outage - Discrete Fading
- Unspecified Outage - Rayleigh Fading

The integral in (3.41) does not have a closed form solution In the following section, the analysis of these formulated problems are presented. The performance of our proposed HDA scheme under different channels is also investigated and compared to those of other schemes in chapter 4.

3.4 Analysis

In this section, the four formulated problems in the last section are analyzed and studied in more details. We start by analyzing the case where maximum tolerable outage probability is specified.

The average distortion expressions (3.34) and (3.41) are functions of several source and channel parameters, as well as the design parameters λ^*. To prove their convexity, the second derivative must satisfy $\frac{\partial^2 E\{D\}}{\partial \lambda^2} > 0$. This is not practical to be performed analytically, due to the large number of parameters and the many conditions they impose on λ which do not yield any meaningful information. We rather consider practical source and channel parameter values, and find the corresponding convexity conditions.

Let us assume that the source follows a first-order Gauss-Markov source, with unit variance $\sigma_X^2 = 1$, and correlation parameter $\rho = 0.9$.

$$\begin{aligned} X_n &= 0.9X_{n-1} + \varepsilon_n \\ \varepsilon &\sim \mathcal{N}(0, {\sigma_\varepsilon}^2) \\ X &\sim \mathcal{N}(0,1) \\ \sigma_\varepsilon^2 &= \left(1-\rho^2\right)\sigma_X^2 = 0.19 \end{aligned} \tag{3.44}$$

The channel is modelled as a Rayleigh channel, with the parameter $\theta = 0.5$, hence resulting in the following fading parameters.

$$\begin{aligned} f_g(g) &= 4ge^{\left(-2g^2\right)} \\ F_g(g) &= \left(1 - e^{\frac{-g^2}{2\theta_g^2}}\right) \\ \mu_g &= 0.5\sqrt{\frac{\pi}{2}} \approx 0.63 \\ \sigma_g^2 &= \frac{4-\pi}{2}\theta_g^2 \approx 0.1 \end{aligned} \tag{3.45}$$

For the case of discrete fading channel, the same fading channel is uniformly discretized to four and eight levels, resulting in the following channel gains and corre-

sponding pmfs.

$$\begin{aligned} G_4 &= [0.1885\ 0.5655\ 0.9425\ 1.3195] \\ Q_4 &= [0.2452\ 0.4412\ 0.2451\ 0.0685] \\ G_8 &= [0.0943\ 0.2828\ 0.4713\ 0.6598\ 0.8483\ 1.0368\ 1.2253\ 1.4138] \\ Q_8 &= [0.0679\ 0.1773\ 0.2348\ 0.2064\ 0.1522\ 0.0929\ 0.0477\ 0.0208] \end{aligned} \tag{3.46}$$

Assuming a maximum tolerable outage probability of 10% results in the following values of g_0.

- 4-Level Discrete Fading: $g_0 = g_1 = 0.1885$
- 8-Level Discrete Fading: $g_0 = g_2 = 0.2828$
- Rayleigh Fading: $g_0 = F_g^{-1}(P_{out}) = \approx 0.23$

By solving the inequality $\frac{\partial^2 E\{D\}}{\partial \lambda^2} > 0$, it can be shown that all the three scenarios result in a convex objective function for the range $\lambda \in [0, 1]$, which is the range of values an allocation factor can take. It should be noted that the computations were performed with the help of Mathematica. To help gain more insight into the behaviour of these objective functions, they were also plotted as a function of λ in MATLAB, as illustrated below.

It can be seen from Fig.3.8-3.10 that the objective functions are convex for the selected values, although this does not mean that they remain convex for all other values as well. We also show the behaviour of the expected distortion for the Rayleigh channel, as an example, for different values of ρ, the source correlation parameter, $\frac{P_T}{N}$, total power to noise ratio, and θ_g, the fading parameter. These figures can provide more information about the convexity of the objective function, as well as more insight into how the optimum power allocation varies based on different source and channel parameters.

In Fig.3.11, the average distortion is plotted against different values of λ for different values of total power to noise ratio $\frac{P_T}{N}$. It can be seen that the distortion function

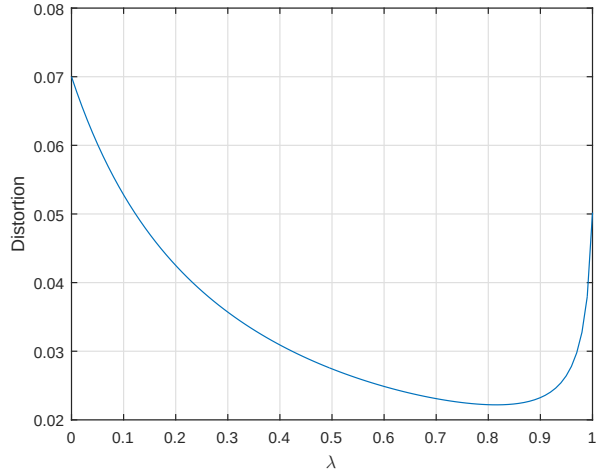

Figure 3.8: Average Distortion Vs. λ: Four-State Fading Channel

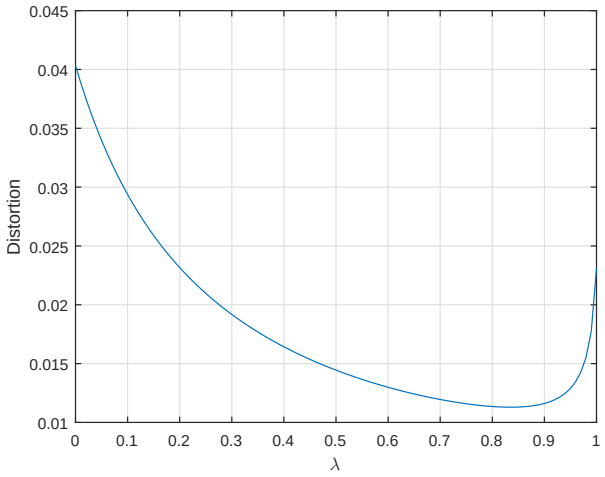

Figure 3.9: Average Distortion Vs. λ: Eight-State Fading Channel

remains convex for all the total CSNR values, even at very high CSNRs. The optimum value of λ shifts more toward the right as total CSNR increases, suggesting that the proposed HDA system won't provide significant average improvement if the power is very high.

When plotted for different values of source correlation parameter, ρ, it can be seen in Fig.3.12 that for highly correlated sources, the system performance at $\lambda = 0$, corresponding to fully analog, is very poor. For a source with a very low value of

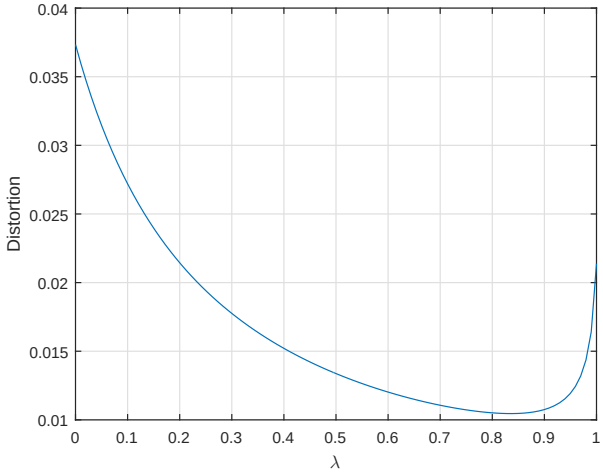

Figure 3.10: Average Distortion Vs. λ: Rayleigh Fading Channel

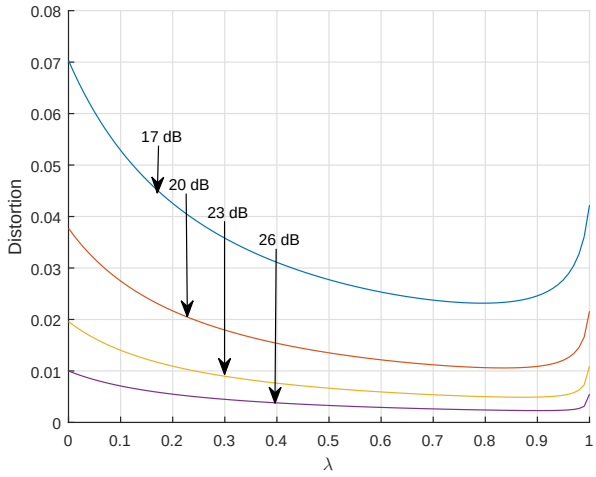

Figure 3.11: Average Distortion Vs. λ for Different Values of Total CSNR

$\rho = 0.3$, the curve is still convex, but the optimum power allocation value occurs at $\lambda = 0.05$. This is consistent with the previously stated fact that for an iid Gaussian source, allocating all the power to analog is the optimum way. It is also interesting to see that the value of ρ does not make a significant performance change when all the power is allocated to digital. By inspecting the distortion expression (3.41), we can observe the insignificant role of ρ when $\lambda = 1$.

