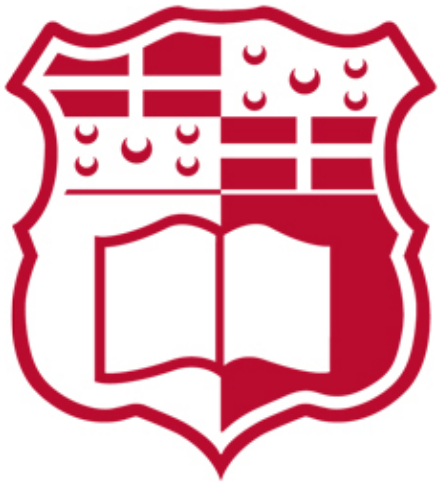
# CROSS-LAYER DESIGN FOR MULTI-VIEW VIDEO PLUS DEPTH TRANSMISSION OVER LTE NETWORKS IN CROWD EVENT SCENARIOS

Mario Cordina

Supervised by Prof. Ing. Carl James Debono



Department of Communications and Computer Engineering
University of Malta

2018

*Dissertation for the degree of Doctor of Philosophy in Communications Engineering*

**Board of Examiners**

Prof. Ing. Saviour Zammit (Chair)

Prof. Ing. Adrian Muscat

Prof. Christian Timmerer (External)

Prof. Giovanni Schembra (External)

To my Parents, loving wife Anabel
and daughter Giulia

“It always seems impossible until it’s done.”

Nelson Mandela

# Declaration

Plagiarism is defined as "the unacknowledged use, as one's own, of work of another person, whether or not such work has been published, and as may be further elaborated in Faculty or University guidelines" (University Assessment Regulations, 2009, Regulation 39 (b)(i), University of Malta).

I, the undersigned, declare that the dissertation submitted is my work, except where acknowledged and referenced.

I understand that the penalties for committing a breach of the regulations include loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Course code: CCE6000

Title of work submitted: Cross-Layer Design for Multi-View Video plus Depth Transmission over LTE Networks in Crowd Event Scenarios.

Date: 1st October 2018

Mario Cordina

Ph. D. Graduate

# Abstract

The development of digital multimedia systems has seen an unprecedented growth in recent years, with immersive video technology taking a central role. An application which is driving interest in this technology is free-viewpoint video which allows viewers to interactively navigate a scene by selecting their preferred viewing position. This is made possible through the generation of novel viewpoints rendered from a small set of texture and depth map views using a view synthesis technique. Meanwhile, according to Cisco, mobile video traffic accounted for 60% of total mobile data traffic in 2016 and this is expected to reach 78% by 2021. This growth in mobile video traffic coupled with the introduction of free-viewpoint video in the mobile ecosystem will have an impact on the user experience especially in crowd event scenarios. Such scenarios are characterised by high uplink user data traffic coupled with excessive uplink signalling overhead caused by channel quality feedback reports.

In this thesis, the high uplink signalling overhead problem is tackled through the design and development of a set of novel Channel Quality Indicator (CQI) feedback reduction schemes. These are based on a User Equipment (UE)-assisted predictive filtering technique and a CQI clustering scheme respectively, where the latter is able to achieve an uplink signalling feedback reduction of 88.2%. Moreover, a cross-layer depth-texture bit rate allocation estimation technique and an enhanced depth map rate control scheme aimed at improving the synthesised view quality is proposed. Furthermore, a content-aware scheduling algorithm based on the widely used modified largest weighted delay first (M-LWDF) packet scheduling scheme is designed and tested in conjunction with the combined CQI feedback reduction schemes mentioned above. Whilst, the content-aware scheduling scheme yields an improvement in both the system performance and visual quality metrics, the use of the feedback reduction schemes has a detrimental effect on the visual quality. For this reason, a lean cross layer technique is designed to adapt the CQI feedback by soliciting CQI reports from individual UEs. This solution has not only improved the texture and synthesised view Peak Signal-to-Noise Ratio (PSNR) quality, approaching that of the content-aware M-LWDF scheme without any CQI feedback reduction applied, but also achieves an uplink feedback signalling overhead reduction of 84.1%.

# Acknowledgments

I would like to thank Prof. Ing. Carl J. Debono for his persistent guidance and prompt availability throughout the course of this study.

My deepest gratitude goes to my parents who instilled a sense of curiosity and persistence. Finally I am also indebted to my wife Anabel for her unwavering support and encouragement in this journey.

# List of Figures

# List of Tables

# List of Publications

1. M. Cordina, C.J. Debono, “A CQI Feedback Clustering Technique for Signaling Traffic Reduction During Crowd Events in 3GPP LTE Systems,” Proc. of the 2017 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'17), October 2017.

2. M. Cordina, C.J. Debono, “Robust Predictive Filtering Schemes for Sub-band CQI Feedback Compression in 3GPP LTE Systems,” IET Communications, vol. 11, no. 11, pp. 1797 - 1807, September 2017.

