Soft Video Delivery by Metadata Reduction over LTE

By

Tino Monroe

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
December 14, 2019

Keywords: metadata reduction, software-defined radio, mobile, cellular, 4G, LTE, USRP

Approved By

Shiwen Mao, Samuel Ginn Endowed Professor of Electrical and Computer Engineering
Thaddeus Roppel, Associate Professor of Electrical and Computer Engineering
R. Mark Nelms, Professor of Electrical and Computer Engineering

Abstract

New methods for media delivery, specifically video, have become increasingly important over the past decade. Along with increased user need for video streaming, video is now a majority of the traffic on the internet. In addition, traditionally the conventional design of the end device selecting the appropriate bitrate according to channel quality does not allow graceful degradation of the quality of the received video in a dynamic network environment. In this thesis, a new method is presented through implementing a metadata reduction approach for soft-video delivering over a Long Term Evolution (LTE) network. The goal of this work is to demonstrate that we can obtain superior video transmission over a mobile wireless channel. The design for this system comes in two parts: the soft-video metadata reduction algorithm and the LTE system implemented using National Instruments' Universal Software Radio Peripheral (USRP) software-defined radio. The experimental data collected validates that the proposed video delivery system outperforms the conventional approach.

Acknowledgments

Firstly, I am very grateful to my advisor Dr. Shiwen Mao for both his guidance and profound knowledge. Most of my inspiration in pursuing this research area was due to his passion for the topic. I also appreciate my other committee members, Dr. Thaddeus Roppel and Dr. Robert Nelms. And thank you to the faculty and staff of both the electrical engineering and computer science departments. All of you were instrumental in my journey as I pursued my graduate education here at Auburn University.

In addition, I would like to thank my research members, particularly, Ticao Zhang, who greatly assisted me with this research topic. I would not have been able to effectively pursue this research project without his help, knowledge, and passion for the topic, which kept me motivated.

Lastly, I want to thank my family members and friends, who have supported me in many ways during my time in Auburn.

This work was supported in part by the NSF under Grants IIP-1822055 and CNS-1702957, the GPS and Vehicle Dynamics Laboratory (GAVLAB), and Wireless Engineering Research and Education Center (WEREC) at Auburn University.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the wireless environment, video delivery has been one of the most important applications to date [25, 26, 27, 28, 29]. Based on a survey done by the network giant Cisco, 75% the entire world's mobile traffic will be comprised of video content [2]. Conventionally, the digital compression and transmission of video are performed separately. For example, the video compression part uses H.264/Advanced Video Coding (AVC) or H.265/High-Efficiency Video Coding (HEVC) codec standards to generate bit stream through quantization and entropy coding [8, 29]. Both a digital modulation and coding techniques are used to transmit the bit stream. And with this conventional source-channel coding separation, problems arise. As per Shannon's separation theorem, source coding and channel coding are used to reduce the redundancy and deal with transmission losses between the sender-receiver pair, respectively [15]. When the channel signal-to-noise ratio (SNR) is low, the quality of the video degrades significantly as a result of bit errors. For instance, if the source transmits a video stream to a single receiver and the quality of the channel is known or can be quantified at the source, then the source will select the best bit rate, forward error control and modulation scheme for the video bit stream [7]. After the bit rate is determined, a video codec will compress and send that video with the selected bit rate. This design has been proven to be very beneficial in many cases for video transmissions. However, if the channel quality becomes worse, the video stream with the determined rate will suffer high bit errors and a big drop in the received video quality. On the other hand, if the channel quality gets improved during the transmission, the video quality does not benefit from the improved channel, since the bit rate is fixed according to the previous channel quality. In the realm of communications, this phenomenon is known as the *cliff effect* [15].

Several recent video delivery schemes have shown promise over the tradition source-channel separation-based ones. The first is SoftCast, a compression architecture for wireless video transmission [3]. In contrast to the separate conventional design, SoftCast is a design that encodes and protects the data from bit errors. First, it starts with a number sequence that represents the video and those same numbers being pixel luminance [5]. A sequence of transformations are done to acquire those signals, which are then transmitted. The main property of this method is that it consists of linear operations. In this way, the transmitted samples are done linearly with respect to the original pixel values. Therefore, an increase in distortion of the transmitted bits is proportional to the priority of the bits on that video sequence. Each end device decodes the quality of the video dependent on the channel quality. SoftCast's end-to-end architecture incorporates the following linear components: Compression, Error Protection, and Resilience to Packet Loss [12].

The second reduction method builds upon the previous two and it will be the target method for this paper [11]. In this paper, a random or "blind" data detection algorithm is proposed. Initially, a 3-dimensional discrete cosine transform (3D-DCT) is applied to the video luminance values to decorrelate the signal. The optimal scaling of the DCT coefficients and proportional power scaling factors are encoded as metadata [3]. What makes this process different from prior work is that, to eliminate improper scaling, the transmitter scales each DCT coefficient optimally. On the side of the receiver, those received signals are squared once recovered. Therefore, once the need for prior information regarding the power scaling factors are eliminated, the system becomes nearly metadata-free, i.e., the total amount of transmitted data is greatly reduced as compared to previous work [3,5].

In [11], the metadata-free scheme is proposed and studied under an analog additive white Gaussian noise (AWGN) channel. In [24], the resource allocation problem associated with the

proposed scheme is investigated. However, both [11] and [24] evaluate the proposed schemes with simulations. In this thesis, we investigate SoftCast wireless video transmission over a wireless channel. In particular, an LTE cellular platform is implemented for SoftCast video transmission experiments. Software-Defined Radio (SDR) is a popular concept for implementing radio equipment in software, using low-cost general-purpose computers and radio frontends. In recent years, it is gaining popularity as a tool to build close-to-reality testbeds for experimental research [4]. And in this instance, since there is an availability of application code for long term evolution (LTE) systems, the testbed for this application can be developed and scale for the appropriated metrics. The SDR platform of choice is the srsLTE open-source platform. This development will be completely over-the-air (OTA) as a realistic wireless environment. The platform currently provides a standard-compliant implementation of a subset of Release 10 LTE, including key elements of the network such as user equipment (UE), evolved node B (eNB), mobility management entity (MME), home subscriber server (HSS), serving gateway (SGW), and packet data network gateway (PGW) on a standard Linux-based computing equipment (Intel x86 PC architectures). The software can be used in conjunction with standard RF laboratory equipment. The RF equipment of choice is National Instruments' (NI) B210 Universal Software Radio Peripheral (USRP), which provides support for 3GPP LTE base station and UE functionalities. The open source software has the functionality allowing various modulation coding schemes (MCS) to be used with its coding architecture. Both QAM (16 and 64) and QPSK are modulation techniques used in IEEE 802.11 and used for the testing environment. The SoftCast videos are transmitted over such an LTE network and the received video quality is evaluated.

1.1 Approach

The motivation for pursing this experimentation project is the need for a most robust video compression algorithm that can keep up with the video delivery demands. In addition, since most of the traffic will be comprised of mobile video, seeing the effects on a real cellular data network would provide useful insights on the performance of the proposed schemes. Since there have been considerable previous research done in the area of video and media compression, it would also be great to conduct a comparison of those schemes regarding performance.

First, a basis for which the experiment needs to be done will be presented, comparing conventional methods to the proposed and making a case for each will be done. The two video delivery methods are both developed to improve the efficiency of the video delivery platform. An analysis of the algorithms will be done over a physical platform. Before doing so, the data retrieved from the calculations will be compared. The performance metrics used in our experimental evaluation include the peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM), and the bit size of the video transmission over the platform.

This will be done over a channel bandwidth of 20MHz with 100 resource blocks. The subcarriers are 1201 for the downlink and 1200 for the uplink. The RF frontend will include two USRPs. Both RF frontends will communicate wirelessly over VERT2450 Dual Band omnidirectional antennas. Two laptops, i.e., HP Zbooks with Intel based i5 processor and USB 3 capable, are used, both on the Linux Ubuntu operating system. These are for emulating the base station (i.e. cell tower) and UE, respectively. As the experimental results will show, the proposed algorithm will dramatically reduce the metadata overhead thus in turn increasing transmission efficiency of the soft videos.

1.2 Layout

The remainder of this thesis is organized as follows. Chapter 2 dives into the background and related works to the topic. In Chapter 3, a technical challenge analysis is done for both the cellular network and video compression. The video delivery system design and implementation will be discussed in Chapter 4. Our experiments and experimental results will be presented in Chapter 5. Chapter 6 concludes this thesis with a discussion of future work.

Chapter 2

Background

To lay the groundwork for the proposed methodology for soft video delivery, first addressed is an understanding of a few key concepts in video compression for wireless communications. The Shannon-Hartley theorem introduced the notion of channel capacity, i.e, information can be transmitted reliably over a communications channel at a certain maximum rate [5]. As a byproduct of this theorem, the communication process can be split between a source coder and a channel coder, statistics of the random source are dealt with at the source coder for compression and the channel coder deals with the statistics of the random channel for reliability [24]. As illustrated in Figure 2.1, at a fixed rate (i.e., bits per second), a data stream exists uniformly between the source and channel coders. By convention, all components in this link are dependent on each other. The goal thereafter is to find the possible balance, breaking the archaic dependence and deploy a new platform. However, over time, this traditional process will soon show to be inefficient with the advent of new technologies and increased data demands.

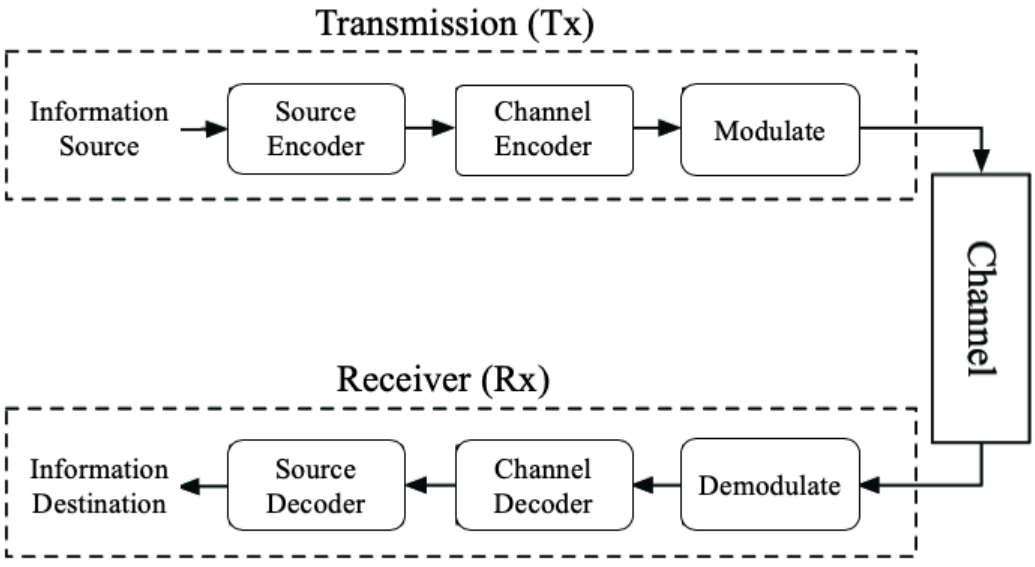


Figure 2.1: Block diagram of a traditional source-channel design.

2.1 Source Coding

In data communications, source coding refers to a reduction of the resources needed to send the source data with either lossless compression or introducing certain distortion as a price to pay [12]. The theory is that the encoder's requirement for decoding the bit stream are normally met and the information loss is normally due to the encoder. However, in the field, many compression algorithms implicitly introduce compression inefficiency by implementing error detection and concealment mechanisms. Though these mechanisms are required for a high-quality transmission, there needs to be a leveraging of the resources to increase the efficiency.

