Opportunistic Real-time Video Transmission over Fading Cognitive Radio Channels using Hybrid Digital Analog Coding

by

Sudipto Debnath

A thesis submitted to The Faculty of Graduate Studies

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

The University of Manitoba

Winnipeg, Canada

August 2020

To my parents and my wife

Abstract

Cognitive radio is a wireless communication paradigm that has achieved greater research attention due to the increasing demand for high-quality multimedia services and scarcity of the available spectrum to support these services. This paradigm allows unlicensed (secondary) users to access the unused spectrum portion of the licensed (primary) users dynamically. However, an unlicensed user faces a high level of uncertainty because of random arrivals of the licensed users, a varying channel quality, and variability in terms of number and capacity of the available channels. On the other hand, fixed-rate source-channel single-layer coding in the traditional wireless communication systems suffers greatly from the *cliff effect* and the lack of scalability. Therefore, most conventional wireless transmission schemes may not perform well in a cognitive radio environment which requires the source encoder to be adaptive to channel variations and the arrival-rate of the licensed users.

This thesis proposes a new approach to reliable transmission of real-time video over cognitive radio channels based on a hybrid digital analog (HDA) joint source-channel (JSCC) superposition coding scheme. The proposed scheme is capable of adapting itself to channel variations, primary-user arrival-rates, and exploits both the high coding gain of a digital video codec and the inherent capability of analog coding to adapt to the time-varying channel signal-to-noise ratio (CSNR) of a wireless channel.

Experimental results demonstrate that the proposed HDA coding approach is far superior to the digital layered coding (DLC), single-layer digital coding, and pure analog coding in terms of decoded video quality, quality scalability, robustness to imperfect channel estimation, and transmission power utilization. It has been observed that HDA coding produces a video quality which is on the average 2.5-4.0 dB higher in peak signal-to-noise ratio (PSNR) compared to single-layer digital coding and as much as 15-19 dB higher compared to pure analog coding respectively. Also, upon close observation of the experimental comparisons between the HDA and DLC systems reveals that the HDA approach achieves up to 2-5 dB improvements in PSNR than DLC system when intensive motion is present between video frames.

Acknowledgment

Firstly, I would like to express my sincere gratitude to my advisor Prof. Pradeepa Yahampath for his patience, encouragement, and continuous support throughout my M.Sc program. I am grateful for his constant efforts to help me to learn, understand, and conduct this research work. I would also like to appreciate his guidance toward writing my thesis.

Secondly, I would like to convey my deepest appreciation to the province of Manitoba for awarding financial support through Manitoba Graduate Scholarship (MGS) for my M.Sc program.

I would like to pay tribute to my parents Madhab Chandra Nath and Suniti Rani Debnath, for their continuous support, affection, hard-works, dedication, and sacrifices that help me to reach this far. I am indebted to my in-laws Dulal Chandra Saha and Radha Rani Saha, for their encouragement to pursue my M.Sc abroad. I am also grateful to my elder sister Susmita Debnath for her continuous guidance that helps me moving forward. Last but not least, I wholeheartedly thank my loving wife, Susmita Saha, whose ongoing mental support and advice help me to withstand during tough times and step towards achieving my goals.

Contents

List of Figures

List of Tables

List of Acronyms

3D-DCT	three-dimensional discrete cosine transform
AM	amplitude modulation
BCH	Bose–Chaudhuri–Hocquenghem
BS	base station
CSI	channel state information
CSNR	channel signal-to-noise ratio
CSU	channel state update
dB	deci-bell
DCT	discrete cosine transform
DLC	digital layered coding
FCC	federal communication commission
FDCT	forward discrete cosine transform
fps	frames per second
GOP	group of pictures
HDA	hybrid digital-analog coding
HP	high-priority
HQAM	hierarchical quadrature amplitude modulation
HVS	human visual system
IDCT	inverse discrete cosine transform
JPEG	joint photographic experts group
JSCC	joint source-channel coding
LP	low-priority
LTE	long term evolution
MCS	modulation and coding scheme
MMSE	minimum mean square error
MSE	mean square error

NALU	network abstraction layer unit
PSNR	peak signal-to-noise ratio
PU	primary user
QAM	quadrature amplitude modulation
QE	quantization error
QoS	quality of service
QP	quantization parameter
RMS	root mean square
SPIHT	set partitioning in hierarchical trees
SSIM	structural similarity index
SU	secondary user
SVC	scalable video coding
UEP	unequal error protection

Chapter 1

Introduction

Throughout the last decade, we have observed rapid developments in wireless communications, especially in the field of multimedia applications. It all started with the first generation cellular-network, where we were introduced to mobile devices with the capability of sending only voice and text wirelessly. This technology soon got immense popularity with people, and parallelly a demand for the capability to send images and videos through cellular networks got increased. As a result, we have seen substantial advancements in cellular technology. We are now using the fourth-generation cellular-network, which supports versatile multimedia applications, and this era will see more breakthroughs than ever in wireless communication technologies. In general, the transmission of multimedia contents via wireless networks consumes a significant amount of bandwidth. Furthermore, as technologies evolve, we are seeing an increasing usage of high definition multimedia services. On the one hand, these new applications will require more bandwidth, and on the other hand, we will have a limited allotment of bandwidth to support these services. It is becoming quite impossible to support the bandwidth requirements of these services as a significant portion of the bandwidth is taken by various other services. However, it has

been seen that different spectrum bands are highly underutilized [5], [6]. Therefore, an approach to improve the spectrum utilization by allowing secondary access of frequency bands already licensed for other existing services is approved by the Federal Communication Commission (FCC) [7]. This idea is facilitated by cognitive radio technology.

1.1 Cognitive Radio Model

Cognitive radio is a wireless communication paradigm that allows unused or underutilized bands of the spectrum (spectrum holes) owned by the licensed users, to become temporarily available for the non-licensed users [8]. A licensed user is known as a primary user (PU), whereas a non-licensed user is known as a secondary user (SU). This paradigm relies on the cognitive radio cycle to facilitate this access, which includes three major parts: sensing, management, and action, as shown in Figure 1.1 [8]. Also, in the cognitive radio environment, the dynamic access to the underutilized spectrum

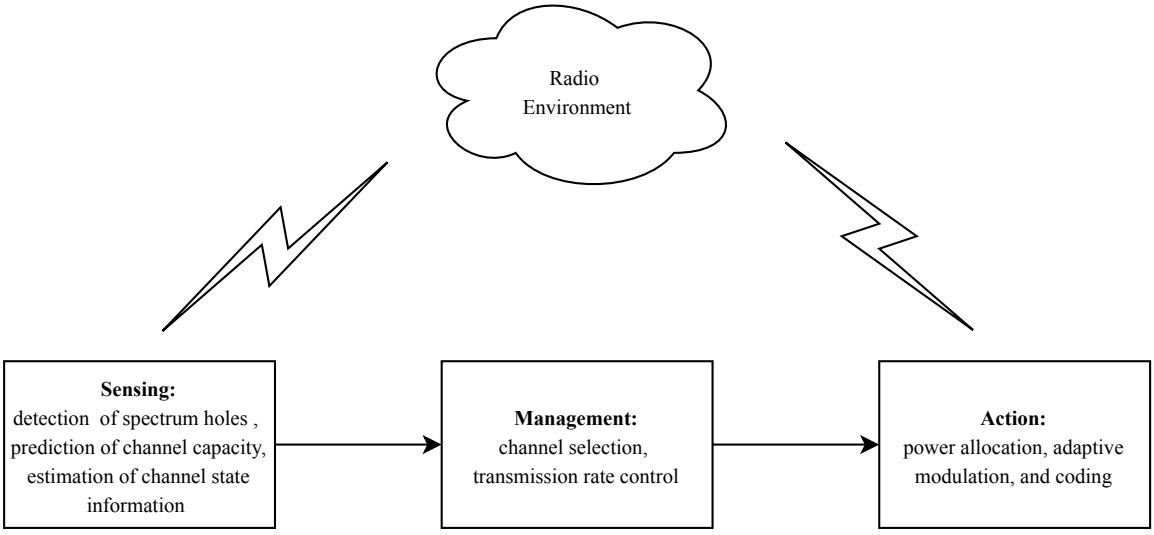


Figure 1.1: Basic cognitive cycle.

can be on an individual or a sharing basis [9]. The secondary access to spectrum based on sharing is permitted by *underlay* and *overlay* paradigms. With *underlay*, SU has the knowledge of interference caused by its transmitter to PU and thus SU can transmit simultaneously with a PU by maintaining the interference caused by itself to an acceptable threshold. The *overlay* paradigm assumes that SU knows the channel gains and the message begin transmitted by PU. With this knowledge, SU can transmit concurrently with PU by assigning part of its transmission power for its own communication and the remainder of the power to assist the PU's transmission [9].

The individual access is granted by the *interweave* paradigm, that is SU can transmit whenever the PU is inactive. A schematic view of the *interweave* paradigm is shown in Figure 1.2 [10].

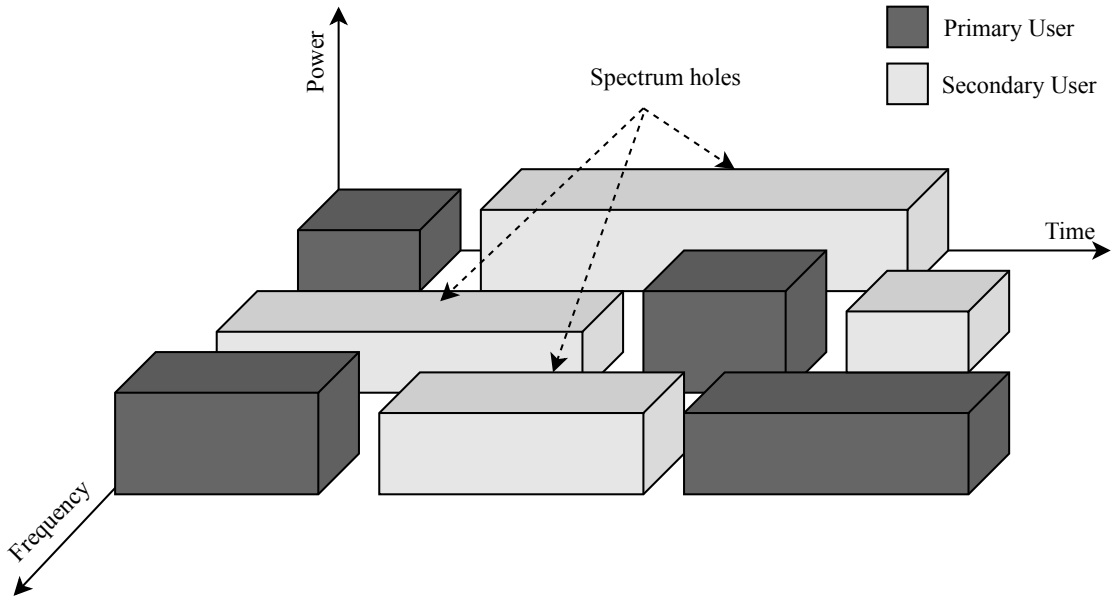


Figure 1.2: A schematic view of *interweave* paradigm.

1.2 Challenges in Real-time Cognitive Radio Communication

In the cognitive radio scenario, the secondary user faces a high level of uncertainty and variability in terms of number and the capacity of the available channels due to the random PU activity. Moreover, unlike conventional wireless communication technology where packets are lost only due to channel congestion and noise, here, losses can also occur due to the arrival of the PU. Packet losses for SUs are quite inevitable in a cognitive radio environment because of the inability to set up a reliable re-transmission link [10]. This inability arises due to the erratic PU arrivals that interrupt the link between transmitter and receiver. Additionally, in a cognitive radio environment, fluctuations of the transmission bit-rate for the SU occur not only due to the channel's fading but also due to the arrival-rate of the PU. This problem necessitates adaptive coding and modulation to realize the maximum channel throughput, given the constraints on the channel bandwidth, transmitter power, and the PU arrival-rate. Therefore, the real-time transmission, e.g., video over cognitive radio by satisfying the underlying constraints while maintaining an acceptable quality of service (QoS), has become an interesting problem.

Also, in the cognitive radio networks, PU activity is highly uncertain. As a result, if we are considering the *interweave* paradigm, it is not logical to rely on just one band or channel because if arrival occurs on that particular band, the SU needs to vacate the band and cannot access it until PU leaves. This uncertain interruption breaks the communication link between the SU's transmitter and receiver, which reduces the reliability of successful transmission. Therefore, it is required to adopt multiple channels during the secondary access so that if one link collapses, then others will help to maintain the secondary link [8]. Thus, the use of multiple channels increases

the reliability of successful transmission. Additionally, it is not required that if we adopt multiple channels, those channels have to be contiguous [8].

The main goal of this research is to investigate a new approach, which is capable of addressing the problems discussed above and ensuring reliable real-time video transmission through secondary access over the time-variant cognitive radio channels.

1.3 Issues in Real-time Multimedia Transmission

Most of the conventional wireless transmission schemes depend on the separate source-channel coding technique. That is, a source signal, e.g., video, is first quantized and coded by a source encoder, and then forwarded to channel encoder before being transmitted over the channel. It should be noted that a wireless channel exhibits fading, which results in fluctuations of the transmission bit-rate supported by the channel. However, in the conventional scheme, the output bit-rate of both the source and channel encoders remains fixed for a given channel. Therefore, this separate fixed-rate source-channel coding creates a performance bottleneck when the channel does not remain static. That is, when channel signal-to-noise ratio (CSNR) starts degrading below a certain threshold, the video quality drops dramatically due to entropy coding's sensitivity to bit-errors and complete breakdown of error-correcting codes which is generated by the fixed-rate channel encoder [11]. On the other hand, when CSNR starts improving and crosses that threshold, the decoded video quality does not improve with increasing CSNR as a fixed-rate source encoder is used and its output bit-rate is typically fixed below the worst-case bit-rate of the channel to increase the likelihood of successful transmission. This phenomenon is known as the *cliff effect* [1].

Typically, in video coding, quantization parameters are not chosen for an entire

video sequence rather they are chosen for a collection of video frames, which is know as group of pictures (GOP). As the adjacent video frames are highly correlated, this approach is adopted to increase the coding efficiency. However, in many wireless situations, GOP duration can be larger than the channel coherence time (the time during which the channel remains stationary). For example, at a video frame rate of 30 fps and a GOP size of 8 frames, the GOP duration is 267 ms. On the other hand, a mobile device operating at 900 MHz and traveling at 80 km/h experiences a channel coherence time (T_c) of about 7 ms. Since the channel bit-rate changes every 7 ms, the encoder will have no knowledge of the total channel bit payload that will be available to encode a GOP. Therefore, efficient real-time video communication over wireless links requires variable-rate video coding to enable the adjustment of transmission bit-rate at every T_c second intervals, according to the observed channel gain for the best quality.

Another issue with conventional wireless video transmission scheme is that it lacks scalability, and therefore, cannot fit all receivers at the same time [12]. This is because, on the one hand, if the video is encoded at the highest possible rate to allow the receiver with the best channel quality to realize the maximum video quality, the receivers with poor channel conditions will not be able to decode the video at all. On the other hand, if the video is encoded at a low enough rate so that all the receivers can decode it properly, the video quality of the best receiver will not be better than that of the worst receiver. Therefore, to overcome this issue, it is necessary to use the layered video coding approach. This coding approach consists of a base-layer and one or more refinements layers. During the transmission, the base layer bit rate is set to a low enough value to guarantee its delivery under the worst expected channel, and the refinement layers are transmitted adaptively based on the short-term channel

state information (CSI) to improve the video quality of the base layer as the channel condition varies. There are two main approaches to layered coding: (1) digital layered coding (DLC) and (2) hybrid digital-analog (HDA) coding.

1.4 Hybrid Digital-Analog Scheme

To obtain the adaptiveness with channel variations and to support scalability, one may think of using a purely analog transmission scheme due to its inherent capability to adapt to the time-varying CSNR of a wireless channel, which results in a decoded video quality that is commensurate with the CSNR. One such scheme is SoftCast [1]. Although this scheme can avoid the *cliff effect* and support scalability, it suffers from drastic degradation in decode video quality when the channel condition worsens [12]. Further, most of the real-time wireless multimedia applications, e.g., videos, require high bandwidth compression, which is difficult to achieve with analog transmission [11]. Therefore, a better solution is to use layered video coding.

With DLC such as scalable video coding (SVC), only the base-layer bit rate is set according to the worst-case channel bit-rate. One or more scalable refinement layers may be sent depending on the actual channel throughput. This way, the SVC encoder operates independently of the instantaneous channel conditions, yet the quality of video decoded at the receiver improves when the channel conditions improve.

On the other hand, in HDA coding, a fixed-rate digital video encoder is used to generate the base layer, which is channel coded and transmitted using a digital modulation scheme such as digital quadrature amplitude modulation (QAM). Additionally, the analog quantization error of the base layer is also transmitted using analog QAM, superimposed on the digital signal (therefore, the transmitted QAM symbol is hybrid digital-analog). The analog layer acts as a single refinement layer which adapts itself

to the channel variations.

If SVC is used, it is necessary to have multiple layers to obtain a smooth decoded video quality. However, it has been shown in previous work that at the same bit rate, state-of-the-art single-layer codecs outperform over the base-layer of an SVC codec, which generates multiple layers [11–13]. For this reason, HDA video coding using a state-of-the-art single layer video codec is preferred as its base-layer may perform better than that of an SVC codec.

1.5 Contributions

This thesis makes the following contributions.

- A new approach to reliable multimedia transmission as a secondary user over a cognitive radio channel based on HDA video coding is presented, which can adapt to channel capacity variations caused by both fading and random primary user arrivals.

- In HDA transmission, an important issue is the allocation of the total available transmitter output power between the analog and digital transmissions. To address this issue, a method to distribute transmitter power between analog and digital transmission is presented.

1.6 Related Work

The related work can be divided into two categories: (1) multimedia transmission over conventional wireless channels, and (2) multimedia transmission over cognitive radio channels.

1.6.1 Multimedia Transmission Over Conventional Wireless Channels

As discussed in section 1.3, the conventional multimedia transmission over wireless channels suffers greatly from *cliff effect* and lack of scalability. Many approaches are presented in the literature to overcome these two issues.

Pure analog transmission schemes are presented in [1], [14] to solve the aforementioned issues. They adopt the three-dimensional discrete cosine transform (3D-DCT) to compress the video frames. Error protection to the transmitted analog video samples is obtained by scaling the magnitude of the transmitted samples using optimized scaling factors subject to the transmission power constraint. Moreover, resilience to packet loss is achieved by using the Hadamard transform.

Pure digital approaches to solve the *cliff effect* and enable scalability can be found in [15], [16]. They couple SVC (e.g., H.264/SVC) with HQAM [11]. However, their proposed scheme cannot completely overcome the *cliff effect* and instead results in a staircase effect [1], [11].

Some literature adopts both digital and analog coding techniques by transmitting hybrid symbols that are obtained by the superposition of digital and analog modulation signals. Such studies can be found on [11–13, 17–20]. Most of these approaches, generate a base-layer using H.264/AVC encoder and a refinement layer (analog) which is the quantization errors. Later, the analog layer is compressed using 3D-DCT, modulated, and superimposed to digitally modulated signals. Previous work shows that HDA coding can, in most cases, outperform digital-only and analog-only coding.

1.6.2 Multimedia Transmission over Cognitive Radio

As the concept of cognitive radio technology is still developing, research work that are primarily focused on real-time multimedia transmission over cognitive radio channels

is very limited [21], [22].

The authors in [8] implement the spectrum pooling concept [5] to establish the SU transmission links and propose a quality metric to define a subset of SU transmission links that are used during transmission. They use the Poisson process to model PU arrivals, and they rely on digital fountain codes [23] to transmit multimedia content over SU links. Their study mainly focuses on achieving maximum effective throughput and spectral efficiency by optimizing the number of SU links. However, they use a fixed channel capacity for all SU links and do not consider the channel variations, which is unrealistic. Further, the source coding strategy they adopt requires a bipartite graph to be transmitted along with the data packets to enable proper decoding of packets, which increases the overhead. Additionally, they do not present any numerical results on the real-time multimedia transmission to demonstrate the effectiveness of their source coding strategy.

A joint source-channel coding (JSCC) approach, which uses frames [24] to transmit multimedia over cognitive radio, can be found on [10]. Like [8], this research models the PU arrival as a Poisson process and considers the *interweave* paradigm to establish SU links. Instead of using fountain codes, this work uses frames to expand the quantized source vectors optimally into multiple descriptions to compensate for packet losses due to PU activity. Also, like [8], this work considers fixed channel capacity, which is not realistic in a cognitive radio environment as here the channel capacity varies with the arrival-rate as well as with fading. Also, this model lacks scalability, which is one of the major requirements in real-time multimedia transmission. Their experimental results show that the decoded image quality directly depends on the arrival-rate, which indicates poor coding efficiency.

The authors in [21] present a new approach to transmit real-time multimedia over

a cognitive radio environment using hierarchical quadrature amplitude modulation (HQAM) under imperfect channel sensing. The use of HQAM adds unequal error protection (UEP) to the transmitted data. For video, they use H.264/MPEG encoder and split the resulting bit-stream into high-priority (HP) and low-priority (LP) bits (for video, bits associated with I-frames are considered as HP bits and rest of the bits are considered as LP bits) which they place into packets by maintaining an equal ratio of HP and LP bits in packets. Later, these packets are modulated using HQAM. The output power of HQAM symbols depends on CSI and PU activity. Therefore, they have presented an optimal power control strategy. Their proposed scheme can transmit multimedia contents of SU concurrently with the PU. However, they do not present any model for PU activity on a channel. Additionally, they use fixed-rate source encoder that results in a *cliff effect*.

Another JSCC approach to transmit images over multiple long term evolution (LTE) cognitive radio channels can be found in [25]. They adopt set partitioning in hierarchical trees (SPIHT) and a Reed-Solomon code to perform source and channel encoding, respectively. To set up the SU links, they rely on the underlay paradigm. Like [21], they consider UEP to protect packet loss that arises due to PU activity. This model estimates PU arrivals using the Poisson process as well. However, the proposed scheme does not adopt scalability. Further, no power allocation strategy is presented, although they consider lossy channels and PU interference during transmission.

A layered coding approach for video transmission over multi-channel cognitive radio networks can be found on [22]. They adopt a utility-based H.264/SVC video transmission scheme that uses UEP. Moreover, they study to improve the video quality by using a flexible sensing-transmission scheme and by providing a level of utility to each different network abstraction layer unit (NALU) that depends on their loss

probability and the contribution towards video quality. Unlike [8, 10, 25], the authors in [22] adopt a Markov chain [26], [27] to model the PU in cognitive radio networks. They also consider the *interweave* paradigm to setup the SU links.

1.7 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 describes the fundamental concepts and background of the work presented in this thesis. Chapter 3 presents the system architecture of the model in detail. Then, the experimental procedures and results followed by the analysis of the outcomes are presented in chapter 4. Finally, chapter 5 concludes the thesis and highlights some possible future work.

Chapter 2

Background

In 1948, Claude Shannon, "the father of information theory", introduced the mathematical model of communication. According to him, any communication system consists of five basic elements as shown in Figure 2.1. However, in a digital communication system, the transmitter and receiver blocks are represented by a pair of source-channel encoders and decoders, respectively. A schematic diagram of a digital communication system is presented in Figure 2.2. The source encoder compresses the input signal based on its statistics and forwards it to the channel encoder as a uniformly-random binary stream. In [28], Claude Shannon stated two fundamental

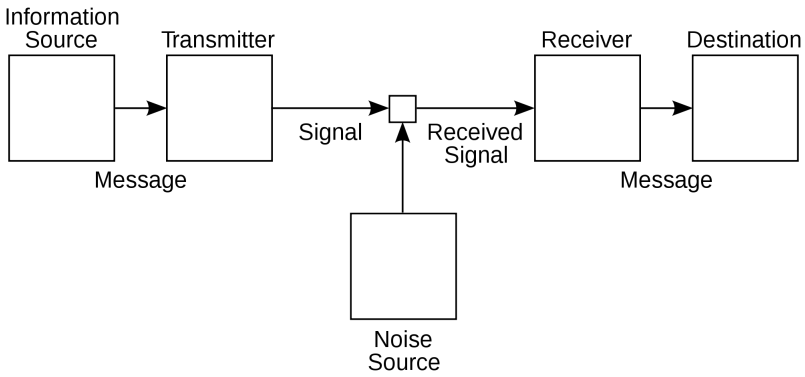


Figure 2.1: Shannons diagram of a general communication system.