Finally, Fig.3.13 depicts the effect of the Rayleigh fading parameter θ_g on the

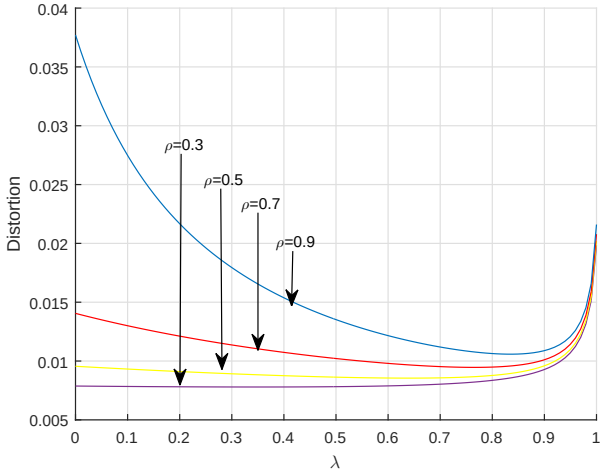

Figure 3.12: Average Distortion Vs. λ for Different Values of Source Correlation

distortion curves. A higher value θ_g means a higher mean fading value, which translates into a higher received CSNR. It's obvious hence that better channel conditions result in lower average distortion. Also as discussed, lower received CSNR shifts the optimum value of λ more to the left, resulting in more power being allocated to the analog part of the system. For the case of unspecified outage, optimization of the

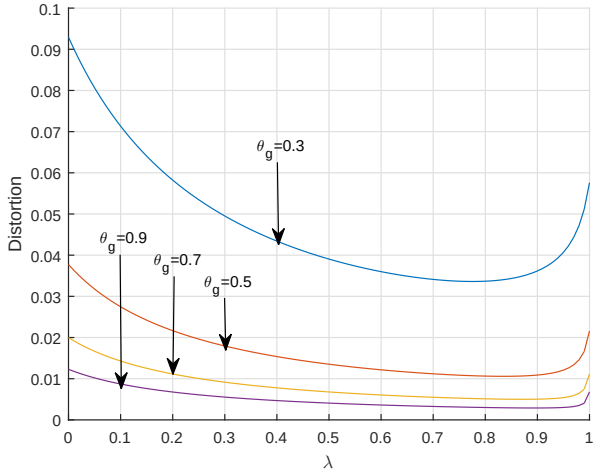

Figure 3.13: Average Distortion Vs. λ for Different Values of Rayleigh Parameter

distortion expressions given by (3.34) and (3.41) involves two variables, as shown in problem formulation (3.43). To examine the convexity of this objective function, the

Hessian matrix $\begin{bmatrix} \frac{\partial^2 E\{D\}}{\partial \lambda^2} & \frac{\partial^2 E\{D\}}{\partial \lambda \ \partial g_0} \\ \frac{\partial^2 E\{D\}}{\partial \lambda \ \partial g_0} & \frac{\partial^2 E\{D\}}{\partial g_0^2} \end{bmatrix}$ was found to be positive definite, assuming the source and channel parameters given by (3.44)-(3.46). A 3-D plot for this distortion expression in dB against λ and g_0 is plotted in Fig.3.14.

It is also interesting to see how each of λ and g_0 affect the average distortion

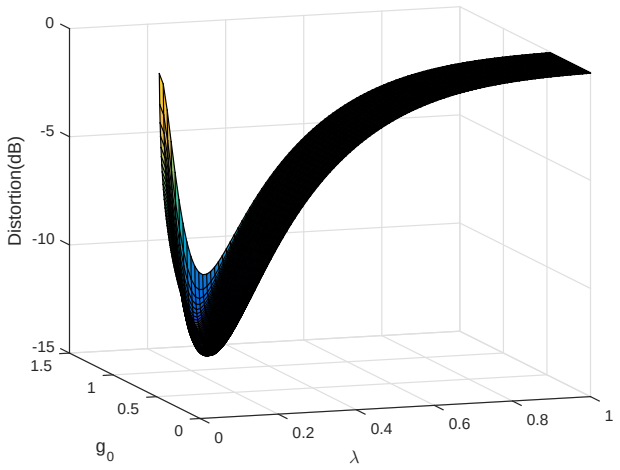

Figure 3.14: Average Distortion Vs. λ and g_0

individually. We can set $\lambda = \lambda_i$ and find the optimum $g_{0i}^* \in [0, 3\theta_g]$ and the corresponding $E\{D(g_0*, \lambda_i)\}$. In Fig.3.15, $E\{D(g_{0i}^*, \lambda_i)\}$ is plotted against $\lambda_i \in [0, 1]$.

Similarly, $E\{D(g_{0_i}, \lambda_i{}^*)\}$ can be computed and plotted against $g_{0_i} \in [0, 3\theta_g]$, as illustrated in Fig.3.16. Together, these two plots show the existence of a global minimum for the selected source and channel parameters over the feasible set determined by the constraints set on λ and g_0.

The optimization constraints in both (3.42) and (3.43) define sets that are convex. We have considered discrete and Rayleigh fading, with specified and unspecified outage probabilities, resulting in four optimization problems. Since the presented objective functions are convex, defined over convex sets, we are dealing with four convex optimization problem with inequality constraints. In the next section, another coding scheme known as Multi-Layer coding is introduced, which can be

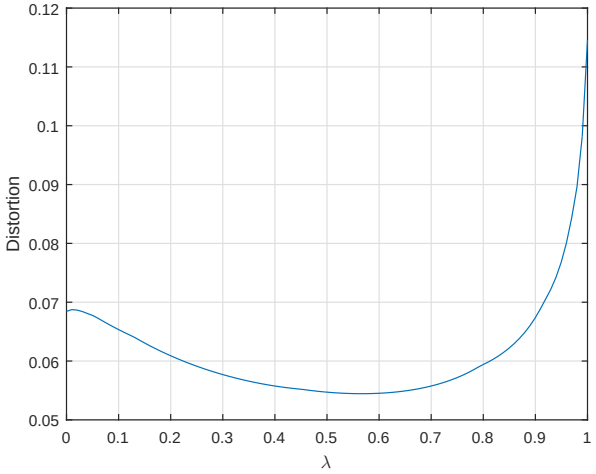

Figure 3.15: Average Distortion Vs. λ at ${g_{0_i}}^*$

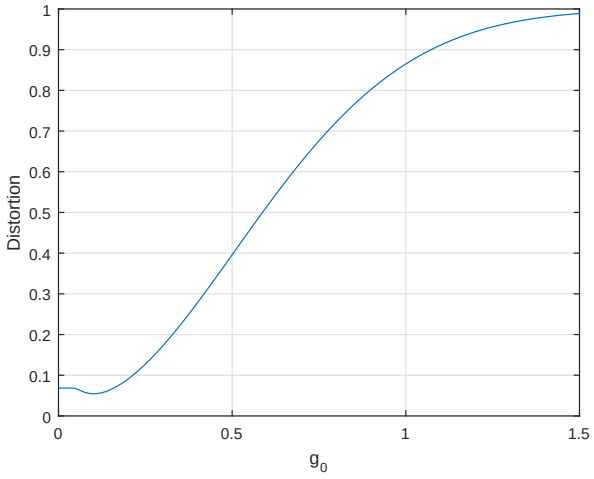

Figure 3.16: Average Distortion Vs. $g0$ at ${\lambda_i}^*$

considered as a rival to our HDA system. ML coding is analyzed and its performance results are compared with those of the proposed HDA system. Optimizing the M-L coding system also results in four convex optimization problems with inequality constraints. Such problems can be solved by different numerical techniques, such as interior-point methods. In this work, though, we opted to solve them using MAT-LAB built-in function *fmincon*, since the focus of this thesis is mainly study of HDA systems. The systems are studied under different scenarios, and their performances

are computed and compared with each other. These results are presented in chapter 4.

3.5 Multi-Layer(ML) Coding

ML coding can be considered as a competitor to HDA coding, as it is also meant to provide better decoding quality to channels with higher CSNR. Similar to HDA, an important factor in designing an ML scheme is the power allocation between the different layers. Inspired by the work in [50], where the number of layers is matched to the number of fading levels, we found the optimum power allocation for two ML schemes, with 4 and 8 levels, corresponding to two considered discretized Rayleigh channels. Similar to our HDA scheme, the two ML coders also incorporate a predictive coder for source compression, and then code the prediction error to different resolutions. The power allocation proposed in [50] could not be applied, however, due to the difference in the source model and the encoder structure.