3. M. Cordina, C.J. Debono, “A Support Vector Machine Based Sub-band CQI Feedback Compression Scheme for 3GPP LTE Systems,” Proc. of the 14th International Symposium on Wireless Communication Systems (ISWCS 2017), pp. 325 - 330, August 2017.

4. M. Cordina, C.J. Debono, “A Cross-layer MV-HEVC Depth-Texture Rate Allocation Estimation Technique in 3GPP LTE Systems,” Proc. of the IEEE International Conference on Multimedia and Expo (ICME 2017), pp. 19 - 24, July 2017.

5. M. Cordina, C.J. Debono, “A Depth Map Rate Control Algorithm for HEVC Multi-view Video plus Depth,” Proc. of the IEEE International Conference on Multimedia and Expo (ICME 2016) - 22nd International Packet Video Workshop (PV), July 2016.

6. M. Cordina, C.J. Debono, “An Adaptive Lagrange Multiplier Technique for Multi-view Video plus Depth Coding,” Proc. of the 2013 Picture Coding Symposium (PCS 2013), pp. 249 - 252, December 2013.

7. M. Cordina, C.J. Debono, “An Adaptive Texture-Depth Rate Allocation Estimation Technique for Low Latency Multi-view Video Plus Depth

Transmission," Proc. of the IEEE Visual Communications and Image Processing Conference (IEEE VCIP 2013), November 2013.

8. M. Cordina, C.J. Debono, "A Novel View-level Target Bit Rate Distribution Estimation Technique for Real-time Multi-view Video plus Depth," Proc. of the 2012 IEEE International Conference on Multimedia and Expo (ICME 2012), pp. 878 - 883, July 2012. - **Top 15% paper**

# List of Abbreviations

| | |
|---|---|
| AGL | Above Ground Level |
| AM | Acknowledge Mode |
| ARQ | Automatic Repeat reQuest |
| AU | Access Unit |
| CABAC | Context Adaptive Binary Arithmetic Coding |
| CAGR | Compound Annual Growth Rate |
| CP | Cyclic Prefix |
| CQI | Channel Quality Indicator |
| CRUX | Customer Received Usage Experience |
| CSV | Conventional Stereo Video |
| CTB | Coding Tree Block |
| CTC | Common Test Conditions |
| CTU | Code Tree Units |
| CUs | Coding Units |
| DCT | Discrete Cosine Transform |
| DIBR | Depth Image Based Rendering |
| DMM | Depth Modelling Modes |
| DRB | Dedicated Radio Bearers |
| DSCP | Differentiated Service Code Point |
| EESM | Exponential ESM |
| EPC | Evolved Packet Core |
| EPS | Evolved Packet System |
| ESM | Effective SINR mapping |
| FDS | Frequency Domain Scheduling |
| FEC | Forward Error Correction |
| FFT | Fast Fourier Transform |
| FMO | Flexible Macro-block Ordering |

| | |
|---|---|
| FN | Free Navigation |
| FVV | Free-Viewpoint Video |
| GOP | Group Of Pictures |
| GP | Gaussian Process |
| HARQ | Hybrid ARQ |
| HD | High Definition |
| HEVC | High Efficiency Video Coding |
| IP | Internet Protocol |
| ISI | Inter-Symbol Interference |
| ISO/IEC | International Organisation for Standardisation / International Electrotechnical Commission |
| ITU-T | International Telecommunication Union |
| JCT-VC | Joint Collaborative Team on Video Coding |
| LCU | Largest Coding Unit |
| LMS | Least Mean Square |
| LTE | Long Term Evolution |
| MAC | Medium Access Control |
| MAD | Mean Absolute Difference |
| MANE | Media Aware Network Element |
| MCS | Modulation and Coding Scheme |
| MEC | Mobile Edge Computing |
| MIMO | Multiple-Input Multiple Output |
| MME | Mobility Management Entity |
| MNO | Mobile Network Operators |
| MPEG | Motion Picture Experts Group |
| MTU | Maximum Transmission Unit |
| MV | Multi-View |
| MVC | Multi-View Coding |
| MVD | Multi-View Video plus Depth |
| MV-HEVC | Multi-view High Efficiency Video Coding |
| MVV | Multi-View Video |
| NAL | Network Abstraction Layer |
| NLMS | Normalised Least Mean Square |

| | |
|---|---|
| Non-GBR | Non-Guaranteed Bit Rate |
| OFDMA | Orthogonal Frequency Division Multiple Access |
| PAPR | Peak-to-Average Power Ratio |
| PDCB | Packet Data Convergence Protocol |
| PDU | Packet Data Units |
| PFS | Predictive Filtering Scheme |
| P-GW | Packet Data Network Gateway |
| PLR | Packet Loss Rate |
| PMI | Pre-coding Matrix Indicator |
| POC | Picture Order Count |
| PPS | Picture Parameter Set |
| PRB | Physical Resource Block |
| PSNR | Peak Signal-to-Noise Ratio |
| PSO | Particle Swarm Optimisation |
| PTI | Pre-coding Type Indicator |
| PUs | Prediction Units |
| QCI | QoS Class Indicators |
| QoS | Quality of Service |
| QP | Quantisation Parameter |
| RBSP | Raw Byte Sequence Payload |
| RE | Resource Element |
| RI | Rank Indicator |
| RLC | Radio Link Control |
| RLS | Recursive Least Square |
| RoHC | Robust Header Compression |
| ROI | Region Of Interest |
| RRC | Radio Resource Control |
| RS | Reference Signals |
| RSRP | Reference Signal Received Power |
| RTP | Real Time Protocol |
| SAW | Stop-And-Wait |
| SC-FDMA | Single Carrier Frequency Division Multiple Access |
| SD | Standard Definition |