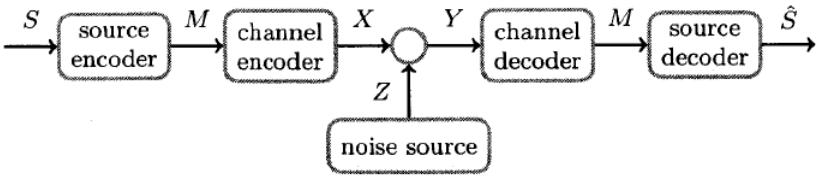


Figure 2.2: A schematic diagram of a digital communication system [1].

problems of digital communication. The first one is to find the exact or close approximate representation of an information source which is obtained by source coding at some given bit-rate and the latter is *channel capacity*, that represents the maximum rate at which reliably transmission over a channel is possible that is achieved by channel coding. Therefore, the purpose of channel encoder, in a digital communication system is to protect the source information from being corrupted by the channel noise. However, in the digital communication system, the exact reproduction of information is hampered not only by channel noise but also due to the presence of the quantization process in source encoder.

To get more insights into the digital communication of a video source, the rest of the chapter is organized according to the flow of Figure 2.2 in the context of video transmission.

2.1 Video Compression Framework

The goal of any source encoder is data compression, that is to reduce the amount of bit-rate necessary to convey the source information either without a loss or with a certain distortion [1]. The source encoders used for real-time video coding fall into the latter category. A schematic view of a modern video encoder, e.g., H.264/AVC [2], is presented in Figure 2.3.

The compression of an information signal is possible because of the presence of

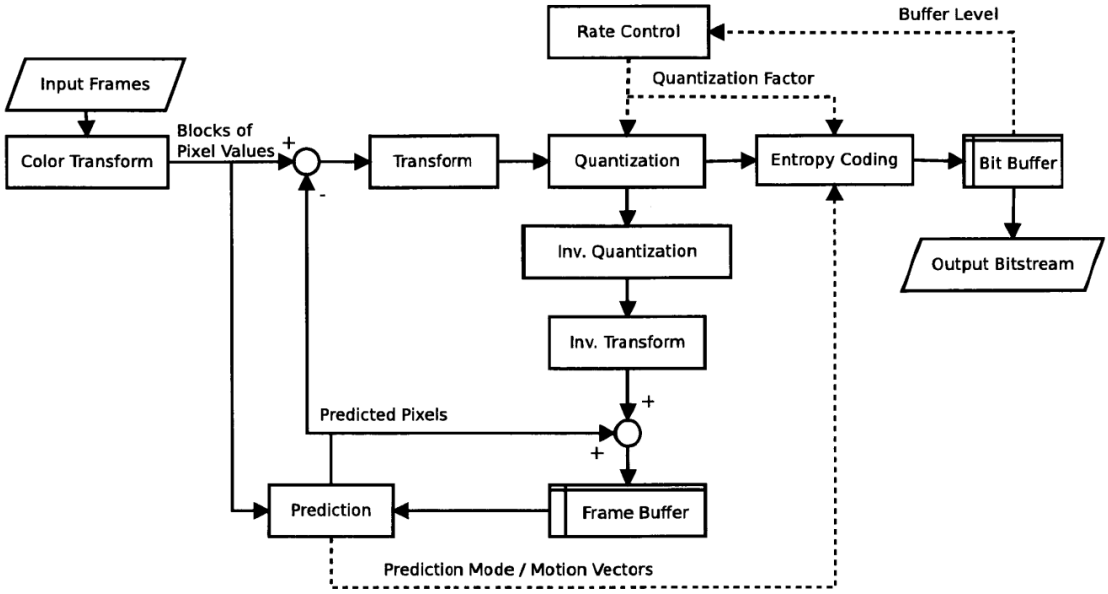


Figure 2.3: Anatomy of a modern video encoder [1].

redundancies. In the case of video a significant amount of compression can be achieved by taking into consideration the spatial, temporal, and psycho-visual redundancies which are described below.

Spatial: In an image, spatial redundancy occurs as pixels in the same frame are highly correlated. This is indicated by sets of pixels that have the close values for luminance and chrominance.

Temporal: In a typical 30 frames per second (fps) video sequence, 30 images are captured in every second; therefore, very little motion of objects present in the video may occur between consecutive frames. As a result, there is a good chance of capturing the same object 30 times with nearly the same sets of pixel values. Moreover, if motion occurs within one second, the object in motion will have the same set of pixels but at a different location (an offset). Thus, significant temporal redundancy can be found in most real-world video sequences.

Psycho-visual: This type of redundancy arises due to the fact that the sensitivity

of the human visual system (HVS) is selective, that is, it does not consider individual pixel values to form an image mentally. The HVS is less sensitive to the changes in chrominance (color) and more sensitive to the changes in luminance (brightness). As a result, the elimination of such quantitative information (pixel values) is possible without disturbing the HVS. Since the quantitative information is eliminated, this process is commonly referred to as quantization.

In a video encoder, by adopting different techniques, these three redundancies can be omitted. The spatial and temporal redundancies are exploited by using a suitable linear transform and predictive coding, while psycho-visual redundancy is taken into account by quantization [1].

A video encoder does not operate on a whole frame or collection of frames; rather, it operates on a block or collection of blocks from the same or different frames to exploit the techniques that are mentioned above. A block is usually a collection of several neighboring pixels. In video coding, such a block is known as a *macro-block*. Three types of frames are used in most video coding algorithms, which were first introduced in MPEG-1 [29]. They are: I- (intra-coded) frames, P- (predictive-coded) frames, and B- (bidirectional-coded) frames.

In the following subsections, we will first briefly present the structure of the macro-blocks and the video frames. Latter we will discuss some key components of a modern video encoder, as shown in Figure 2.3.

2.1.1 Macro-block

A macro-block typically consists of several 8×8 blocks of luminance and chrominance pixels. To form a macro-block, we need six 8×8 blocks, and among them, we have four 8×8 luminance blocks, one block of Cb, and one block of Cr. Therefore, we get

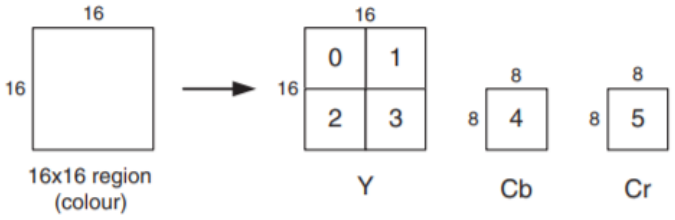


Figure 2.4: Basic structure of a macro-block (4:2:0) [2].

a 16×16 pixel region where we have 16×16 pixel luminance components and two 8×8 chrominance components. A macro-block is formed in this manner because the HVS is more sensitive to luminance than chrominance (more details can be found in section 2.1.3). A schematic view of a macro-block is presented in Figure 2.4.

2.1.2 Video Frames

A video encoder typically uses a combination of I-, P-, and B-frames to compress a video file. To encode an I-frame, temporal correlations are not considered, which means an I-frame is encoded without any reference from the past frames. Each I-frame is treated as a separate digitized picture and coded block-by-block similar to the joint photographic experts' group (JPEG) [30]. Therefore, the compression rate of a I-frame is quite low compared to frames which exploits temporal correlations [31]. As I-frames are not predicted from any other frames, they contribute less error in the decoder. However, the P- and B-frames are predicted from the I-frames. Therefore, I-frames are typically repeated after every 3 to 12 frames. The number of frames between successive I-frames is known as a GOP [30].

2.1.3 Color Transform

In computer systems, colors are represented by the RGB model, which is founded on the trichromacy of human vision [32]. Therefore, any color video sequence has three input data channels: red, green, and blue. Since a high correlation exists among these three channels and human eyes are less sensitive to chroma components than luma; therefore, the first step of any modern video encoding algorithm is to perform color transform [32]. This transformation is obtained by introducing a new color space (YCbCr) using the equations [2]

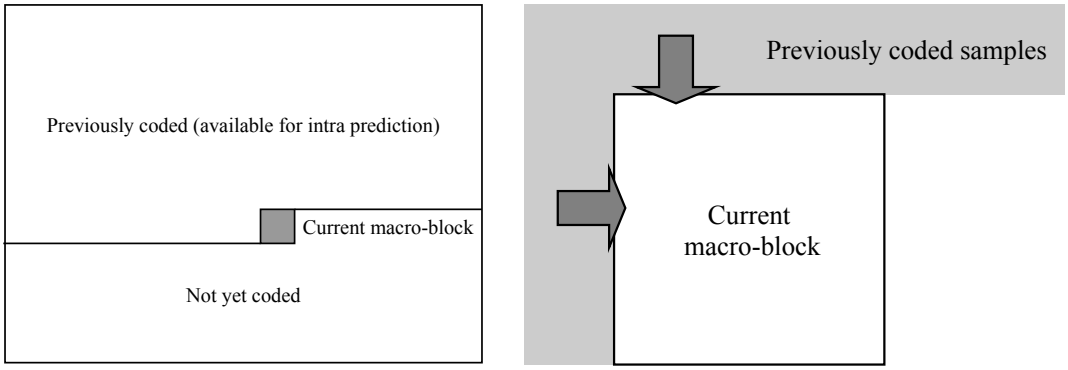
$$\begin{aligned} Y &= k_r R + k_g G + k_b B, \\ Cb &= B - Y, \\ Cr &= R - Y, \end{aligned}$$

where k's are the weighting factors, Y represents the luminance, and Cb and Cr represent chrominance.

However, as human eyes are more sensitive to variations of luminance than chrominance, chrominance components are subsequently sub-sampled usually at a rate of 2 [32]. For video encoding, a popular sub-sampling format is $4:2:0$. In this format, from a 16×16 color region, as shown in Figure 2.4, only one set of chrominance components is taken against four sets of luminance components.

2.1.4 Predictive Coding

In predictive coding, redundancy is reduced by forming a prediction of the data and subtracting this prediction from the current data. In a video coding, this prediction is usually formed from previously coded frames (a temporal prediction) or previously



(a) Intra prediction: available samples (b) Intra prediction: spatial extrapolation

Figure 2.5: A schematic of intra prediction mode.

coded image samples in the same frame (a spatial prediction) [2]. The subtraction process in predictive coding generates a residual or a prediction error. As a result, the more accurate the prediction, the smaller the energy contained in the residual and vice-versa [2]. Typically, predictive coding can be divided into two modes: Intra prediction and inter prediction. In practice, these modes are operated over macro-blocks.

Intra Prediction: To predict, intra prediction mode does not require any reference frame; rather, it uses previously encoded macro-blocks of the current frame. This mode is used only to encode I-frames. A schematic view of intra prediction is presented in Figure 2.5.

Inter Prediction: Inter prediction mode uses a reference frame to predict a given frame. This mode is used to reduce temporal redundancy that is presented between two frames. Typically, P- and B-frames are encoded using inter prediction mode. Inter prediction mode includes motion estimation, motion compensation, the computation of motion vectors to predict the residuals.

Motion estimation is the process where for each macro-block in the current frame, a search is performed in the reference frame (can be previously encoded frame from

the sequence and maybe before or after the current frame in display order [2]) to find a 16×16 region in the reference frame that closely matches the current macro-block. However, during motion estimation, the whole reference frame is not searched. Rather, a search area is considered, which is located within ± 15 pixels in both horizontal and vertical directions of the macro-block being encoded [31]. Figure 2.6 illustrates the motion estimation technique.

The closeness of a match or distance is measured by computing either the sum of absolute differences or by the sum of squared differences between the corresponding pixels in a 16×16 region and a macro-block [31]. If the distance from the current macro-block to the closest 16×16 region in the reference frame is greater than some prespecified threshold, then the macro-block is declared uncompensable and is encoded without the benefit of prediction [31]. On the other hand, if the distance is below the threshold, then a motion vector is computed.

The motion vector is the relative location of the best-matching 16×16 region in the reference frame. It is obtained by subtracting the coordinates of the upper-left

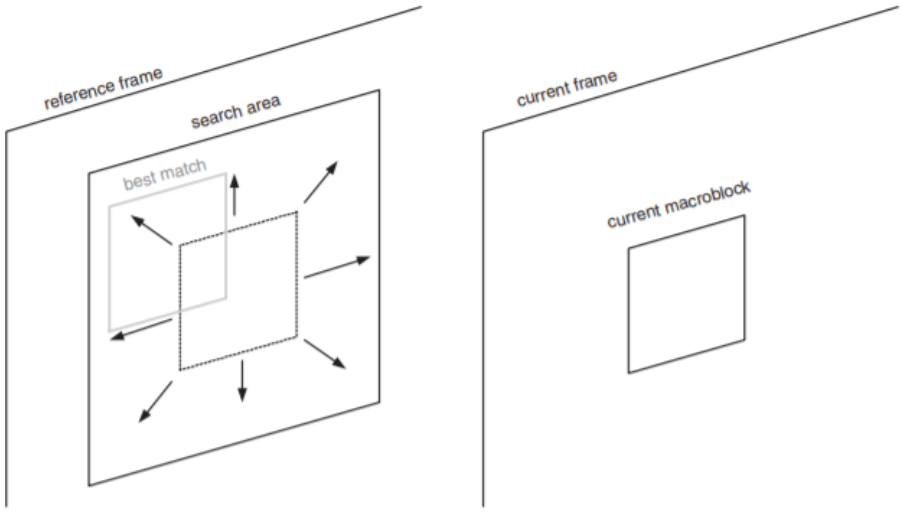


Figure 2.6: Motion estimation [2].

corner pixel of the macro-block from the coordinates of the upper-left corner pixel of the 16×16 region.

2.1.5 Transform Coding

The purpose of transform coding is to apply a decorrelating transform to compand the energy of the signal down to a few components [32]. Therefore, it is possible to take a subset of coefficients from a set of transform coefficients, which contains enough details to reconstruct the signal, and thus introduce compression [2]. However, in the context of video coding, only the energy compacting property of the transform coding is exploited.

Transform coding can be performed only if the components of the source signal are correlated. As pixels within a macro-block are highly correlated, a high correlation may exist within a residual data of a macro-block if the prediction is not very good. As a result, in video coding, transform coding is performed over the residual data of macro-blocks. The most popular transform is the discrete cosine transform (DCT), which is applied over 8×8 blocks of residual data. An inverse operation of DCT is performed at the decoder side and it is known as the inverse DCT (IDCT). The forward DCT (FDCT) and IDCT are defined as

$$Y(x,y) = C(x)C(y) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X(i,j) \cos \frac{(2j+1)y\pi}{2N} \cos \frac{(2i+1)x\pi}{2N},$$

$$X(i,j) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} C(x)C(y)Y(x,y) \cos \frac{(2j+1)y\pi}{2N} \cos \frac{(2i+1)x\pi}{2N},$$

respectively. Here $C(\zeta) = \sqrt{\frac{1}{N}},$ if $\zeta = 0,$ and $C(\zeta) = \sqrt{\frac{2}{N}},$ if $\zeta > 0$, $Y(x,y)$ is the

DCT coefficient in row x and column y of the DCT matrix, and $X(i, j)$ the intensity of the pixel in row i and column j.

2.1.6 Quantization

Quantization is the process of mapping a continuous random variable $X \in \mathbb{R}$ to a finite set $y_i = \{y_1, \ldots, y_N\}, y_i \in \mathbb{R}$. Therefore, it becomes possible to represent the random variable with a finite number of bits. If X is a single random variable then the process is called *scalar quantization*, and if X is a random vector then the process is called *vector quantization*.

In H.264 video coding standard, a simple divide-and-truncate operation is performed over the DCT coefficients, which is equivalent to a uniform scalar quantizer [2], [31]. The quantized value of the variable *coeff* is then given by

$$Q_{coeff} = round(coeff/QP),$$

where QP is the quantizer step size. In the H.264 video encoder, QP is always set to an integer value. If QP is set to a higher value, there will be more coefficients with the zero value, which results in a higher compression at the cost of poor decoded video quality. On the other hand, if it is set to a lower value, then there will be more coefficients with non zero values, thus results in less compression, which guarantees better-decoded video quality. However, QP can be different for each DCT coefficient according to the human psycho-visual model [32]. This strategy is adopted to quantize the coefficients belonging to low spatial frequencies more accurately rather than the coefficients corresponding to high spatial frequencies.

2.1.7 Entropy Coding

If a source with a finite alphabet of size 2^n has a non-uniform distribution over the symbols, then with entropy coding, it is possible to allot fewer bits than n per symbol but without a loss of any information [32]. Such reduction in bits is obtained by providing shorter codewords to the symbols that have higher probabilities of appearing in the coded stream and vice-versa. Claude Shannon in his 1948 paper [28], describes entropy as the optimal bit-rate for a source with a distribution $p_X(x)$ over its symbols [32]. The entropy is given by

$$H(x) = \sum_i -p_X(x_i) log_2 p_X(x_i).$$

In practice, the quantized DCT coefficients, motion vectors, and quantization factors are compressed using an entropy coder.

2.1.8 Rate Control and Bit Buffer

In order to control the output bit-rate, rate control and bit buffer units work cooperatively. The function of the bit buffer is to keep the output of the encoder fixed according to the transmission rate [31]. Therefore, if the bit buffer starts filling up at a rate faster than the transmission rate, it gives feedback to the quantizer through the rate control unit to increase the value of QP. On the other hand, if the bit buffer sees a decrease in the output rate lower than the transmission rate, it asks to increase the value of QP. In practice, QP is multiplied by the *quantization factor* which helps video coder to trade the bit-rate of for distortion [31], [32].

2.2 Error Correcting Codes

Error-correcting codes are used in digital communication to protect the data bits from inevitable bit errors that occur due to channel noise. The purpose of the error-correcting codes is to add redundancy to the data stream in such a way so that data bits can be transmitted error-free even with some bit errors occur. In the following subsection, we will discuss the two error-correcting codes, which are also implemented in this research work.

2.2.1 Reed-Solomon Codes

Reed-Solomon codes are systematic block codes and are a subset of Bose-Chaudhuri-Hocquenghem (BCH) codes. In systematic block codes, a sequence message bits is divided into blocks and each block is appended a set of parity bits. Each block is made up of multiple symbols that are protected by parity symbols and thus forms a self-contained codeword. Such a block is illustrated in Figure 2.7.

A Reed-Solomon code is denoted by $RS(n, k)$ with m-bit symbols [33]. That is, the encoder takes k data symbols of m-bits and add parity symbols to form a codeword of length n symbols where the maximum codeword length is $n = 2^m - 1$. Therefore,

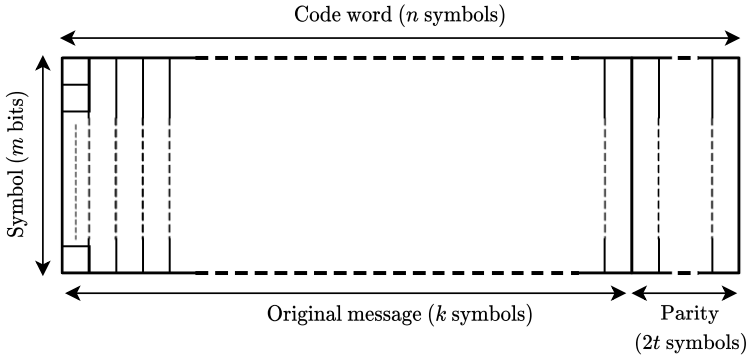


Figure 2.7: Reed-Solomon code definition [3].

for each block, there are $n - k$ parity symbols. A Reed-Solomon encoder has the capability of correcting up to t erroneous symbols in a block, where $2t = n - k$. A symbol is considered erroneous when at least 1 bit in that symbol is in error.

For example, consider a Reed-Solomon code $RS(255, 239)$ with 8 bits/symbol. Therefore, it takes 239 data bytes and produces a codeword of length 255 bytes. For this code $n = 255, k = 239$ and $2t = n - k = 16$ and therefore, $t = 8$ which indicates the code can correct up to 8 erroneous symbols in a block. That is, the code may correct 8 bits, when 1 bit alters on each 8 symbols in worst case and correct 64 bits for 8 complete byte errors in the best case [33].

To construct a (n, k) Reed-Solomon code, we need to define a generator polynomial $g(x)$ which is defined as

$$g(x) = (x + \alpha^b)(x + \alpha^{b+1}) \dots (x + \alpha^{b+2t-1}),$$

where $b = 0$, αs are the elements of the Galois field $GF(2^m)$ and the value of α is usually considered to be 2 [3]. Now to define the finite field $GF(2^m)$, primitive polynomial $p(x)$ is used, which is a polynomial of degree m with no factors, and thus is irreducible. For a Galois field of particular size, $p(x)$ is predefined. For $GF(2^4)$, the polynomial is in the form $p(x) = 1 + x + x^4$. To construct the all non-zero elements of the Galois field, it is assumed that primitive element α is a root of $p(x)$ so that $p(\alpha) = 0$ [3]. Thus we get

$$\alpha^4 + \alpha + 1 = 0,$$

or,

$$\alpha^4 = \alpha + 1. \tag{2.1}$$

Higher powers of α can be obtained using (2.1). That is, to get α^5 and α^6 we can write

$$\alpha^5 = \alpha.\alpha^4 = \alpha(\alpha + 1) = \alpha^2 + \alpha,$$

and,

$$\alpha^6 = \alpha.\alpha^5 = \alpha(\alpha^2 + \alpha) = \alpha^3 + \alpha^2$$

respectively. In this way, we can obtain the complete field as shown in Table 2.1, where binary form is obtained by substituting the value of α.

Table 2.1: The field elements of $GF(2^4)$ with $p(x) = x^4 + x + 1$.

Index form	Polynomial form	Binary form	Decimal form
0	0	0000	0
α^0	1	0001	1
α^1	α	0010	2
α^2	α^2	0100	4
α^3	α^3	1000	8
α^4	$\alpha + 1$	0011	3
α^5	$\alpha^2 + \alpha$	0110	6
α^6	$\alpha^3 + \alpha^2$	1100	12
α^7	$\alpha^3 + \alpha + 1$	1011	11
α^8	$\alpha^2 + 1$	0101	5
α^9	$\alpha^3 + \alpha$	1010	10
α^{10}	$\alpha^2 + \alpha + 1$	0111	7
α^{11}	$\alpha^3 + \alpha^2 + \alpha$	1110	14
α^{12}	$\alpha^3 + \alpha^2 + \alpha + 1$	1111	15
α^{13}	$\alpha^3 + \alpha^2 + 1$	1101	13
α^{14}	$\alpha^3 + 1$	1001	9
α^{15}	1	0001	1
α^{16}	α	0010	2

Consider a $RS(15, 11)$ code, where $t = 2$. Therefore, the generator polynomial $g(x)$ becomes

$$\begin{aligned} g(x) &= (x + \alpha^0)(x + \alpha^1)(x + \alpha^2)(x + \alpha^3) \\ &= (x + 1)(x + 2)(x + 4)(x + 8) \\ &= x^4 + 15x^3 + 3x^2 + x + 12. \end{aligned}$$

Encoding- To encode the message, k message symbols are considered to be a polynomial $M(x)$ of order $k - 1$, so that

$$M(x) = M_{k-1}x^{k-1} + \ldots + M_1x + M_0,$$

where each of the coefficient is an m-bit message symbol, thus an element of $GF(2^m)$ [3]. At first, $M(x)$ is shifted by x^{n-k} and then divided by $g(x)$, which results in

$$\frac{M(x) \times \ x^{n-k}}{g(x)} = q(x) + \frac{r(x)}{g(x)},$$

where $q(x)$ and $r(x)$ are the quotient and remainder of the division respectively. Once $r(x)$ is obtained, the codeword is formed as

$$T(x) = M(x) \times \ x^{n-k} + r(x).$$

The idea behind such encoding is that, if we divide $M(x)$ by $g(x)$, it should be divisible without any remainder. However, if we get a remainder, it indicates one or more error has occurred at $T(x)$.