In an ML coding scheme, the base layer is the most coarse description of the source, decodable at all CSNRs. The receivers experiencing the second lowest CSNR will be able to decode the two lowest layers, and so on. In general , for an iid Gaussian source X, and for an AWGN fading channel with M states, each having a channel gain of g_i and a corresponding probability of q_i, the average distortion bound of an ML code, denoted as D_{ML}, can be found through the following sets of equations, where R_i is the rate associated with the i^{th} coding level, and D_i is the decoding

distortion at the i^{th} CSNR.

$$\begin{aligned} R_1 &= \frac{1}{2} log_2 \left(1 + \frac{g_1^2 P_1}{N}\right) \\ R_2 &= \frac{1}{2} log_2 \left(1 + \frac{g_2^2 P_2}{N + g_2^2 P_1}\right) \\ &\cdot \\ &\cdot \\ &\cdot \\ R_i &= \frac{1}{2} log_2 \left(1 + \frac{g_i^2 P_i}{N + g_i^2 \sum_{j=1}^{i-1} P_j}\right) \qquad i > 1 \end{aligned} \tag{3.47}$$

$$\begin{aligned} D_1 &= \sigma_X^2 2^{-2R_1} \\ D_2 &= \sigma_X^2 2^{-2(R_1+R_2)} \\ &\cdot \\ &\cdot \\ &\cdot \\ D_i &= \sigma_X^2 2^{-2\left(\sum_{j=1}^{i} R_j\right)} \qquad i > 1 \end{aligned} \tag{3.48}$$

$$D_{ML} = \sum_{j=1}^{M} q_j D_j \tag{3.49}$$

For a source described by (3.44), the ML coder combined with predictive source coder can be used in a similar way as our proposed HDA scheme. For such a system, the equations (3.47) and (3.48) should be modified, since the signal being coded is not the source X, but the prediction error e. The prediction error variance was derived and presented in (3.22), and is repeated here below, with the quantization rate R_Q replaced by the rate of the base layer R_0, which is designed for the CSNR corresponding to g_0. In a discrete fading channel with channel gains $g_1, g_2, ..., g_M$, where $g_j > g_k$ if $j > k$, and with corresponding probabilities $q_1, q_2, ..., q_M$, g_0 can only be assigned to one of those M values. Similar to our HDA system, this assignment

is dependent of the outage probability, whether it is limited to a specific ceiling, or if it is allowed to vary.

$$\begin{aligned} R_0 &= \frac{1}{2} log_2 \left(1 + \frac{g_0^2 P}{N}\right) \\ \sigma_e^2 &= \frac{\sigma_\varepsilon^2}{1 - \rho^2\ 2^{-2R_0}} \end{aligned} \tag{3.50}$$

Replacing σ_X^2 by σ_e^2, the average distortion for the predictive ML coder is given by the following equation.

$$D_{ML_{Predictive}} = \frac{\sigma_\varepsilon^2}{1 - \rho^2\ 2^{-2R_Q}} \left(q_1.2^{-2R_1} + q_2.2^{-2(R_1+R_2)} + ... + q_M.2^{-2(R_1+R_2+...+R_M)}\right) \tag{3.51}$$

Finding the minimum average ML distortion involves the optimum distribution of the total power P_t between the M layers, in addition to finding the value of g_0, if there is no limitation on the outage probability. It should be noted that some layers can be allocated no power, and hence resulting in zero rate. There can be two problem formulations, the first of which, presented below, is for the case where the outage ceiling is specified.

$$\begin{aligned} &Min_{P_1,P_2,...,P_M} E\left\{D_{ML}\right\} \\ &s.t.\ 0 \leq P_i \leq P_t \qquad i = 1, 2, ...M \\ &and\ \sum_{i=1}^{M} P_i = P_t \end{aligned} \tag{3.52}$$

If there is no upper bound on the outage probability, however, the optimum amount of outage must be determined as well.

$$\begin{aligned} &Min_{P_1,P_2,...,P_M,g_0} E\left\{D_{ML}\right\} \\ &s.t.\ 0 \leq P_i \leq P_t \qquad i = 1, 2, ...M\ ; \\ &\quad \sum_{i=1}^{M} P_i = P_t \\ &and\ 0 \leq g_0 \leq 3\theta_g \end{aligned} \tag{3.53}$$

As previously explained, the bounds on g_0 are due to the fact that around 99% of occurrences of a Rayleigh distributed random variable with parameter θ_g lie in the interval $[0, 3\theta_g]$.

In the following sections, the performance of our HDA scheme will be illustrated and compared with those of the ML-predictive scheme, purely digital, and purely analog schemes. The results are obtained for the source and channel specifications given by (3.44) and (3.45), for both cases of specified and unspecified outage, and for both discrete and continuous fading channels.

Chapter 4

Results

In this chapter, we present the results obtained on the performance of our proposed HDA scheme, comparing it with its digital counter part, multi-layer coding, as well as pure digital and analog coding schemes. Both continuous and discrete fading channels are considered, where in the former, the fading distribution is assumed to be Rayleigh, and in the latter, the discrete fading levels and their corresponding probabilities are derived from a discretized Rayleigh distribution. In our comparison, we study two important metrics: the average and the instantaneous signal to distortion ratios (SDR). Average SDR is defined as the ratio between the source variance and the optimum average distortion, which we aim at minimizing.

$$\overline{SDR} = \frac{\sigma_X^2}{E_g\left\{D^*\right\}} \tag{4.1}$$

Instantaneous signal to distortion ratio, on the other hand, reflects the optimized system performance at each possible fading fading level (i.e. each CSNR), and is basically a function of the channel gain.

$$SDR(g) = \frac{\sigma_X^2}{D^*(g)} \tag{4.2}$$

For the given source and channel, the two mentioned performance metrics, average and instantaneous SDRs, are found and compared for our proposed HDA scheme, in addition to multi-layer (ML) coding, pure digital coding, and analog transmission.

4.1 Specified Outage

We first present the results for the case where maximum allowed outage is specified. This outage is mapped to a channel gain g_0, which is the minimum channel gain required for the receivers to experience in order to be able to decode the digital data. We analyze the average and the instantaneous signal to distortion ratios (SDR), as previously described by (4.1) and (4.2), and now specialized as follows.

$$\overline{SDR} = \frac{\sigma_X^2}{E_g\left\{D \mid g_0 = g_0^*, \lambda = \lambda^*\right\}} \tag{4.3}$$

$$SDR(g) = \frac{\sigma_X^2}{D(g)|_{g_0=g_0^*,\lambda=\lambda^*}} \tag{4.4}$$

In the remainder of this section, a set of plots are presented that demonstrate the performance of our HDA scheme under both discrete and Rayleigh fading channels, where the source and the channel models are given by (3.44) and (3.45), respectively. Also included are the performance results of the ML-predictive scheme, purely digital, and purely analog systems.

In Fig. 4.1, the average SDR is computed and plotted for different values of the total CSNR, $\frac{P_T}{N}$, where the channel is a four-state fading channel. It can be seen that the HDA-predictive scheme results in the highest average SDR for all the values of $\frac{P_T}{N}$, achieving around an additional 0.6 dB over the ML-predictive scheme designed with 4 layers. Furthermore, the ML-predictive scheme beats the purely digital scheme, which is essentially a one-layer ML scheme. Analog understandably performs the poorest, since it is incapable of taking advantage of the source correlation. The same pattern is repeated for the 8-level fading channel, as seen in Fig. 4.2. It can also be seen that when the same Rayleigh fading channel is discretized to eight levels rather than four, all schemes achieve better results. Evaluation of the other metric, the instantaneous SDR, can provide further insight into the superiority of our HDA scheme. In order to compute the instantaneous SDR, the schemes are optimized for

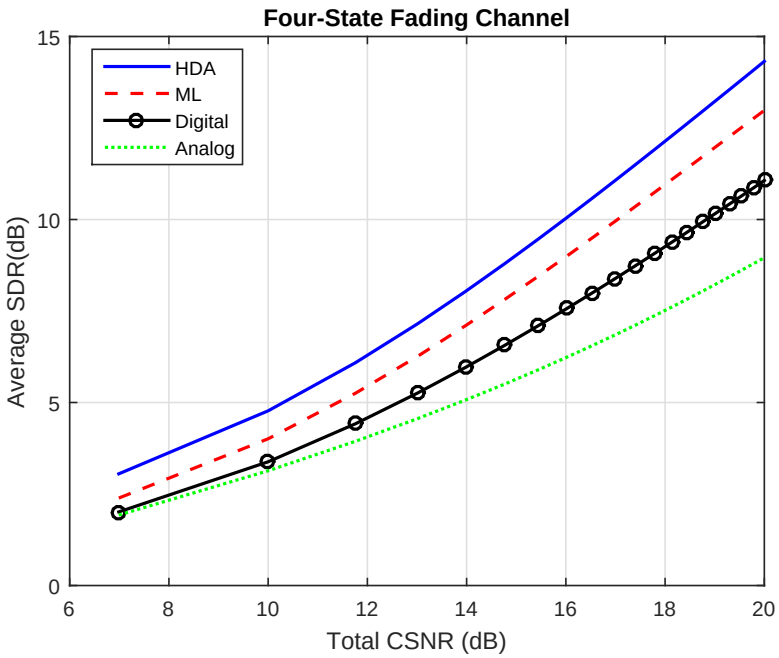

Figure 4.1: Average SDR Vs. Total CSNR: Four-State Fading Channel

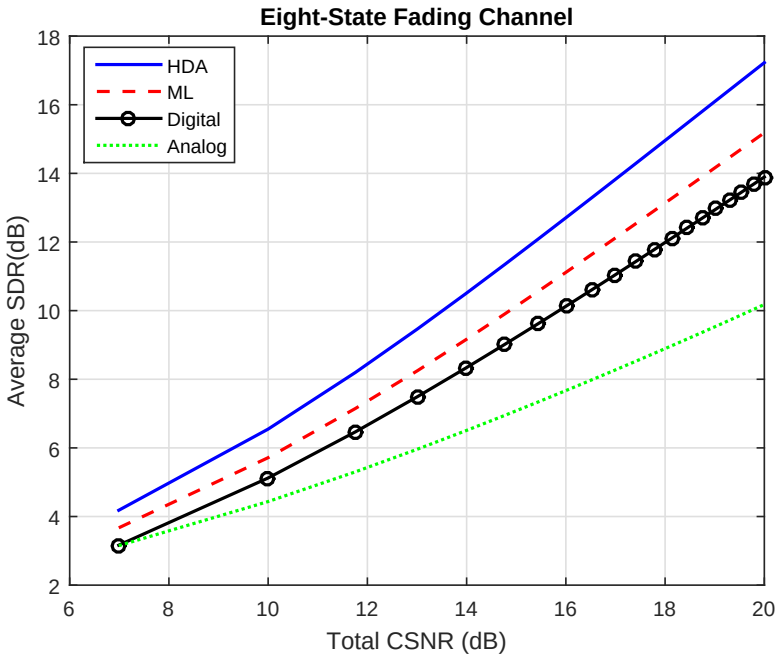

Figure 4.2: Average SDR Vs. Total CSNR: Eight-State Fading Channel

the case where $\frac{P_T}{N} = 100$ or 20 dB and maximum allowable outage is 10%. HDA,

ML, and analog schemes will behave differently at different occurances of the channel gains, or fading levels. This is where the instantaneous CSNR at the receiver comes from. For the case of channel gain of 1, this received CSNR is equal to 20 dB, but at different values of the channel gain, it can take many other values. In Fig. 4.3 and Fig. 4.4, the instantaneous SDR is plotted against the instantaneous CSNR for the two discrete fading channels of four and eight levels, respectively.