2.1.1 Color Transform and Subsampling

The ubiquitous way to delineate color in computer systems is the RGB model, based on human vision's trichromatism. Therefore, the data that is sent has the three colors: red, blue, and green. Due to the sensitive nature of the human eye to luminance over color, the encoding channels are not done independently. In video compression algorithms, color transforms are done first, projecting 3-D pixels onto a different color space [5].

In the acronym YUV, Y represents the luminance of the color spaces in that dimensions and UV represent the dimensions of encoded color [5]. Since luminance is less intensive on the human eyes than color variation, at a rate of two, the chroma channel's data are subsampled. Take for instance, the classic 4:2:0 layout represents a 16 X 16 block of pixels by a 16 X 16 Y block, 8 X 8 U block and an 8 X 8 V block. So, at this rate, subsampling half the number values are needed. The videos used for this experiment are in the .y4m (Y4M) or .yuv (YUV) format to conform to the classic layout presented. Both YUV and Y4M are lossless formats for storing raw video sequences. The main difference between the two lossless formats is that Y4M actually contains

information about the resolution of the picture, scan type, frame rate and the color space, while the YUV does not [5].

2.1.2 Predictive Coding

Independent coding on a 16 X 16 block is not done, even though they are on macroblocks. The block to be encoded becomes possible based on the prediction of the macroblocks from the previous model. Thereafter, the difference of the original and predicted is encoded. As in Figure 2.2, a reference data must always be available at the decoder for reconstruction, in order for data prediction to happen. Therefore, the pixel values are stored at the frame buffer so the decoder can have them ready. Notice that only until the reconstruction of the reference data, the decoder can decode the predicted pixels. This issue brings to light the limits of bit stream data ordering. Intra-frame predictive coding uses reference pixels within the frame. Directions of prediction as defined for the H.264 system, includes up to nine modes (horizontal, vertical, diagonal, etc.) in horizontal prediction, predicted pixel values match those direct left neighbors [16].

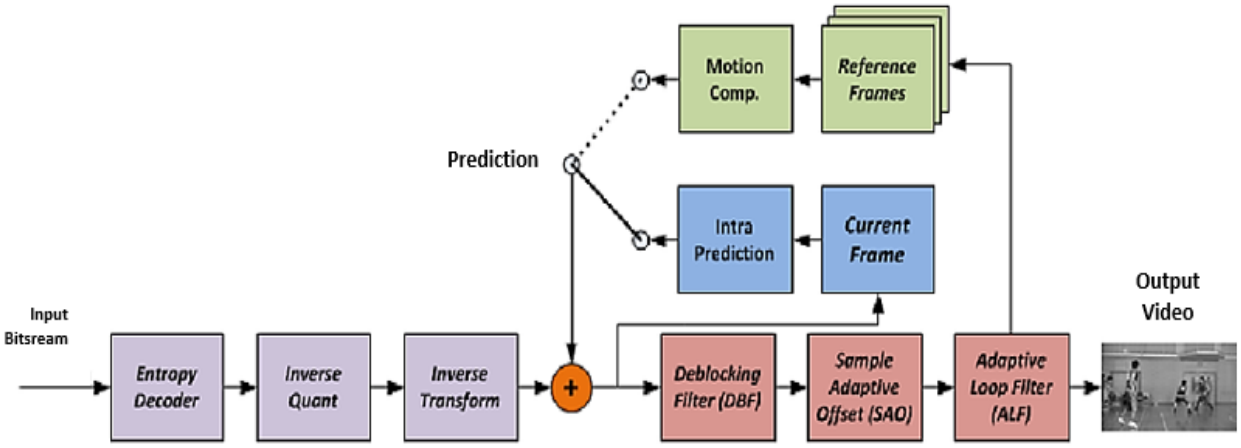


Figure 2.2: Block diagram of the HEVC decoder (Figure courtesy of [8]).

The use of reference pixels pulled from surrounding frames is the main method of inter-frame coding. Even though reference pixels are used to capture similarities of an object or scene, offset vectors are ultimately used, not the exact spatial coordinates [8]. Next, the motion vector

comes from the best fit reference in order for prediction of the macroblock. To increate accuracy and efficiency, multiple frames from the ordered group can be used. All references weighted in a linear combination. This is crucial due to the nature of cross-fading and encoding fading.

Take an instance where each frame corresponds with a specific reference, such as MPEG-2. There are three frame types used: I-frames, P-frames, and B-frames. I-frames are the least compressible but do not require other video frames to decode, so they are not predicted at all. P-frames can use data from the previous I- or P-frame frames to decompress. And finally, B-frames can use both previous and forward I- or P-frames frames for data reference to get the highest amount of data compression [8]. In H.264, distinction between the frame types can be made per macroblock, 2-16 weighted reference frames can be used, B-frames can be used for prediction, and estimated motion is done independently for parts of the macroblock [3].

2.1.3 Adaptive Quantization

Using a divide and truncation function, the DCT coefficients become quantized. Approximation is proportional to the divisor and in turn effects how coarse it becomes. Quantized high spatial frequencies quantized normally preform coarsely than those at lower frequencies and the average value of the whole frame is quantized finely [20]. In other words, the after prediction of inter-frame samples oftentimes do not resemble the natural image due to uniform quantization.

After the quantizer phase, the product of a quantization factor is taken to get the decreased distortion. The lower quality means a coarser approximation. This also means very little bits were required for encoding. With the case of streaming, there needs to be a required constant bit rate, because once the quantization factor increases, this signals that the stream has reached its limit [9]. In essence, it is very difficult to predict the bit rate that is complimentary to the quantization factor, hence making this process complex.

2.2 Channel Coding

After a sequence of bits are taken from the source coder, they are then mapped to a codeword then sent onwards down the channel. Thereafter, the decoder analyzes the original sequence from the received distorted codewords. Channel coders address the channel model and while source coders address the stochastics of the source model [9].

2.2.1 Wireless Channel

Most wireless systems make use of the multiplexing to allow multiple concurrent transmissions in multiple frequency ranges or channels with efficiency propagation over-the-air. However, before analog interpolation, up-conversion at the transmitter, and down conversion sampling at the receiver, digital processing has to happen through all these phases [7]. For this reason, the discrete-time baseband model of the radio frequency channel (RF) channel is commonplace in the communication system design. A stationary wireless channel can be modeled as a linear time-invariant (LTI) system [7].

Hence, the impulse or frequency response can be used to characterize the channel set forth by the hardware and environment. In theory, the response is instantaneous, varying over time, and is ultimately unknown. The Doppler shift also plays a role. This is a phenomenon were a signal influenced by non-linear and time dependent distortion during the digital and analog processing at the transmitter or receiver [16]. Even more so, other signals, such as interference or circuit noise, increases the likelihood of distortion of signal at the receiver. Lastly, instrumental to a mobile wireless channel is fading over space and frequency. These are just a few of the constraints considered when developing a system for wireless transmission of data.

2.2.2 Equalization and Orthogonalization

To mitigate against inter-symbol interference (ISI), two methods are implemented: equalization and orthogonalization. The effect equalization has on ISI is the value of estimating the channel response at the receiver by decoding previous symbols, repeating the coding, and nullifying any interference on the channel through filtering to make the response to that frequency flat [14]. On the other hand, orthogonalization prevents ISI by ensuring that it is not possible for two symbols to interfere with each other. There is always a guard interval inserted between symbols that also ensures limiting the possibility of any ISI on the channel. A pulse-shaping filter at the transmitter can also ensure that the channel is compliant with the Nyquist ISI parameter [14].

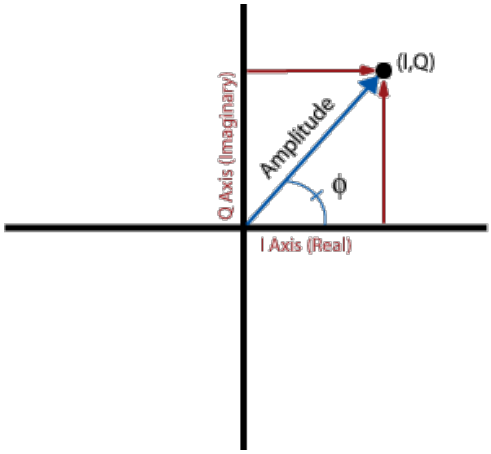


Figure 2.3: Diagram showing the I/Q relationship.

In the OFDM spectrum, symbols can also be modulated in terms of the amplitude and phase of each subcarrier. It is through the I/Q components that digital modulation of data can be mapped. As shown in Figure 2.3, the I-component and Q-component are common ways of referring to the in-phase and quadrature components that conveys some sort of information. These signals are relative and carries no real meaning on their own. However, by convention, the I-component is comprised of a cosine waveform and the Q-component representing the sine wave. The main benefit of this scheme is that they are amplitude modulated and not frequency or phase modulated.

This is important when it relates to the scaling of the DCT coefficients and implementing the proposed soft video transmission algorithm. Since there are various phases a signal can appropriate in the spectrum, the summation of the two aids in the positioning that is directly between the I-signal and phase of the Q-signal. The quadrature amplitude modulation (QAM) integrates directly into the OFDM architecture and the preferred wireless environment for the experiment shown in later chapters.

2.2.3 Digital Modulation

In general, modulation refers to process of modifying a sinusoid by mixing a signal with it to produce a new signal. By this token, the electromagnetic (EM) spectrum is divided into separate channels for RF use; this is called frequency division multiplexing (FDM) which modulation facilitates. Modulation is also applied within the information signal. An orthogonal frequency division multiplexing (OFDM) symbol can also be modulated in terms of the amplitude and phase of each subcarrier [16]. Each of those subcarriers can be modulated because each cosine and sine pair represent each subcarrier on a specific frequency. Identifiably the quadrature amplitude modulation (QAM) integrates directly into the OFDM architecture. Ideally, the math is that each complex number has a pair of real numbers for each subcarrier. Meaning, each OFDM symbol has N complex values for N subcarriers.

A channel must be able to map bits to multiple channel symbols to satisfy the requirements for the source-channel system. In practice, there are a finite alphabet of codewords, i.e., vectors in multi-dimensional space R^n, where n is the block-length. For each corresponding codeword, a sequence of bits gets mapped to that location. For instance, to achieve rate of k/n bits per channel dimension the encoder takes k bits where it would need 2^k codewords from the codebook [21]. On the receiving end, the original bit-sequence needs to be recovered from the distortion it receives.

Conditional probability $p(data|received\ codeword)$ [21] is done to get the maximum likelihood (ML) of the original bit-sequence. If the channel is additive white Gaussian noise (AGWN), it is identified that the ML decoder locates the codeword with the minimum mean square error (MMSE) from the received distorted vector, minimum distance from Euclidean distance in R^n. Subsequently, codewords in the codebook are spread out as far as possible in R^n to reduce decoding or encoding errors. Also, the average power $E[X^2]$ of the codebook must conform to the hardware and channel specification of the maximum allotted power. In fact, while n goes to infinity, the AWGN channel capacity is derived by the method of sphere packing the R^n space.