| | |
|---|---|
| SDN | Software Defined Network |
| SDU | Service Data Units |
| SINR | Signal-to-Interference-and-Noise Ratio |
| SM | Super Multi-View |
| SPS | Sequence Parameter Set |
| SRB | Signalling Radio Bearers |
| SRST | Single Media Transport |
| TB | Transport Blocks |
| TM | Transparent Mode |
| TTI | Transmission Time Interval |
| TUs | Transform Units |
| TxD | Transmit Diversity |
| UDP | User Datagram Protocol |
| UE | User Equipment |
| UHD | Ultra High Definition |
| UM | Unacknowledged Mode |
| URQ | Uniform Rate Quantisation |
| VCEG | Video Coding Experts Group |
| VCL | Video Coding Layer |
| VoIP | Voice over IP |
| VPS | Video Parameter Set |
| VSP | View Synthesis Prediction |

# Table of Contents

# Chapter 1 Introduction

## 1.1 Context and Motivation of the work

The past decade has seen a rapid development in digital multimedia systems owing to the increasingly demanding consumer needs of higher resolution, improved realism and better interactivity. Whilst video quality has progressively improved from the legacy standard definition (SD) to 4K/8K ultra high definition (UHD) format, the deployment of 3DTV (3-dimension television), aimed at improving realism through the introduction of depth perception, experienced a slowdown within the consumer electronics industry in recent years [1]. This is generally attributed to the fact that earlier systems allowed depth perception from a fixed set of viewpoints whilst viewers are required to wear glasses without any form of interactivity. However, the advent of auto-stereoscopic displays coupled with advances in the multi-view video plus depth (MVD) representation and coding standards reinvigorated the interest in immersive technology [2]. In fact, the use of immersive video techniques has already captured the interest of the video advertising industry which estimated an increase of 46% in advertisement views when using immersive video technology compared to the standard video technology [3]. Meanwhile, an application which is fast gaining traction in recent years and continues to drive interest in immersive technology is free-viewpoint video (FVV) [4, 5].

### 1.1.1 Free-Viewpoint Video

In an FVV system based on the MVD representation, viewers have the facility to interactively navigate a scene by selecting either existing or novel viewpoints rendered from a small set of texture and depth map views using a view synthesis technique [6]. Such an application has not only attracted consumers seeking a new, interactive and immersive experience but has also garnered commercial interest in sporting and other entertainment events [7, 8].

The coding of the MVD representation plays a pivotal role in the FVV end-to-end transmission chain and has received significant research attention. This led the standardisation bodies, namely the International Organisation for

Standardisation/International Electrotechnical Commission (ISO/IEC) Motion Picture Experts Group (MPEG) and the International Telecommunication Union (ITU-T) Video Coding Experts Group (VCEG), to jointly develop and release multi-view video coding (MVC) standards starting with H.264/MVC in 2009 and subsequently H.265/Multi-view High Efficiency Video Coding (MV-HEVC) and H.265/3D-HEVC in 2014 and 2015 respectively. Both MV-HEVC and 3D-HEVC can effectively be used in the compression of the MVD representation, however, the latter achieves better coding efficiency, owing to the introduction of coding tools specifically designed for this representation, at the expense of increased computational complexity. Meanwhile, the research and commercial interest in FVV and similar immersive technologies is expected to grow further as the standardisation bodies are preparing for a "beyond H.265" standard featuring super multi-view (SM) and free navigation (FN) coupled with 360-degree video applications to be released by 2020 [9].

### 1.1.2 Mobile Data Capacity Crunch

According to the Cisco Visual Networking Index Forecast report published in September 2017 [10], global mobile data traffic, shown in Figure 1.1, is expected to reach 17EB per month in 2018 and increase to 48EB per month by 2021 with a compound annual growth rate (CAGR) of 42%. Meanwhile, in 2016, mobile video traffic accounted for 60% of total mobile data traffic and is expected to increase to 78% by 2021. This surge in mobile video data traffic, together with the introduction of FVV in the mobile ecosystem is expected to compound the pressure on mobile network operators (MNO) to deliver a superior customer-received usage experience (CRUX) [11], given the capacity constraints of current 4$^{th}$ generation Long Term Evolution (LTE) wireless cellular networks.