Decoding- The receiver performs a division on $R(x)$ by $g(x)$, where $R(x)$ is the received codeword. If $T(x) == R(x)$, then $R(x)/g(x)$ will produce no remainder. If

a remainder is produced, then it is assumed $R(x)$ is in the form

$$R(x) = T(x) + E(x),$$

where $E(x)$ is the error polynomial and $E(x) = E_{n-1}x^{n-1} + \ldots + E_1x + E_0$. Each of the coefficients is an element of $GF(2^m)$ field and thus represents an m-bit error value. The position of the error value is determined by the degree of x for that term [3]. The decoder knows $g(x)$ and its roots, and therefore, can solve $R(x)$ at every root to generate a system of equations to determine the erroneous coefficients of $R(x)$, and their values. Therefore, the decoding process of a Reed-Solomon code involves solving simultaneous equations. In practice, erroneous coefficients are obtained by the Berlekamp-Massey algorithm, and the corresponding values are obtained using the Forney algorithm [33]. Once errors are detected, they are corrected at the decoder end, and then parities are discarded to obtain the data symbols.

2.2.2 Convolutional Codes

Convolutional codes are different from block codes such as Reed-Solomon codes because a convolutional code contains memory, whereas block codes are memory-less. Also, unlike the Reed-Solomon code, the information bits do not directly appear in the coded bitstream. A convolution encoder is denoted by (n, k, m), where n indicates the number of output bits, k is the number of input bits, m is memory order of the encoder, and the rate of the encoder is k/n. At any time, the n outputs of a convolutional encoder depend on the k input bits and the previous m input blocks.

For an $(n, 1, m)$ convolutional encoder, if the input sequence is in the form $\boldsymbol{u} =$

$(u_0, u_1, \ldots, u_l, \ldots)$, the encoder will produce n output sequences as

$$\begin{aligned}
\boldsymbol{v}^1 &= (v_0^{(1)}, v_1^{(1)}, \ldots, v_l^{(1)}, \ldots) \\
\boldsymbol{v}^2 &= (v_0^{(2)}, v_1^{(2)}, \ldots, v_l^{(2)}, \ldots) \\
&\vdots \\
\boldsymbol{v}^n &= (v_0^{(n)}, v_1^{(n)}, \ldots, v_l^{(n)}, \ldots),
\end{aligned}$$

and the coded sequence becomes

$$\boldsymbol{v} = (\boldsymbol{v_0}, \boldsymbol{v_1}, \ldots, \boldsymbol{v_l}, \ldots),$$

where $\boldsymbol{v_l} = (v_l^{(1)}, v_l^{(2)}, \ldots, v_l^{(n)}, \ldots)$. Typically, the code is specified by a set of n generator sequences of length $m + 1$. The generator sequences are constructed as

$$\begin{aligned}
\boldsymbol{g}^{(1)} &= (g_0^{(1)}, g_1^{(1)}, \ldots, g_m^{(1)}) \\
\boldsymbol{g}^{(2)} &= (g_0^{(2)}, g_1^{(2)}, \ldots, g_m^{(2)}) \\
&\vdots \\
\boldsymbol{g}^{(n)} &= (g_0^{(n)}, g_1^{(n)}, \ldots, g_m^{(n)}),
\end{aligned}$$

where

$$g_m^{(n)} = \begin{cases} 1, & \text{if there is connection from input to a output} \\ 0, & \text{otherwise.} \end{cases}$$

The output sequence, therefore, can be thought of as the discrete convolution of

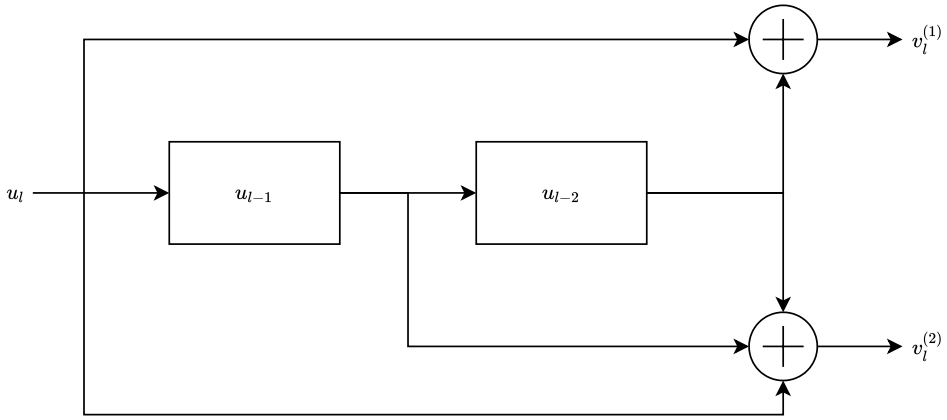


Figure 2.8: Block diagram of 1/2-rate convolutional encoder.

the input sequence $\boldsymbol{u}$ and the generator sequence $\boldsymbol{g}^{(i)}$. That is,

$$\begin{aligned}\boldsymbol{v_l}^{(i)} &= \boldsymbol{u} * \boldsymbol{g}^{(i)} \\ &= \sum_{j=0}^{m} u_{l-j} g_j^{(i)} \\ &= u_l g_0^{(i)} + u_{l-1} g_1^{(i)} + \ldots + u_{l-m} g_m^{(i)},\end{aligned}$$

where $1 \leq \ i \leq \ n$.

Figure 2.8 illustrates a $(2, 1, 2)$ convolutional encoder, which is a 1/2-rate encoder with the order of memory 2. Since, here $m = 2$, we have three generator sequences, and hence we can define the output as

$$\begin{aligned}v_l^1 &= u_l g_0^{(1)} + u_{l-1} g_1^{(1)} + u_{l-2} g_2^{(1)} \\ v_l^2 &= u_l g_0^{(2)} + u_{l-1} g_1^{(2)} + u_{l-2} g_2^{(2)},\end{aligned}$$

where, u_{l-1} and, u_{l-2} are the previous input bits. From Figure 2.8, we can get

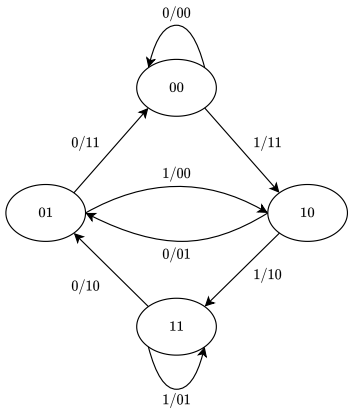


Figure 2.9: State diagram of 1/2-rate convolutional encoder.

$\boldsymbol{g}^{(1)} = (1, 0, 1),$ and, $\boldsymbol{g}^{(2)} = (1, 1, 1)$. Therefore, we can rewrite $v_l^1,$ and, v_l^2 as

$$v_l^1 = u_l + u_{l-2} \tag{2.2}$$

$$v_l^2 = u_l + u_{l-1} + u_{l-2}. \tag{2.3}$$

Let us assume $\boldsymbol{u} = (0, 1, 1, 0, 1)$. Then according to (2.2), and (2.3) the coded sequence becomes $\boldsymbol{v} = (00, 11, 10, 10, 00)$. In practice, a trellis diagram is used to both encode and decode a convolutional code. To get the idea of a trellis, first we need to construct a state diagram as the convolutional encoder can be viewed as a finite-state machine (see Figure 2.9) and the corresponding trellis diagram is presented in Figure 2.10. To encode a sequence of information bits, one just needs to trace a path through the trellis, which is highlighted by a black line in the figure. After a whole sequence is encoded, the encoder goes to all zero state. Therefore, 2^m bits are added with the encoded sequence, and thus the output rate of the encoder becomes $(l + m)n$. To decode the received sequence, a popular method is to use the Viterbi algorithm over the same trellis.

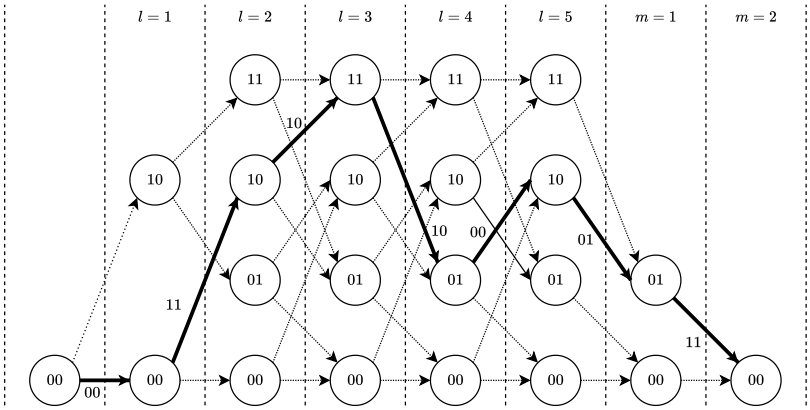


Figure 2.10: Trellis representation of 1/2-rate convolutional encoder.

2.3 Quadrature Amplitude Modulation

Quadrature amplitude modulation (QAM) requires modulating both the amplitude and phase of the carrier signal. This modulation process requires two carrier signals, which are 90° apart from each other. Therefore, it is possible to transmit two signals simultaneously under the same carrier frequency without mutual interference. Since there is a 90° phase difference between two carrier signals, the resulting signals are known as in-phase and quadrature signals. QAM can be used for both analog and digital transmissions. In the following subsections, we will briefly discuss the QAM process for both digital and analog signals.

2.3.1 Analog QAM

Consider two analog signals $I(t)$ and $Q(t)$. Now if we use these two signals to modulated the quadrature carriers and add them together, the result is

$$\begin{aligned} X(t) &= I(t)\cos(2\pi f_c t) - Q(t)\sin(2\pi f_c t) \\ &= x_A(t)\cos[2\pi f_c t + \phi(t)], \end{aligned} \tag{2.4}$$

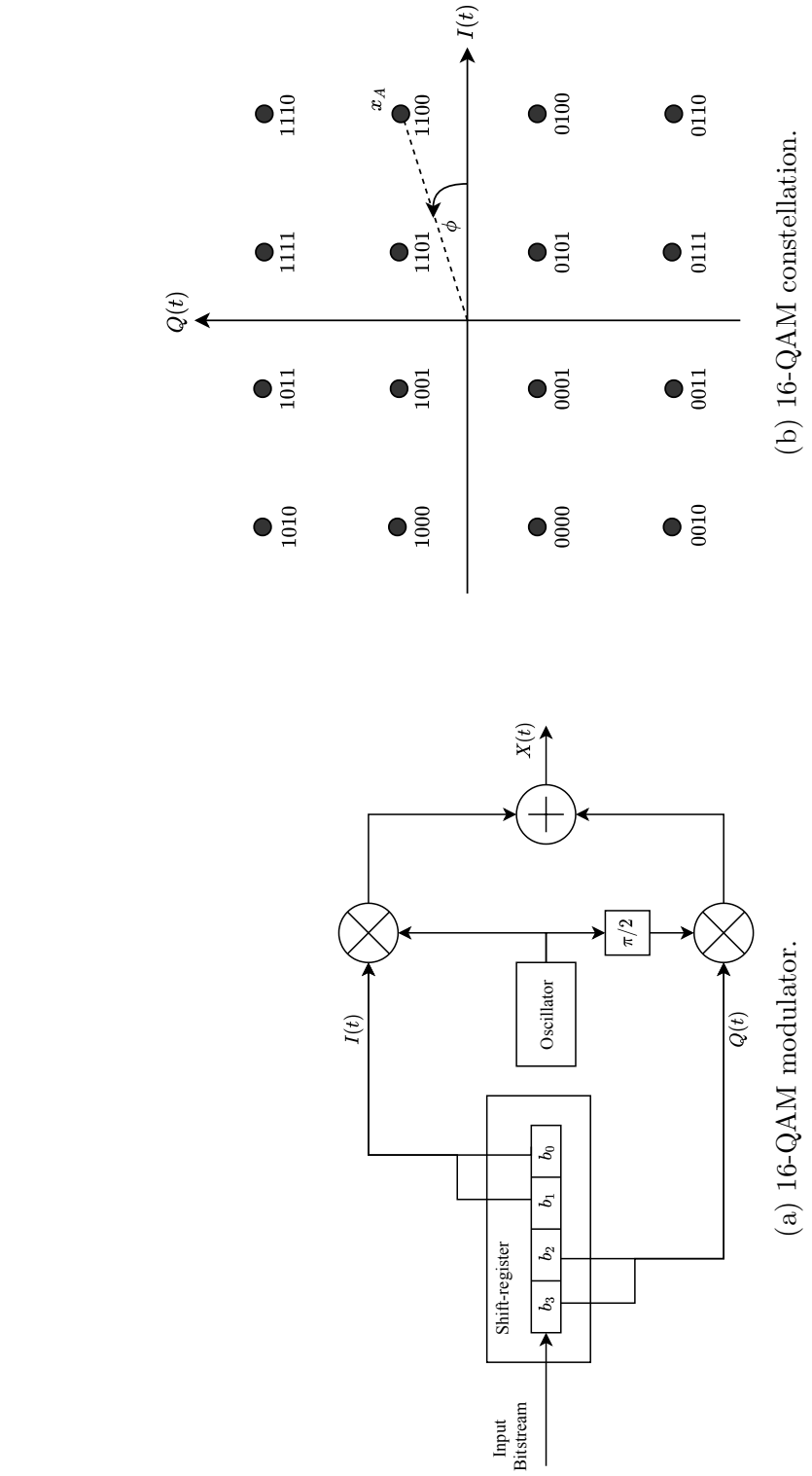
where, $x_A(t) = \sqrt{I(t)^2 + Q(t)^2}$ and $\phi(t) = \tan^{-1}(\frac{Q(t)}{I(t)})$. Therefore, it possible to perform simultaneous amplitude and phase modulation of a carrier by amplitude modulating two quadrature carriers of the same frequency.

2.3.2 Digital QAM

In the digital communication system, by QAM, we refer to multilevel QAM or M-QAM. In digital M-QAM, both $I(t)$ and $Q(t)$ signals take discrete value. A block diagram of 16-QAM and corresponding symbol mapping are illustrated in Figure 2.11a and Figure 2.11b, respectively. As can be seen, the amplitude and phase of the modulated signal depend on the shift-register state. Therefore, with M-QAM, M shift-register states are defined, and thus we have M combinations of phase and amplitude, which are referred to as QAM symbols. Each QAM symbol can carry n bits, where $n = log_2 M$.

2.4 Wireless fading Channel

It is necessary to understand the characteristics of a communication channel before designing and implementing any transmission scheme. A communication channel can be wireline or wireless. In this thesis, we consider communication over the wireless channel. In the simplest case, one may assume the signal to traverse ideal free space. Free space modeling of a wireless channel considers a direct path between the transmitter and receiver, which is free from any object that might attenuate the transmitted signal. However, in the practical wireless channels, where the signal propagates within the atmosphere and near the ground, the free space model is inefficient in terms of predicting the channel behavior and system performance [34]. In such a channel the signal interacts with many objects present in the path between the



(a) 16-QAM modulator.

(b) 16-QAM constellation.

Figure 2.11: Digital QAM.

transmitter and receiver and thus can be reflected, diffracted, and even scattered. As a result, random fluctuations occur in the amplitude, phase, and angle of arrival of the signal received at an antenna [35]. This phenomenon is known as multipath propagation, and the fluctuations in multipaths give rise to the term multipath fading. In the following subsections, we will discuss the characteristics of fading channels.

2.4.1 Fading

Typically, by fading, we refer small-scale fading, which is used to describe the rapid fluctuations of amplitudes, phases, or multipath delays of a radio signal over a short period or travel distance. Since we have multiple paths for the transmitted signal to propagate towards the receiver, we have two or more versions of the transmitted signal which arrive at the receiver at slightly different times, interfere with each other, and thus cause fading. This interference can vary widely in amplitude and phase over time.

Small-scale fading gives rise to two types of variations in the received signal: (1) time spreading of the underlying digital pulses within the signal, (2) a time-variant behavior of the channel due to motion between transmitter and receiver [35]. Time spreading of the signal due to multipath is captured by channel coherence bandwidth in the frequency domain and by the root mean square (RMS) delay spread in the time domain. On the other hand, the time-variant behavior of the channel is characterized by the channel coherence time in the time domain and by the Doppler-shift in the frequency domain.

2.4.2 Delay Spread and Coherence Bandwidth

As we receive the signal from multiple paths, the propagation time of the received signal will be different for each path. Therefore, a single transmitted pulse will be spread in time when it arrives at the receiver. This effect, which spreads out the signal, is captured by the parameter τ_s. In wireless communication, delay spread is visualized using a multipath power profile, and such a profile is illustrated in Figure 2.12. Typically, the RMS metric is used to measure the delay spread and is defined by

$$\sigma_{\tau_s} = \sqrt{\bar{\tau}_s - \bar{\tau}_s^2}$$

where,

$$\bar{\tau}_s = \frac{\sum_k P(\tau_{s_k})\tau_{s_k}}{\sum_k P(\tau_{s_k})},$$

$$\bar{\tau^2}_s = \frac{\sum_k P(\tau_{s_k})\tau_{s_k}^2}{\sum_k P(\tau_{s_k})},$$

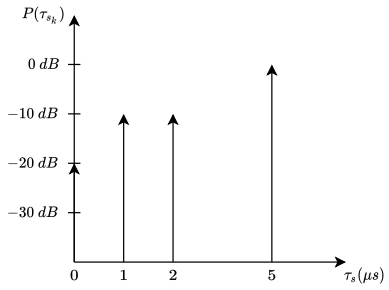


Figure 2.12: Power profile of multi-path delay spread.

$k = 1, 2, \ldots, N_k$, and σ_{τ_s} is known as RMS delay spread.

The coherence bandwidth of a channel is the frequency domain dual of the RMS delay spread and is defined by [4]

$$B_c \approx \frac{1}{5\sigma_{\tau_s}}.$$

Depending on coherence bandwidth B_c a channel can be characterized as flat fading or frequency-selective fading channel. When $B_s << B_c$, the channel is called flat fading channel and the channel is frequency-selective fading channel when $B_s > B_c$, where B_s is the signal bandwidth.

In a flat fading channel, the multipath components of the transmitted signal arrive within a very short fraction of a symbol time. Therefore, such channel can be approximated as a single tap having no delay (i.e., a single delta function with $\tau_{s_k} = 0$) and the input/output relationship of the channel can be expressed as multiplication [4], [36], that is

$$y(t) = c(t)x(t),$$

where $c(t)$ is the time-variant impulse response of the channel.

2.4.3 Doppler-shift

Consider a scenario where the transmitter is a base station (BS), and the receiver is driving a car. Therefore, motion presents between the transmitter and receiver (see Figure 2.13). Depending on the direction of driving, the angle between the mobile receiver and the received signal may vary. The angle can be positive if the receiver drives towards the BS or can be negative if drives in the opposite direction. This

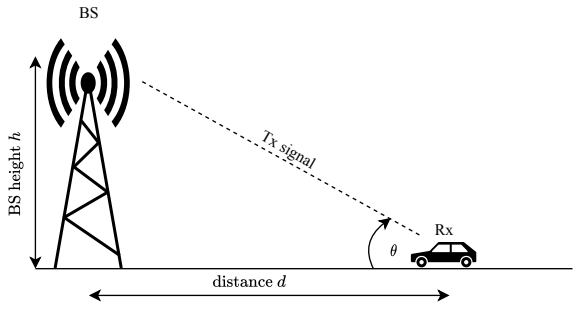


Figure 2.13: Doppler shift.

motion leads to a shift in the frequency of the radio waves, which is received by the mobile receiver. This shift is known as Doppler shift and is defined by

$$f_d = \frac{v f_c}{c} \cos\theta,$$

where, v is the relative velocity of the receiver with respect to transmitter, f_c is the frequency of the transmitted signal, and c is the speed of the light. With this shift, the frequency of the received signal becomes $f_c + f_d$ when the receiver drives towards BS and $f_c - f_d$ when drives away from BS. This is valid, if we assume there presents only a direct path between the transmitter and receiver as shown in Figure 2.13. However, in a multipath scenario, there may exist many paths between the transmitter and receiver. Thus the signals traveling along different paths can have different Doppler shifts, corresponding to different rates of change in phase. In that case, we will have Doppler spread instead of simple Doppler shift.

2.4.4 Coherence time

Coherence time is the time domain dual of Doppler spread and defined by

$$T_c = \frac{0.423}{f_m},$$

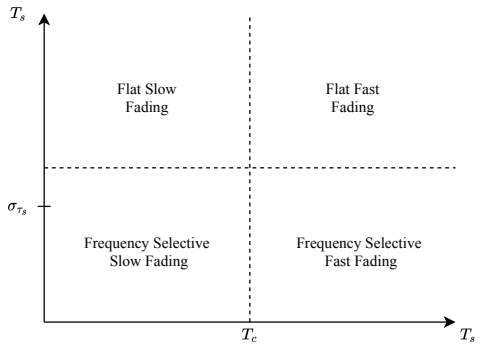


Figure 2.14: Matrix illustrating type of fading experienced by a signal as a function of symbol period T_s [4].

where f_m is the maximum Doppler-shift. It is a statistical measure of time duration over which a channel remains static, and it captures the time-variant nature of the channel due to motion. If two signals are received by a time separation greater than T_c, then these two signals are affected differently by the channel [4]. Depending on channel coherence time T_c, a channel can be fast-fading or slow fading channel. A slow-fading channel is defined by

$$T_s << T_c,$$

where T_s is the symbol duration, and a channel is a fast-fading channel when

$$T_s > T_c.$$

The relationship between different fading based on multipath and motion parameters is presented in Figure 2.14.

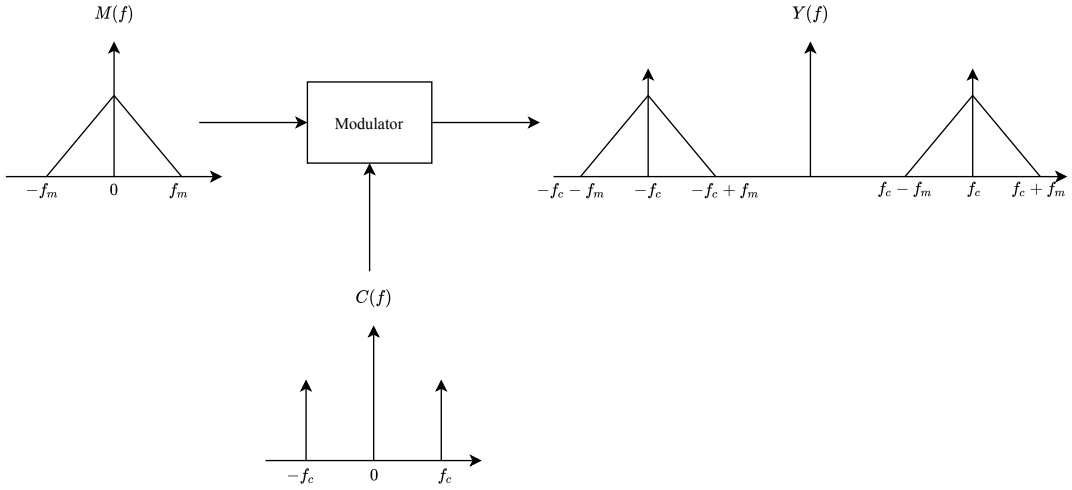


Figure 2.15: Frequency domain representation of AM process.

2.5 Complex Baseband Representation of Signals

In wireless communication system, any information signal such as audio, video is transmitted by modulating a higher frequency carrier signal. After modulation, the resulting signal is known as passband signal and the information signal at the input of the modulator is known as baseband signal. Typically, the baseband signals have much lower frequency energy compared to the carrier signal because that is property of information bearing signals produced by a source. An amplitude modulation (AM) process is illustrated in Figure 2.15, where $M(f)$, $C(f)$, $Y(f)$ represent spectra of baseband signal, high-frequency carrier signal, and passband signal respectively.

This inclusion of f_c will create unnecessary complications in theoretical analysis of communication systems. Therefore, a method of characterizing a communication system, which is independent of f_c is desired. Such characterization can be obtained by the complex baseband representation of the signals.