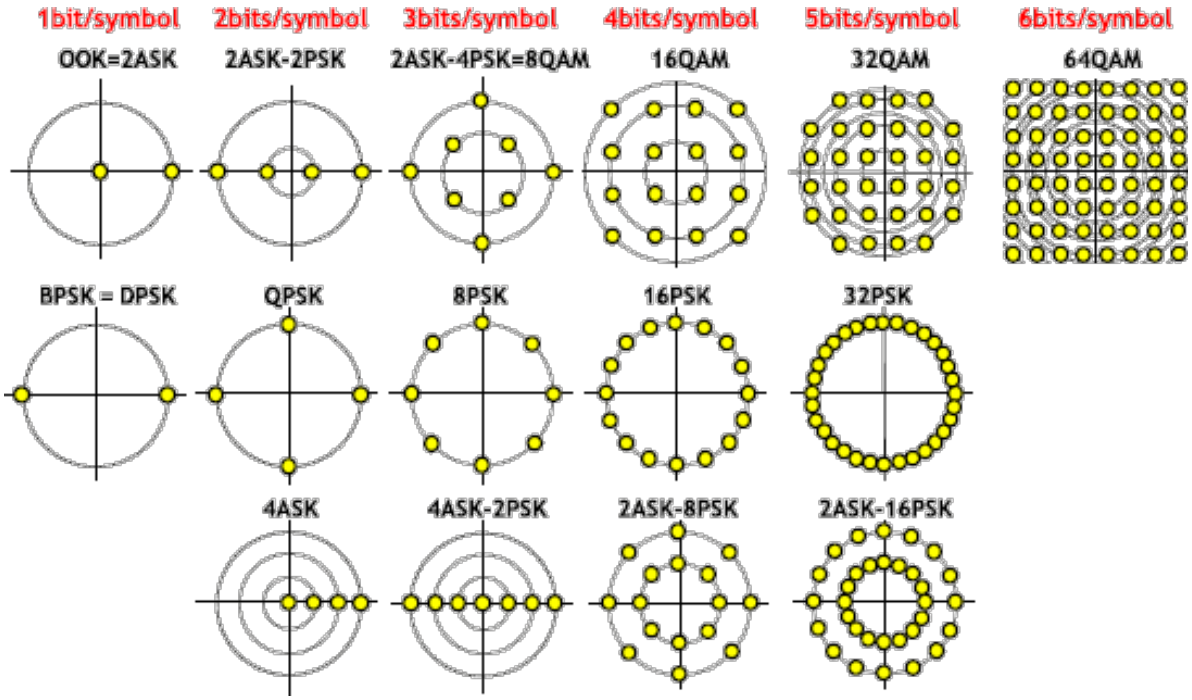


Figure 2.4: Channel Codewords of BPSK, QPSK, 16-QAM, and 64-QAM and equalized symbols (Figure courtesy of [16]).

Regarding the curious case of compression, the codebook actually does not achieve capacity at any SNR. However, the conventional two-dimension design is able complete coding by an efficient search. In the context of QAM, the codebook is referred to as constellation. Common constellations are BPSK, QPSK, 16-QAM, 64-QAM, illustrated on Figure 2.4. Finite-block-length constellations can never guarantee correct decoding, but the probability of bit error

can be competed [16]. Notice that the higher order of a constellation, each symbol increases in bit density, ultimately requiring a lower noise variance and SNR to achieve the same probability of bit error. Therefore, the need for a more robust and dynamic decoding technique is needed as will be discussed in later chapters.

2.3 The Long Term Evolution Network

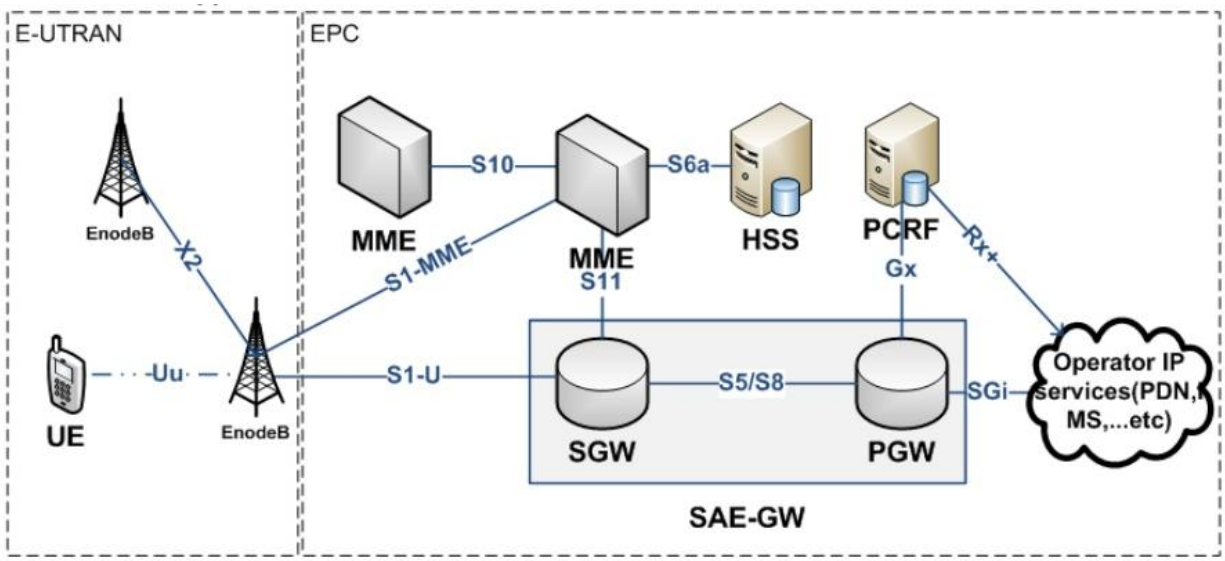


Figure 2.5: Basic architecture of an LTE cellular network (Figure courtesy of [31]).

The current generation of mobile technology is LTE, which succeeds the outdated 3G. The new network still follows the universal mobile telecommunications service (UMTS) technology and guidelines. It has a max data rate of 100 Mbit/s on the downlink and 50 Mbits/s on the uplink for a normal 20 MHz channel. The key advantages are low latency, cost efficiency, and higher data speeds. On the technological side, it incorporates OFDM (orthogonal frequency division multiplexing) for downlink, SC-FDMA (single carrier - frequency division multiple access) for uplink, MIMO (multiple input multiple output), and SAE (system architecture evolution) [19].

The LTE network has four vital components. First, the Evolved Node B (eNB or eNodeB) is central to the radio access network (RAC) of LTE. The eNB is responsible for a

multifaceted set of tasks, such as handling radio communications for multiple devices, allocates resource management, and processes handover decisions between cells. The Serving Gateway (SGW) routes and forwards user data packets and provides mobility anchors for all users during inter-cell handovers. When data for a UE is in the idle state, the SGW does termination of the path and paging for the UE [19]. In addition, it stores the network routing information and bearer service for each UE on the network. The Packet Data Network Gateway (PDN-GW) is the pathway for traffic for the UE, assuring connectivity UE to data packet outside of the network. It maintains Internet connectivity and quality of service for UEs.

The LTE EPC component is tasked with Mobility Management Entity (MME) in the network architecture. The MME is central to the EPC for signaling. MME's main roles are initiating paging and authentication for UEs. It also acts as a gateway to connect to the UEs that are within a cell. Traditionally a complex set of security protocols would need to be in place for this to happen. However, since the entire end devices are on the same *tun_enb* network, the EPC manages all allocation of resources. In addition, it allocates user location and appropriately select the complementing gateway during the initial registration process. Multiple MMEs even work together in peak hours offload network demands. The EPC is the anchor point for the intra-LTE mobility. These involve cases such as handover between eNodeBs, between LTE and legacy 3GPP accesses. It is logically linked to the packet data network gateway. The EPC component is vital for both registration and handover for the cellular network.

2.4 Software Defined Radio and Network

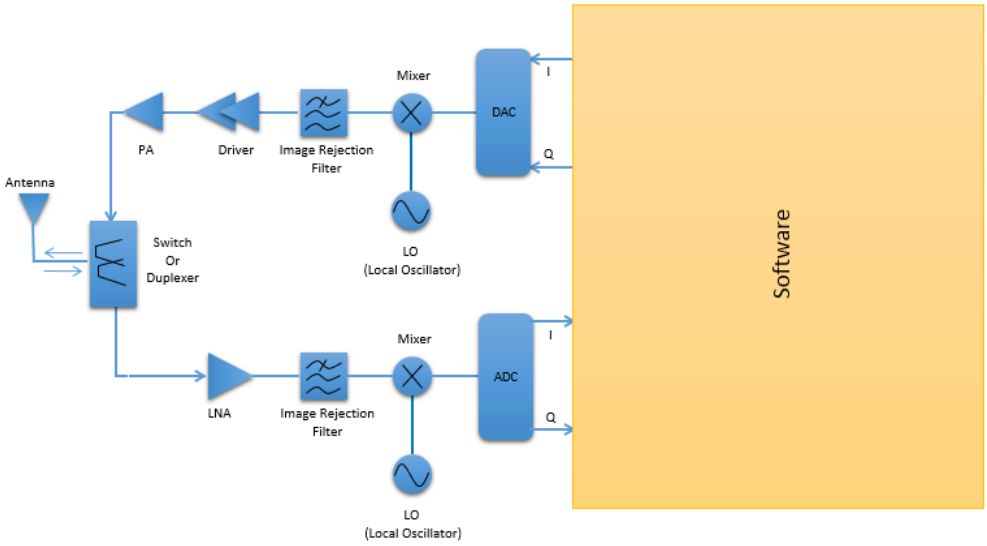


Figure 2.6: Common architecture of the current SDR system (Figure courtesy of [31]).

The chosen platform for developing the cellular system will incorporate the use of a software defined radio (SDR) device. SDR is a wireless technology that is mainly controlled by software to enable components that makes it adaptable, dynamic, manageable, and cost-effective [4]. In the common systems shown in Figure 2.7, the SDR architecture separates the network control and forwarding functions so that the network control is programmed with the appropriate infrastructure and the higher layer handles applications. In its system design, the data resource block is implemented alongside the combination of software or hardware switches components, that operates the forwarding packets based on polices native to the radio device [13].

A common platform for implementing a structure such as this is National Instruments' Universal Software Radio Peripheral (USRP) as shown in Figure 2.8. For the purposes of this experimentation, USRP B210 will be used. For prototyping an LTE network, the USRP B210 provides a fully integrated, single-board, platform with continuous frequency coverage from 70 MHz – 6 GHz. It combines the AD9361 RFIC direct-conversion transceiver providing up to 56MHz of real-time bandwidth, an open and reprogrammable Spartan6 FPGA, and fast

SuperSpeed USB 3.0 connectivity with convenient bus-power. The onboard FPGA performs several DSP operations, real signals in analog domain and complex basebands in the digital domain. In most use-cases, these complex samples are transferred to/from applications running on a host processor, which performs DSP operations. The code for the FPGA is open-source and can be modified to allow high-speed, low-latency operations to occur in the FPGA. Full support for the USRP Hardware Driver (UHD) software allows you to immediately begin developing with GNU Radio or cellular base stations [13].

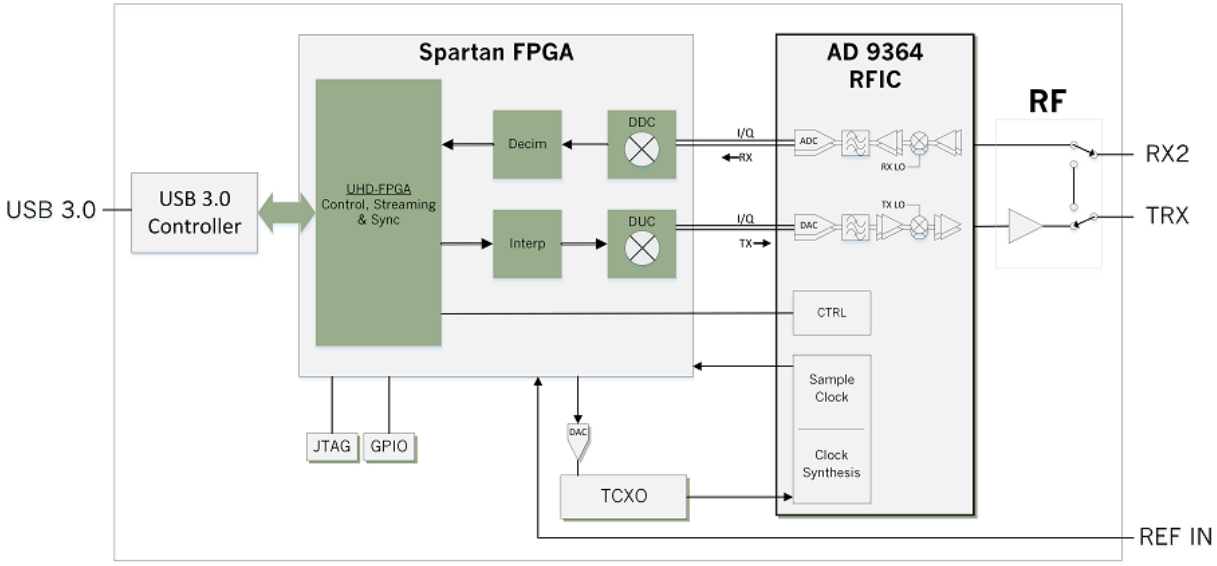


Figure 2.7: Schematic of the USRP B210 (Figure courtesy of [30]).