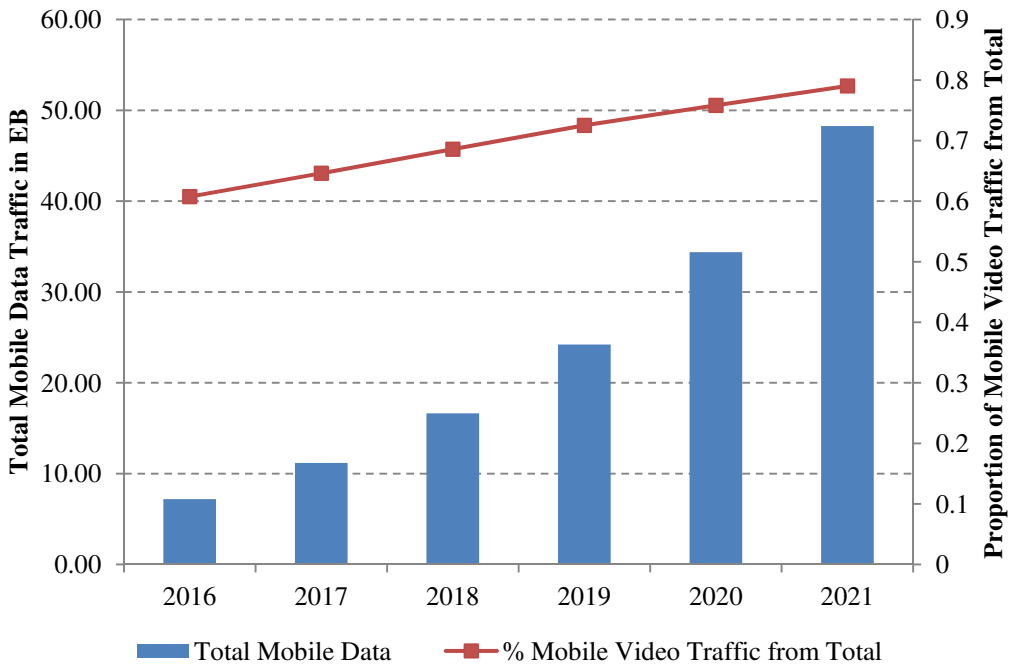



Figure 1.1: Mobile data traffic and proportion of mobile video traffic

Meanwhile, apart from this expected growth, MNO need to design and optimise radio access networks for crowd events such as sports events, mass entertainment gatherings and public demonstrations. These scenarios are characterised by a large number of subscribers, usually in the thousands, confined in a relatively small geographical area thereby leading to a high traffic density zone. With reference to the traffic profile collected from a live LTE commercial network of a radio sector covering a crowd event scenario, shown in Figure 1.2, crowd events exhibit a significantly higher proportion of uplink data traffic relative to the total traffic, when compared to a normal business day, due to the higher user propensity to share multimedia content. In addition to this high uplink data traffic, each user equipment (UE) in the radio sector is required to send channel quality indicator (CQI) feedback reports to the eNB, to support its scheduling functions, leading to high uplink signalling overhead. In order to address this problem and avoid the risk of uplink control channel congestion, MNOs typically increase the resources for the uplink control channel at the expense of lower uplink user data channel capacity thereby leading to lower uplink throughputs. However, this runs counter with the fact that the traffic profile in such events is uplink dominated. Thus, it is necessary to design low complexity CQI feedback reduction techniques such that precious uplink resources for the user data are not sacrificed, particularly in crowd event scenarios.