Let us assume $s(t)$ is a real-valued passband signal. The spectrum of $s(t)$ can be

obtained as

$$S(f) = \mathcal{F}\{s(t)\},$$

where $\mathcal{F}\{.\}$ denotes Fourier transform. Since $S(-f) = S^*(f)$ is valid then by analyzing the positive half of the spectrum $S(f)$ we can retrieve the source information [37]. Therefore, we can define

$$S_+(f) = 2u(f)S(f), \tag{2.5}$$

where $u(.)$ is the unit step function [38]. The equivalent time domain representation of (2.5) is

$$\begin{aligned} s_+(t) &= \mathcal{F}^{-1}\{S_+(f)\} \\ &= \mathcal{F}^{-1}\{2u(f)S(f)\} \\ &= \mathcal{F}^{-1}\{2u(f)\} * \mathcal{F}^{-1}\{S(f)\}, \end{aligned} \tag{2.6}$$

where $\mathcal{F}^{-1}\{.\}$ is the inverse Fourier transform. From [37], we know

$$\begin{aligned} \mathcal{F}^{-1}\{2u(f)\} &= 2\{\frac{1}{2}\delta(t) + j\frac{1}{2\pi\ t}\} \\ &= \delta(t) + j\frac{1}{\pi\ t}. \end{aligned} \tag{2.7}$$

From (2.6) and (2.7), we get

$$\begin{aligned} s_+(t) &= s(t) + j\frac{1}{\pi\ t} * s(t) \\ &= s(t) + j\hat{s}(t) \end{aligned}$$

where

$$\begin{aligned}\hat{s}(t) &= \frac{1}{\pi\, t} * s(t) \\ &= \frac{1}{\pi}\int_{\infty}^{-\infty} \frac{s(\tau)}{t-\tau} d\tau \\ &= \mathcal{H}\{s(t)\},\end{aligned}$$

$\mathcal{H}\{.\}$ denotes Hilbert transform, and $s_+(t)$ is know as the complex pre-envelop signal. The frequency response of $\mathcal{H}\{.\}$ is

$$\mathcal{H}\{f\} = \begin{cases} -j, & f > 0 \\ 0, & f = 0 \\ j, & f < 0. \end{cases}$$

Let us assume the spectrum of $\hat{s}(t)$ is $\hat{S}(f)$. Now, if we plot $S_+(f)$, we will see $s_+(t)$ is still a passband signal (see Figure 2.16).

To make $s_+(t)$ a baseband signal, we need to shift $S_+(f)$ down to origin. The

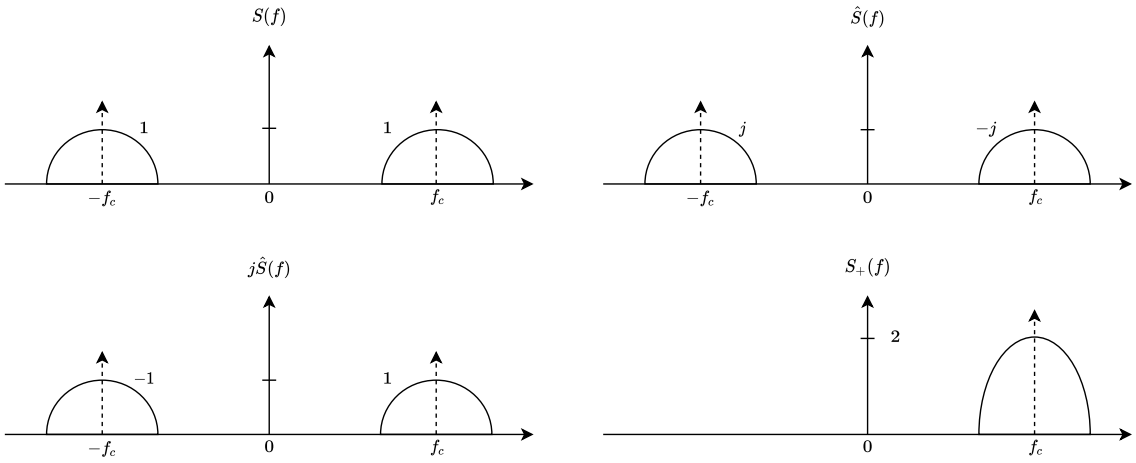


Figure 2.16: Spectrum of different signals.

corresponding baseband signal is defined as

$$S_b(f) = S_+(f + f_c) \tag{2.8}$$

and therefore, the equivalent time domain expression of (2.8) becomes

$$s_b(t) = \mathcal{F}^{-1} S_b(f) = s_+(t) e^{-j2\pi f_c t},$$

which leads to

$$s_b(t) = [s(t) + j\hat{s}(t)] e^{-j2\pi f_c t}. \tag{2.9}$$

Since $s_b(t)$ does not contain f_c, it is known as the complex baseband representation of passband signal $s(t)$. Now, if we rewrite (2.9) as

$$s(t) + j\hat{s}(t) = s_b(t) e^{j2\pi f_c t} \tag{2.10}$$

then by taking $\mathbf{Re}[s_b(t) e^{j2\pi f_c t}]$, we will get the passband signal $s(t)$. As a result, $s_b(t)$ is also known as complex envelop of $s(t)$ [38]. In general $s_b(t)$ is complex valued [38], therefore, we can define

$$s_b(t) = x(t) + jy(t).$$

Now, if we substitute $s_b(t)$ in (2.10) and equate real and imaginary parts on each side,

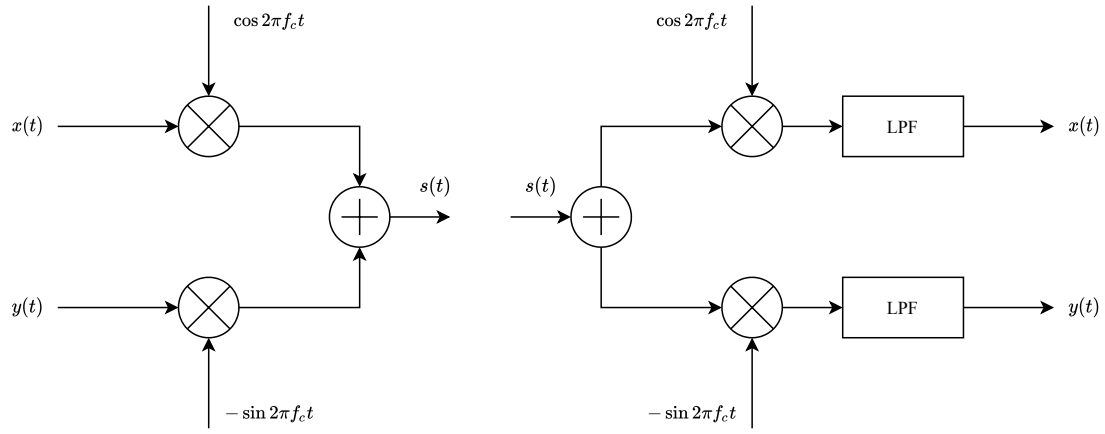


Figure 2.17: Conversion from baseband to passband and vice-versa.

we get

$$s(t) = x(t)\cos 2\pi f_c t - y(t)\sin 2\pi f_c t$$

$$\hat{s}(t) = x(t)\sin 2\pi f_c t + y(t)\cos 2\pi f_c t$$

This representation of $s(t)$ suggests a practical way to generate the complex envelop of passband signal. Figure 2.17 illustrates the conversion from baseband to passband and vice-versa.

Chapter 3

Proposed System Architecture and Design Procedure

This chapter proposes a new end-to-end HDA coding framework for reliable and efficient transmission of real-time video over cognitive channels. In contrast to a conventional design which considers fixed-rate source and channel encoders, this framework adopts a JSCC approach to overcome the *cliff effect* typically presents in a conventional design. The reliable and efficient transmission is ensured by using a transmission scheme which adapts the bit-rate to channel conditions based on the PU arrival-rates on the cognitive channels. An important feature is the use of analog refinement layer superimposed on the digital base-layer which guarantees a decoded video quality commensurate with the instantaneous channel quality.

3.1 Problem Description

Let $\boldsymbol{S}$ be a sequence of video frames modeled as a sequence of real-valued random vectors. The main goal of this study is to compress $\boldsymbol{S}$ and then transmit it over

cognitive channels through secondary access so that the distortion between $\boldsymbol{S}$ and $\hat{\boldsymbol{S}}$ is minimized subject to constraints on PU arrival-rate, transmission-rate, total transmission power, and channel bandwidth. It is assumed that the duration of a GOP (T_{GOP}) is greater than the channel coherence time (T_c) and, therefore, CSI is updated many times during the transmission of a GOP and the power allocation is repeated at regular time intervals which is referred to as channel state update (CSU) period. However, in *interweave* cognitive radio environment, each wireless sub channel is licensed to a PU and it is assumed that SU can access the channel only when the PU is inactive. Therefore, a sensing mechanism is required which periodically senses the cognitive radio environment, and set-up the secondary transmission link by considering the PU activities. It is not possible to determine the maximum channel payload capacity for the entire duration of a GOP when transmitting a video over the fading cognitive radio channels. This is because, in a secondary access scenario, the available channel payload capacity depends on random PU activity. One approach to reliable transmission is to maintain the encoder output bit-rate below the worst-case bit-rate of the channel. But this will not only result in a poor decoded video quality but it will also fail to benefit from higher channel capacity available during the periods of low fading. Another approach is to use layered video coding rather than fixed-rate coding such as HDA coding, which consists of a base-layer and an analog refinement layer. The refinement layer is superimposed on the base-layer prior to transmission. If we can set the base-layer bit-rate in such a way that it guarantees the delivery of the base-layer even in the the worst case scenario, then by transmitting the quantization error (QE) samples as a refinement layer can improve the decoded video quality. To get the benefit of HDA coding in a cognitive radio environment, we need to set the base-layer bit-rate according to CSI and the PU arrival-rate so that

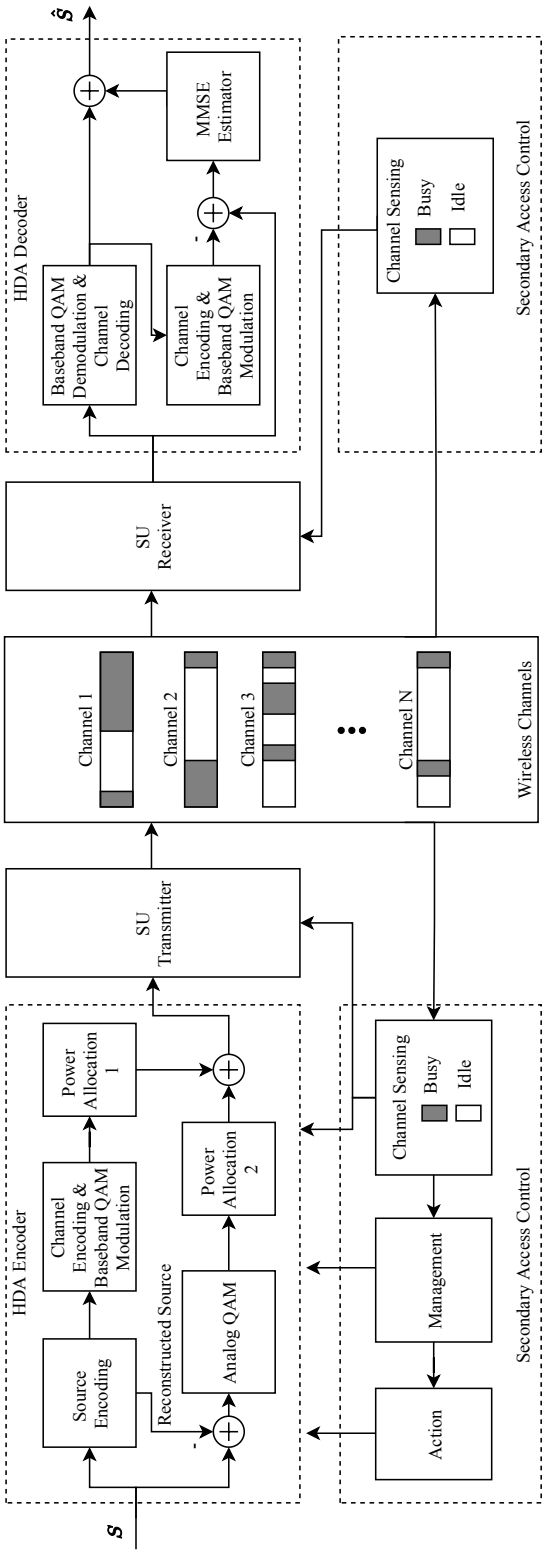


Figure 3.1: System architecture.

a certain probability of reliable transmission can be guaranteed. Once the base-layer is received error-free, depending on the number of received analog QE samples, the video quality can be improved. Thus, it is possible to achieve a higher decoded video quality and reliability at the same time by proper utilization of the available channel resources in a cognitive radio environment with HDA coding. A block diagram of the proposed system is presented in Figure 3.1.

3.2 Wireless Channel Model

In this thesis, we assume N_c primary channels, each with bandwidth BW, are available through secondary access to an SU for transmitting video. It is also assumed that random power gains of the different channels are independent, but available to both transmitter and receiver as CSI. Moreover, it is assumed that random channel gains are updated every T_{CSU} seconds, and each primary channel is licensed to a single PU.

3.3 PU Traffic Model

It is known that in a cognitive radio environment, the PU arrivals are more frequent and less correlated [8]. Therefore, to model PU traffic in such environment, Poisson process has been commonly used in [8,10,25]. To model the random PU traffic in this thesis, we also consider a Poisson process with arrival-rate λ_i, where, $i = 1, 2, \ldots, N_c$. The inter-arrival time τ_i is exponentially distributed with mean arrival time $1/\lambda_i$. We assume that upon arrival of a PU on a particular channel, it occupies the channel for a while to complete its task and then departs. In queuing theory [39], the occupancy time can be well modeled by the service time μ_i which is also exponentially distributed. Both τ_i and μ_i are considered as independent random variables.

3.4 Secondary Access Model

In this thesis, we consider real-time video transmission using *interweave* secondary access model [8], [10], [22]; that is, the SU and PU cannot access the channel concurrently. The secondary access to a channel depends on the PU's usage statistics of that channel. These usage statistics are captured by τ_i and μ_i discussed in section 3.3. The SU can access the ith channel for a duration of τ_i. At the end of this duration, the ith channel is reoccupied by the PU and the secondary transmission link breaks. However, a μ_i duration, thereafter, the channel becomes vacant again. Therefore, it is necessary to introduce a secondary access model that can be used to sense the vacant slots, initiate channel selection, and perform data transmission. Figure 3.2 illustrates the secondary access scenario.

Since a single video transmission session involves the transmission of multiple GOPs, it is assumed that the secondary link setup and data transmission occur per GOP basis. The GOP frame structure is shown in Figure 3.3.

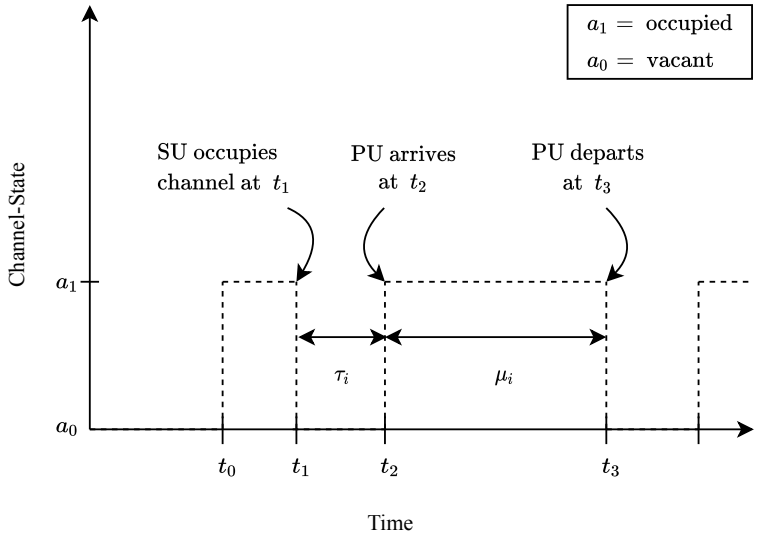


Figure 3.2: Secondary access over a cognitive channel.

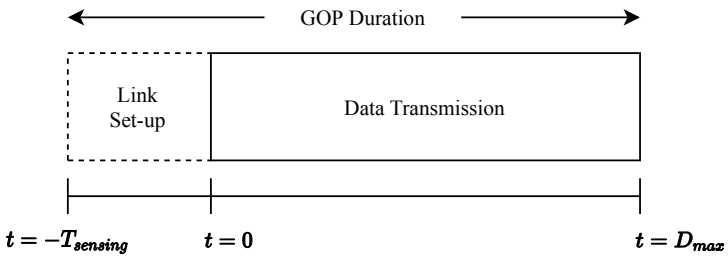


Figure 3.3: GOP frame structure.

While most of the previous work such as [8], [10], adopts a fixed sensing-transmission scheme, we consider the flexible scheme that is proposed in [22]. The fixed sensing-transmission scheme consists of a fixed link setup period followed by a fixed data transmission period. In this setup, if a channel is occupied by a PU after the sensing period, the SU will not access the channel. The channel will be sensed again after a D_{max} duration, which results in a waste of channel resources as the channel can be vacant before D_{max}. In the flexible sensing-transmission scheme, although the link setup period is fixed, the transmission duration is adaptively determined according to the PU activities, channel sensing results, and the channel quality [22].

The link setup period of a GOP frame includes three operations: sensing, management, and action.

3.4.1 Sensing

The secondary link setup begins at $t = -T_{sensing}$ with sensing of spectrum holes [8]. During the sensing period, a set of N_v vacant channels out of N_c is selected using the parameter $a(i)$, wherein $a(i) = 0$ indicates the availability of channel i and $a(i) = 1$ indicates otherwise. Also, this sensing period includes two more important tasks: prediction of the channel capacity and the estimation of CSI. The latter is beyond the scope of this study and assumed to be available to both secondary transmitter

and receiver. To predict the channel capacity, we consider the arrival-rate of the PU along with a predefined probability $P_{success}$, which indicates the probability of successful transmission of B bits for the duration of an entire GOP by the SU.

Channel Capacity- It is assumed that a GOP is transmitted using fixed-length packets, where each packet contains N_m QAM symbols. Let R_i be the rate in bits/QAM symbol, where $i = 1, 2, \ldots, N_c$. Therefore, each packet can carry $R_i N_m$ bits. Now, if we can find out how many packets can be transmitted with given probability $P_{success}$ within a GOP duration, then it is possible to determine B. Let's assume, the duration of a packet is T_p and we want to transmit η_i packets before a PU arrival occurs on channel i. To ensure, η_i packets get transmitted before a PU arrival, the arrival should occur after $\eta_i T_p$. Since the PU arrival occurs after a τ_i duration, the probability of successful transmission of η_i packets is $P_{success} = P(\eta_i T_p \leq \tau_i)$. Given that τ_i follows an exponential distribution we get,

$$P(\tau_i \geq \eta_i T_p) = \int_{\eta_i T_p}^{\infty} \lambda_i e^{-\lambda_i t} dt, \tag{3.1}$$

where λ_i is the PU arrival-rate of the ith channel. By solving (3.1), we get

$$\eta_i = \frac{-\ln(P_{success})}{\lambda_i T_p}.$$

It should be noted that a higher value of $P_{success}$ increases the reliability but decreases the decoded video quality. Given η_i, we obtain

$$B_i = \eta_i R_i N_m,$$

where B_i is the estimated channel capacity of the i-th channel.

3.4.2 Management

The management unit performs two very specific tasks: best channel selection and transmission rate control. Once the information regarding vacant channels and their capacities are forwarded by the sensing unit to management unit, the management unit first figures out the channel k among the N_v ($N_v \leq N_c$) candidate channels using (3.2)

$$k = \arg\max_i \{B_i : \ i = 1, \ldots, N_v\}, \tag{3.2}$$

as the cognitive radio channel for transmission of the current GOP and hence $B = B_k$, where B is the estimated capacity of the selected channel.

Transmission rate control is achieved by using Algorithm 1 which depends on a set of modulation and coding scheme (MCS) presented in Table 3.1. Typically, all modern multimedia transmission standards depend on such MCS set. This MCS set is defined in [40], but the approximate threshold CSNR value of each MCS which corresponds to a bit-error rate of less than 10^{-7} has been estimated using Monte-Carlo simulation [41]. In Algorithm 1, I_{MCS} represents the MCS no. presented in Table 3.1, P_D is the power required for digital transmission, P_T is the total power

Algorithm 1 Transmission rate control

```
Initialize: I_MCS = 8, P_D = 0, P_T = 1
Get CSNR of the selected channel
while true do
    Compute P_D = ThresholdSNR (I_MCS) / CSNR    ▷ ThresholdSNR obatins from Table 3.1
    if 0 < P_D ≤ P_T then
        return I_MCS
        break
    else
        I_MCS ← I_MCS − 1
Using I_MCS, get an appropriate MCS set from Table 3.1.
```

Table 3.1: MCS Set Defined in The IEEE802.11a Standard

MCS no.	Modulation	Channel Coderate	bits/symbol	Threshold SNR (dB)
1	BPSK	1/2	0.5	7
2	BPSK	3/4	0.75	9
3	QPSK	1/2	1	10
4	QPSK	3/4	1.5	13
5	16-QAM	1/2	2	14
6	16-QAM	3/4	3	20
7	64-QAM	2/3	4	24
8	64-QAM	3/4	4.5	26

budget, and it is assumed that CSNR is available to both the transmitter and the receiver as CSI.

3.4.3 Action

After secondary link is established, the next step is to prepare the coded video data and quantization errors for transmission. Now, consider the HDA QAM tranmission scheme, shown in Figure 3.4. Let D_i and V_i be zero mean complex random variables which denote digital and analog QAM symbols respectively. Let g_i be the complex gain and W_i be the zero mean complex white Gaussian noise of the ith channel. In HDA coding, analog QAM symbols are superimposed on digital QAM symbols to form a hybrid symbol as shown. Let X_H be the hybrid symbol which is transmitted over the cognitive radio channel, and Y_i be the received signal. Therefore, we get

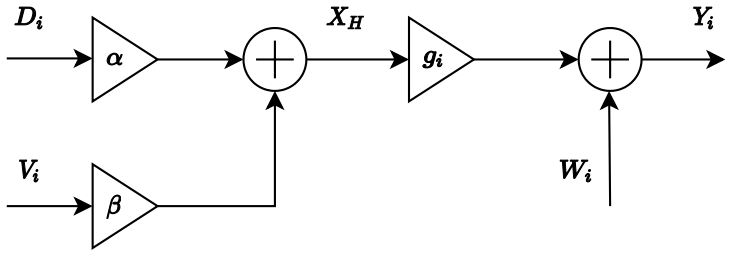


Figure 3.4: Power allocation blocks of HDA QAM transmission system.

$$Y_i = g_i X_H + W_i.$$

After performing channel equalization, we get

$$\tilde{Y_i} = Y_i/g_i = \alpha D_i + \beta V_i + \frac{W_i}{g_i}, \tag{3.3}$$

where α and β are the digital and analog power allocation factors respectively. The equalized channel output $\tilde{Y_i}$ is then fed to the channel decoder (see Figure 3.1). From (3.3), it can be seen that the total noise at the input to the channel decoder for the ith channel will be the sum of the interference from the analog signal and the channel noise. Therefore, the CSNR at the input of the channel decoder becomes

$$\gamma_i = \frac{\alpha^2 |D_i|^2 |g_i|^2}{\beta^2 |V_i|^2 |g_i|^2 + |W_i|^2}. \tag{3.4}$$

In order to ensure a low decoding error probability, it is required to allot sufficient power to the base-layer so that γ_i is above a certain threshold which depends on the MCS (I_{MCS}) chosen for the i-th channel. However, the power required to ensure a given minimum CSNR at the receiver is often less than the total available power as the available number of MCS is finite. Therefore, any remaining power can be used to transmit analog symbols carrying quantization errors. The power allocation scheme that is used in this thesis is presented in section 3.5.

3.5 Power Allocation

Our HDA coding scheme requires that the base-layer get transmitted error free. Therefore, we first allot sufficient power to digital QAM symbols. Let us consider, $\alpha = \sqrt{P_D}$ and $\beta = \sqrt{P_A}$, where P_D and P_A are the amounts of power allotted to

digital and analog QAM symbols respectively. If we consider the transmission of the base-layer only, then the CSNR at the input of the channel decoder is

$$\gamma_i = \frac{P_D |g_i|^2 |D_i|^2}{|W_i|^2}. \tag{3.5}$$

Without a loss of generality, we assume that $E\{|D_i|^2\} = 1$ and $E\{|W_i|^2\} = 1$. Then from (3.5) we get

$$P_D = \frac{\gamma_i}{|g_i|^2}, \tag{3.6}$$

where $0 < P_D \leq P_T$ and P_T is the total transmission power budget.