Chapter 3

Soft Video Delivery and Technical Challenges

3.1 Distortion

Source coding and channel coding for error resilience are blocks of functionality appropriated by the sender. Distortion, $E\left[D_{R,k}\right]$, is a parameter that is commonly used as a performance metric. This distortion metric depends on the source coding parameter such as quantization step size, predictions, and the error concealment used at the decoder level [6]. For instance, if S is the source coding parameters and C is the channel coding parameter, $s = \{s_1, \dots, s_m\} \in S^M$ and $c = \{c_1, \dots, c_m\} \in C^M$ represent the vectors of these parameters. The bit rate constraint becomes [6]:

$$\min \{s \in S^M, c \in C^M\}^{E[D(s,c)]} \approx R(s,c) \leq R_{0,} \tag{3.1}$$

where *R(s,c)* is the total bits used for source and channel coding for *M* group of pictures (GoP) at R_0, the bit rate constraint. Even though (3.1) illustrates a clean slate constraint, the problem is how to leverage error resilient source decoding, channel coding, and error concealment jointly versus the traditional separation scheme.

These limitations of separation are particularly important for wireless channels. Particularly in scenarios where the user is mobile and the SNR starts to vary and show unpredictability, distorting target data capacity and channel code capabilities [8]. One way to mitigate such scenarios is to control the selection of the channel-rate. For instance, with broadcast channels, diverse in nature due to multitude of devices and SNRs often times invalidate separation theorem [22]. To that effect, one proposed suggestion is to identify results that are decreasing latency due to complexity and increase the distortion by a small magnitude. Thereafter, if the source and

channel codes can be optimized in unison, the method of joint source-channel coding (JSCC) could provide a solution.

For OTA video transmission with a joint source-channel coding scheme, the list of components covers modulation and demodulation, data rate adaptation, packet scheduling, and power adaptation [24]. The adaptation components are considered channel-coding parameters. However, source coding parameters span areas such as mode selection, packetization, intra-MB refreshment rate, and entropy coding mechanisms [22]. Optimal performance can be achieved if both source and coding parameters can be optimally chosen.

3.2 Linear Error Protection

By convention, error protection codes convert video data to a different bit sequence. However, the traditional process corrupts the quantitative properties of the transmitted video, which in turn limits the scaling that can be done from the sampled pixels. For this reason, the first step is to overcome this challenge by preserving that value.

For instance, take SoftCast's approach for example (because the proposed method is a derivative and has a similar approach of tackling this problem). First, the DCT components are scaled to magnitude per frame to warp channel noise [12]. If a system sends an I/Q value of 1.5 over a wireless channel, then the value gets scaled up by 10 to produce 10.5. Then the receiver reconstructs this value and scales it down back to the correct value of 1.5. The issue is that the amplification of power increases strain on the power budget of the system. Traditionally, hardware systems have a fixed power budget, when exhausted, intrudes problems. The new method needs to know how to adapt and be meticulous in scale selection [18].

$$g_i = \lambda_i^{\frac{1}{4}} \left(\sqrt{\frac{P}{\sum_i \sqrt{\lambda_i}}} \right) \tag{3.2}$$

As with the SoftCast method, the requirement is finding a satisfactory scaling factor for each chunk of video data (i.e., discrete cosine transform coefficients). It is imperative to model each $x_i[j]$ value as a random variable. The means are transmitted as metadata after a zero-mean distribution is done after each chunk. A variance is captured by the information given by the mean form each chunk. On each chunk the variance, λ_i, is computed in order to find an optimized way to minimize the reconstruction error in equation (3.2). For every distribution, D_i, there is only one scaling factor g_i. The encoder in this system becomes linear since both the DCT and the error protection method are linear operations.

3.3 Metadata

The encoder also sends a negligible amount of data to assist in inverting received signals in the form of metadata. The bitmaps associated with the discarded chunks, the mean, and variances of each chunk are sent by the encoder [3]. On the backend, the scaling factors ae calculated from this information. The DCT and Hadamard matrices are not transmitted because they are justified. Sometimes run length encoding is used to compress the chunks and all metadata is compressed by Huffman coding [15]. In essence, the metadata overhead is .014 bits/pixel after applying the Reed-Solomon code.

To guarantee success the metadata has to be delivered to the receivers correctly. This can be done by sending it over 16-QAM, BPSK, or PSK modulation at half rate convolution code to protect it against channel errors. This is because these have the lowest FEC code in the 802.11 bitrate [8]. The method of distributing metadata across all GoP is used to ensure low probability of packet loss. Thus, each video data has a packet head consisting of the bit rate followed by the

metadata. Ultimately, it is possible to send metadata and soft video data in the same packet because each OFDM symbols use different FEC code and modulation.

3.4 Preliminaries of the Proposed Method

In continuation of Section 2.2.3, correcting errors in the current separate coding scheme could be solved by adopting a new method. As demonstrated by SoftCast, JSCC has the capabilities of solving the problem [3]. Under this methodology, the integration of compression, data protection and transmission is key. A 3D-DCT is first done to de-correlate the signal from the luminance value. Next, to achieve a reduction in the end-to-end distortion, power allocation of the DCT coefficients are done, and the relevant data are encoded as metadata. The whitening process through the Hadamard matrix alleviates packet loss. Lastly, in the constellation plot, the values are mapped, modulated, and then transmitted. Next, a linear square error (LLSE) at the decoder is used. Considering all the operations are linear in nature, the cliff effect problem can be effectively solved since the quality of the video now gracefully varies with the channel quality.

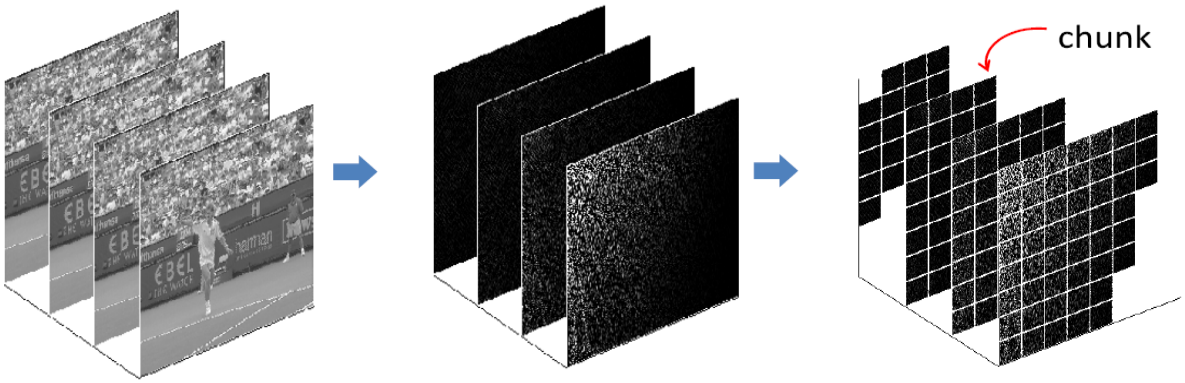


Figure 3.1: The flow off the diagram left to right: (i) a GoP of 4 frames, (ii) 3D-DCT, and (iii) chunk layering (Figure courtesy of [3]).

From SoftCast's framework, the DCT coefficients after each chuck division, as in Figure 3.1, have an applied scaling factor. There is ultimately a relationship between chunk sized and distortion. A small chuck means a large overhead, and a large chunk increases the distortion. For

instance, an 8 X 8 = 64 chunk division produces an acceptable metadata range of .005 bits/pixel [3]. Conversely, under this chunk size the PSNR range is still too large. Therefore, the aim would be to achieve proficient video transmission without overloading on metadata.

Chapter 4

The Proposed Soft Video Transmission System

4.1 End-to-End Architecture of the Proposed Method

For the blind detection method to be implemented, new encoding, decoding, and distortion schemes needs to be done. First a 3D-DCT on the GoPs are applied to obtain the DCT coefficients, which are placed in a matrix form of the same size. In the orthogonal axis, low frequency sections are compressed and signals decorrelated. For the previous matrix, we get a division of chunks. This method differs from the MPEG architecture shown in Figure 2.2, where a division of blocks on the picture is done, and then a 2D-DCT is performed on each block [17].

Assume each frame is divided into m^2chunks and such that $N = Fm^2$ chunks represent the GoP. Furthermore, each chunk is fitted with height $h = H/m$ and width $w = W/m$. Notice the *j*-th DCT coefficient in the *i*-th chunk is representative in $s_i[j], i = 1,2, \dots, N$, where $s_i[j]$ is scaled with a factor g_i for noise reduction [11]. The resulting value of g_i becomes metadata and is digitally sent to the receiver.

For the wireless transmission, the scenario envisioned assumes an additive white Gaussian noise (AWGN) channel. After demodulation, the receiver receives $y_i[j] = g_i s_i[j] + n_i[j]$, where $n_i[j]$ is AWGN with variance σ_n^2 [11]. This process in matrix form is **Y** = **GS**+**N**, where **Y**, where **G** is a diagonal matrix of power scaling factors, **S** is the stacked DCT coefficients matrix, and **N** is the AWGN noise matrix [11]. The job of the receiver is to reconstruct the DCT coefficients through the linear least square estimator (LLSE) (4.1) where Λ_s is the diagonal matrix where λ_i is the variance of the *i*-the chunk and Σ is the diagonal matrix that represents the variance of the AWGN entries. Such that equation (4.2) is the scalar equivalent.

$$\widehat{\boldsymbol{S}} = \boldsymbol{\Lambda_s G^T (G \Lambda_s G^T + \Sigma)^{-1} Y} \tag{4.1}$$

$$\hat{s}_i[j] = \frac{g_i \lambda_i}{g_i^2 \lambda_i + \sigma_n^2} (g_i s_i[j] + n_i[j]) \tag{4.2}$$

Distortion is taken into account by looking at the expected mean square error (MSE) (4.3) and forming a new distortion with scaled coefficients of the proportional chunks (4.4). In (4.4) g_i is the power scaling factor where if it is large enough the distortion is minimized.

$$\mathrm{MSE_s} = \mathbb{E}[(s_i[j] - \hat{s}_i[j])^2 \approx \frac{\sigma_n^2}{N} \sum_{i=1}^{N} \frac{1}{g_i^2} \tag{4.3}$$

$$\mathrm{MSE}_s \approx \quad \frac{\sigma_n^2 hw}{NP} \left(\sum_{i=1}^{N} \frac{1}{\lambda_i^{.5}} \right)^2 \tag{4.4}$$

If the signal to noise ratio (SNR) can be calculated as $\rho = \frac{P}{Nhw\sigma_n^2}$ and then the distortion of the DCT coefficient can written as (4.5). There is an inequality represents the when the chunk size is close to one and we have minimum distortion. However, if g_i is large then the transmission overhead will be large. Therefore, a method of reducing the overhead will be implemented.

$$\mathrm{MSE_s} \approx \frac{1}{\rho} \left(\frac{1}{N} \sum_{i=1}^{N} \mathrm{E}[s_i[j]^2]^{.5} \right)^2 \geq \frac{1}{\rho} (E[|s_i|])^2 \tag{4.5}$$

Before the overhead can be reduced, an MMSE filtering in (4.2) is done with the values of λ_i of all coefficients metadata is sent to the receiver. Take this example, a video sequence transmission where the resolution is 352 × 288, 811, 008 DCT coefficients will accompany it by the product of its dimensions [11]. Based on this resolution and DCT, 5.8 bits/pixel is produced in metadata [11]. For these modulated symbols and that bits/pixel rate, the channel performance will degrade in attempt to meet the power demand. For this reason, DCT, scaling, and MMSE filtering are deployed in the new scheme [5]. Even though the chunk can be lowered, the quality of the video

degrades significantly as seen in Chapter 5. And this can also have a negative effect on the recovering of DCT coefficients.