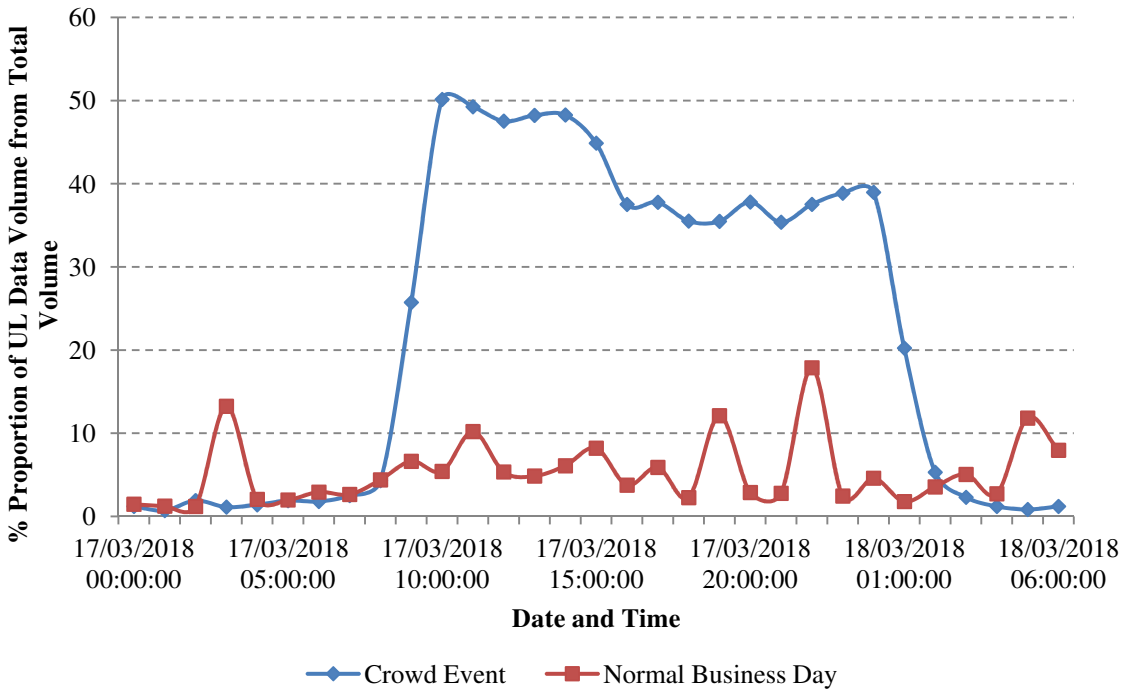



Figure 1.2: Proportion of uplink data volume from total - Normal business day vs. Crowd event traffic profile

As crowd events constitute an attractive use case for the deployment of FVV systems, the study of low latency MVD video transmission over an LTE network in such scenarios is particularly relevant. Apart from the typical challenges associated with video transmission over wireless networks, the relatively high number of active users per radio sector coupled with the limited channel state information impinge on the key quality of service parameters, namely delay and packet loss. Hence, the use of content-awareness in both the packet scheduling and CQI feedback reduction techniques is essential.

Meanwhile, considering the channel bandwidth constraints of commercial LTE radio access networks, the video encoders are generally equipped with a rate control scheme to ensure that the encoded bit stream abides with strict target bitrates. In the case of MVD video content, rate control needs to be applied across both the texture and depth map views. Although the depth map views are strictly used for view rendering, encoding these views at a low bit rate results in a significant degradation of the synthesised view quality [12]. Moreover, the same rate control scheme is generally applied to both the texture and depth map views without any consideration of the depth map characteristics. Thus it is imperative that the depth-texture view bit allocation and rate control for the depth map views are tuned appropriately.

## 1.2 Main Objectives of this work

The focus of this work is on the mitigation of the high uplink signalling overhead present in crowd event scenarios, whilst minimising view quality distortion, through the use of lean cross-layer design for the transmission of low latency MVD video content over an LTE network. To this end, the main objectives for this work are:

- Develop a novel, low complexity scheme to reduce the uplink signalling overhead present in crowd event scenarios thereby leaving more resources for the uplink user data traffic.
- Develop a low complexity technique to estimate the optimal depth-texture view bit allocation for the transmission of low latency MVD video that improves the synthesised view quality.
- Exploit depth-map characteristics to enhance the standard rate control scheme.
- Study the performance of standard Quality of Service (QoS)-aware packet scheduling schemes for the transmission of MVD video in crowd event scenarios and explore enhancements that can be applied to such schemes with the aim of improving the quality of the delivered MVD video.
- Establish the impact of CQI feedback reduction scheme on both the system level performance and quality of the delivered MVD video and explore the use of cross-layer design to mitigate any video quality degradation introduced by the CQI feedback reduction scheme.

## 1.3 Contribution of this work

The main contributions of this work are split into three main areas:

1. The high uplink signalling overhead present in crowd event scenarios reduces the capacity for uplink user data traffic. This problem is addressed through:
    - A novel UE-assisted predictive filtering scheme based on the LMS and NLMS adaptive filters located in both the UE and eNB;

- A reduced complexity UE-assisted predictive filtering scheme able to reduce computational load by up to 35% when compared to the LMS based predictive filtering scheme;
- A novel CQI clustering coupled with a cluster leader election technique exploiting the slow varying radio channel conditions experienced by static users.

2. The bit rate allocation for the depth map and texture views rate control algorithm, in the respective encoders, is either set to a fixed proportion of the total available bit rate or estimated through the use of techniques based on pre-encoding or pre-estimation of model parameters, making it unsuitable for real time encoding. Moreover, the rate control scheme applied to the texture views is typically re-used for the depth map views without consideration of the depth map characteristics. The contribution in this area are:

    - A novel cross-layer MV-HEVC depth-texture bit rate allocation estimation technique based on a formulated statistical model using both the image characteristics and the total available bit rate for the MVD.
    - An enhanced R-λ model based rate control scheme for the depth map views using a refined key frame bit allocation and R-λ model coupled with an adaptive Lagrange multiplier and QP clipping function.

3. The use of the CQI feedback signalling reduction schemes result in degradation of the MVD video quality thereby affecting the user experience. Moreover, the standard QoS-aware packet scheduling schemes for the transmission of MVD video in a crowd event scenario were found to offer the same level of performance. The contribution in this area are:
    - A content-aware packet scheduling scheme based on a well established QoS-aware packet scheduling scheme utilising RLC queue based weighted factors coupled with an MVD packet prioritisation scheme.
    - A novel CQI feedback reduction scheme based on the combined use of CQI clustering and UE-assisted predictive filtering scheme.