From (3.3) it can be seen that the analog refinement layer appears as noise to the base-layer, therefore, if we allot P_D for the base-layer and simply use rest of the power $P_A = P_T - P_D$ for transmitting the analog refinement layer, then γ_i will decrease, which will increase the decoding error probability. To ensure that γ_i is kept above the threshold CSNR of the MCS being in use, we need to add simultaneously more power to the digital QAM symbols. In order to meet this requirement, we rewrite (3.4) as

$$\frac{(P_D + P'_D)|g_i|^2}{P_A |g_i|^2 E\{|V_i|^2\} + 1} \geq \gamma_i, \tag{3.7}$$

where P'_D is the penalty that is added to P_D to compensate for the increased noise due to the analog refinement layer. Since, it is considered that the complex symbol V_i is formed by combining a pair of transformed QE coefficients (see section 3.6.2),

we get $E\{|V_i|^2\} = 2\sigma_{v_i}^2$. Therefore, we can rewrite (3.7) as

$$\frac{(P_D + P'_D)|g_i|^2}{P_A|g_i|^2 2\sigma_{v_i}^2 + 1} \geq \gamma_i. \tag{3.8}$$

After allocation of P_D the remaining unused power is

$$P_{rem} = P_T - P_D. \tag{3.9}$$

This remaining power is then allotted to analog QAM symbols. Therefore, we get

$$P_{rem} = P_A + P'_D. \tag{3.10}$$

From (3.9) and (3.10), we get

$$P'_D = P_T - P_D - P_A. \tag{3.11}$$

Using (3.11) , we can rewrite (3.8) as

$$\frac{(P_T - P_A)|g_i|^2}{P_A|g_i|^2 2\sigma_{v_i}^2 + 1} \geq \gamma_i,$$

and therefore,

$$P_A \leq \frac{P_T - P_D}{1 + 2\sigma_{v_i}^2 \gamma_i}.$$

The maximum power that can be used for analog transmission is therefore,

$$P_A = \frac{P_T - P_D}{1 + 2\sigma_{v_i}^2 \gamma_i},$$

where $0 \leq P_A < P_T$.

3.6 Proposed HDA codec

The encoder-decoder pair of the HDA codec is illustrated in Figure 3.5. We can divide the HDA coding process into two parts: digital coding and analog coding.

3.6.1 Digital coding scheme

The digital encoder compresses the input video sequence, controls the bit-rate, and then performs channel encoding and complex baseband QAM modulation.

An important parameter in encoding a GOP is QP, which controls the output bit-rate of the H.264/AVC video encoder. Therefore, the digital encoder first focuses on finding an appropriate QP. As we are considering secondary access over a licensed channel, it is assumed that QP is selected by considering the constraints imposed by the PU arrival-rate, mean CSNR, and channel code-rate. The algorithm that is used in this thesis to find QP and encode a GOP is shown in Algorithm 2. In this algorithm, V_{fr} and N_f represent the video frame-rate and the number of frames in a GOP, respectively. After encoding, the base-layer data (H.264 bitstream) is forwarded to the channel encoder. The code-rates presented in Table 3.1 are those of the rate- 1/2 punctured convolutional code specified for OFDM-PHY in IEEE802.11a [40]. However, it has been found that occasional bit-errors do occur despite the error-correction by the convolutional coders and thus the base-layer doesn't decode properly. Therefore, the base-layer data is further protected by an outer code before begin passed to the channel encoder. We use Reed-Solomon code $RS(255, 239)$ with 8-bit symbols specified in [42] for this purpose.

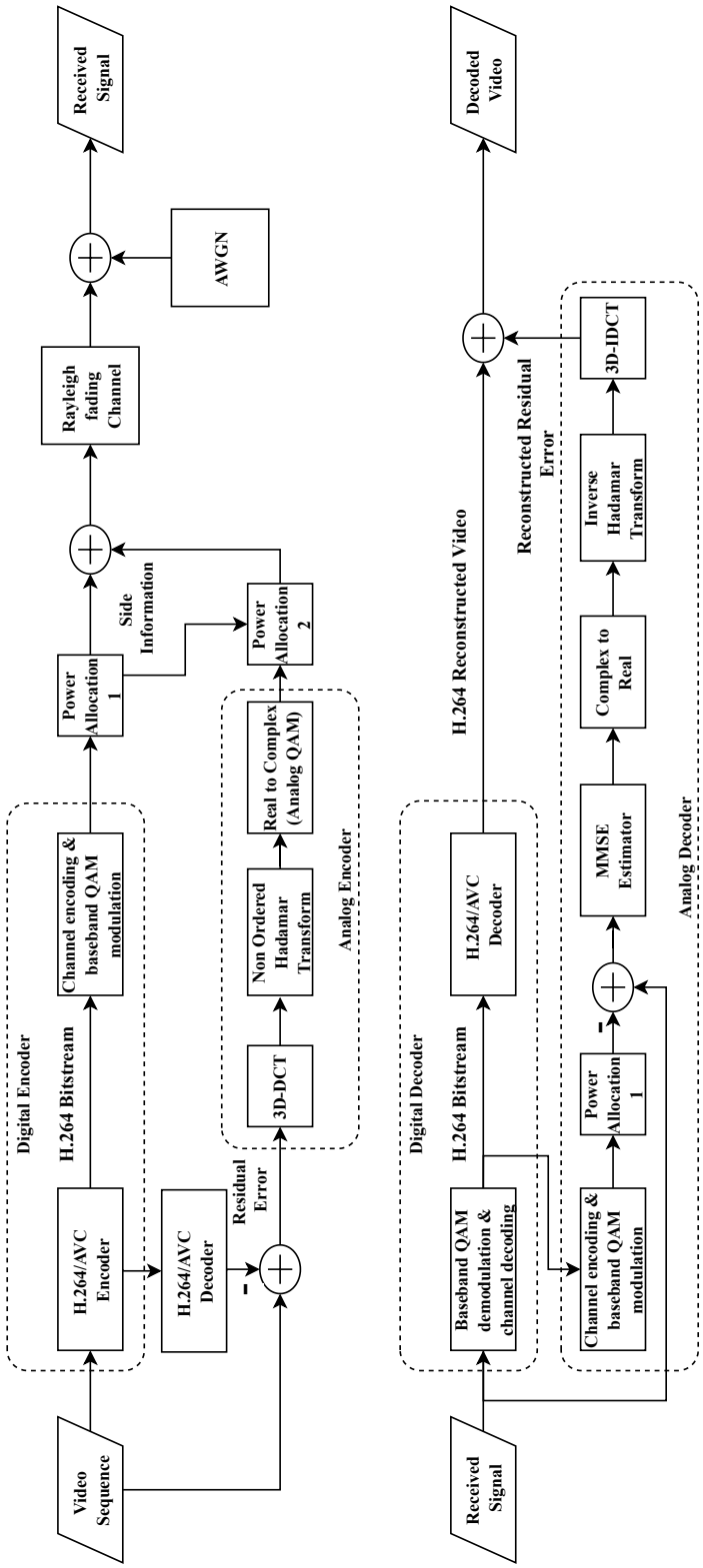


Figure 3.5: The encoder and decoder of HDA system.

Algorithm 2 Video bit-rate control

Initialize: $B_{cap} = 0$, $B_{rate} = 0$, $a = 0$, $QP = 0$, $m = QP_{min}$, and $n = QP_{max}$
Get B and mean CSNR of the selected channel
Using Algorithm 1 get the desired channel code-rate.
Using the code-rate, find out how many extra bits (B_{extra}) will be add to the input
$B_{cap} \leftarrow (B - B_{extra})$, and $B_{rate} \leftarrow \frac{B_{cap} \times V_{fr}}{N_f}$
while true **do**
    $a \leftarrow round(\frac{m+n}{2})$, and $QP \leftarrow a$
    Compress GOP and calculate the output bit-rate B_{output}
    if $B_{output} > B_{rate}$ **then**
        $m \leftarrow a + 1$
    else if $B_{output} < PB_{rate}$ **then** ▷ P indicates percentage (%)
        $n \leftarrow a - 1$
    else if $PB_{rate} < B_{output} \leq B_{rate}$ **then**
        break
return QP, B_{output}

3.6.2 Analog coding scheme

The first step of analog coding is to reconstruct the digitally encoded video and generate QE samples. Once we have the H.264 bitstream, using an H.264 decoder local to the transmitter, it is possible to decode the video. Earlier in section 3.1, we assume $\boldsymbol{S}$ as a sequence of real-valued random vectors. However, as we now focus on video transmission, and the basic unit of encoding and decoding is a GOP. Therefore, $\boldsymbol{S}$ here represents a GOP of N_f consecutive video frames. Now, if we assume $\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ is the reconstructed base-layer then QE sequence of a GOP is defined as $\boldsymbol{S} - \hat{\boldsymbol{S}}_{\boldsymbol{b}}$. However, in the context of HDA coding QE is know as the residual error, and we will use these two term interchangeably. This residual error is then further processed by the analog encoder as shown in Figure 3.5.

Both the original and the reconstructed video sequences have intra- and inter-frame redundancies. Some of these redundancies can also appear in the residual sequence particularly at low bit-rates. Therefore, it can be very inefficient in terms

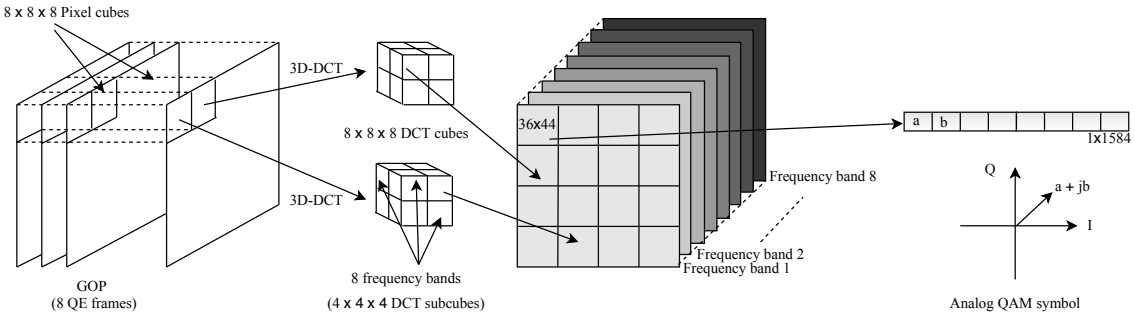


Figure 3.6: Mapping of QE samples of a GOP to analog symbols.

of bandwidth to send the residual error samples directly over the channel. It is possible to reduce these redundancies by applying 3D-DCT to the residual error sequence as it exploits both space and time correlations among consecutive frames. In earlier research it has been found that 3D-DCT is a very effective method to compress video in analog transmission of video [32]. In this thesis, we consider color video in YCbCr format, where Y represents the luminance component, and the other two represent chrominance components. Therefore, all of these three components need to be processed before transmission. The complete process is illustrated in Figure 3.6, which is applied to each of the components individually. At first, 3D-DCT is performed over $8 \times 8 \times 8$ non-overlapping cubes of QE pixels. Then like [41], we divide each $8 \times 8 \times 8$ coefficient cube into $4 \times 4 \times 4$ sub-cubes, which can be considered as different frequency sub-bands. Later, we collect all the sub-cubes from different DCT cubes that fall under the same frequency sub-band to form a frequency band. Thus we have eight frequency bands for each of the three components.

Each of these frequency bands is then divided into non-overlapping 36×44 rectangular sub-blocks, which we refer to as analog packets. Therefore, we have 768 analog packets in total for all three components in a GOP. These packets are then whitened by non-ordered Hadamard transform so that they will appear as white noise to digital symbols. Later, the variances of analog packets are computed to organize

them in descending order. This approach is adopted so that we can transmit as much high variance packets as possible before having to vacate the channel due to the possible arrival of the licensed PU on the channel. Due to the energy compaction property of the DCT, the packets carrying quantization errors with higher variances contain the most important information required to reconstruct the video. Thus this packet ordering strategy ensures the best possible video quality for a given number of transmitted analog packets. When modulating, each packet is converted to an array of size 1584 coefficients, and two consecutive coefficients are taken as in-phase and quadrature components to form one analog QAM symbol (see Figure 3.6).

Another important component of the analog coding scheme is to estimate DCT coefficients from the noisy QAM symbols at the decoder side. In this thesis, we consider a linear minimum mean square (MMSE) estimator for this purpose which is presented in section 3.7.

3.7 Linear MMSE Estimator

The problem at hand is to estimate a complex random variable V_i given the observed value of another complex random variable Y_a (see Figure 3.7). In linear estimation, the estimated value of $\hat{V_i}$ is modeled as a liner function of Y_a. Therefore we have,

$$\hat{V_i} = aY_a + b, \tag{3.12}$$

where the optimal values of the coefficients a and b have to be determined by minimizing the mean square error (MSE) of the estimator. The MSE of this estimator is given by

$$\begin{aligned} e &= E\{(|V_i| - |\hat{V}_i|)^2\} \\ &= E\{(|V_i| - a|Y_a| - b)^2\}. \end{aligned}$$

With the above error measure the optimal estimator becomes

$$(a^*, b^*) = \arg\min\ E\{(|V_i| - |\hat{V}_i|)^2\}.$$

To solve the minimization problem, we let

$$\begin{aligned} \frac{\partial}{\partial a}[E\{(|V_i| - a|Y_a| - b)^2\}] &= 0 \\ E\{2(|V_i| - a|Y_a| - b)(-|Y_a|)\} &= 0 \\ -2E\{|V_i||Y_a|\} + 2aE\{|Y_a|^2\} + 2bE\{|Y_a|\} &= 0 \end{aligned} \tag{3.13}$$

and,

$$\begin{aligned} \frac{\partial}{\partial b}[E\{(|V_i| - a|Y_a| - b)^2\}] &= 0 \\ \therefore b^* &= E\{|V_i|\} - aE\{|Y_a|\}. \end{aligned} \tag{3.14}$$

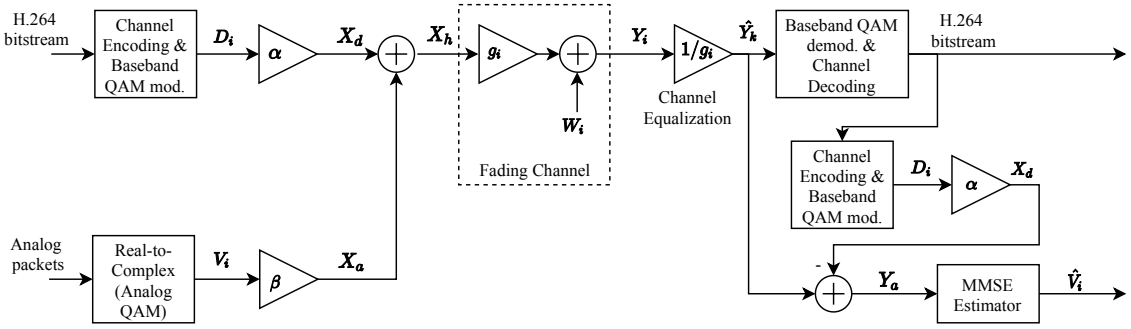


Figure 3.7: An HDA QAM transmission system.

Now, by substituting b in (3.13) we get

$$aE\{|Y_a|^2\} + [E\{|V_i|\} - aE\{|Y_a|\}]E\{|Y_a|\} = E\{|V_i||Y_a|\}$$
$$a^* = \frac{E\{|V_i||Y_a|\} - E\{|V_i|\}E\{|Y_a|\}}{E\{|Y_a|^2\} - E^2\{|Y_a|\}}. \tag{3.15}$$

we can rewrite (3.15) as

$$a^* = \frac{Cov(V_i, Y_a)}{\sigma^2_{Y_a}}. \tag{3.16}$$

Therefore, from (3.12), (3.14) and (3.16) we get

$$\hat{V_i} = \frac{Cov(V_i, Y_a)}{\sigma^2_{Y_a}}[Y_a - E\{|Y_a|\}] + E\{|V_i|\}. \tag{3.17}$$

In order to get the value of $Cov(V_i, Y_a)$ and $\sigma^2_{Y_a}$ we need to consider Figure 3.7 from which we get

$$X_d = \alpha D_i$$
$$X_a = \beta V_i$$
$$Y_i = g_i(X_d + X_a) + W_i$$
$$\hat{Y_i} = X_d + X_a + \frac{W_i}{g_i}$$
$$Y_a = X_a + \frac{W_i}{g_i}. \tag{3.18}$$

Now, by substituting X_a to (3.18) we get

$$Y_a = \beta V_i + \frac{W_i}{g_i}, \tag{3.19}$$

and therefore,

$$\begin{aligned}
\sigma^2_{Y_a} &= E\{|Y_a|^2\} - E^2\{|Y_a|\} \\
&= E\{(\beta|V_i| + \frac{|W_i|}{|g_i|})^2\} - E^2\{(\beta|V_i| + \frac{|W_i|}{|g_i|})\} \\
&= E\{\beta^2|V_i|^2\} + 2E\{\beta|V_i|\frac{|W_i|}{|g_i|}\} + E\{\frac{|W_i|^2}{|g_i|^2}\} - E^2\{\beta|V_i|\} - E^2\{\frac{|W_i|}{|g_i|}\} \\
&= \beta^2 E\{|V_i|^2\} + \frac{2\beta}{|g_i|}E\{|V_i||W_i|\} + \frac{1}{|g_i|^2}E\{|W_i|^2\} - \beta E^2\{|V_i|\} - \frac{1}{|g_i|}E^2\{|W_i|\}.
\end{aligned}$$

Since V_i and W_i are independent, $W_i \sim \mathcal{N}(0,1)$, and $E\{|V_i|\} = 0$ (see section 3.4.3), we get

$$\sigma^2_{Y_a} = \beta^2\sigma^2_{V_i} + \frac{1}{|g_i|^2}. \tag{3.20}$$

We also have

$$\begin{aligned}
Cov(V_i, Y_a) &= Cov(V_i, \beta V_i + \frac{W_i}{g_i}) \\
&= Cov(V_i, \beta V_i) + Cov(V_i, \frac{W_i}{g_i}) \\
&= \beta E\{|V_i|^2\} - \beta E\{|V_i|\}E\{|V_i|\},
\end{aligned}$$

and therefore

$$Cov(V_i, Y_a) = \beta\sigma^2_{V_i}. \tag{3.21}$$

Now, by substituting the values in (3.20) and (3.21) in (3.17) we get

$$\hat{V_i} = \frac{\beta\sigma^2_{V_i}|g_i|^2}{\beta^2\sigma^2_{V_i}|g_i|^2 + 1}Y_a,$$

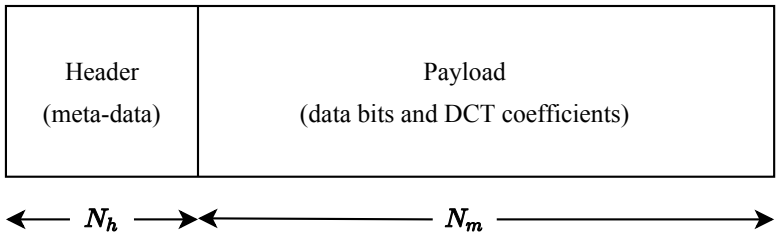


Figure 3.8: HDA Packet structure.

which defines the linear MMSE estimator for analog symbols V_i.

3.8 Packet Structure

As mentioned in section 3.4.1, the GOPs are transmitted using fixed length packets of size N_m symbols/packet. In HDA transmission system, it is necessary to transmit certain amount of meta-data to aid decoding the packet, which includes quantized values of the power allocation factors α and β, the means and variances of analog packets, and the sequence numbers indicating the locations of analog packets in the frequency bands. It is assumed that the headers are also transmitted using fixed length packets, but they contain N_h symbols each, where $N_h << N_m$. However, analog symbols are not transmitted in the header. Since, meta-data are crucial for decoding, they are sent by using the lowest possible MCS at the start of every data packets to ensure the maximum reliability. A schematic of the HDA packet structure is shown in Figure 3.8.

Chapter 4

Simulation Results and Analysis

In this chapter, we present the test results obtained using a complete implementation of an HDA codec for transmission of videos over cognitive radio channels as described in chapter 3. To compare the performance of our model, pure analog coding and digital layered coding (DLC) are also implemented. This chapter is organized as follows: first we discuss the assumptions and parameters that we adopt to perform the simulations followed by a brief discussion on pure analog and DLC codecs. Then the rest of the chapter covers the simulation results and the analysis of the outcomes.

4.1 Assumptions

The following assumptions are made for simulation purposes:

- Channel parameters such as fading gains, mean CSNR are available to both the secondary transmitter and the receiver as CSI.
- Arrival-rates and service time are fixed for each distinct cognitive channel and are known to the secondary transmitter.

- The transmitter senses the vacant channels at the start of each GOP while the receiver is continuously sensing the secondary link to sense the possible PU arrivals during each reception of a packet. Therefore, after the sensing period, if an arrival occurs on the selected channel during the transmission of packets belonging to one GOP, the whole GOP is considered as lost if the data packets already received by the receiver is not sufficient for decoding the current GOP.

- The channel state remains unchanged for the duration of a packet, and hence, CSI is updated many times during the transmission of a GOP.

- The codebooks and normalizing constants are available locally to both the sender and receiver to encode and decode the mean and variance of the analog packets as well as the power allocation factors of digital and analog packets.

4.2 Parameterization

4.2.1 Video Source and Format

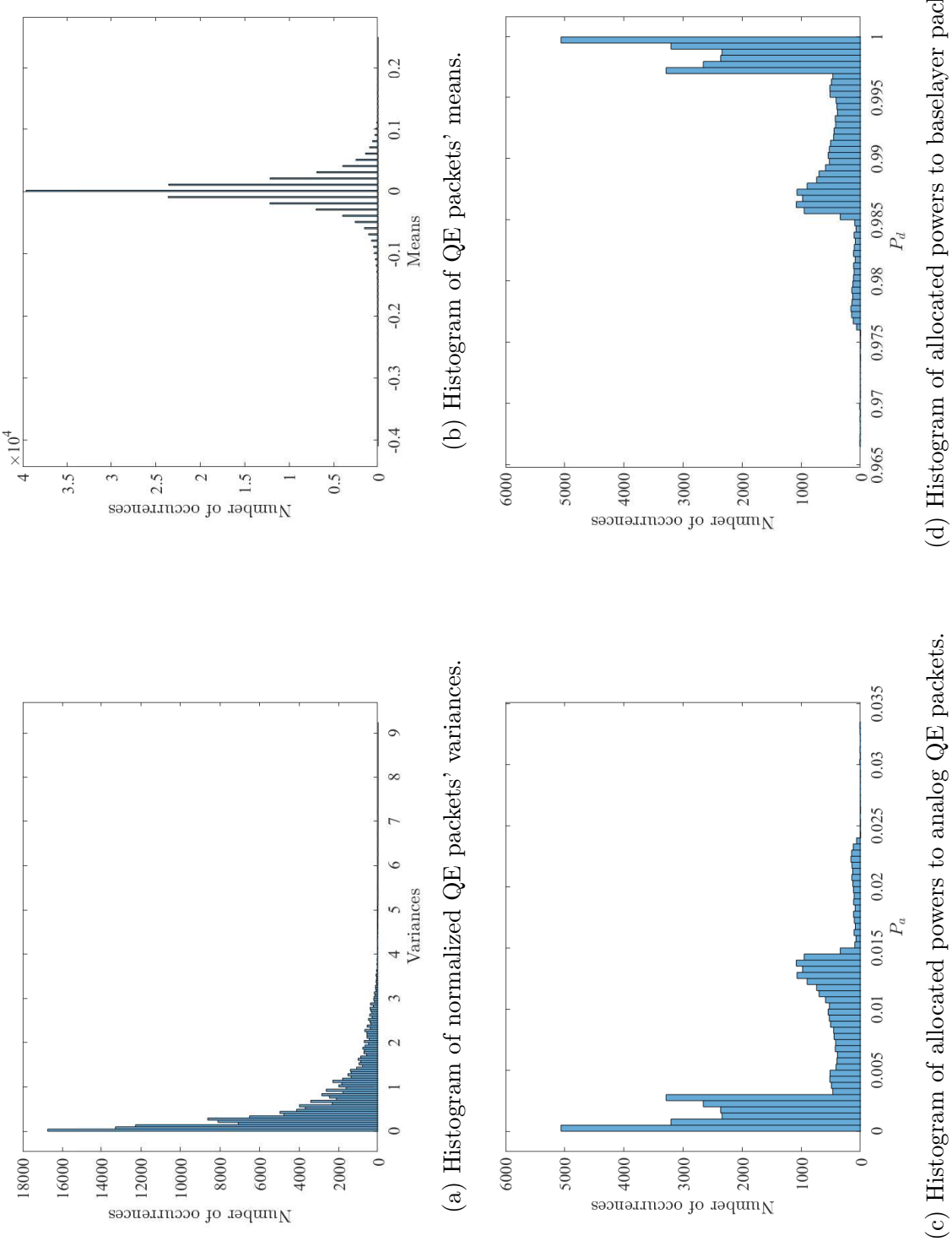
We consider a number of standard CIF (352×288) color video sequences which are widely used in literature and they are: *Football*, *Foreman*, *Coastguard*, *Stefan*, and *Tabletennis*, where the first consists of 360 frames, and the rest has 296 frames each. These video sequences are rich in motion activity and scene changes. To perform the simulations, we merge these five video sequences one after another to form a single video sequence that contains 1544 video frames in total. As mentioned in chapter 3, to process the video sequence for transmission, we use YCbCr 4:2:0 color video format. In this format, only Y frames have the full spatial resulation while Cb and Cr have a sub-sampled resolution of 176×144.