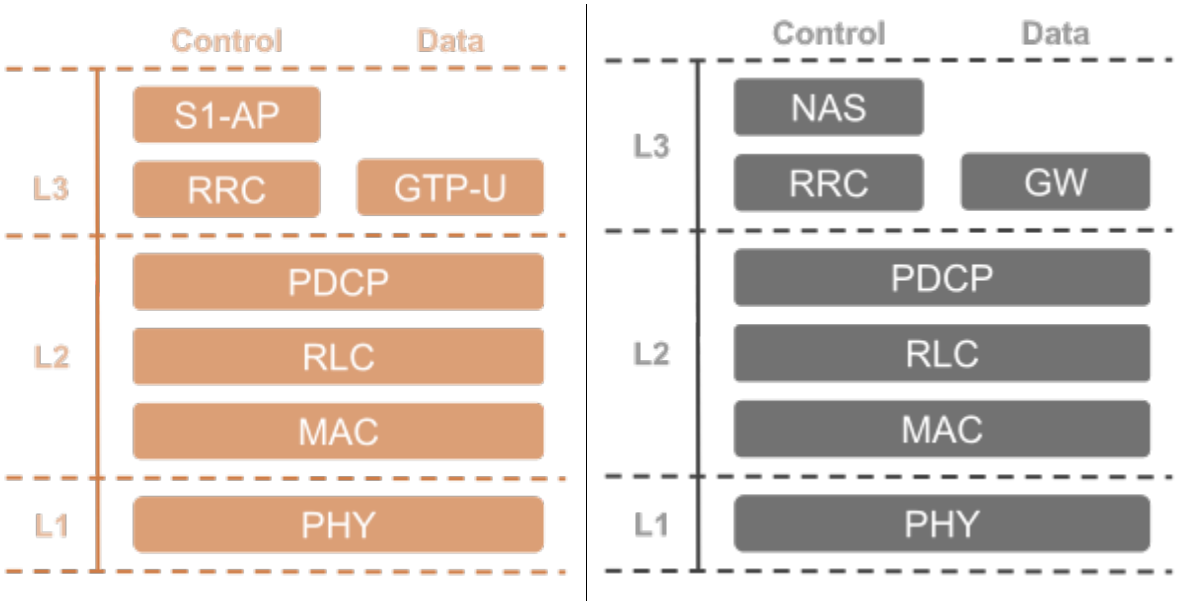
The traditional signal energy distribution (SED) signal detection algorithm leverages the information of an approximated g_i[20]. Conversely, this proposed algorithm decodes the signal without prior knowledge of g_i. The SED modeling-based method has the DCT coefficients scaled after learning from references, which sometimes introduce distortion. The scaling is done at the encoder prior in the proposed method.

$$MSE_s \approx \left(\frac{4}{\rho} + \frac{3}{\rho^2}\right)\left(\mathbb{E}[|s_i|]\right)^2 \tag{4.6}$$

The decoder only needs to know the constant and scalar value along with the total power for transmission from the DCT coefficients. This is all done with little cost per GoP. In essence, this proposed signal recovery method is near metadata-free with little overhead. The expected theoretical MSE of the proposed blind detection algorithm is (4.6) [24]. The ρ represents the SNR and $\mathbb{E}[|s_i|]$ the absolute value of the DCT coefficients in one GOP. The MSE depends on both the channel condition (ρ) and the video content (s_i).

4.2 The srsLTE System

The LTE system of choice will be the open-source srsLTE software suite [23]. The reason for selecting this system is that it incorporates the three vital components of an LTE network. E-UTRAN Node B (eNB), evolved packet core (EPC), and a user equipment (UE) are all the software components that is installed on the laptops and integrated on the RF frontends. Figure 4.1 shows the architecture for both the eNB and UE, with some minor differences at layer three. The following is a summary of how srsLTE manages the cellular system.



(a) (b)

Figure 4.1: The three-layer protocol stack for the srsENB (a) and the srsUE (b).

At the eNB, Figure 4.1 (a), Physical (PHY) layer carries all information from the MAC over the air interface. It is responsible for link adaptation and power control. The Medium Access Control (MAC) layer multiplexes data between one or more logical channels into Transport Blocks (TBs) which are passed to/from the PHY layer. The Radio Link Control (RLC) layer can operate in one of the three modes: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). The RLC achieves multiple logical channels or bearers for each connected UE.

The Packet Data Convergence Protocol (PDCP) layer is responsible for ciphering of control and data plane traffic, integrity protection of control plane traffic, duplicate discarding and in-sequence delivery of control and data plane traffic to and from the RRC and GTP-U layers respectively. The Radio Resource Control (RRC) layer manages control plane exchanges between the eNodeB and connected UEs. It generates the System Information Blocks (SIBs) broadcast by the eNodeB and handles the establishment, maintenance and release of RRC connections with the

UEs. Above the RRC, the S1 Application Protocol (S1-AP) layer provides the control plane connection between the eNodeB and the core network (EPC) [23].

Management Entity (MME) in the core component. Messages from the MME to UEs are forwarded by S1-AP to the RRC layer, where they are encapsulated in RRC messages and sent down the stack for transmission. The GPRS Tunneling Protocol User Plane (GTP-U) layer inside srsENB provides the data plane assembly between the eNodeB and the core network (EPC). The GTP-U layer joins to the Serving Gateway (S-GW) in the core network. Data plane IP traffic is condensed in GTP packets at the GTP-U layer and these GTP packets are burrowed through the EPC [23]. That IP traffic is then extracted from the tunnel at the Packet Data Network Gateway (P-GW) and passed out into the network.

At the UE side, as shown in Figure 4.1 (b), the Physical (PHY) layer transmits all information from the MAC OTA interface. It is responsible for link adaptation, power control, cell search and cell measurement for all cells. The Non-Access Stratum (NAS) layer organizes control plane switches between the UE and entities within the core network (EPC). It controls PLMN selection and organizes network attachment procedures, changing identification and authentication information with the in the register system. The NAS is responsible for starting and upholding IP connectivity between the UE and the PDN gateway within the EPC. The Gateway (GW) layer within srsUE is responsible for the formation and upkeep of the TUN virtual network kernel interface, simulating a networking device within the Linux OS. The GW layer permits srsUE to run as a user-space application and operates with data plane IP packets. All three components work in consortium to create the cellular network.

4.3 srsLTE Cellular Parameters

The cellular system will be operated with the Frequency Division Duplexing (FDD) mode. This means the uplink and downlink data channels are separated by frequency. In the realm of LTE, the selection of frequencies is not to be labored over. This is the job of the Evolved-UTRA Absolute Radio Frequency Numbers (EARFCN). Normally, if there is an EARFCN set to 3400, a calculation will spit out the frequency values for the uplink and downlink channels. Though this is just an example, there are multiple EARFCN's for uplink and downlink. Also, the channel bandwidth and the Number of Physical Resource Blocks (NRB) are directly related. More resource blocks require more channel bandwidth and an increase in the sampling rate on the SDR side, which in our case is 11.52 MHz. The relationship between the resources are summarized in Table 1 below.

Table 1: List of cellular channel resources and relationships.

Bandwidth	**Resource Blocks**	**Subcarriers (downlink)**	**Subcarriers (uplink)**
1.4 MHz	6	73	72
3 MHz	15	181	180
5 MHz	25	301	300
10 MHz	50	601	600
15 MHz	75	901	900
20 MHz	100	1201	1200

In our experiments, the resource block of 100 is the set parameter, which translates to 20MHz bandwidth. Selecting the band is related to the ERFCN and at 3400, it is in Band 7. At this

band, there is an uplink range of 2500–2570MHz (70MHz wide) and downlink range of 2620–2690MHz (also 70MHz wide), with a duplex spacing of 120MHz.

4.4 SDR and Algorithm Integration

Figure 4.2 illustrates a shared physical layer that allows for the branching of direct OFDM symbols or normal encoding/decoding video scheme. One of the conventional schemes is the SVC-M, which ultimately has to do a hierarchical demodulation at the receiver. And MPEG4 goes through a QAM demodulation but still has to follow through to the deinterleaver and Viterbi decoder before the output is processed. The latter route is the JSCC implementation on the physical system where it bypasses all onboard encoder-decoder pair. Instead does so within the algorithm. The scaled symbols are calculated using the preferred methods of metadata reduction. It is then sent over that OFDM scheme in order to process the codewords to data on the backend of the algorithm. This is vital in overcoming the cliff effect of traditional HEVC systems. This is due to the process of using the proposed method. The goal is for lower latency and lower overhead with regards to metadata transmission.

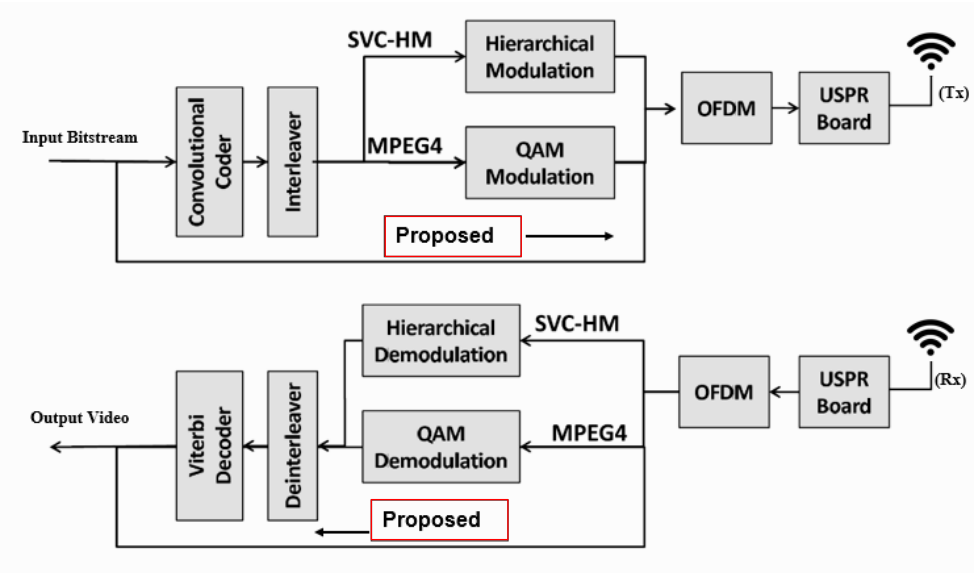


Figure 4.2: Block diagram for the proposed method integrated into the traditional system.