- A novel lean cross-layer technique allowing the eNB to solicit CQI feedback from individual UEs thereby adapting the CQI feedback based on high priority packet loss rate and weighted head of line packet delay metrics.

## 1.4 Chapter Organisation

With reference to Figure 1.3, this work is organised into three main sections with each section divided into two chapters.

The first section focuses on the design, implementation and testing of novel CQI feedback reduction schemes. This starts with chapter 2 giving a detailed technical review of the key elements in an LTE radio access network followed by a literature review of CQI feedback signalling reduction techniques. The new CQI feedback reduction schemes are then presented in chapter 3.

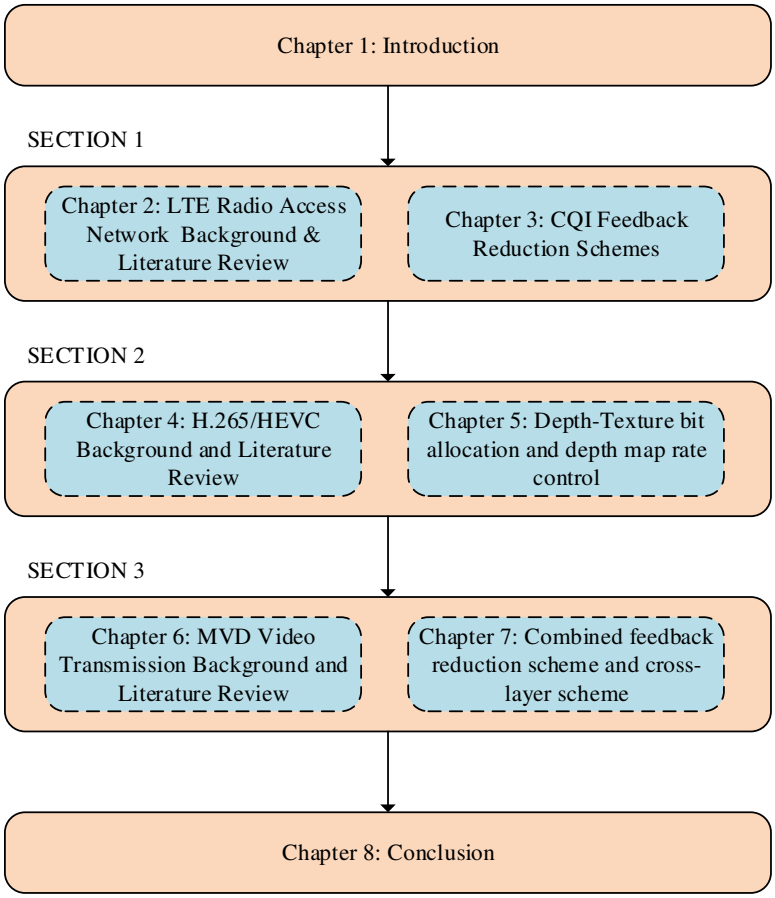



Figure 1.3: Chapter Organisation

The second section focuses on the MVD video encoder residing in the application layer. This section starts with chapter 4 giving an overview of the latest standard video encoder, namely High Efficiency Video Coding (HEVC) together with its Multi-View (MV)/3D extensions and proceeds to a review of work aimed at the depth-texture view bit allocation problem and rate control schemes. In chapter 5, we introduce a new cross-layer MV-HEVC based depth-texture bit rate allocation estimation scheme for an FVV application aimed to maximise the rendered view quality. Cognizant of the fact that the rate control scheme used for texture is generally utilised for the depth map views, we explore potential improvements in this area and propose an enhanced depth map rate control scheme to improve the rendered view quality. Both the cross-layer depth-texture bit rate allocation estimation scheme and the enhanced depth map rate control scheme, introduced in this section, are independent of the CQI feedback reduction schemes presented in the first section.

The third section focuses on the MVD video transmission over LTE network in a crowd event scenario where the CQI feedback reduction schemes and the optimal depth-texture bit allocation technique, presented in the previous sections, are combined with the proposed content-aware scheduling scheme. This starts with chapter 6 giving a comprehensive overview of the video bit stream packetisation coupled with the relevant packet scheduling algorithms and cross-layer schemes. In chapter 7, we first establish the performance of several quality of service (QoS) aware packet scheduling algorithms in a crowd event scenario followed by the introduction of an enhanced content-aware scheduling scheme based on a well established packet scheduling algorithm. This is then followed by a study to establish the impact of a new combined CQI clustering and predictive filtering scheme on both system performance metrics and downlink transmitted view quality. In order to alleviate the impact of the proposed scheme, a cross-layer technique is utilised to adapt the CQI feedback with the aim of improving the system level performance metrics and minimise view quality degradation, whilst reducing the uplink signalling overhead and yielding additional capacity for the uplink user data. The improvement achieved by this solution has been verified using both objective and subjective testing methodologies.

Finally, chapter 8 provides a summary of the key results presented in this thesis together with an outlook of potential future research avenues.