4.2.2 Cognitive Radio Channels

To simulate the cognitive radio environment, we consider five flat slow-fading wireless channels and each of which is licensed to a PU. Therefore, the number of cognitive channels $N_c = 5$. The channel traces are generated for all five channels using Clarke's model [4] in baseband with maximum Doppler shifts of 85 Hz, 81 Hz, 87 Hz, 78 Hz, and 89 Hz respectively. The sampling rate of CIF video at 30 fps is 4.56×10^6 pixels/second, and hence the bandwidth of a CIF video is less than 2.3 MHz. In our experiments, we set the channel bandwidth to 2 MHz. Additionally, we set $P_{success} = 0.90$, which is the desired probability of successful secondary transmission of packets. This relatively high probability ensures that the baselayer of a GOP gets fully transmitted before an arrival occurs, and thus increases the reliability. The arrival-rates (λ) of PU on five different channels are set to $10T_{GOP}$, $5T_{GOP}$, $7T_{GOP}$, $4T_{GOP}$, and $3T_{GOP}$ respectively. On the other hand, departure-rate is set to $15/T_{GOP}$ for all five channels. Such values are chosen so that the effect of random PU arrivals can be experienced during experiments as the duration of whole video sequence is around 52 seconds.

4.2.3 HDA System

To generate the baselayer of the HDA codec, like [11–13,41], we use JM 19.0 reference software for H.264/AVC video codec in high profile. As mentioned in chapter 3, the HDA system requires the transmission of certain amount of meta-data for decoding the received packets, which include the quantized values of the power allocation factors for both digital and analog packets, the mean and variances of each analog packets, and the sequence numbers indicating the locations of each analog packets in the frequency band. The power allocation factors are quantized using 5 bits each while



(a) Histogram of normalized QE packets' variances.

(b) Histogram of QE packets' means.

(c) Histogram of allocated powers to analog QE packets.

(d) Histogram of allocated powers to baselayer packets.

Figure 4.1: Histograms of means, variances, and allocated powers.

6 bits are used to quantize mean and variance of each analog packets. On the other hand, 10 bits are used to transmit the location of each analog packets. Therefore, the header of a packet consists of 32 data bits in total. The headers are transmitted with fixed MCS ($I_{MCS} = 1$) to ensure maximum reliability. It should be noted that the analog QAM symbols are not transmitted in the packet header.

Design of Quantizers for Means, Variances, and Power Allocation factors- To design the quantizers, we simply rely on Lloyd-Max scalar quantizer [43] as it is easy to implement. Figure 4.1 illustrates the histogram of means and variances of analog packets and power allocation factors of digital and analog packets. Although the histograms that are presented here are generated for a channel with mean CSNR = 23 dB, through the experiments it has been found that with other mean CSNR values, these variables show quite similar distribution patterns. With such non-uniform distributions of these variables, the MSE of the implemented quantizers remains within 10^{-4}. However, with each mean CSNR value the mean and variance of the analog packets and the power allocation factors of both digital and analog packets differ significantly, thus we have to generate codebooks for each different mean CSNR.

4.2.4 Delay and Other Requirements

It has been found that for video transmission the delay requirement is typically between $150 - 400$ ms [44]. Therefore, like [8], we set $D_{max} = 200$ ms which is the maximum tolerable delay. Also, the sensing time $T_{sensing}$ is set to 10 ms according to [8]. Several other parameters which are used for simulations are presented in Table 4.1.

Table 4.1: Other simulation Parameters

Parameter	Symbol	Value
Video frame rate		30 fps
Video resolution (bits/YCbCr-pixel)		24
GOP size (frames)	L_{GOP}	8
GOP duration	T_{GOP}	267 ms
QAM video data symbols/packet	N_m	792
QAM meta data symbols/packet	N_h	64
Packet duration	T_p	428 μs
Packets per GOP	N_p	622
QAM symbol duration	T_{sym}	0.5 μs

4.3 Digital Layered and Pure Analog video coding

The encoder-decoder pair of DLC system is presented in Figure 4.2. In this system, both the baselayer and refinement layer are generated using H.264/AVC encoder. To make this system comparable with HDA system, the baselayer of the DLC system is generated similar to that in HDA system with $P_{success} = 0.90$. On the other hand, the refinement layer of the DLC system, is generated with much higher bit-rate than the baselayer by setting $P_{success} = 0.85$. This value of $P_{success}$ for refinement layer is chosen so that the total number of transmitted packets in both HDA and DLC systems are comparable. Moreover, to decode the refinement layer, it is required that all the packets must be transmitted before a random PU arrival occurs. Through the experiments it has been seen that, with a much lower $P_{success}$, the system ends up with partial reception of the refinement layer packets and thus resulting in a decoded video quality according to base-layer.

The pure analog video coding is inspired by SoftCast scheme [14], [1], where 3D-DCT is performed over the original video frames, followed by non-ordered Hadamard transform and analog QAM. The GOP size is considered same as in the other two investigated schemes. Like our QE analog packets in the HDA system, here, the analog packets are prioritized based on their variances, and higher variance analog

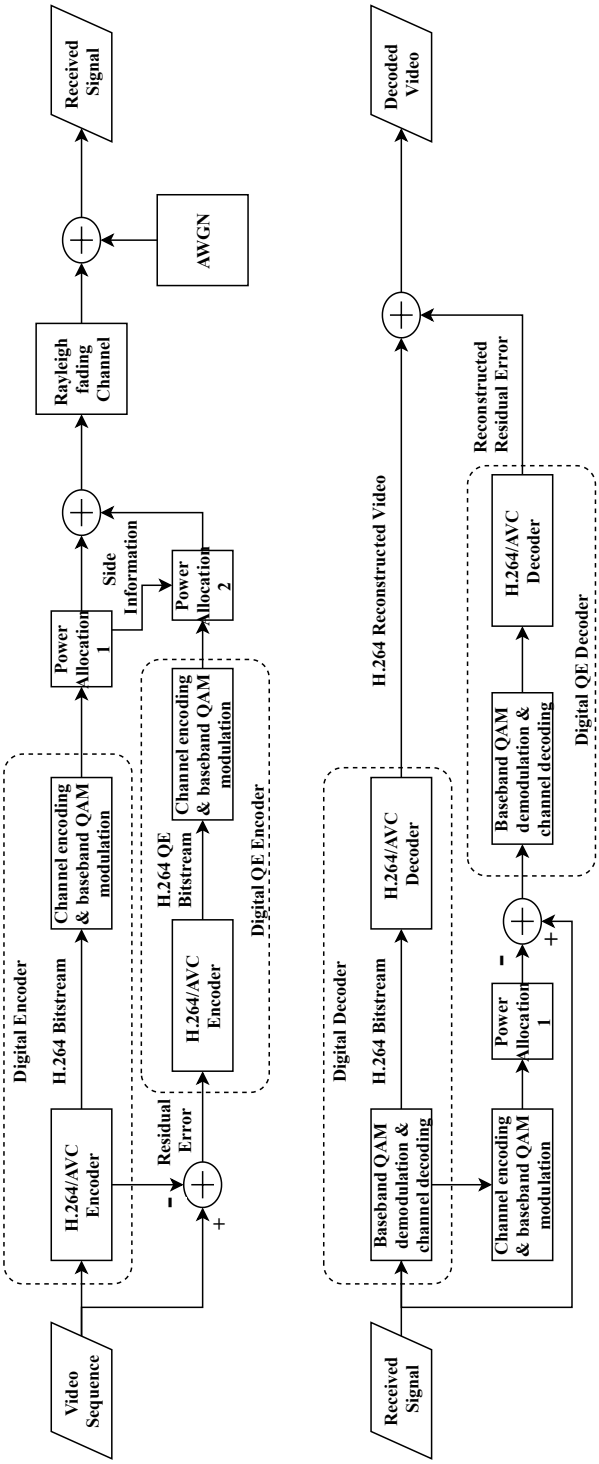


Figure 4.2: The encoder and decoder of DLC system.

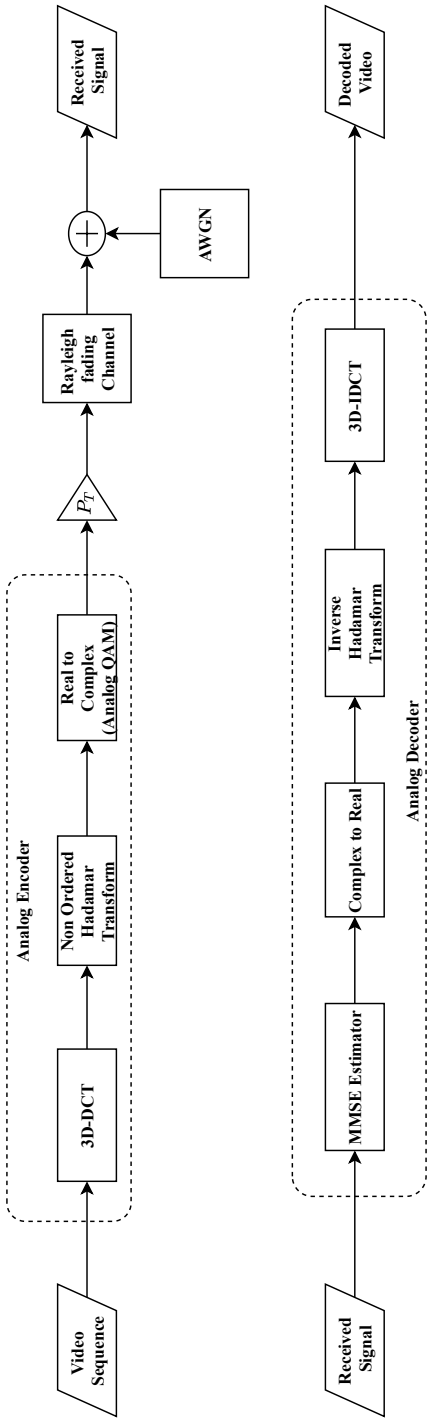


Figure 4.3: The encoder and decoder of pure analog system.

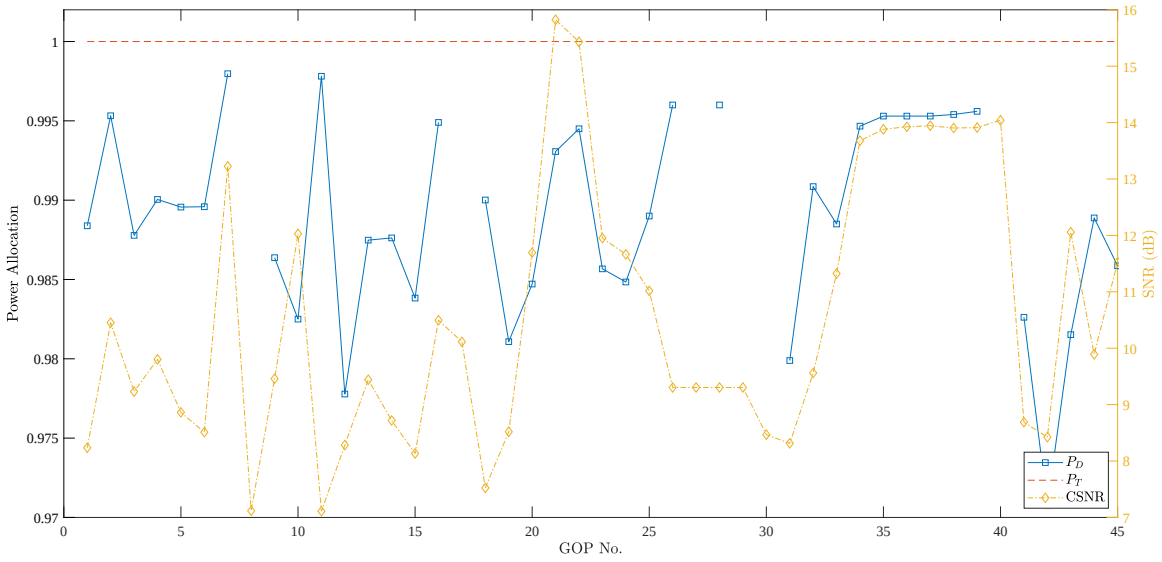
packets are transmitted with the total available transmission power P_T at first. The encoder-decoder pair of pure analog system is illustrated in Figure 4.3.

4.4 Results and analysis

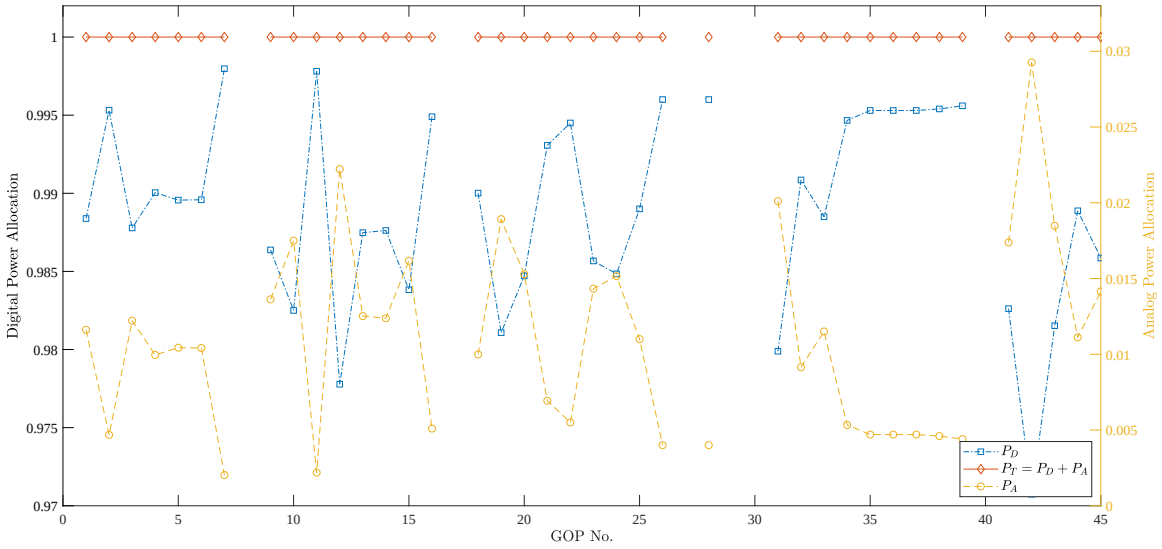
In this thesis, we present a new approach to reliable multimedia transmission as a secondary user over fading cognitive radio channels based on HDA video coding. To evaluate the performance of the HDA system, we investigate single-layer digital coding (base-layer) and analog coding along with DLC from secondary user point of view. In the experiments, all the schemes are tested at multiple different mean CSNR values within the range 7 dB to 25 dB and all of these simulations are conducted in *MATLAB* environment.

4.4.1 Transmission Power Utilization

In wireless communication, one of the scarce resources is transmission power. Therefore, proper utilization of the transmission power is one of the main goals of any wireless transmission system. To understand, how HDA and other investigated systems utilize transmission power, let us refer to Figure 4.4 and Figure 4.5. These graphs are generated by observing the power allocation to each packets as the short-term CSNR fluctuates and finally averaged over the duration of a GOP. In our experiments, we set our total transmission power budget P_T to 1. From Figure 4.4a and Figure 4.5a, it can be seen that with pure (single-layer) digital transmission we are not able to utilize 100% of the available transmission power at all times. This problem arises as the digital transmission is required to follow a fixed set of MCS schemes (see chapter 3).

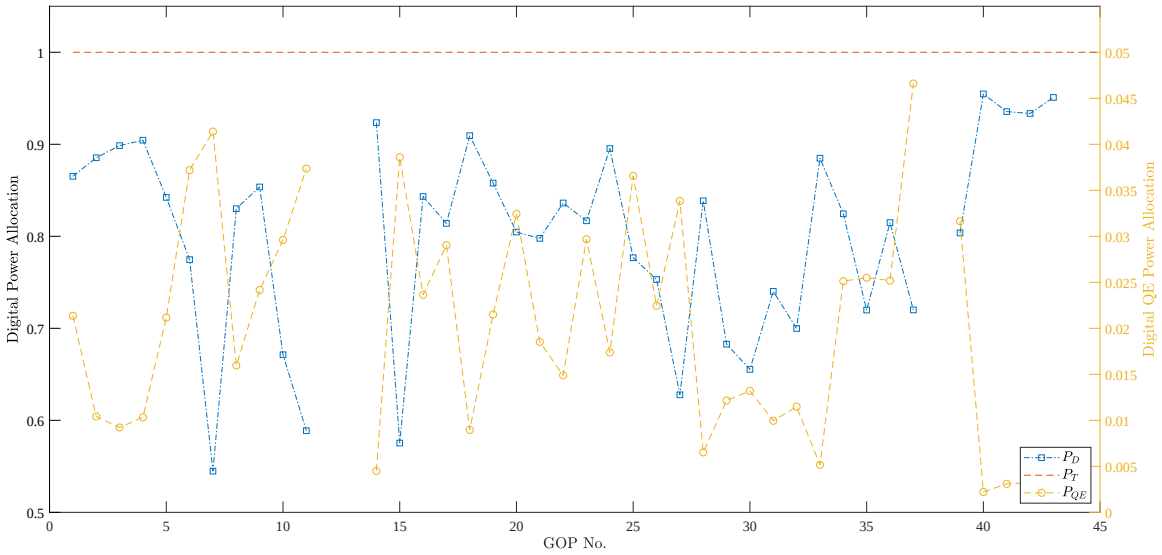


(a)



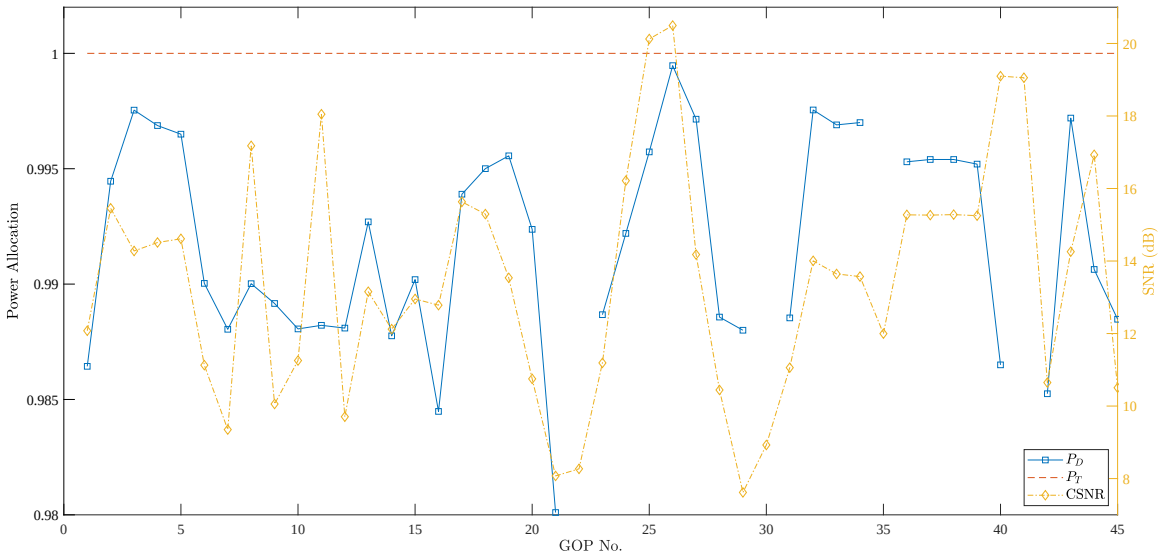
(b)

Figure 4.4: The fraction of total power allocated to (a) digital transmission, (b) HDA transmission, (c) DLC transmission during each GOP of *Football* sequence: mean CSNR = 10 dB. Also shown is the experimentally found mean CSNR during each GOP.



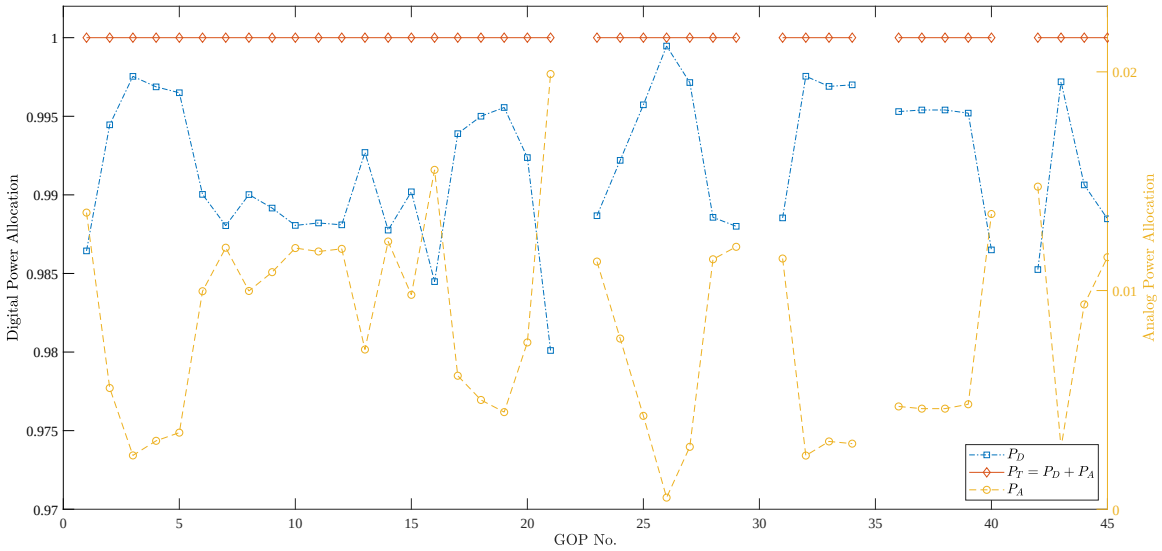
(c)

Figure 4.4: The fraction of total power allocated to (a) digital transmission, (b) HDA transmission, (c) DLC transmission during each GOP of *Football* sequence: mean CSNR = 10 dB. Also shown is the experimentally found mean CSNR during each GOP (cont.).