Chapter 5

Experimental Study

5.1 Experiment Setup and Metrics.

Multiple software and hardware components need to be assembled to complete this set-up. The two vital components are the laptops and USRP B210s. The performance metric that is used are the peak-signal-to-noise ratio (PSNR) (5.1) and the Structural Similarity Index (SSIM) (5.2) [22].

$$PSNR = 10\log_{10}\left(\frac{\left(2^B-1\right)^2}{MSE}\right) \tag{5.1}$$

$$SSIM(x,y) = \frac{(2\,\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2\,)} \tag{5.2}$$

Hardware: Two laptops HP Zbooks with Intel based i5 processor and USB 3 interfaces, and two NI's USRP B210s. These laptops must be installed with Ubuntu 18.04.02 and MATLAB. Four VERT2450 Dual Band 2.4 to 2.48 GHz and 4.9 to 5.9 GHz omni-directional vertical antenna, at 3dBi Gain, are used by the transmitter and receiver.

Test Videos: Monochrome video sequences at 20 frames per second. The chosen images are *akiyo and stefan*. Since the samples are in the .y4m format, they will be converted to the yuv format for better calculations [10].

Modulation and Wireless settings: the srsLTE suite using FDD frames built in an OFDM framework. The uplink frequency will be 2565MHz and downlink frequency will be 2685MHz. The clock rate will be 30.72MHz and sampling frequency will be 11.52 MHz. Through an AWGN channel, the digital modulation scheme is configured to 16QAM. Convolution, coding, FEC components are enabled in the system. Four VERT2450 Dual Band 2.4 to 2.48 GHz and 4.9 to 5.9 GHz omni-directional vertical antennas at 3dBi Gain are used.

Figure 5.1: The configuration of the sender (left) as the eNB and the receiver (right) as the UE.

Figure 5.1 represents the complete physical experiment LTE testbed. Both USRP are USB powered and have light indicators representing the system in the ‘on’ state. The data being sent is from the left where we see the graphical user interface computing the data of the available symbols and those equalized. Control of the channel includes creating a TCP socket on both end devices, one acting as the server and the other the client. A few tests are done to confirm operation of the cell towers including network attachment, EPC registration and iperf throughput. Bit rates on both ends are observed to signal the process of data transfer. A follow up spectrometer test is done to confirm the appropriate signal strength coming from the antennas. Even though the VERT2450 antenna may come with a base set of specifications, the experimentation in the field is always shown to underperform.

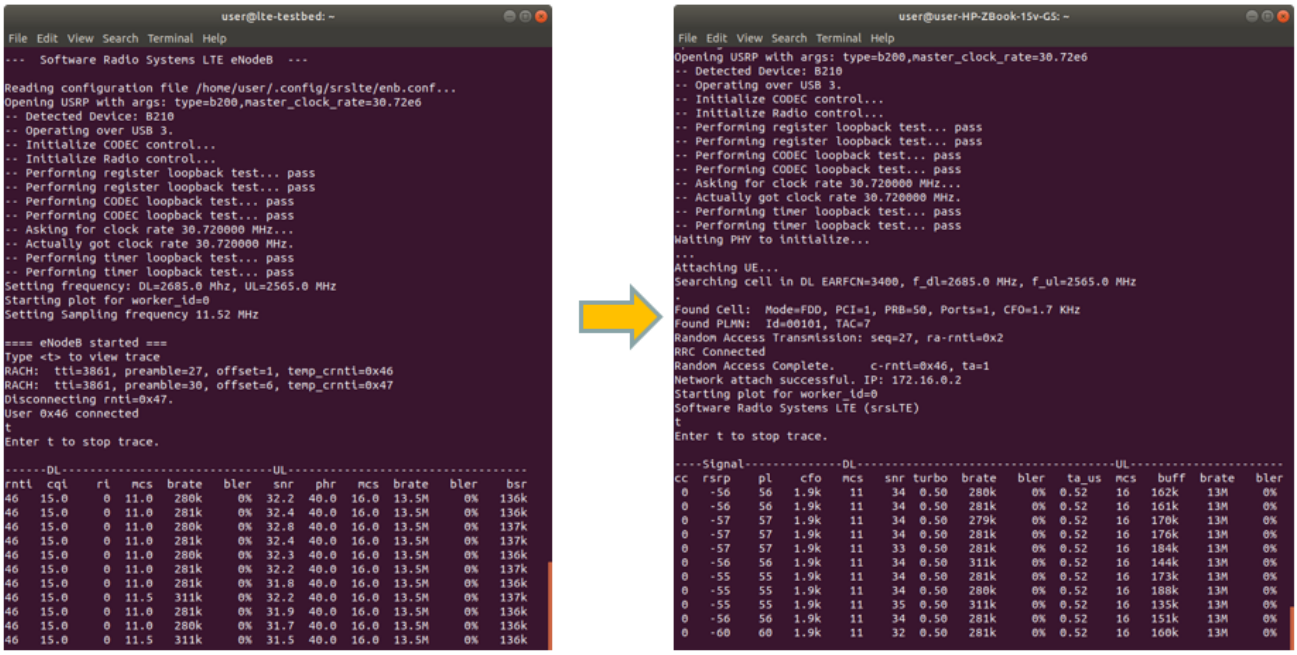


Figure 5.2: The left image is from the eNB and the right image is the UE configuration window.

One of the most useful components of the srsLTE systems is the onboard diagnostics and metrics watcher as seen in Figure 5.2. The radio network temporary identifier (RNTI) is a type of UE ID for the traffic between UE and eNB lower layers. However, it is more accurately described as the UE ID and the downlink control information (DCI) ID because each CRC bits of DCI are distorted by each RNTI. CQI stands for Channel Quality Indicator. As the name implies, it is an indicator carrying the information on the communication channel quality. Rank indicator (RI) is an indicator showing how well multiple antennas work. In addition, for the applications of the experiment we are only using one antenna. MCS (Modulation Coding Scheme) is related to Modulation Order which can be QPSK or QAM based on the translation table for each MCS that is in the LTE system. Block Error Rate (BLER) is a ratio of the number of erroneous blocks to the total number of blocks transmitted on a digital circuit. In LTE BLER is to determine the in-sync or out-of-sync indication during radio link monitoring (RLM). Power headroom indicates how

much transmission power left for a UE to use in addition to the power being used by current transmission. It is described as:

$$Power\ Headroom = UE\ Max\ Transmissio\ Power - PUSCH\ Power = P_{max} - P_{push}.$$

The other known variables on the table are signal-to-noise ratio (SNR) and bitrate (Brate). All changes to the system are made in the *enb.conf* and *ue.conf* files.

The characteristics of the physical channel during transmission are shown in Figure 5.3. There are four quantities measured through the GUI platform: channel response, channel response, physical uplink shared channel (PUSCH), and physical uplink control channel (PUCCH) equalized symbols. A successful connection of the two base stations are represented through the completion of the equalized symbols of the 16-QAM.

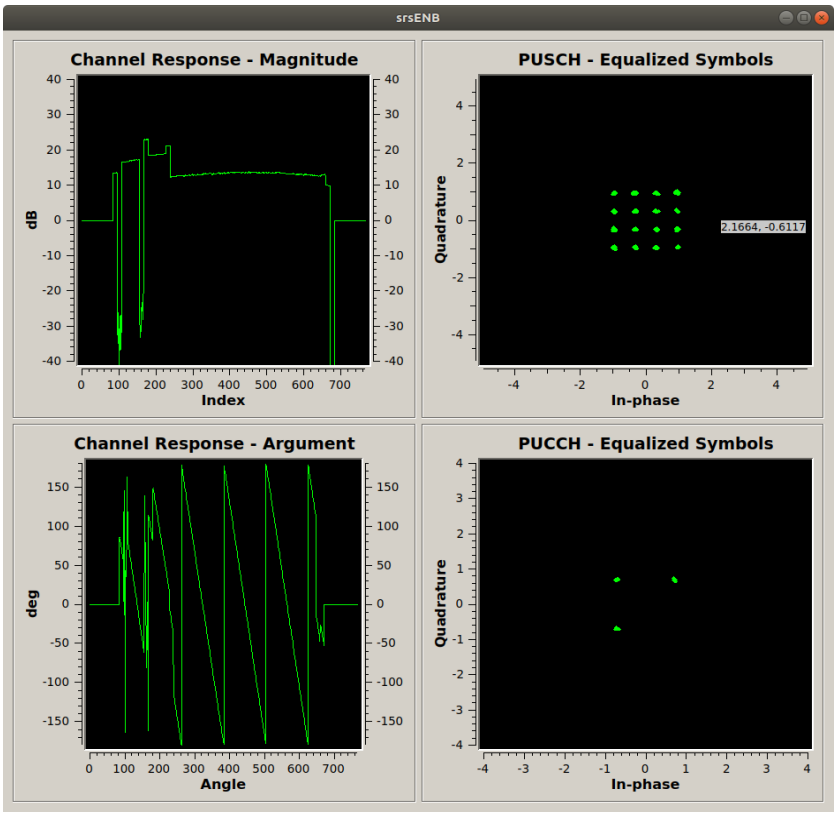
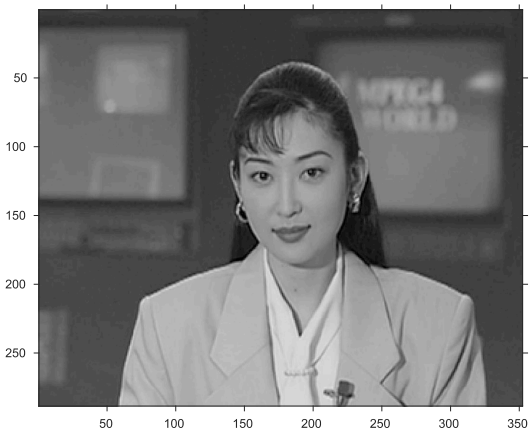


Figure 5.3: Different analog/digital channel measurements and equalized symbols (eNB)

The sending of symbols on the shared channel is representative of the control areas those are sent to in the PUCCH. Not all of the resources in-phase are used to do so since the codewords required very little data due to the nature of reduced metadata. Conversely, when the channel does experience and increased amount of traffic, most, if not all, of the PUCCH symbols are equalized. Signaling a very important process, all the resource blocks are in use.

5.2 Overhead Reduction Test

In the following test scenarios, results of the various videos simulated over an AWGN based on the metrics discussed in the previous section are presented. This test is done in two scenarios. The first video is uniform in nature and the second video requires more resource use.

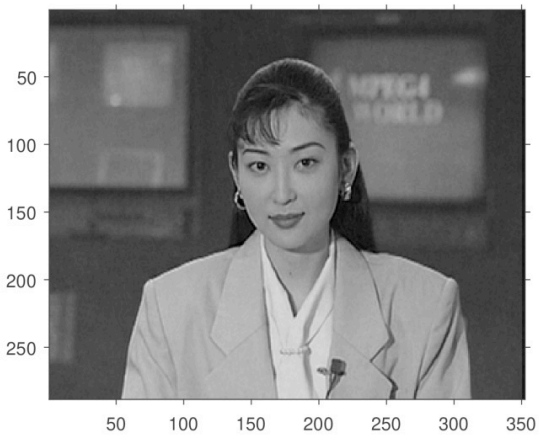


(a) Original

(b) Chunk: 16x16, PSNR: 40.75, SSIM: .9692

(c) Chunk: 8x8, PSNR: 37.67, SSIM: .9473

(d) Chunk: 2x2, PSNR: 29.08, SSIM: .6112



(e) Proposed, PSNR: 37.8677, SSIM: .9458

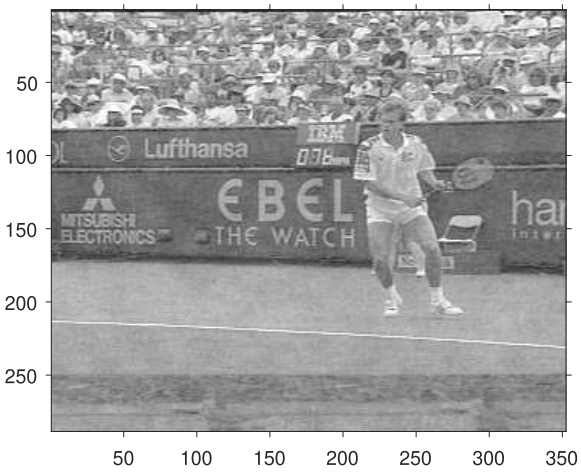
Figure 5.4: Decoded video frames from the experimented schemes from video *akiyo* at 10 dB SNR over LTE.