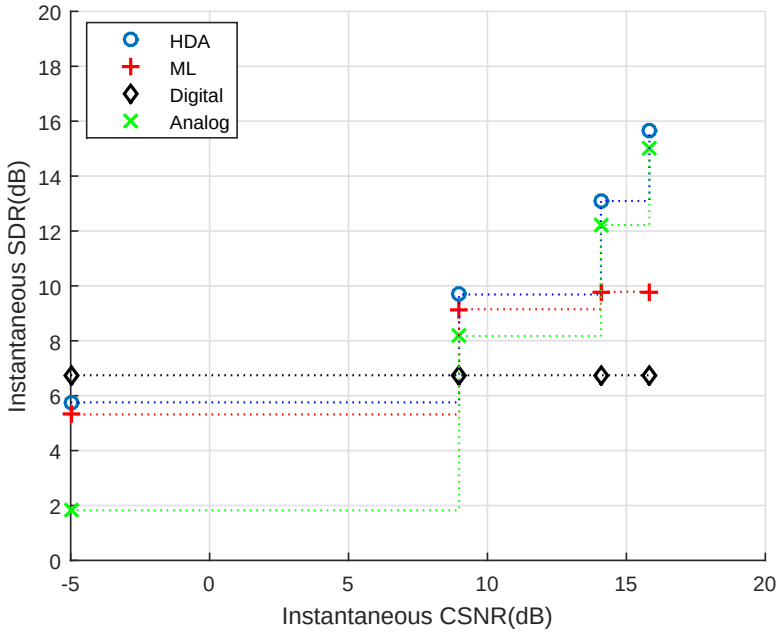

Figure 4.3: Instantaneous SDR Vs. Instantaneous CSNR: Four-State Fading Channel

It can be seen that in the case of 4-level fading channels, there are only four different possible received qualities. Similarly, the 8-level fading channel results in discrete instantaneous SDR of eight values. Channels experiencing higher received CSNR, corresponding to better channel conditions, experience a higher signal quality. In figure (4.3), none of the HDA, ML and digital schemes result in any outage-all receiver are able to decode the data. This cannot happen in a continuous fading channel, though, since there are instances that the CSNR drops to very low values. In the case of 8-level channel, as shown in Fig. (4.4), the poorest channel quality

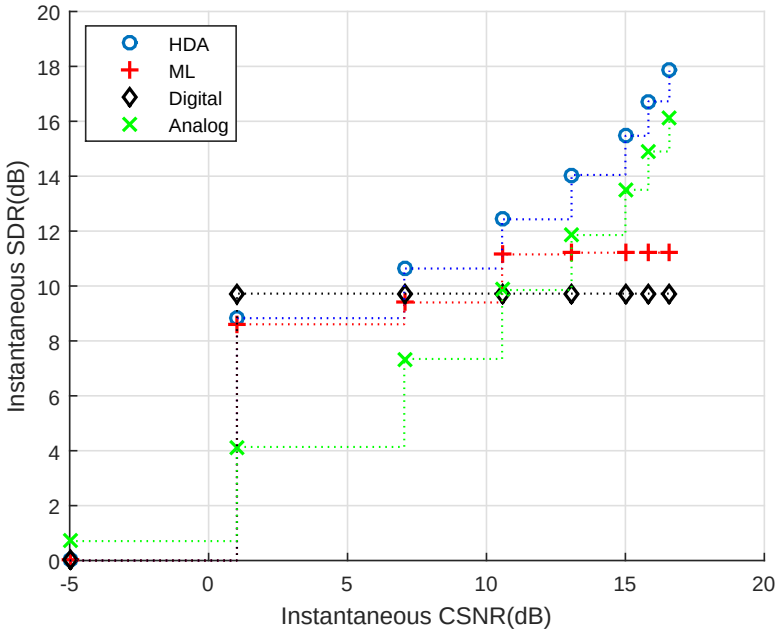

Figure 4.4: Instantaneous SDR Vs. Instantaneous CSNR: Eight-State Fading Channel

does result in outage. The analog scheme, though, never results in complete outage, as it is in general a feature of analog transmission in general. In both figures, HDA-predictive strictly outperforms its counterpart ML-predictive. It only results in lower performance than digital at very low CSNRs, corresponding to deep fades, and compared to analog, always has higher SDR values, except in the case of outage. The other strong feature of HDA-predictive is the graceful degradation it achieves, where unlike the digital scheme, its instantaneous SDR degrades smoothly with the degradation of the channel quality. The ML-predictive scheme does seem to achieve graceful degradation as well, but as it will be shown soon, this is not exactly the case.

In figures (4.5) and (4.6), the average SDR and the instantaneous CSNR are plotted for the case of Rayleigh fading. In this channel, the channel gain amplitudes are not discrete any more, going all the way down to zero. Hence, in such a channel, which is actually a good model for urban environments [4], there must always be some

outage in digital and ML schemes. The same is true for HDA, of course, due to the presence of its digital base layer. When designing an ML scheme, the channel must be considered as discrete, however, since the number of layers are limited, and each layer must be designed for a specific CSNR. Therefore, in this comparison, the same Rayleigh fading channel has been discretized to four and eight levels to compute the performance of ML-4 and ML-8 schemes.

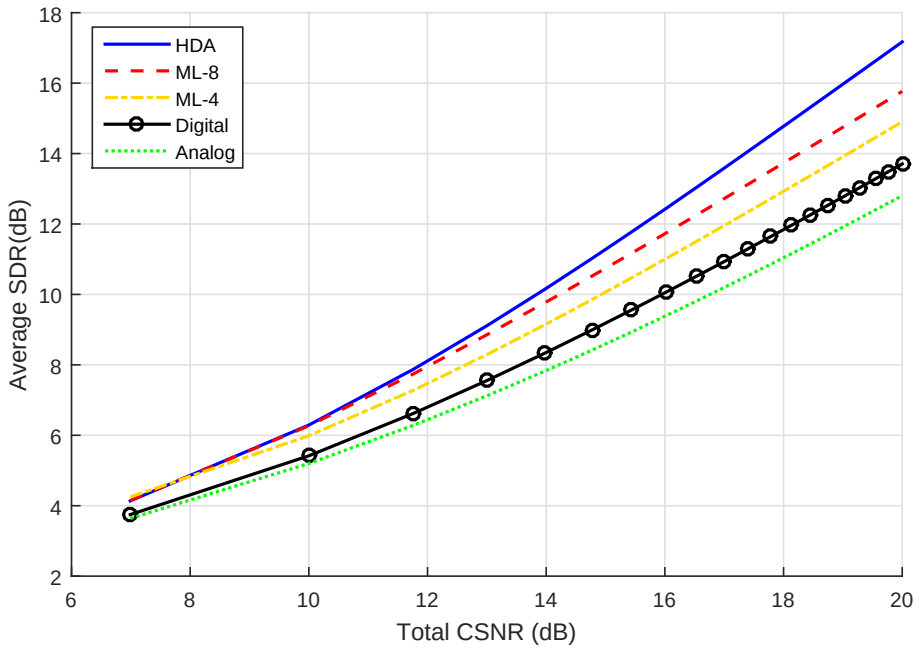

Figure 4.5: Average SDR Vs. Average CSNR: Rayleigh Fading Channel

It is observed that HDA-predictive scheme results in the highest average SDR for the entire range of total CSNRs, followed by ML-8 and ML-4 schemes. The gap between the curves increases at higher total CSNRs. A deeper insight can be taken away from figure (4.6), showing the instantaneous performance of these schemes. We see that except for analog, all other schemes result in some outage, as expected. The outage for HDA-predictive and Digital is the same amount, but for ML-4 and ML-8, the outage probability can only be set to certain values. We can also see that the digital scheme provides two levels of quality only, either outage, or a certain