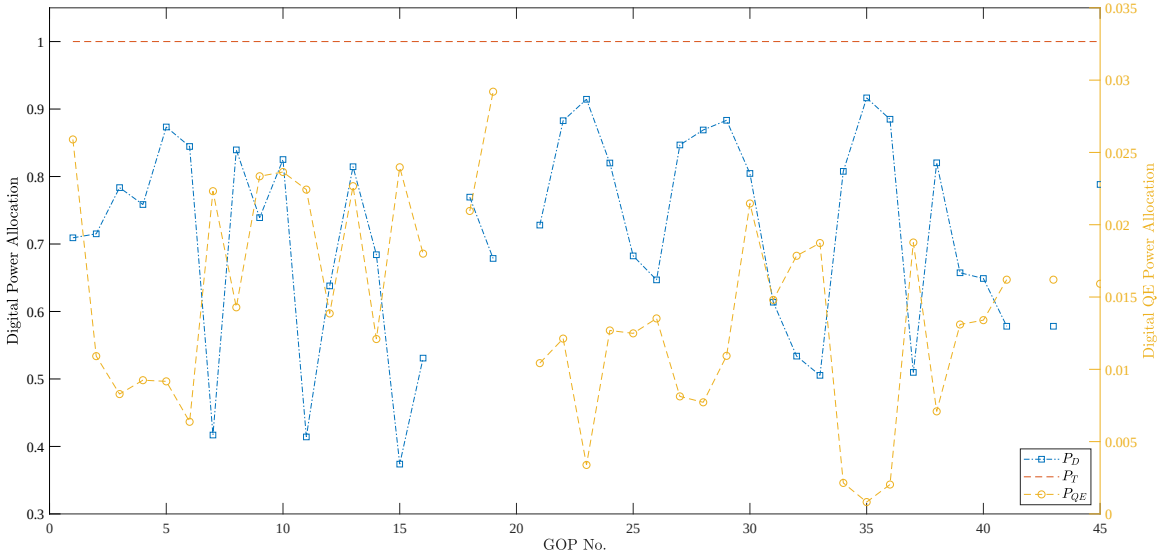


(a)

Figure 4.5: The fraction of total power allocated to (a) digital transmission, (b) HDA transmission, (c) DLC transmission during each GOP of *Football* sequence: mean CSNR = 15 dB. Also shown is the experimentally found mean CSNR during each GOP.



(b)



(c)

Figure 4.5: The fraction of total power allocated to (a) digital transmission, (b) HDA transmission, (c) DLC transmission during each GOP of *Football* sequence: mean CSNR = 15 dB. Also shown is the experimentally found mean CSNR during each GOP (cont.).

It is also noticeable that there exists an inverse relationship between the CSNR and the digital power allocation which can be easily verified from (3.6). When the short-term CSNR is low we require more power be allocated to digital packets so

that packets can be transmitted error free and vice-versa. From Figure 4.4b and Figure 4.5b, it is noticeable that with HDA transmission 100% power utilization is possible as after allocating power to digital packets the remaining power is used for transmitting the analog packets and thus total power budget is used to improve the decoded video quality. However, from Figure 4.5c and Figure 4.4c, it can be seen that DLC is not capable of utilizing 100% of the transmission power budget. This is simply because it is a fully-digital video transmission system with a digital QE layer. As a result, with the remaining power it is not possible to satisfy the threshold SNR for any available MCS set for the digital QE layer and thus it has to be sent after the complete transmission of the digital packets. Here, we skip the plotting of transmission power utilization by pure analog transmission as all the time P_T is used to transmit analog packets and thus 100% power utilization is possible with pure analog transmission.

4.4.2 Quality of the Decoded Videos

The qualities of the decoded video by HDA, DLC, digital, and pure analog systems at the secondary receiver end are compared, both objectively and subjectively. The peak signal-to-noise ratio (PSNR) as well as the structural similarity index (SSIM) [45] are used for objective quality comparisons. Although all of our investigated systems use YCbCr video format for encoding and transmitting video, the PSNR is computed between original and decoded sequences in the RGB color space by taking the average of the three color-components. However, the SSIM is computed in monochrome color-space as this measurement consider only the luminance components of an image [45]. While Figure 4.6-4.8 illustrate the PSNR measurements, Figure 4.9-4.11 show the SSIM measurements of our composite test sequence. The first thing noticeable from

all of these figures is that random GOP drops occur through out the decoded test video sequence at the receiver side. This is because of random PU arrivals on a particular channel during transmission of the baselayer of a GOP. It is also noticeable that while the digital systems including HDA system are more prone to GOP loss, the pure analog system has barely seen such loss. The reason behind this is that, with the reception of certain amount of high variance analog packets before a PU arrival, the pure analog decoder can decode the whole GOP. Since we have random GOP losses, it is hard to compare each of the systems on a GOP basis.

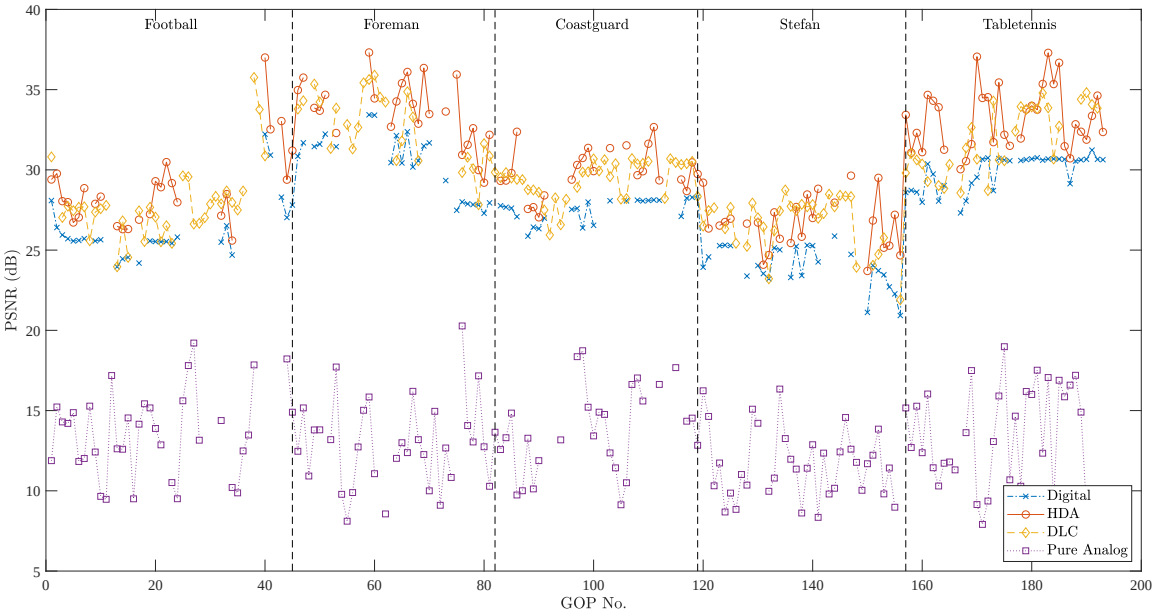


Figure 4.6: Objective quality comparisons in RGB color-space of video transmitted using different investigated methods at 10 dB mean CSNR. PSNR is averaged over frames in a GOP.

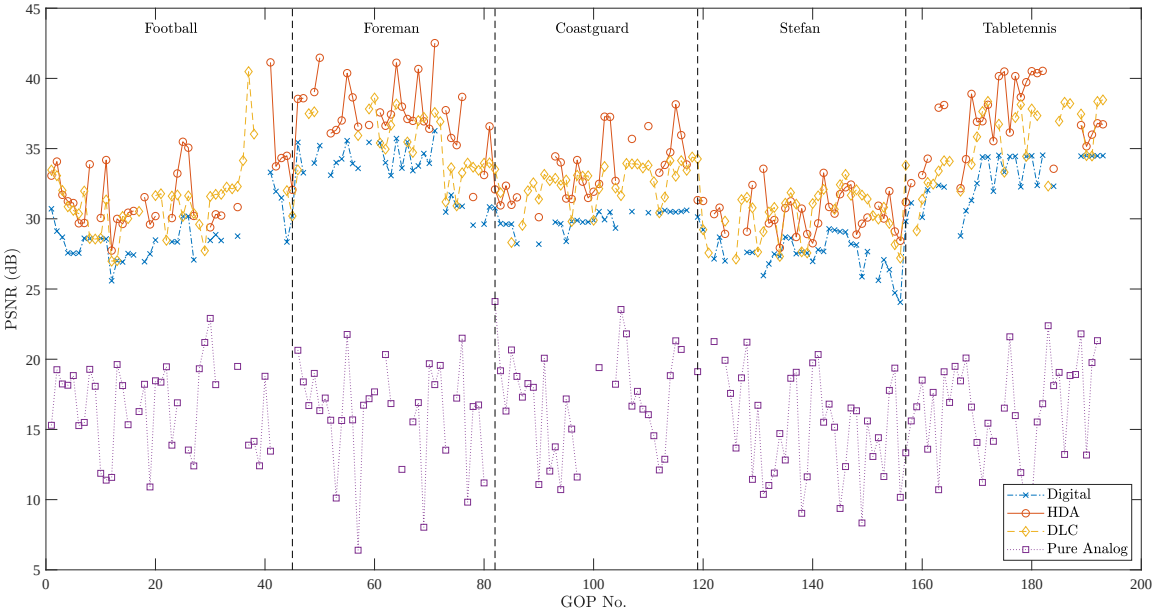


Figure 4.7: Objective quality comparisons in RGB color-space of video transmitted using different investigated methods at 15 dB mean CSNR. PSNR is averaged over frames in a GOP.

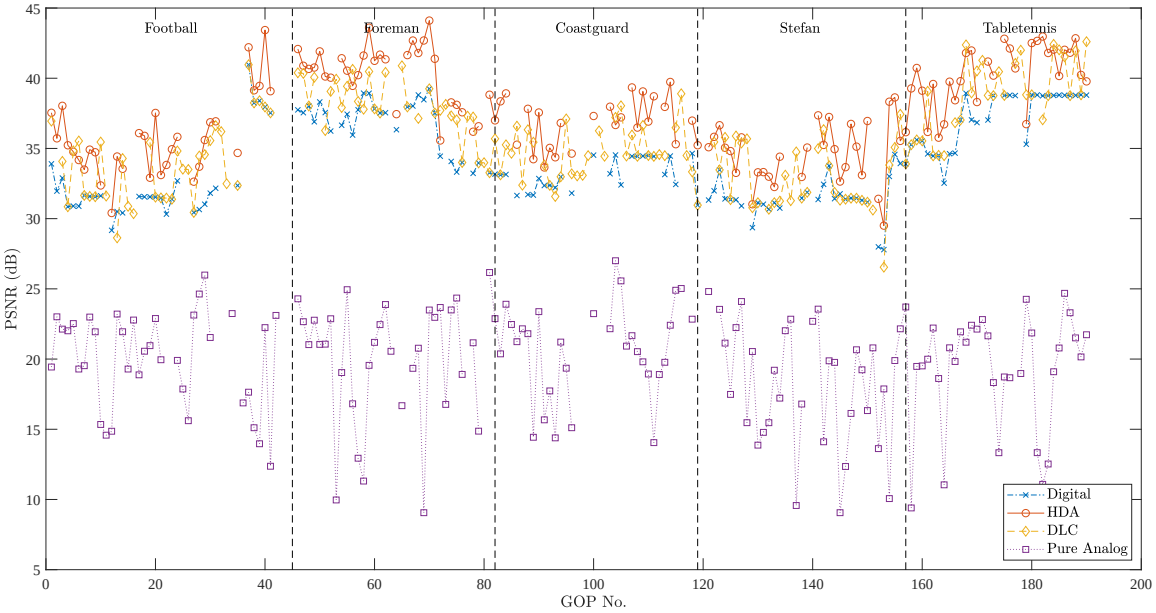


Figure 4.8: Objective quality comparisons in RGB color-space of video transmitted using different investigated methods at 20 dB mean CSNR. PSNR is averaged over frames in a GOP.

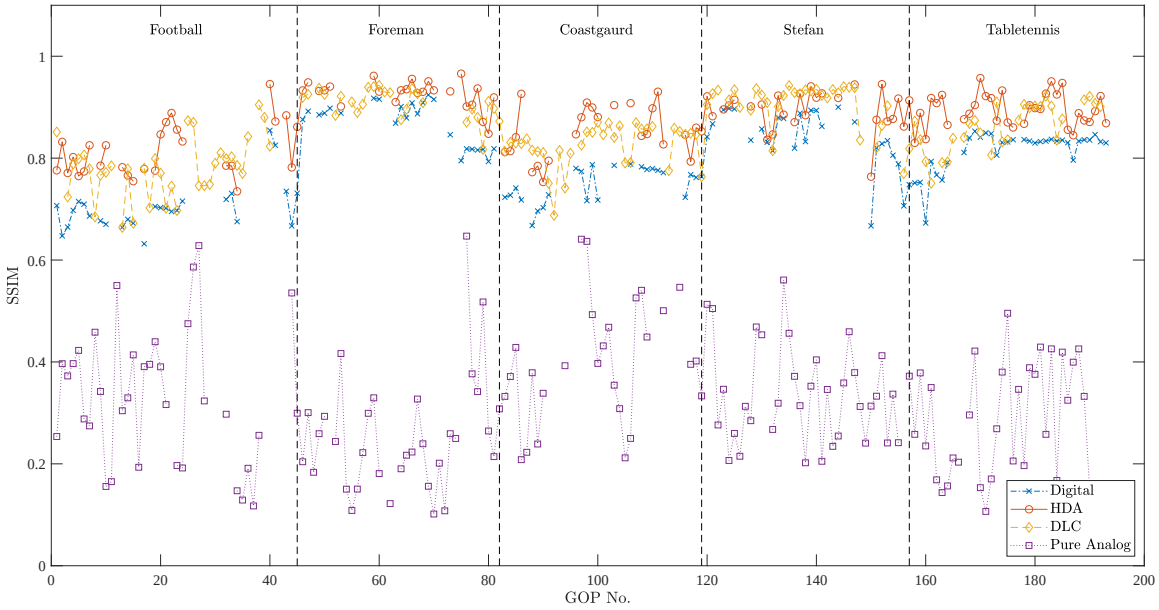


Figure 4.9: Objective quality comparisons in monochrome color-space of video transmitted using different investigated methods at 10 dB mean CSNR. SSIM is averaged over frames in a GOP.

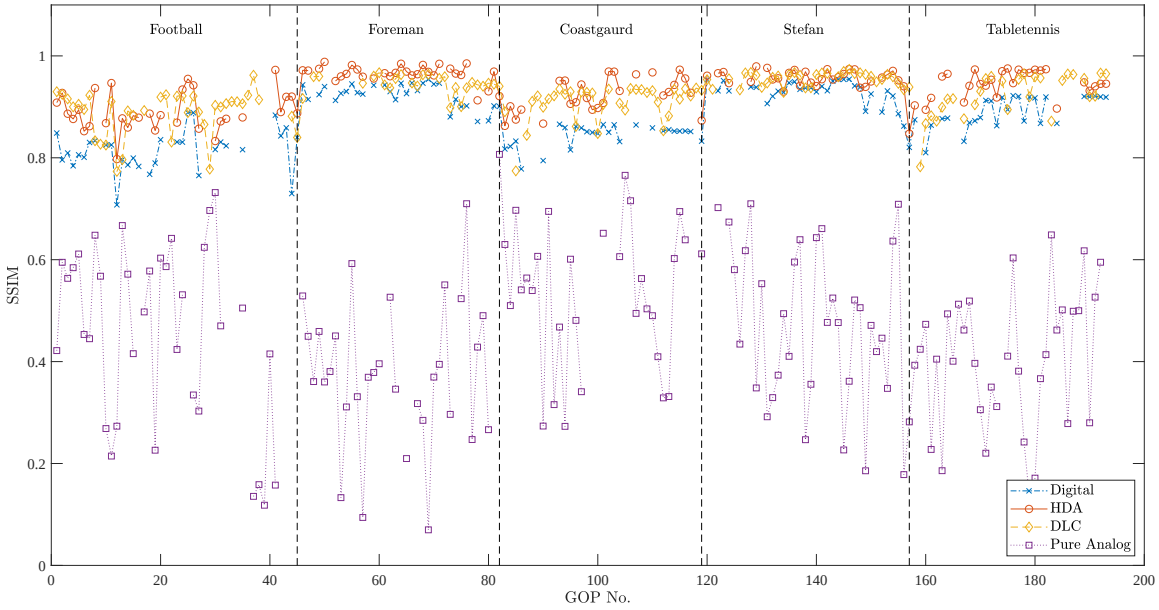


Figure 4.10: Objective quality comparisons in monochrome color-space of video transmitted using different investigated methods at 15 dB mean CSNR. SSIM is averaged over frames in a GOP.

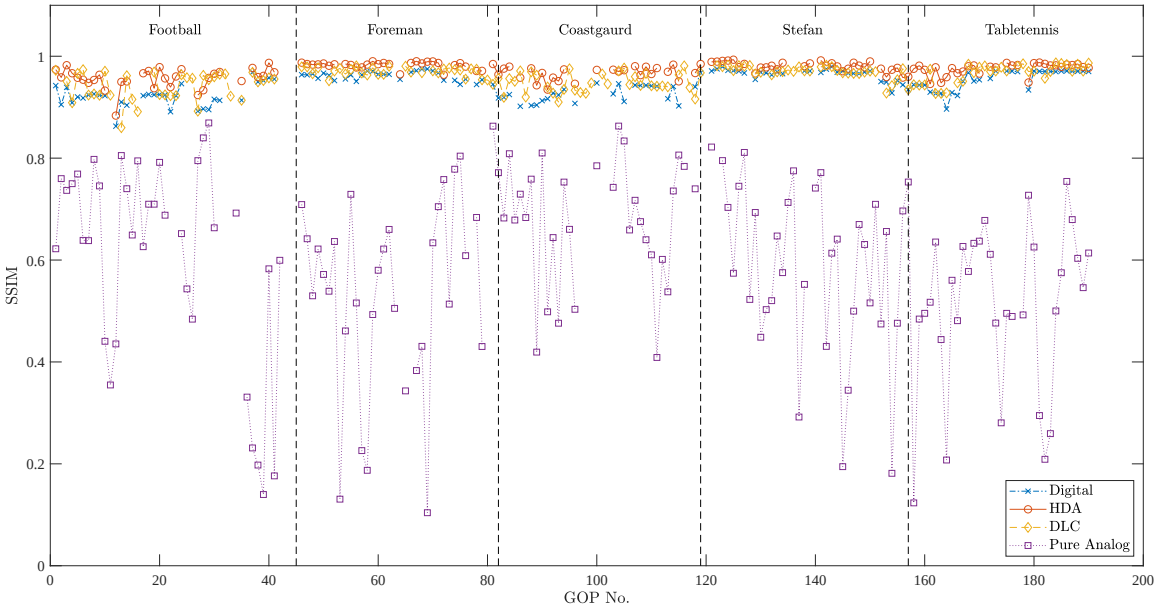


Figure 4.11: Objective quality comparisons in monochrome color-space of video transmitted using different investigated methods at 20 dB mean CSNR. SSIM is averaged over frames in a GOP.

Upon close observation of different video segments, it can be seen that HDA coding outperforms DLC, sometimes by as much as 2-5 dB in PSNR, when the content has intensive motion activity, whereas both systems tend to perform similar during low inter-frame variations. For example, we can consider *Foreman* segment (which mostly shows rapid changes in facial expression and body movements of a person) and *Tabletennis* segment (focuses on the abrupt movements of a ping-pong ball). Moreover, on average 2.55-3.82 dB and 15.39-18.79 dB increment in PSNR of the decoded GOPs have been seen with the HDA system while compared to the single-layer digital system and pure analog system respectively.

The number of GOPs in HDA system that outperforms DLC system in terms of PSNR, increases dramatically with increasing mean CSNR of the channel. This is due to the adaptability of the analog refinement layer of the HDA system. One the other hand, the DLC system lack such adaptability as it uses digital refinement layer.

Since it uses digital refinement layer, with increasing mean CSNR of the channel, the bit-rate of the refinement layer increases significantly, thus it requires more packets to transmit all bits of the refinement layer than what is required at low mean CSNR, which in turns arises the partial reception of the refinement layer. As a result, the quality of the decoded GOP depends only on the baselayer in DLC system especially during high mean CSNR of the channel. From Figure 4.11, it can be seen that the SSIM of both these systems become quite comparable at high mean CSNR of the channel. It is also noticeable from Figure 4.6−4.11 that both HDA and DLC systems outperform single layer digital coding and analog coding. Even single layer digital coding is far superior than pure analog coding. This is quite obvious, as the 3D-DCT alone cannot match the high coding gain achievable with the state-of-the-art H.264/AVC codec.

Now for subjective comparison, let us refer to Figure 4.12 and Figure 4.13. It can be seen that for both 10 dB and 20 dB mean CSNR channels, HDA coding preserves the more visual details of original sequence while comparing with rest of the investigated methods. Also, it is noticeable that at low mean CSNR the single-layer digital system has significant amount of *block-artifacts* which arises due to the low bit-rate encoding with H.264/AVC encoder. On the other hand, the decoded HDA video sequence has barley such artifacts even at low mean CSNR. With DLC the *block-artifacts* is more frequent than HDA especially at low mean CSNR. However, these artifacts visible less in both decoded DLC and single-layer digital sequences when the channel mean CSNR is improved significantly.

(a) Original

(b) HDA: mean CSNR 10 dB

(c) HDA: mean CSNR 20 dB

(d) DLC: mean CSNR 10 dB

(e) DLC: mean CSNR 20 dB

Figure 4.12: Frame 17 of *Football* sequence and its coded version at two different mean CSNR.

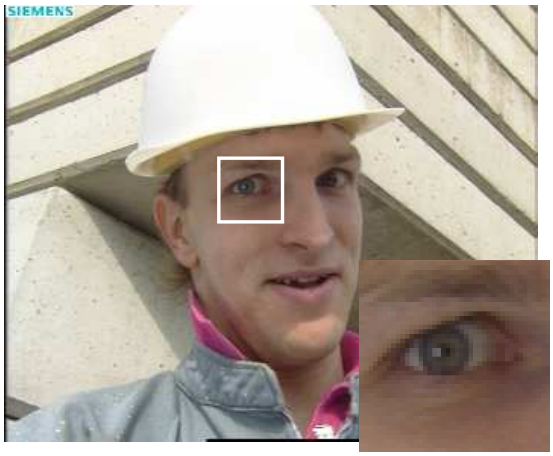
(f) Digital: mean CSNR 10 dB

(g) Digital: mean CSNR 20 dB

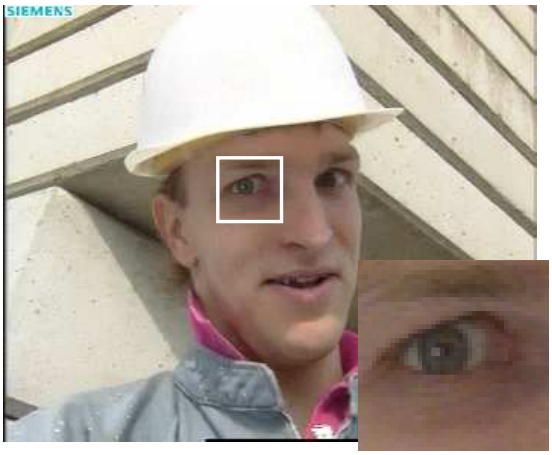
(h) Pure analog: mean CSNR 10 dB

(i) Pure analog: mean CSNR 20 dB

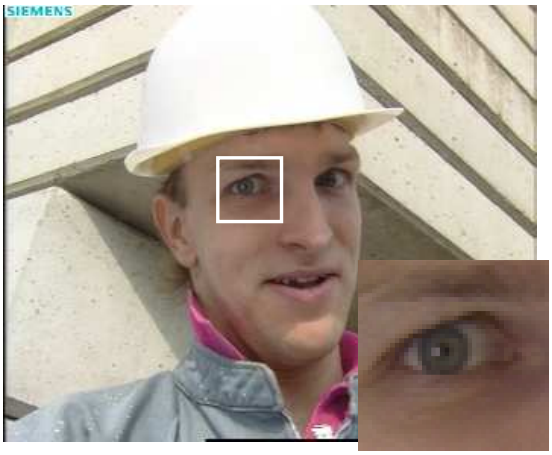
Figure 4.12: Frame 17 of *Football* sequence and its coded version at two different mean CSNR (cont.).