In scenario one, Figure 5.4 compares the visual quality of the images through the various compression methods for *akiyo*. Starting with the original image (a), we can look at the impact degradation of the channel quality might have on the video quality. A few key points to keep in mind is the relationship between the PSNR and SSIM [22]. The chunk-based method for the SoftCast scheme are done over 16 X 16, 8 X 8, 2 X 2 with the corresponding SSIMs .9692, .9473, .6112 achieved, respectively. Whereas the SSIM for the proposed method is .9458. The quality *akiyo* at 16 X 16 is closer in similarity to the original. However, the price is the increased overhead with a higher PSNR. The other reason why the quality of (b) is comparable to the original is due to the allocation of resources in this image. The YUV color space in these frames is vastly smaller than that which you would find in a blockbuster motion picture film. The background is static and almost of one color. The woman represented here is uniform with the environment, very little discord between the two. Even though the size of this video is almost the same as the others experimented. The lower end of 2 X 2 (d) was chosen to show how much distortion could happen in a frame that encounters some form of channel quality degradation. When compared to the

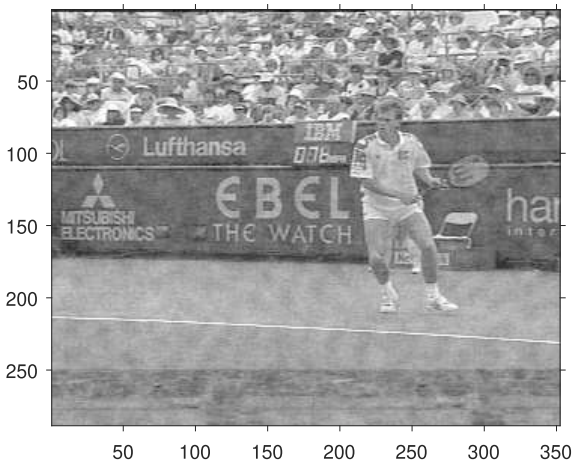
superior quality of either (b) or (e), we can automatically see how much more efficient the algorithm is in this scenario.

In scenario two, Figure 5.5 shows how much more complex it is to reconstruct a video that incorporates many elements. The chunk-based methods for the SoftCast scheme are done over 16 X16, 8 X 8, 2 X 2 with the corresponding SSIMs .8744, .8418, .5900 achieved, respectively. Starting with the (a) it can be seen that these images have a lot more components. It ranges from the man playing tennis to the rows of spectators in the stands. The level of granular complexity is increased in this video frame set. Comparing (b) to (a) the quality is seen to have taken a large drop. As compared to the SSIM of Figure 5.4, there is about a 6 - 9% difference in quality. The distortion is greater due to increase in diverse elements that have to be reconstructed as supposed to figure 5.4. Even with the increased complexity of this video, the range for the SSIM stays within a reasonable range. However, once it reaches the lowest compression rate at 2 X 2, the quality of the video drops below the 60 percent range. As compared with its relative counterpart in figure 5.4 there is only a 2% change between the two, with *stefan* on the lower end. Another point of discussion is the use of the codec and efficient calculation under the pressure of time. One reason for this is the chosen platform, the YUV format. In the traditional MPEG or high definition, the computation time would have decreased considerably. Whitening process is an added layer of efficiency that keeps the duty to a minimum. This adds a layer of uniformity, especially with it comes to more complex videos across all digital solutions. Compressing a chosen color scheme is a little different in terms of the selected algorithm. A consistent method of measuring videos in black and white is based on looking solely at numerical positives. Even though the SSIM for the blind method (e) is lower than 16x16 chunk method (b), the overall metadata overhead is lower. These numerical positives assist in validating the results of this novel process.

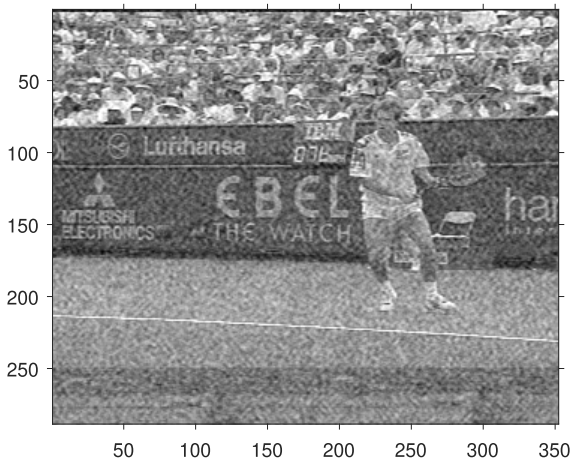
(a) Original



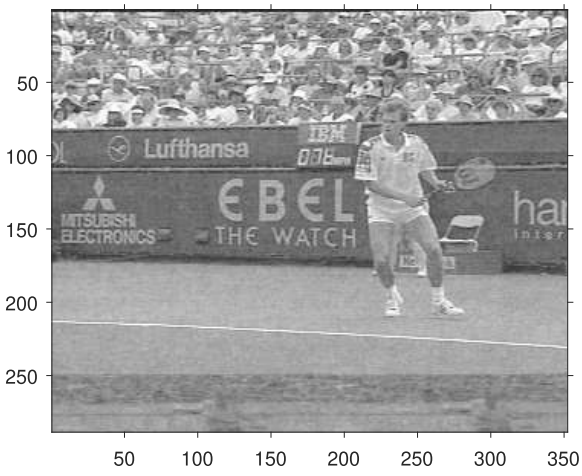
(b) Chunk: 16x16, PSNR: 31.49, SSIM: .874



(c) Chunk: 8x8, PSNR: 29.86, SSIM: .8418



(d) Chunk: 2x2, PSNR: 23.62, SSIM: .5900



(e) Proposed, PSNR: 29.27, SSIM: .8284

Figure 5.5: Decoded video frames of the experimented schemes from video *stefan* at 10 dB SNR over LTE.

From Figures 5.6 and 5.7, it can be seen that the proposed method performs much better, offering a gain of 2 dB generally (note that a 0.5 dB gain is worth pursuing). Looking at *akiyo's* data, the graph is overall uniform. The proposed method is able to increase the quality of the video linearly over time as with the other SoftCast method. The graph incorporates three different levels of quality. On the lower end, 2 X 2, this represents the data rate being bottlenecked at a lower bit rate, ultimately defaulting to chosen quality. Even though the video can be compressed at a larger chunk as shown in the experiment, the 16 X 16 normally is representative of a superior quality of the delivery channel. This is not always the case with the average userbase. For this reason, the 8 X 8 chunk is chosen to represent the average medium.

The performance results for the video *stefan* show some interesting observations. The proposed method started out on the lower end of the PSNR spectrum. This goes back to the use of resources. It seems that the scaling process before decoding, some pertinent information on the video becomes lost. This could be that shrinking the metadata size on a bad quality consequentially affects it even more. From 10 dB – 16 dB the quality of the video starts to equalized into a closer quality range. This proves that the quality of the video, when operated over the proposed algorithm, increases the PSNR of those chosen video sequences.

The main reason is that the proposed method recovers the coefficients from the amplitude of the received signals. There are come close similarities among the performance of the sequence *stefan*. But the consistent result is that the quality of the video changes gracefully. The SNR for the video increases the proposed method is able to select from a higher realm of SNR values with very little channel noise. More resources are available to select from to recover the transmitted signals. After a certain point, after 40 dB PSNR, the quality of the video becomes negligible. For this reason, the range for tracking changes in the video quality caps at 51 dB for most cases.