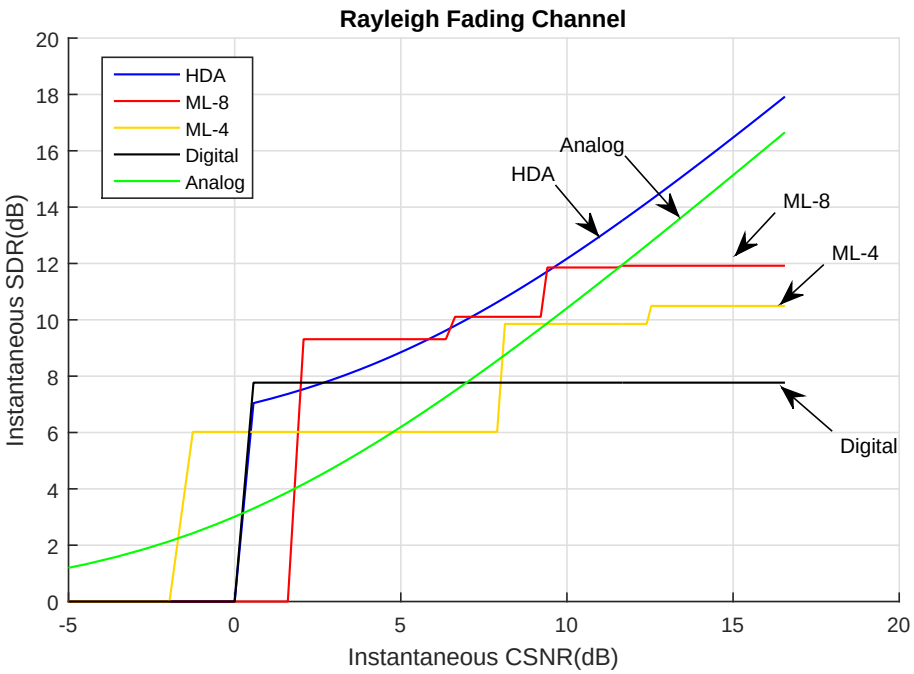

Figure 4.6: Instantaneous SDR Vs. Instantaneous CSNR: Rayleigh Fading Channel

level designed for a user with the weakest channel that is able to decode the data. The ML schemes improve on the digital schemes, providing improvements at higher CSNRs, but this improvement happens only at the discrete designed levels, resulting in stair case curves. Analog provides graceful degradation, outperforming the other schemes at very high CSNRs, but achieving very low quality at low CSNRs. The HDA-predictive scheme however, provides a true graceful degradation, having a shape similar to analog, while providing higher signal qualities than the rest of the schemes at most non-outage CSNRs.

In the next section, the same results presented here are going to be shown for the case where the outage ceiling is not specified.

4.2 Unspecified Outage

For the case where the outage level is free to change, there are two variables that can play a role in minimizing the end to end distortion: the parameter g_0, corresponding to the outage level, and λ, representing the power allocation factor. For the case of a discrete fading channel with four and eight levels of channel quality, the following figures show the average SDR against total CSNR.

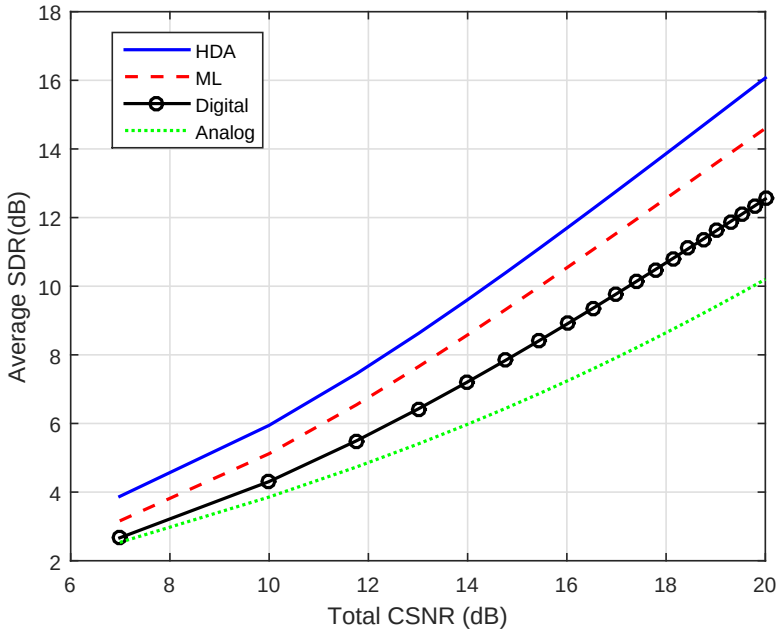

Figure 4.7: Average SDR Vs. Average CSNR: Four-State Fading Channel

We can observe a very similar pattern as those observed in Fig. 4.1 and 4.2, where HDA-predictive provides the highest average signal to distortion ratio, followed by the ML-predictive scheme, digital and analog. By comparing Fig.4.7 with 4.1, and also 4.8 with 4.2, it can be seen that the case where the outage ceiling is not limited does result in a higher performance. In addition, the case where the channel is discretized to more levels causes a higher signal quality.

Similarly, the instantaneous performance of the schemes, as shown in 4.9 and 4.10 for four-state and eight-state channels, reconfirm the results shown in Fig.4.3 and

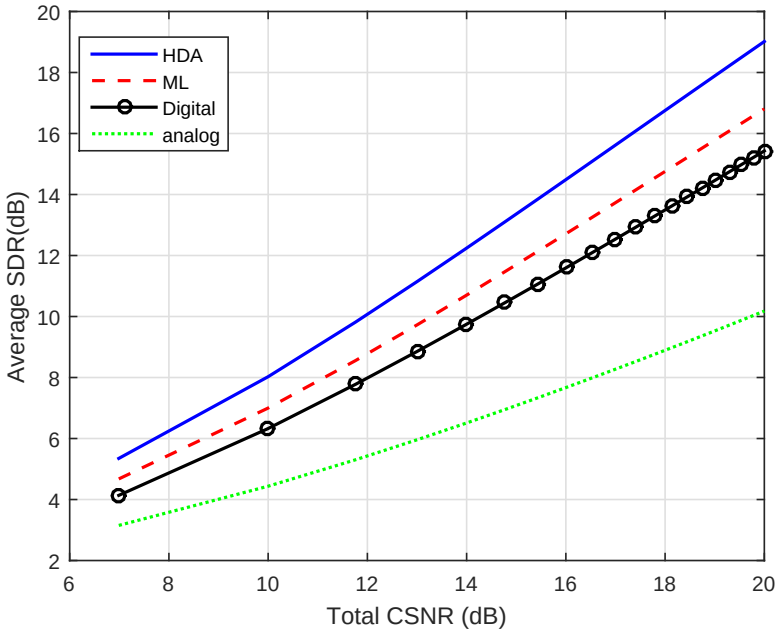

Figure 4.8: Average SDR Vs. Average CSNR: Eight-State Fading Channel

4.4.

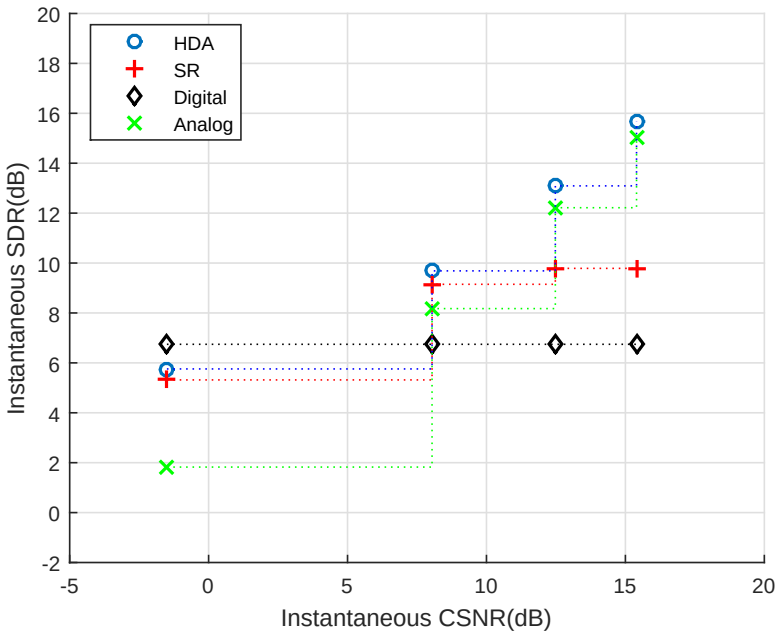

Figure 4.9: Inst SDR Vs. Inst CSNR: Four-State Fading Channel

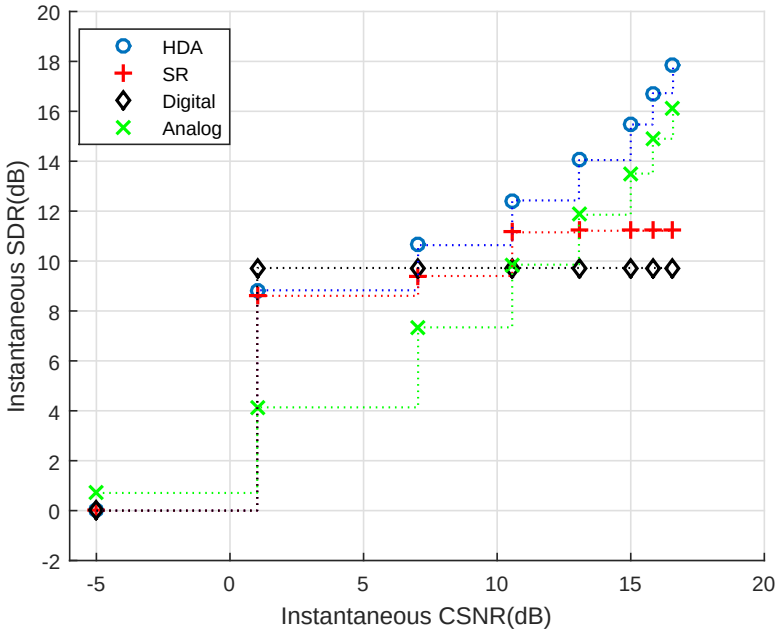

Figure 4.10: Inst SDR Vs. Inst CSNR: Four-State Fading Channel

Finally, a comparison of the studied schemes for a Rayleigh channel is presented in Fig.4.11 and 4.12. Similar to the case of unspecified outage, HDA-predictive provides a higher average SDR for the Rayleigh fading channel. We can see that each scheme selects a different outage level to achieve their best performance, but HDA-predictive offers the highest signal quality. The two ML schemes results in the stair case phenomenon, by providing a limited number of signal quality levels according to the channel CSNR. The number of the stairs in an ML scheme with M layers though, need not be equal to M, as some layers might not be allocated any power. The digital scheme, which is essentially an ML scheme with one layer, provides only one level of quality to all users, when it does not result in outage.