# Chapter 2 LTE Radio Access Network

This chapter presents a technical review of the LTE radio access network together with the background necessary for the design and development of the new CQI feedback signalling reduction schemes. It gives:

- A description of the LTE radio protocol architecture including radio resource management and link adaptation;
- An overview of the channel feedback report types together with a description of the CQI feedback process and signalling reduction techniques.

## 2.1 LTE Radio Protocol Architecture

The LTE radio protocol architecture, shown in Figure 2.1, resides in both the UE and E-UTRAN. As shown in Figure 2.1, the protocol stack is split into two; namely the control plane and the user plane. The user plane transports the IP data packets generated at the application layer to and from the evolved packet core (EPC) whilst the control plane transports both access stratum (between UE and eNB) and non-access stratum (between UE and EPC) signalling.

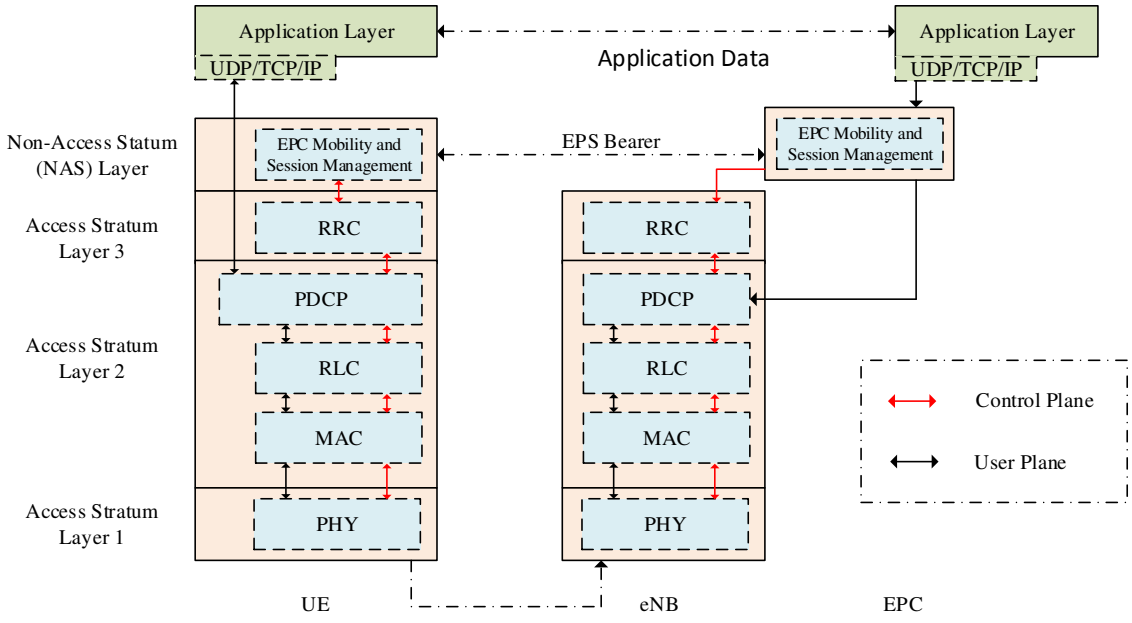



Figure 2.1: LTE Radio Protocol Architecture

The radio resource control (RRC) management entity, residing at layer 3 of the access stratum protocol stack, manages all the signalling between the UE and the eNB including the setup of dedicated radio bearers (DRB) for data transmission [13]. Moreover, it also facilitates the transport of non-access stratum signalling through signalling radio bearers (SRB) established between the UE and the eNB. Both the signalling and user plane traffic in the access stratum flows through the layer 2 entities made up of the packet data convergence protocol (PDCP), radio link control (RLC) protocol and the medium access control (MAC) protocol prior to transmission over the air by the physical layer.

### 2.1.1 PDCP Layer

The functions of the PDCP layer, shown in Figure 2.2, include the implementation of a sequence numbering scheme to aid in-sequence delivery and missing/duplicate packet detection [14]. Moreover, security features such as integrity protection and ciphering are implemented in the PDCP layer to ensure secure delivery of data packets over the air interface. In addition to these functions, the PDCP layer performs robust header compression (RoHC) [15] for the transport of Internet Protocol (IP) / User Datagram Protocol (UDP) / Real Time Protocol (RTP) traffic, such as Voice over IP (VoIP), thereby improving the air interface efficiency.