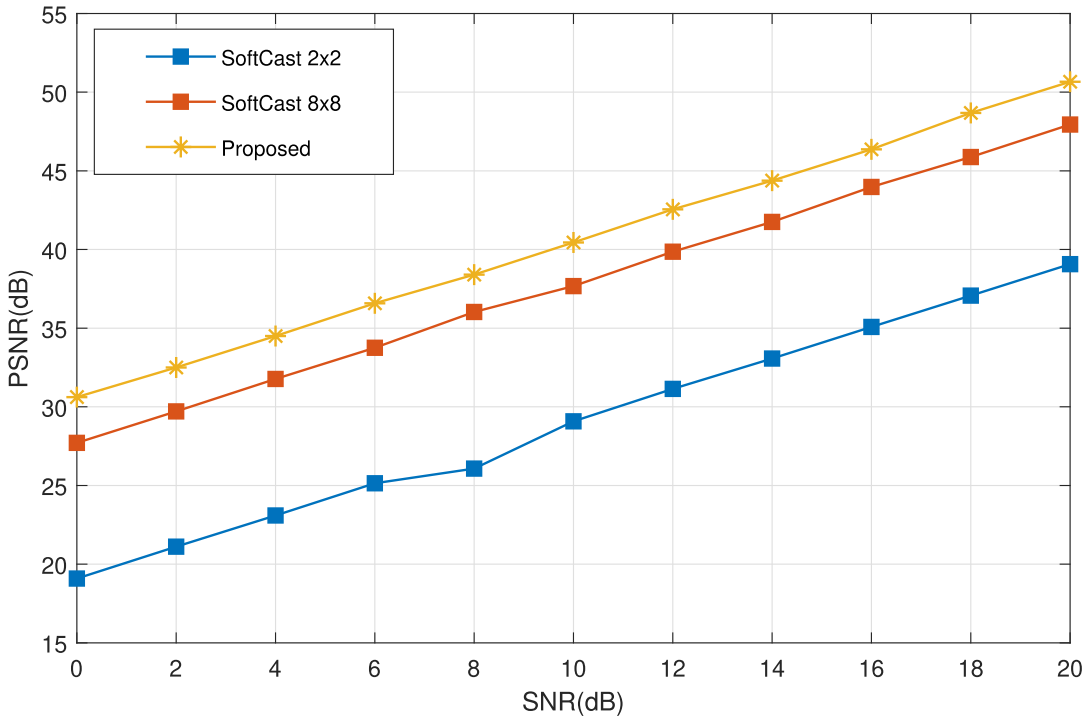


Figure 5.6: PSNR Performance for sequence *akiyo*

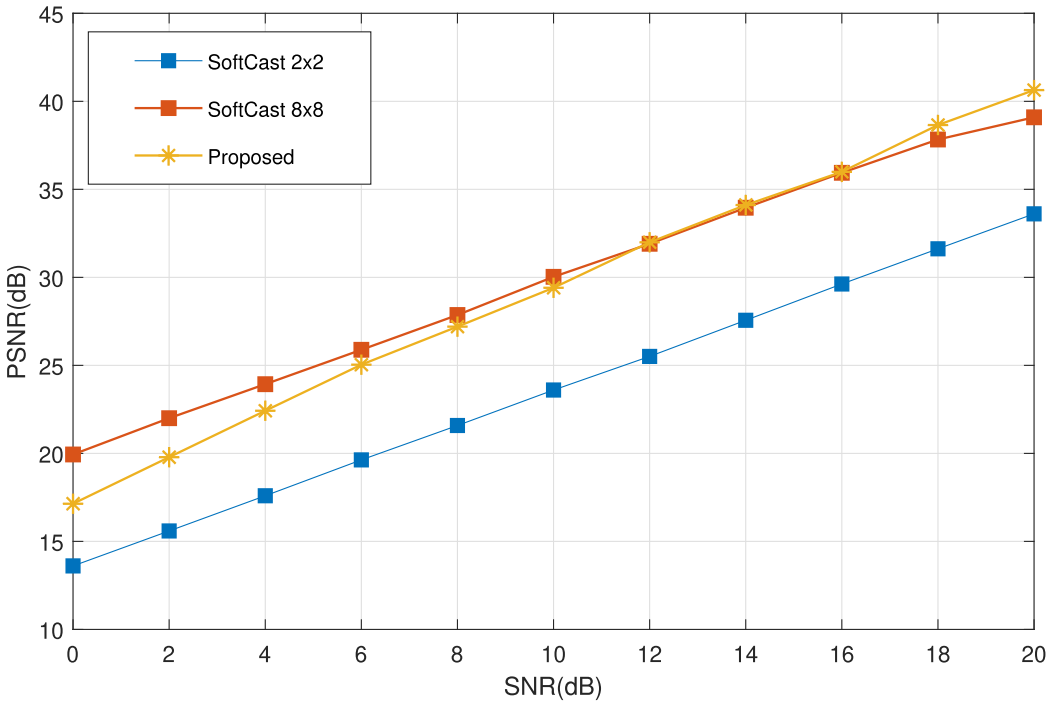


Figure 5.7: PSNR Performance for sequence *stefan*

5.3 Comparison with Traditional Solutions

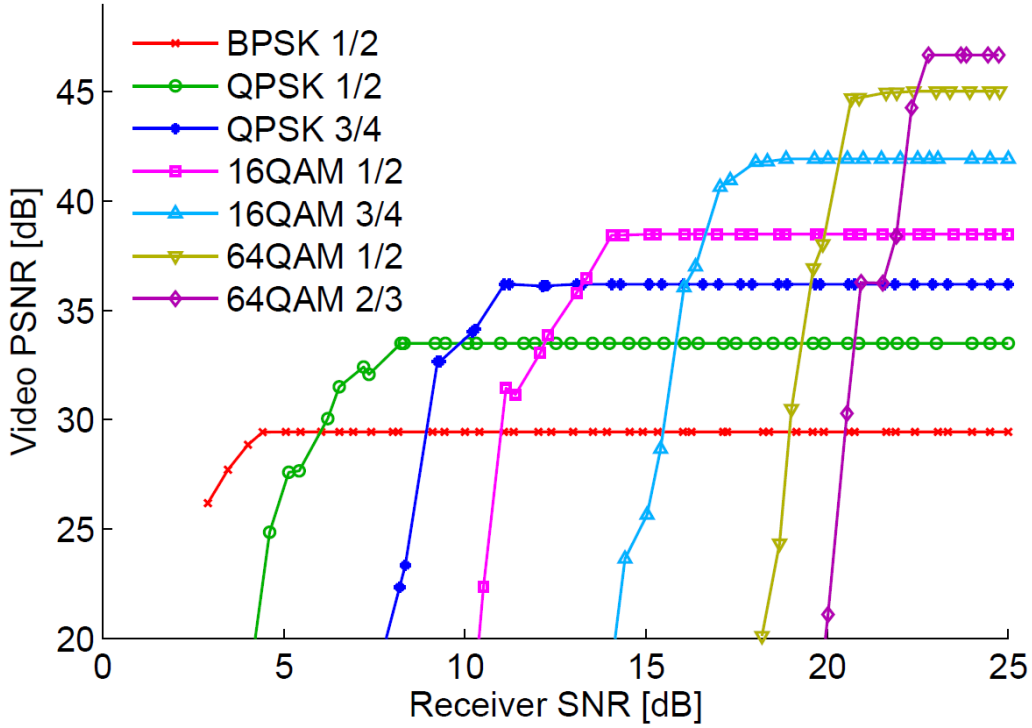


Figure 5.8: Comparison of video quality with other digital solutions (Figure courtesy of [3]).

As in Figure 5.8, it is shown that there is a point in the video transmission where the design plateaus sharply and where the quality can no longer be improved. These design curves represent the cliff effect discussed earlier. On the srsLTE platform, modeling of various schemes can be accomplished through multiple system alterations within the code. With this tradition scheme, the higher the modulation, the higher the video quality. Most of the conventional systems max out at around 48 dB. As seen from the experiments, the proposed method is even able to achieve an additional 2 dB over the max threshold with 51 dB. In addition, as seen by the metrics in the figures, as compared with the linear nature of the proposed methods, this method of video deliver is inferior in terms of performance. The proposed method can be adopted to increase performance of video delivery over LTE.

Chapter 6

Conclusion and Future work

6.1 Conclusion

In this thesis, we presented a new method of wireless video transmission and compared it with several baseline methods already used in the field. Moreover, we aimed to see if a stable LTE network could be created where the proposed compression method could be integrated. It was found that srsLTE was able enough to provide the appropriate amplitude modulation to carry out the soft video transmissions. Since these were low data-rate computations, the wireless channel was able to handle the transmissions without much distortion. The test was done across two different video sequences, using both the baseline SoftCast scheme and the proposed method, both implemented on the NI USRP SDR platform. It was through the experiments that we found the proposed method worked great in terms of minimizing the metadata. However, the quality of the video could be impacted based on the intensity of pixel usage. Overall, the proposed method shows high promise for soft video delivery.

6.2 Future Work

6.2.1 Mobile Application

One key area that would be great to investigate is implementing the system on a cellular phone through a mobile application. This is more interesting given the wide available of smartphones. In addition, registering the phone with a subscriber identity module (SIM) could provide more insight on how phone actually behaves in the field.

6.2.2 Multiple UE deployment

It is also interesting to investigate the system performance under a multiple UE deployment scenario. In such case, the radio resources are shared by the UEs, where the problem of resource

allocation and optimization arises [24]. It is also interest to construct a multi-cell LTE network, and to transmit soft videos over the LTE network to other cells. It could also assist in the development of understanding on hand-over and on how to better use the video delivery resources in the multi-cell environment.

6.2.3 Cellular Handover

The handover process in telecommunications is when a call or data session is transferred from one cell base station to another when the user moves. And this is all done without disconnecting the user session. All cellular services are based on the mobility aspect, which involves handovers to aid in the finding better cell performance. There are two types of handovers, soft and hard handover. As with the nature of video transmission in this process, there is always a performance hit. It would be interesting to see the impact of the handover process on soft video delivery.

References

[1] S. Kokalj-Filipovic, E. Soljanin, and Y. Gao, "Cliff effect suppression through multiple-descriptions with split personality," in *Proc. IEEE ISIT 2011*, St. Petersburg, Russia, Aug. 2011, pp. 948–952.

[2] Cisco visual networking index: Forecast and methodology 2008-2013. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ ns537/ns705/ns827/white-paper-cll-481360_ns827_Networking_SolutionsWhitePaper.html, June 2009.

[3] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," in *Proc. ACM SIGCOMM'10*, New Delhi, India, Aug./Sept. 2010, pp.449–450.

[4] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, "Cactus: A hybrid digital-analog wireless video communication system," in *Proc. ACM MSWiM'13*, Barcelona, Spain, Nov. 2013, pp. 273–278.

[5] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. ACM MobiCom'11*, Las Vegas, NV, Sept. 2011, pp. 289–300.

[6] R. Xiong, J. Zhang, F. Wu, J. Xu, and W. Gao, "Power distortion optimization for uncoded linear transformed transmission of images and videos," *IEEE Trans. Image Process*., vol. 26, no. 1, pp. 222–236, Jan. 2017.

[7] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust uncoded video transmission over wireless fast fading channel," in *Proc. IEEE INFOCOM'14*, Toronto, Canada, Apr./May 2014, pp. 73–81.

[8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol*., vol. 13, no. 7, pp. 560–576, July 2003.

[9] K.-H. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1283–1290, Dec. 1976.

[10] "Xiph.org video test media [derf's collection]." [Online]. Available: https://media.xiph.org/video/derf/

[11] T. Zhang and S. Mao, "Metadata-reduction for soft video delivery," *IEEE Networking Letters*, vol. 1, no. 2, pp. 84–88, June 2019.

[12] C.-L. Huang and S. Liang, "Unequal error protection for MPEG-2 video transmission over wireless channels," *In Signal Process.: Image Commun*., vol. 19, no. 1, pp. 67–79, Jan. 2004.

[13] Z. Chen and J. Wu, "LTE physical layer implementation based on GPP multi-core parallel processing and USRP platform," in *Proc. EAI ChinaCom*, Maoming, China, Aug. 2014, pp.197–201.

[14] X. Fan, R. Xiong, D. Zhao, and F. Wu, "Layered soft video broadcast for heterogeneous receivers," *IEEE Trans. Circuits Syst. Video Technol*., vol. 25, no. 11, pp. 1801–1814, Nov. 2015.

[15] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Trans. Circuits Syst. Video Technol*, vol. 24, no. 2, pp. 331–345, Feb. 2014.

[16] B. Barmada, M. Ghandi, and E. Jones, "Prioritized transmission of data partitioned h. 264 video with hierarchical QAM," *IEEE Signal Processing Letters*, vol. 12, no. 8, p. 577–580, Aug. 2005.

[17] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, "ParCast+: Parallel video unicast in MIMO-OFDM WLANs," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2038–2051, Nov. 2014.

[18] D. Liu, et al., "Cost-distortion optimization and resource control in pseudo-analog visual communications," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3097–3110, Nov. 2018.

[19] A. K. Sadek, T. Kadous, K. Tang, H. Lee, and M. Fan, "Extending LTE to unlicensed band-Merit and coexistence," in *Proc. IEEE International Conference on Communications Workshop (ICCW)*, London, UK, June 2015, pp. 2344–2349.

[20] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Scalable video multicast for MU-MIMO systems with antenna heterogeneity," *IEEE Trans. Circ. Syst. Video Technol*., vol. 26, no. 5, pp. 992–1003, May 2016.

[21] C. Schertz, A. Week, "Hierarchical modulation- the transmission of two independent DVB-T multiplexes on a single frequency," *EBU Technical Review*, Mar. 2003.

[22] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans on Image Processing*, vol. 11, no. 9, pp. 1043–1052, Sept. 2002.

[23] "srsLTE," 2015. [Online]. Available: https://github.com/srsLTE/

[24] T. Zhang and S. Mao, "Joint Power and Channel Resource Optimization in Soft Multi-View Video Delivery," *IEEE Access Journal*, vol. 7, pp. 148084–148097, 2019. DOI: 10.1109/ACCESS.2019.2946607.

[25] S. Mao, *Video Streaming over Cognitive Radio Networks: When Quality of Service Meets Spectrum*. New York, NY: Springer Science+Business Media, 2014.

[26] Z. He, S. Mao, and T. Jiang, "A survey of QoE driven video streaming over cognitive radio networks," *IEEE Network*, Special Issue on Quality-of-Experience (QoE)-Aware Design in Next-Generation Wireless Networks, vol.29, no.6, pp.20–25, Nov./Dec. 2015.

[27] M. Amjad, M. H. Rehmani, and S. Mao, "Wireless multimedia cognitive radio networks: A comprehensive survey," *IEEE Communications Surveys and Tutorials*, vol.20, no.2, pp.1056-1103, Second Quarter 2018.

[28] Z. Wang, S. Mao, L. Yang, and P. Tang, "A survey of multimedia big data," *IEEE/CIC China Communications*, vol.15, no.1, pp.155–176, Jan. 2018.

[29] T. Zhang and S. Mao, "An overview of emerging video coding standards," *ACM GetMobile Magazine*, vol.22, no.4, pp.13–20, Dec. 2018.

[30] "USRP B2x0 series" 2018 [Online]. Available: http://files.ettus.com/manual/index.html

[31] "SDR" 2016 [Online]. Available: http://www.sharetechnote.com/