4.3 Simulation Results

In this section, the designed HDA-predictive scheme is used in a communication

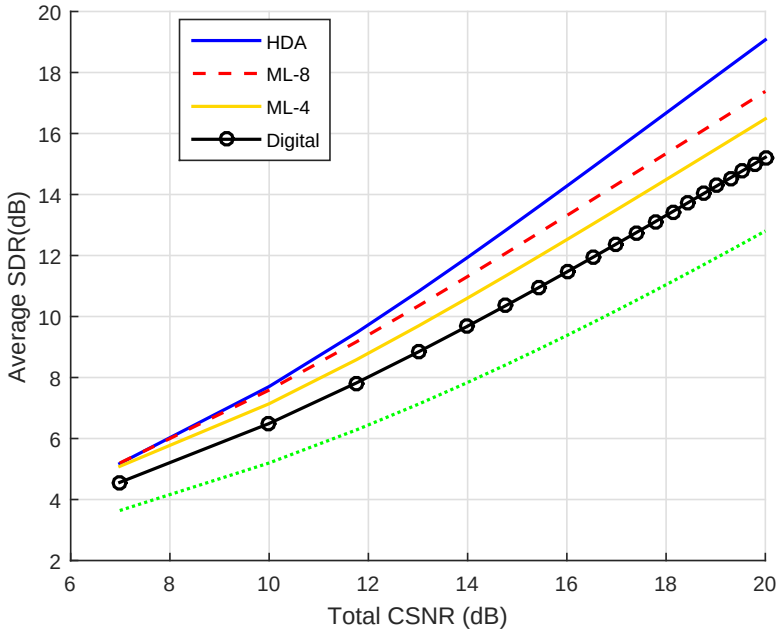

Figure 4.11: Average SDR Vs. Average CSNR: Rayleigh Fading Channel

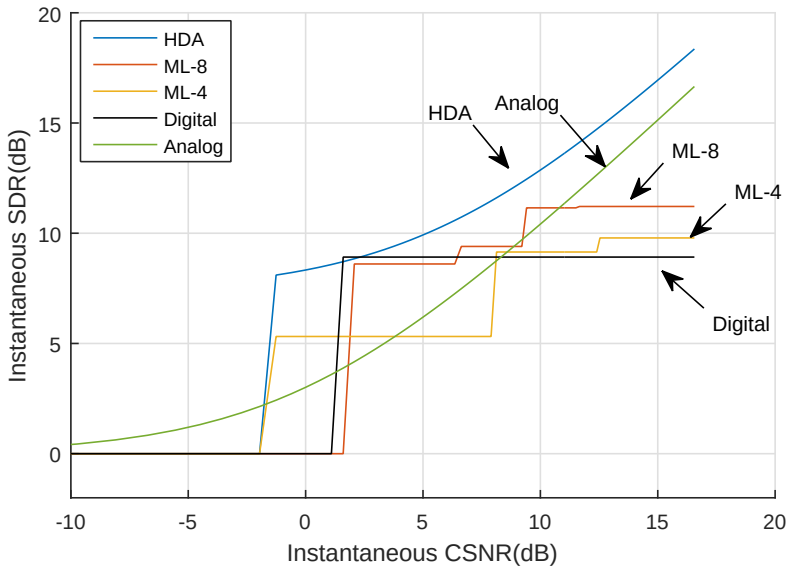

Figure 4.12: Inst SDR Vs. Inst CSNR: Rayleigh Fading Channel

system to compare the practicality of the method, and whether or not it can actually outperform the conventional predictive digital coding. A large sample of data from a first-order Gauss-Markov source, following (3.44), was generated in MATLAB.

This was done by randomly generating 100000 uncorrelated bits, following Gaussian distribution, resulting in a white Gaussian data source. The data was then passed through a first-order all-pole filter, with the following transfer function.

$$H(z) = \frac{1}{1 - \rho z^{-}1} \tag{4.5}$$

In order to source code the data, a first-order predictor filter with parameter $a = \rho$ is used. The prediction error is then fed into a uniform quantizer in cascade with an entropy encoder, a combination that is known to perform well for Gaussian sources [6], while at the same time, being easy to implement. This is as opposed to the use of a vector quantizer which can achieve a nearly optimum performance for high dimensional code spaces, but with the cost of increased complexity. The distortion expression for such combination of a uniform quantizer and an entropy coder, for a Gaussian source with variance σ_S^2, is given by the following.

$$D = \frac{\pi e}{6}\sigma_S^2 2^{-2H} \tag{4.6}$$

The parameter H is the entropy of the output bit stream resulting from the quantizer. This expression is obtained from the relationship between a uniform quantizer step size Δ and its resulting distortion, in addition to the relationship between the input and output entropy of the entropy encoder.

$$\begin{aligned} D &\approx \frac{\Delta^2}{12} \\ H(Q(X)) &= h(X) - log_2(\Delta) \\ h(X) &= log_2\left(\sigma\sqrt{\pi e}\right) \end{aligned} \tag{4.7}$$

Notice that $h(X)$ is the differential entropy of a continuous Gaussian random variable [3]. Comparing this distortion expression to the rate-distortion function of a Gaussian source, we can see that such an encoding scheme increases the distortion by a factor of $\frac{\pi e}{6}$ or 1.53 dB. It should be noted that equation (4.6) does not hold for very small rates. The distortion value actually will be larger than the source

variance σ_S^2 for rates below 0.255, and is not very accurate for slightly higher rates. Therefore, in our simulations, we only studied the cases where the rate was greater than 0.5.

An arithmetic encoder was used for the purpose of entropy coding, since it is a powerful lossless code which can compress a bit stream very close to the source entropy for a sufficiently long data block. The error obtained from the scalar quantization is the analog signal that is sent over the channel.

Typically, the source data stream after compression is channel encoded, where the addition of redundancy in forms of channel codes provides protection against corruption caused by data transmission through a communication channel. Channel coding hence lowers the effective transmission rate, distancing it from the optimal Shannon rate. In our simulation setting, we opted to omit channel encoding, since we were interested to observe the performance of the HDA-predictive scheme under different power allocation factors, including the optimum one. This would essentially translate into different transmission rates, and hence would require designing multiple channel codes, making the simulation test significantly more complicated. Since the purpose of this study is not to test out channel coding schemes, we opted to omit channel coding, operate at the Shannon capacity, and assume the channel does not cause any errors in the transmitted data. This in fact is not an unrealistic assumption, since powerful codes such as LDPC and turbo codes are able to reach close to the Shannon bound, while achieving arbitrarily low probability of transmission error [7]. The analog signal, however, is passed through a Rayleigh fading AWGN channel, as modelled in (3.45), with the Rayleigh parameter $\theta_g = 0.5$, power to noise variance ratio of $\frac{P_T}{N} = 100$, and maximum allowable outage of 10%. The fading was assumed to be slowly-varying, with the channel gains remaining constants over a block of data being transmitted. We used the MATLAB function *raylrnd* to generate the Rayleigh-distributed channel gains, and *randn* to generate WGN samples. At the receiver, the corrupted analog signal is passed through a linear minimum

mean-squared error filter, as described in (3.14). The recovered digital signal, which is the quantized prediction error samples, is used to obtain the quantized source samples. This is done by using a second predictor filter at the receiver side. Finally, the recovered analog signal samples are summed up with the reconstructed digital samples to form the final source sample estimates. This process was explained more thoroughly in chapter 3, and illustrated in Fig. 3.4 and Fig. 3.6.

In the first part of the simulation process, different power allocation factors, λ, were tried out, and the corresponding average end-to-end distortion values were computed, where time average was used as an estimation of the ensemble average. This was done to see if there is indeed an optimum value of λ. It should be noted again that $\lambda \in [0, 1]$, where $\lambda = 0$ corresponds to the case where all the power is allocated to analog, while $\lambda = 1$ means that the system is purely digital. The result is illustrated in Fig. 4.13. It can be seen that the optimum power allocation factor occurs at $\lambda^* = 0.87$, indicating that allocating 87% of the power to the digital part, and the remaining 13% to analog will result in the lowest average distortion. The reason that values smaller than 0.6 were not considered for λ is to have the digital rate greater than 0.5. It can be seen that operating at λ^* improves the average distortion by approximately 1 dB, compared to the fully digital system.