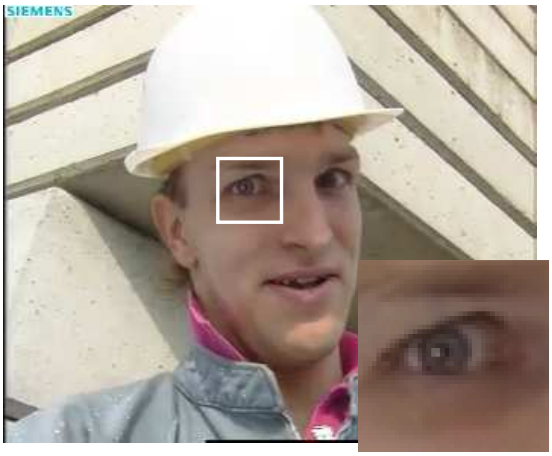
(a) Original



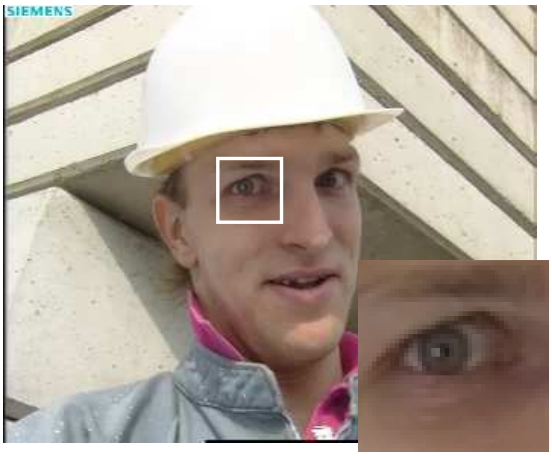
(b) HDA: mean CSNR 10 dB



(c) HDA: mean CSNR 20 dB



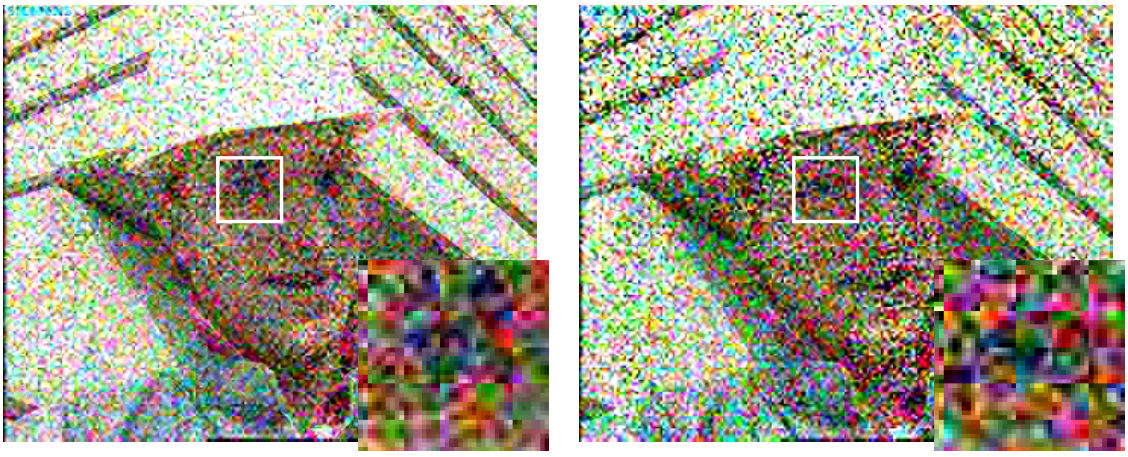
(d) DLC: mean CSNR 10 dB



(e) DLC: mean CSNR 20 dB

Figure 4.13: Frame 112 of *Football* sequence and its coded version at two different mean CSNR.

(f) Digital: mean CSNR 10 dB (g) Digital: mean CSNR 20 dB



(h) Pure analog: mean CSNR 10 dB (i) Pure analog: mean CSNR 20 dB

Figure 4.13: Frame 112 of *Foreman* sequence and its coded version at two different mean CSNR (cont.).

4.4.3 Quality Scalability

Quality scalability measures how gracefully the reconstructed video quality of the receiver varies when there is a variation of the channel quality over time [11]. It reveals the adaptability and robustness of any wireless transmission scheme with channel variations. Quality scalability is high when the reconstructed video quality varies smoothly along with channel variations [11]. Figure 4.14 and Figure 4.15 present the

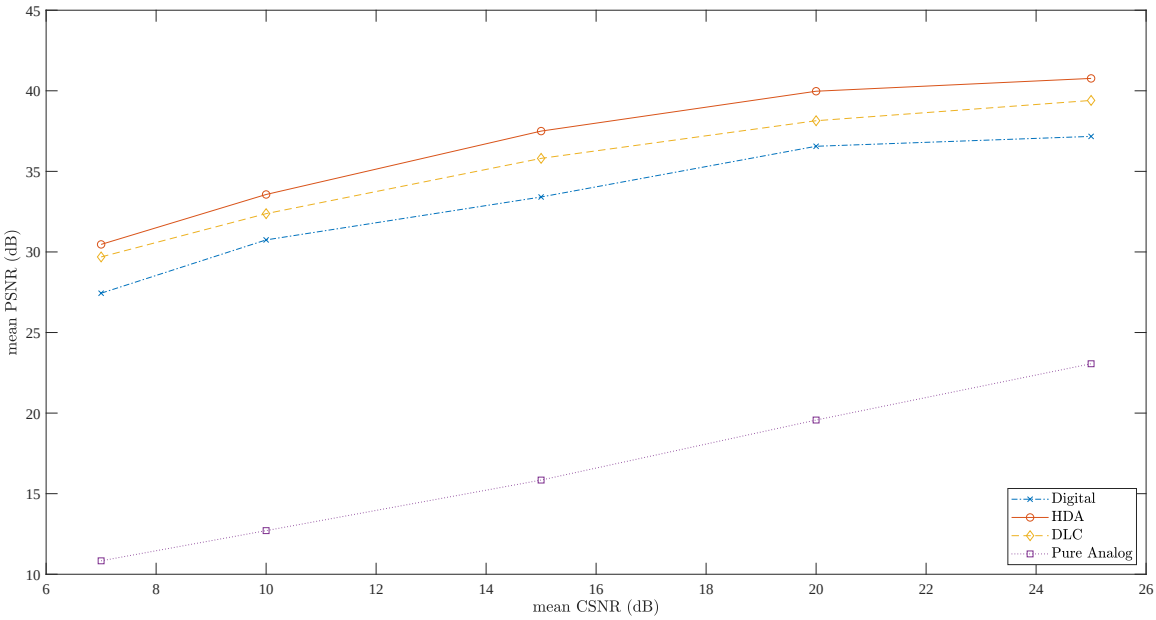


Figure 4.14: Quality scalability comparison in terms of PSNR. Results are obtained using the *Foreman* sequence.

quality scalability comparisons of HDA coding, DLC, digital and pure analog coding.

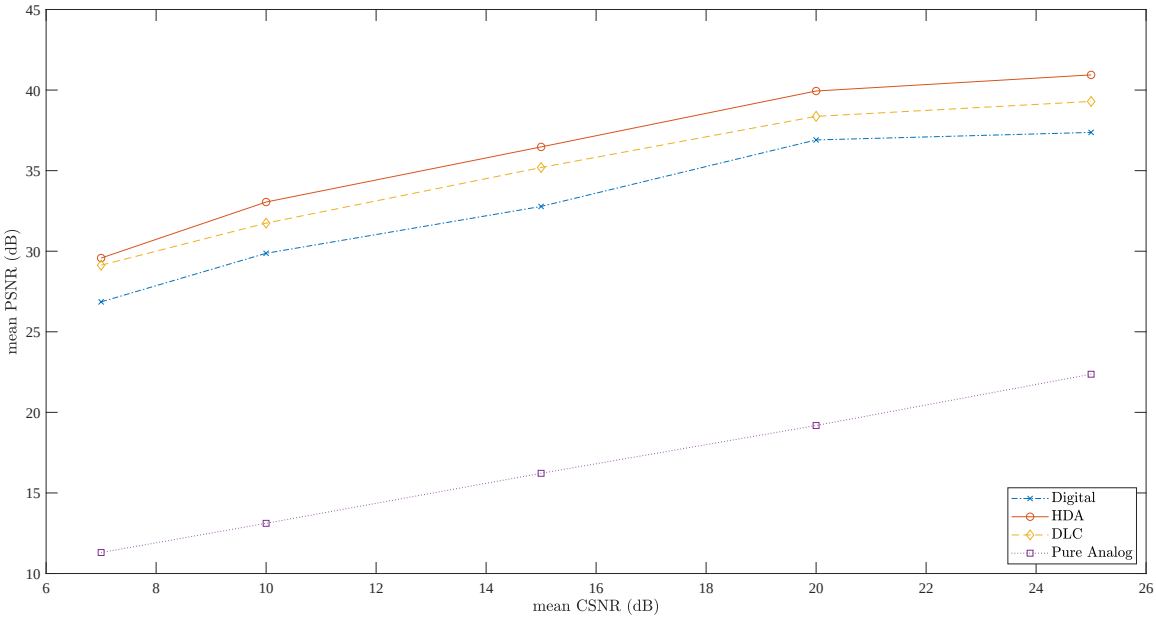


Figure 4.15: Quality scalability comparison in terms of PSNR. Results are obtained using the *Tabletennis* sequence.

Experimental results suggest that HDA coding has the higher quality scalability than the rest of the methods. This is because HDA coding exploits both the advantages of digital coding and analog coding. Upon complete error free reception of the baselayer, the addition of analog refinement layer which adapts itself with channel variations such superior adaptability and robustness is achieved.

4.4.4 Robustness Analysis Under Imperfect CSI

In real-life, estimation of CSI is sometimes erroneous. Therefore, prior to implementing any wireless transmission system in real-life, it is necessary to measure the robustness of that system during imperfect estimation of CSI. The results that are presented in Figure 4.6−4.15 are under the assumption that both the secondary transmitter and the receiver perfectly estimate the CSI. To measure the robustness of the investigated systems, we consider that while the actual mean CSNR of the channels

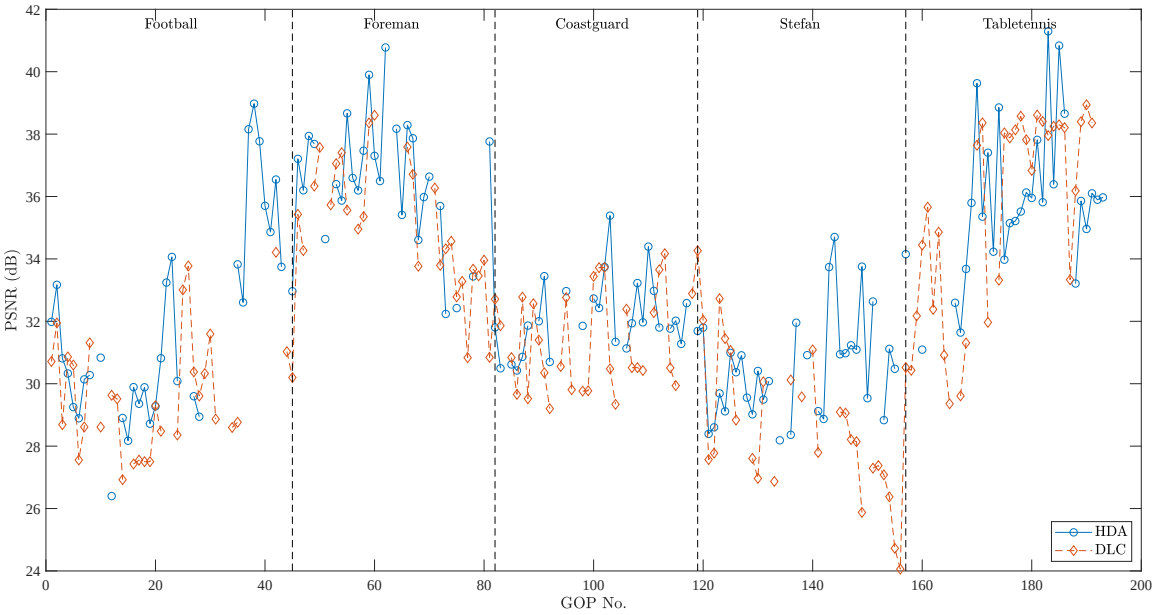


Figure 4.16: Objective quality comparisons in RGB color-space of video transmitted using HDA coding and DLC with imperfect CSI. PSNR is averaged over frames in a GOP.

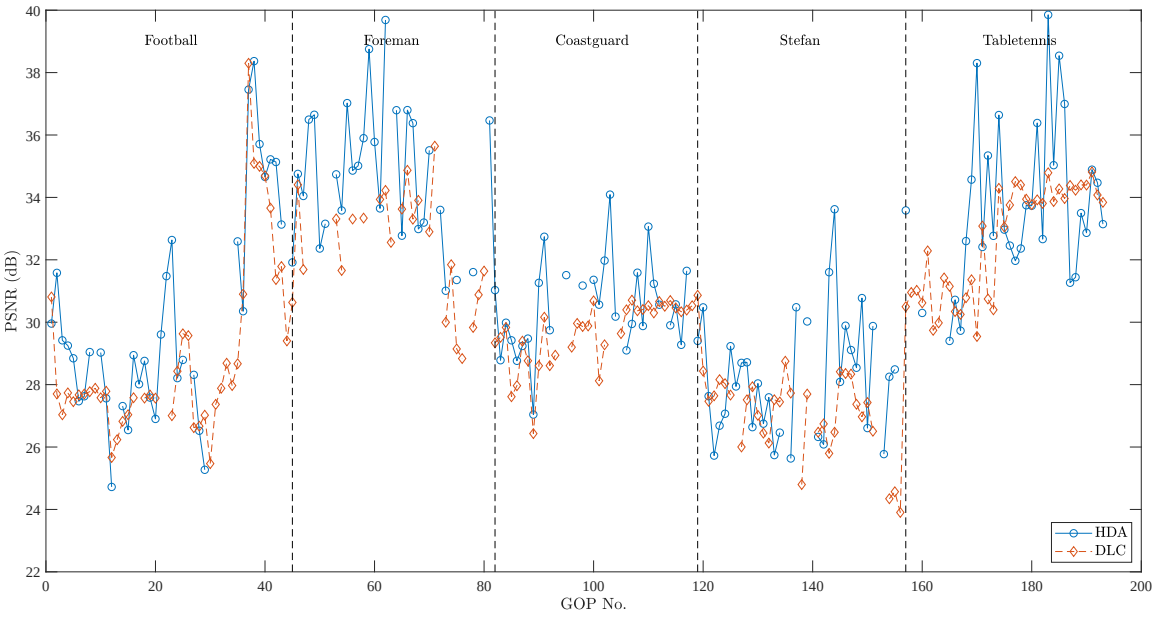


Figure 4.17: Objective quality comparisons in RGB color-space of video transmitted using HDA coding and DLC at 12 dB mean CSNR and perfect CSI. PSNR is averaged over frames in a GOP.

is 12 dB, due to erroneous estimation, the mean CSNR of the channels appear as 15 dB to our systems. Figure 4.16 shows the comparison between HDA coding and DLC in terms of PSNR under imperfect CSI and Figure 4.17 presents the same comparison with perfect CSI. Here, we knowingly omitted the pure analog and single-layer digital systems due to their poor performances in comparision to HDA and DLC systems as presented in previous figures. Experimental results shows that both HDA and DLC transmission systems are robust to imperfect estimation of CSI and like perfect estimation case HDA outperforms DLC many times during transmission of composite test sequence. Upon close observation of the decoded video sequence in imperfect estimation case, it is noticeable that on average up to 3 dB improvement in PSNR is seen while compared to perfect estimation of CSI with HDA coding. On the other hand, it is nearly 1.8 dB with DLC. However, due to imperfect estimation, the packets are transmitted at much low bit-rates than what we expect for channels with

mean CSNR 15 dB. As a result, both HDA and DLC transmissions are disturbed more often due to random PU arrivals than compared to the case with perfect CSI estimation.

Chapter 5

Conclusion and Future Work

In recent times, on the one hand, we have seen an increasing demand for high-quality multimedia services that require large channel bandwidths, and on the other hand, being a scarce resource, the spectrum portion that can support such requirement is almost emptying. Therefore, accessing the underutilized or unused portions of the allotted spectrum of different services opportunistically through cognitive radio technology has started evolving. Keeping in mind that cognitive radio technology is an emerging technology in wireless communications, in this thesis, we have presented a new approach to reliable real-time multimedia transmission as a secondary user over time-variant cognitive radio channels based on HDA coding, which exploits the advantages of both digital and analog coding. Specifically, we have presented an HDA JSCC framework that facilitates the dynamic secondary access, considers the variations of both random channel gains and random PU arrivals over time, prioritizes transmission of high variance QE analog packets, distributes transmission power between digital and analog components, and thus ensures the best possible decoded video quality with high reliability. Experimental results demonstrate that with motion intensive video the HDA system produces a decoded video quality up to 2-5

dB higher PSNR compared to its close competitor, the DLC system. Additionally, the quality of the decoded GOPs in the HDA system outperforms single-layer digital coding by 2.55-3.82 dB on the average in terms of PSNR and pure analog coding by as much as 15.39-18.79 dB.

Visual artifacts in decoded video are more frequent in DLC and single-layer digital systems than the HDA system. Furthermore, the HDA coding preserves better visual details compared to the other methods investigated here and it has been found that the HDA system has a higher quality scalability as well. Moreover, it has been seen that the HDA system is more robust under imperfect estimation of CSI than the DLC system. Experimental results show that unlike digital methods the HDA system, can utilize the 100% of the total available transmitter output power. Thus the total power budget is used to improve the decoded video quality, which makes the HDA system far superior to the DLC and the single-layer digital systems. All of these experimental outcomes suggest that the HDA coding as proposed is a better candidate for reliable multimedia transmission as a secondary user in a cognitive radio system than traditional digital-only coding.

Although the proposed approach shows very promising results, there is room for further improvement. One such improvements is the inclusion of more than one cognitive radio channel for transmitting each GOP. In this way, the streaming of analog packets will divide among multiple channels which will increase the chance of more analog packets being received compared to in a single channel system (the proposed approach guarantees the successful delivery of maximum 4/5 th of the total analog packets if no PU arrival occurs) and thus will enhance the quality of the decoded video further. However, multi-channel transmission will require a more complex channel sensing strategy and a power allocation scheme than what is presented in this thesis.

In this thesis, we assume that the mean CSNR, fading gains, PU arrival-rate and service time are known to both the transmitter and the receiver beforehand as a part of CSI. Furthermore, we assume that these parameters are fixed for each GOP, which is unrealistic. As a result, we cannot properly utilize the available channel capacity. The inclusion of processes for estimating the short-term values of these parameters will help utilize the full channel capacity and thus will contribute to further improvement of the decoded video quality.

Bibliography

[1] S. Jakubczak and D. Katabi, "Softcast: one-size-fits-all wireless video," in *Proceedings of the ACM SIGCOMM 2010 conference*, 2010, pp. 449–450.

[2] I. E. Richardson, *The H. 264 Advanced Video Compression Standard*. John Wiley & Sons, 2011.

[3] C. Clarke, "Reed-Solomon error correction. white paper WHP 031," *British Broadcasting Corporation Research and Development*, 2002.

[4] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[5] D. Čabrić, S. M. Mishra, D. Willkomm, R. Brodersen, and A. Wolisz, "A cognitive radio approach for usage of virtual unlicensed spectrum," in *14th IST mobile and wireless communications summit*, June 2005.

[6] P. Kolodzy, "Spectrum policy task force: Findings and recommendations," *International Symposium on Advanced Radio Technologies (ISART)*, vol. 5, no. 12, pp. 459–465, 2003.

[7] Federal Communications Commission (FCC), "Spectrum policy task force: Report," Nov. 15, 2002, ET Docket 02-135.

[8] H. Kushwaha, Y. Xing, R. Chandramouli, and H. Heffes, "Reliable multimedia transmission over cognitive radio networks using fountain codes," *Proceedings of the IEEE*, vol. 96, no. 1, pp. 155–165, 2007.

[9] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[10] A. Kedia, "Adaptive joint source-channel coding of real-time multimedia for cognitive radio," Master's thesis, University of Manitoba, 2014.

[11] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 331–345, 2013.

[12] ——, "Hybrid digital-analog scheme for video transmission over wireless," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*. IEEE, 2013, pp. 1163–1166.

[13] P. Yahampath, "Digital-analog superposition coding for OFDM channels with application to video transmission," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1802–1806.

[14] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proceedings of the 17th annual international conference on Mobile computing and networking*, 2011, pp. 289–300.

[15] C. Hellge, S. Mirta, T. Schierl, and T. Wiegand, "Mobile TV with SVC and hierarchical modulation for DVB-H broadcast services," in *2009 IEEE International*

Symposium on Broadband Multimedia Systems and Broadcasting. IEEE, 2009, pp. 1–5.

[16] M. M. Ghandi and M. Ghanbari, "Layered H. 264 video transmission with hierarchical QAM," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 451–466, 2006.

[17] B. Tan, H. Cui, J. Wu, and C. W. Chen, "An optimal resource allocation for superposition coding-based hybrid digital–analog system," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 945–956, 2017.

[18] Y. Wang, F. Alajaji, and T. Linder, "Design of vq-based hybrid digital-analog joint source-channel codes for image communication," in *Data Compression Conference*. IEEE, 2005, pp. 193–202.

[19] N. Fan, Y. Liu, and L. Zhang, "Hybrid digital-analog video transmission based on H. 264/AVC and ParCast in MIMO-OFDM WLANs," in *2014 21st International Conference on Telecommunications (ICT)*. IEEE, 2014, pp. 395–399.

[20] J. Shen, L. Yu, and H. Li, "Hybrid digital-analog scheme for video transmission over fading channel," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2016, pp. 1582–1585.

[21] C. Ye, G. Ozcan, M. C. Gursoy, and S. Velipasalar, "Multimedia transmission over cognitive radio channels under sensing uncertainty," *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 726–741, 2015.

[22] R. Yao, Y. Liu, J. Liu, P. Zhao, and S. Ci, "Utility-based H. 264/SVC video streaming over multi-channel cognitive radio networks," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 434–449, 2015.

[23] J.-P. Wagner, J. Chakareski, and P. Frossard, "Streaming of scalable video from multiple servers using rateless codes," in *2006 IEEE international Conference on Multimedia and Expo*. IEEE, 2006, pp. 1501–1504.

[24] V. K. Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 203–233, 2001.

[25] L. Pham, "Joint source-channel coding for image transmission over underlay multichannel cognitive radio networks," Master's thesis, Rochester Institute of Technology, 2019.

[26] C. Guo, T. Peng, Y. Qi, and W. Wang, "Adaptive channel searching scheme for cooperative spectrum sensing in cognitive radio networks," in *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009, pp. 1–6.

[27] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, 2008.

[28] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[29] International Organization for Standardization/International Electrotechnical Commission and others, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbit/s- part 2: Video," *ISO/IEC 11172*, 1993.

[30] F. Halsall, *Multimedia Communications: Applications, Networks, Protocols, and Standards*. Pearson, 2001.

[31] K. Sayood, *Introduction to Data Compression, (Morgan Kaufmann Series in Multimedia Information and Systems)*, 3rd ed. Elsevier, 2005.

[32] S. K. Jakubczak, "Softcast: exposing a waveform interface to the wireless channel for scalable video broadcast," Ph.D. dissertation, Massachusetts Institute of Technology, 2011.

[33] M. Riley and I. Richardson. (1998) An introduction to Reed-Solomon codes: principles, architecture and implementation. [Online]. Available: http://www.cs.cmu.edu/~guyb/realworld/reedsolomon/reed_solomon_codes.html

[34] F. Petré, "Unraveling the BFWA propagation environment," in *Broadband Fixed Wireless Access*. Springer, 2006, pp. 29–59.

[35] B. Sklar, "Rayleigh fading channels in mobile digital communication systems .I. Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.

[36] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of communication systems: modeling, methodology and techniques*. Springer Science & Business Media, 2006.

[37] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-hill New York, 2001.

[38] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-hill New York, 2000.

[39] J. Sztrik, *Basic Queueing Theory: Foundations of System Performance Modeling*. GlobeEdit, 2016.

[40] "Supplement to IEEE standard for information technology, part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High speed physical layer in the 5GHz band," *IEEE Std 802.11 a-1999*, 1999.

[41] P. Yahampath, "Video coding for OFDM systems with imperfect CSI: A hybrid digital-analog approach," *Signal Processing: Image Communication*, vol. 87, 2020, to appear.

[42] European Telecommunications Standard Institute, *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television*, ETSI EN 300 744 V1.6.1 (2009-01). 2009.

[43] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer Science & Business Media, 2012, vol. 159.

[44] 3GPP TSG-SA Codec Working Group, "Error resilience in real-time packet multimedia payloads," July 1999, Rep. TSGR1#6(99)804. [Online]. Available: ftp://www.3gpp.org/tsg_ran/WG1_RL1/TSGR1_06/Docs/Pdfs/r1-99804.pdf

[45] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.