Next, for different values of total CSNR, $\frac{P_T}{N}$, the simulation was run to find the average distortion for the HDA system, operating at λ^*. The obtained average SDR values are shown in Fig. 4.14, alongside with the theoretical performance curve for the HDA system, as well the simulation results for a purely digital system. It can be seen that the HDA-Simulation average SDR follows a similar pattern to that of HDA-Theoretical, while being approximately 2 dB lower. This is expected, as the source coder used in the simulation, uniform quantizer followed by arithmetic encoder, can not reach the optimal performance level. The simulated HDA system, however, outperforms the simulated digital system, which again confirms that $\lambda^* \neq 1$.

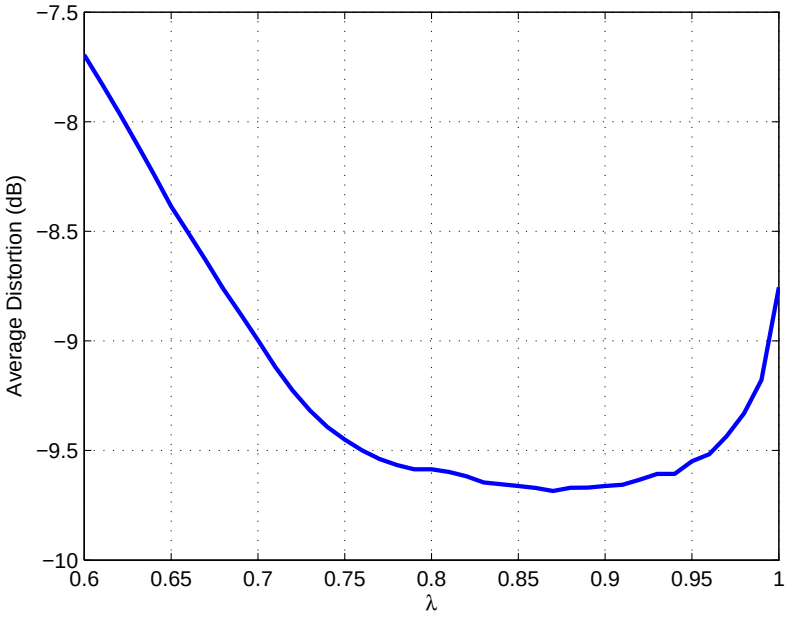

Figure 4.13: Average Distortion Vs. λ

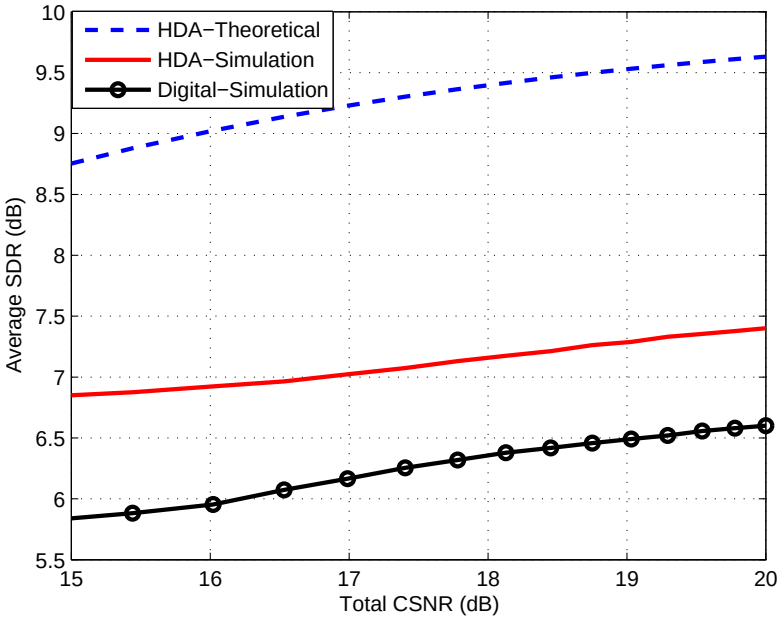

Figure 4.14: Average Distortion Vs. Total CSNR

In Fig. 4.15, the instantaneous SDR is plotted against the Instantaneous CSNR for HDA-Simulated, HDA-Theoretical, and Digital-Simulated systems. Again, it is

observed that the simulated HDA system does not match the theoretical case. It does, however, achieve a graceful degradation in the non-outage parts, a feature that is one of the great advantages of HDA systems. It also outperforms the digital simulated systems for the majority of CSNRs, with the performance difference increasing as the channel conditions improve, i.e. higher instantaneous CSNR. The digital system exhibits the threshold effect by providing two levels of quality only, outage or a constant value.

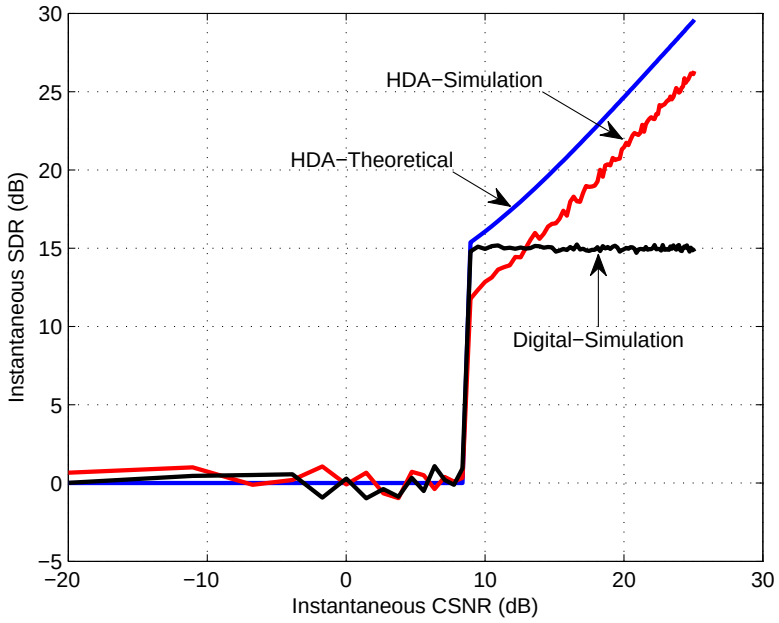

Figure 4.15: Instantaneous SDR Vs. Instantaneous CSNR

Chapter 5

Discussion and Conclusion

In this work, a Predictive Hybrid Digital-Analog coding system was proposed and studied for coding correlated sources. HDA systems can combat the problems associated with digital coding systems, while maintaining many of their advantages. The most important drawbacks found in digital coding are the levelling-off effect, where the signal quality does not improve as the channel condition improves, and the threshold effect, where the signal decoding process breaks down if the channel CSNR falls below the code design point. These problems don't occur in analog systems, however, which is the motivation behind the proposal of HDA systems, where part of the signal is digitally coded, and the other part is transmitted using analog techniques. HDA systems can eliminate levelling-off effect, and can potentially mitigate the threshold effect. In this work, we incorporated a predictive source coder into an HDA coder to be able to code correlated sources using HDA systems as well, a problem that is not studied thoroughly in the literature.

Due to the presence of two coders in an HDA system, digital and analog, the resources must be divided between these two parts. These resources are the available transmission power and bandwidth. A technique called superposition coding is used, where the digital and analog signals at the encoder side are summed up and transmitted over the entire bandwidth, and at the receiver side, separated by the digital

decoder, by treating the analog signal as noise. This comes at the the cost of reduced digital rate, but proves to be optimum for Gaussian sources. Since we focused on correlated Gaussian sources, we opted for superposition coding, and hence avoided the problem of bandwidth splitting between the digital and analog coders. The transmission power, however, must be divided between the two, and it must be done in an optimal way. The optimality criterion we selected was the average end to end decoding distortion. We hence formulated an optimization problem, where the optimum power allocation is found to minimize the average distortion. The source was modelled as a first-order Gauss-Markov source, and the channel as an AWGN fading channel.

We carried out a series of analyses to observe the behaviour of the proposed predictive HDA operating at the optimum power allocation point. The results were categorized into two groups, one where there is no upper bound on the allowable outage probability, and the other one with a specified maximum tolerable outage level. For each category, three cases of fading were considered: 4-level and 8-level discrete fading, in addition to continuous Rayleigh fading. Two performance metrics were computed for each case,one being the average end to end distortion,or equivalently, average signal to distortion ratio (SDR), and the second one being the instantaneous SDR. Average SDR was the metric that we aimed at minimizing, whereas instantaneous SDR shows how the code works under different channel conditions. In order to subjectively evaluate the performance of the predictive HDA system, we evaluated the two metrics for a similar coding technique, a rival to HDA, multi-layer (ML) coding. An ML coder transmits a coarse description of the source intended for users experiencing weak channel conditions, while sending additional details, or refinement information about the source, at multiple levels for stronger users. A predictive source coder was also incorporated into the ML coder to have a fair comparison. In addition, purely analog and digital systems were also evaluated. The results for all the cases consistently proved the superiority

of the HDA-predictive scheme, whether for the average or the instantaneous SDR, while avoiding the levelling-off effect. The ML-predictive scheme does provide the stronger users with higher signal quality, but this improvement in the performance is not smooth like its HDA counterpart; it is rather in a stair-case shape. Increasing the number of layers reduces the performance jumps between these stair case levels, but at the time, increases the complexity of the system design significantly. HDA can actually be considered a generalization of the ML scheme, where slightest improvement in the instantaneous CSNR improves the instantaneous SDR, while being considerably less computationally expensive to design. An ML code with M layers must divide the power optimally between all the M levels, while in the design of an HDA system, the power must be divided between two layers only.

Simulations were also carried out, where a uniform quantizer in cascade with an arithmetic encoder was used to source code the prediction error. This combination is easy to implement and very effective for Gaussian sources, except at very low rates (We used rates higher than 0.5 bits per sample.)In one simulation setting, different levels of power allocation were tested to find the optimum one, showing that at that point, the system performance is superior to that of a purely digital system. Average and Instantaneous SDRs were also found for the implemented HDA-predictive scheme, and the results were compared to the theoretical SDR curves for the HDA-predictive coder, as well as a simulated purely digital system using the same source coder. While beating the digital system, the implemented HDA-predictive scheme did not reach the theoretical SDR values. This was predicted, since the theoretical curves are based on asymptomatic values, which the combination of uniform quantizer and entropy coder cannot reach. The simulations did show, however, that the HDA architecture can be successfully implemented and can outperform digital systems.

Overall, it can be said that HDA coding is a very promising technique, with a great potential to be adapted for broadcasting due to the graceful performance improve-

ment/degradation it provides. There is certainly lots of room for further research into this area. A possible next step to further this work can be to study the cases where there is excess channel bandwidth compared to the source bandwidth. In that case, a joint optimal bandwidth splitting and power allocation must be considered, since superposition coding will no longer be optimal. This will of course increase the design complexity. Another interesting study could be to pay more attention to the content of the analog signal. An example could be splitting the signal in bandwidth, and transmitting the quantization error for the high-energy content only. This can also be accomplished by using transform coding, with a joint power and bit allocation between the signal components for both digital and analog codes. In fact, the quantization error does not necessarily have to be the only candidate for analog transmission.

Another important step that could facilitate the adaption of HDA schemes in the industry standards is the study of more practical sources and codes. More practical codes must be developed, tailored for common audio and video sources.

References

[1] R.G. Gallager. *Principles of Digital Communication*. Cambridge University Press, 2008.

[2] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[3] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. A Wiley-Interscience publication. Wiley, 2006.

[4] A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[5] C.E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[6] A. Gersho and R.M. Gray. *Vector quantization and signal compression*, volume 159. Springer Science & Business Media, 2012.

[7] J. Proakis and M. Salehi. *Digital Communications*. McGraw-Hill International Edition. McGraw-Hill, 2008.

[8] T.J. Goblick Jr. Theoretical limitations on the transmission of data from analog sources. *Information Theory, IEEE Transactions on*, 11(4):558–567, 1965.

[9] J. Ziv. The behavior of analog communication systems. *Information Theory, IEEE Transactions on*, 16(5):587–594, 1970.

[10] U. Mittal and N. Phamdo. Hybrid digital-analog (hda) joint source-channel codes for broadcasting and robust communications. *Information Theory, IEEE Transactions on*, 48(5):1082–1102, 2002.

[11] W. H.R. Equitz and T.M. Cover. Successive refinement of information. *Information Theory, IEEE Transactions on*, 37(2):269–275, 1991.

[12] L. Lastras and T. Berger. All sources are nearly successively refinable. *Information Theory, IEEE Transactions on*, 47(3):918–926, 2001.

[13] M.D. Trott. Unequal error protection codes: Theory and practice. In *Proc. IEEE Information Theory Workshop*, page 11, 1996.

[14] S. Vembu, S. Verdu, and Y. Steinberg. The source-channel separation theorem revisited. *Information Theory, IEEE Transactions on*, 41(1):44–54, 1995.

[15] A. Goldsmith. Joint source/channel coding for wireless channels. In *Vehicular Technology Conference, 1995 IEEE 45th*, volume 2, pages 614–618. IEEE, 1995.

[16] M. Gastpar, B. Rimoldi, and M. Vetterli. To code, or not to code: Lossy source-channel communication revisited. *Information Theory, IEEE Transactions on*, 49(5):1147–1158, 2003.

[17] F. Jelinek. *Probabilistic information theory: discrete and memoryless models*. McGraw-Hill, 1968.

[18] N. Farvardin. A study of vector quantization for noisy channels. *Information Theory, IEEE Transactions on*, 36(4):799–809, 1990.

[19] N. Farvardin and V. Vaishampayan. On the performance and complexity of channel-optimized vector quantizers. *Information Theory, IEEE Transactions on*, 37(1):155–160, 1991.

[20] K. Zeger and A. Gersho. Pseudo-gray coding. *Communications, IEEE Transactions on*, 38(12):2147–2158, 1990.

[21] K. Ramchandran, A. Ortega, K.M. Uz, and M. Vetterli. Multiresolution broadcast for digital hdtv using joint source/channel coding. *Selected Areas in Communications, IEEE Journal on*, 11(1):6–23, 1993.

[22] Z. He, J. Cai, and C. Wen C. Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(6):511–523, 2002.

[23] S.B.Z Azami, P. Duhamel, and O. Rioul. Joint source-channel coding: Panorama of methods. In *Proceedings of CNES workshop on Data Compression*, pages 1232–1254, 1996.

[24] F. Hekland. A review of joint source-channel coding. *Norwegian University of Science and Technology (NTNU)*, 2004.

[25] T.M. Cover. Broadcast channels. *Information Theory, IEEE Transactions on*, 18(1):2–14, 1972.

[26] A. El Gamal and Y.H. Kim. *Network information theory*. Cambridge university press, 2011.

[27] J. Kiirner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. Info. Theory*, 23:60–64, 1977.

[28] C. Tian, S. Diggavi, and S. Shamai. Approximate characterizations for the gaussian source broadcast distortion region. *Information Theory, IEEE Transactions on*, 57(1):124–136, 2011.

[29] W.F. Schreiber. Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions. *Proceedings of the IEEE*, 83(6):958–981, 1995.

[30] S. Shamai, S. Verdú, and R. Zamir. Systematic lossy source/channel coding. *Information Theory, IEEE Transactions on*, 44(2):564–579, 1998.

[31] N. Phamdo and U. Mittal. A joint source-channel speech coder using hybrid digital-analog (hda) modulation. *Speech and Audio Processing, IEEE Transactions on*, 10(4):222–231, 2002.

[32] Z. Reznic, M. Feder, and R. Zamir. Distortion bounds for broadcasting with bandwidth expansion. *Information Theory, IEEE Transactions on*, 52(8):3778–3788, 2006.

[33] M. Skoglund, N. Phamdo, and F. Alajaji. Design and performance of vq-based hybrid digital-analog joint source-channel codes. *Information Theory, IEEE Transactions on*, 48(3):708–720, 2002.

[34] Y. Linde, A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, 28(1):84–95, 1980.

[35] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: Turbo-codes. *Communications, IEEE Transactions on*, 44(10):1261–1271, 1996.

[36] M. Skoglund, N. Phamdo, and F. Alajaji. Hybrid digital-analog source-channel coding for bandwidth compression/expansion. *Information Theory, IEEE Transactions on*, 52(8):3757–3763, 2006.

[37] Y. Wang, F. Alajaji, and T. Linder. Hybrid digital-analog coding with bandwidth compression for gaussian source-channel pairs. *Communications, IEEE Transactions on*, 57(4):997–1012, 2009.

[38] M.P. Wilson, K. Narayanan, and G. Caire. Joint source channel coding with side information using hybrid digital analog codes. *Information Theory, IEEE Transactions on*, 56(10):4922–4940, 2010.

[39] M.H.M. Costa. Writing on dirty paper (corresp.). *Information Theory, IEEE Transactions on*, 29(3):439–441, 1983.

[40] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *Information Theory, IEEE Transactions on*, 22(1):1–10, 1976.

[41] V.M. Prabhakaran, R. Puri, and K. Ramchandran. Hybrid digital-analog codes for source-channel broadcast of gaussian sources over gaussian channels. *Information Theory, IEEE Transactions on*, 57(7):4573–4588, 2011.

[42] S. Jakubczak and D. Katabi. Softcast: one-size-fits-all wireless video. *ACM SIGCOMM Computer Communication Review*, 41(4):449–450, 2011.

[43] L. Yu, H. Li, and W. Li. Wireless scalable video coding using a hybrid digital-analog scheme. *Circuits and Systems for Video Technology, IEEE Transactions on*, 24(2):331–345, 2014.

[44] G. Caire and K. Narayanan. On the distortion snr exponent of hybrid digital–analog space–time coding. *Information Theory, IEEE Transactions on*, 53(8):2867–2878, 2007.

[45] M. Rungeler and P. Vary. Hybrid digital analog transform coding. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5109–5113. IEEE, 2013.

[46] A. Papoulis and S.U. Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[47] B. Abraham and J. Ledolter. *Statistical methods for forecasting*, volume 234. John Wiley & Sons, 2009.

[48] J.G. Proakis and D.G. Manolakis. *Digital Signal Processing*. Pearson Prentice Hall, 2007.

[49] R.G. Gallager. Low-density parity-check codes. *Information Theory, IRE Transactions on*, 8(1):21–28, 1962.

[50] C.T.K. Ng, D. Gündüz, A.J. Goldsmith, and E. Erkip. Distortion minimization in gaussian layered broadcast coding with successive refinement. *Information Theory, IEEE Transactions on*, 55(11):5074–5086, 2009.