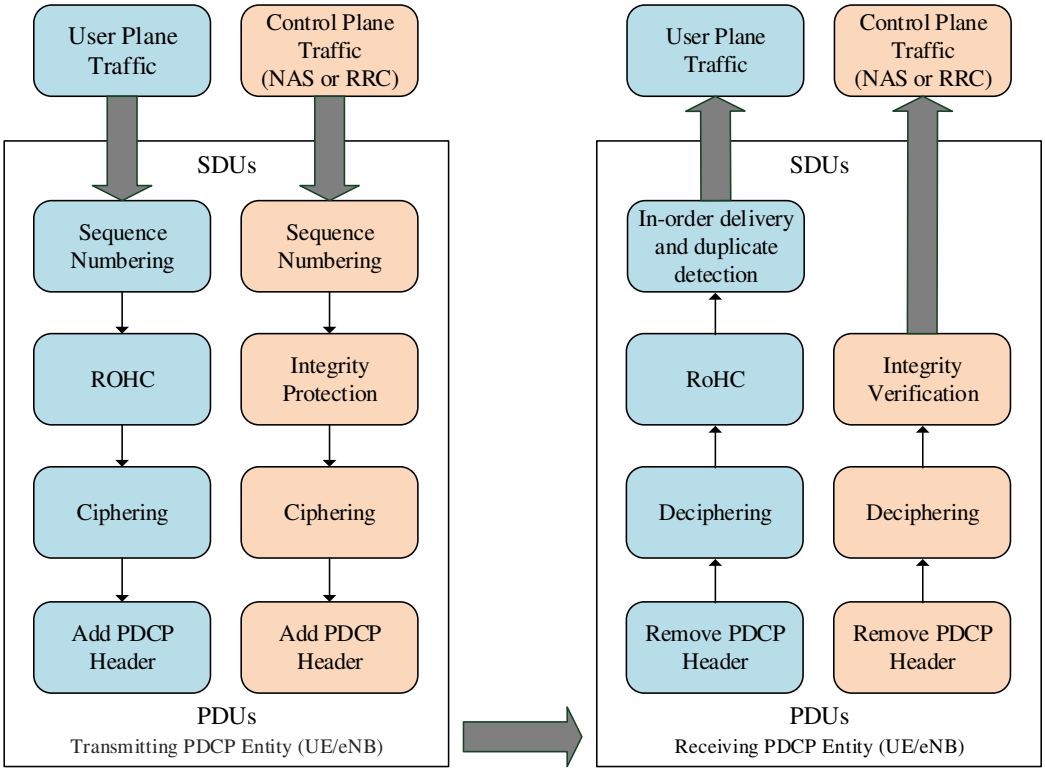



Figure 2.2: PDCP Layer Operation

### 2.1.2 RLC Layer

The RLC layer receives service data units (SDU) from the PDCP entity and transforms them into RLC packet data units (PDU) by performing segmentation and concatenation depending on the RLC mode of operation [16]. This process is conducted in order to fit the RLC SDUs into the transport blocks (TB) available in the MAC layer based on the underlying physical channel. The RLC protocol offers three different modes of operation namely Transparent Mode (TM), Unacknowledged Mode (UM) and Acknowledge Mode (AM). The RLC TM mode of operation, shown in Figure 2.3, is simply a transmission buffer to store the RLC SDUs until a transmission opportunity is signalled from the MAC layer. This is strictly used to transport common signalling channels responsible for paging, system information block transmission, or initial RRC connection establishment [17].

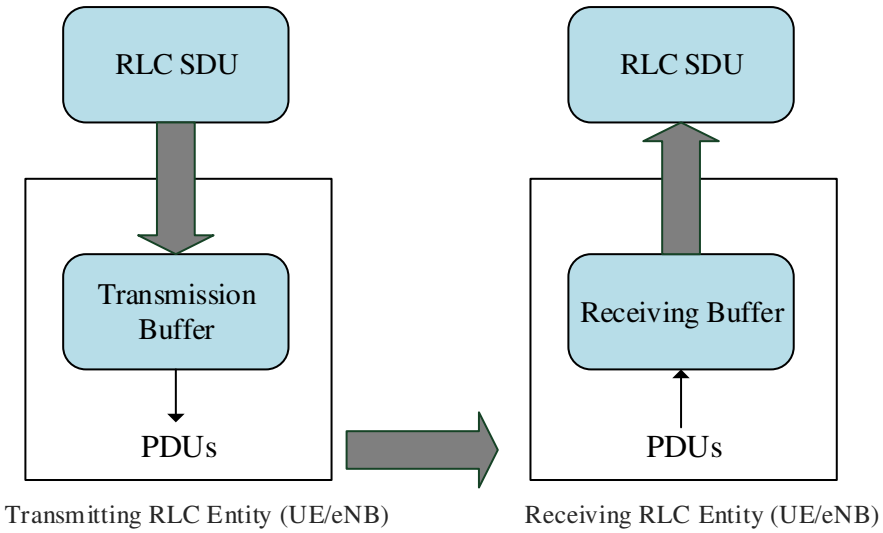



Figure 2.3: RLC TM Operation

As shown in Figure 2.4, both RLC UM and RLC AM implement segmentation and concatenation coupled with re-ordering and duplicate detection of RLC PDUs. In addition, the RLC AM operates an Automatic Repeat reQuest (ARQ) re-transmission mechanism to recover any missing PDUs over the air interface. These modes can be used for both control and user plane RLC PDUs whilst the mapping of the corresponding SRB or DRB is done by the eNB using the QoS Class Indicators (QCI). This allows the RLC layer to tailor its services to the radio bearers based on the QoS requirements.

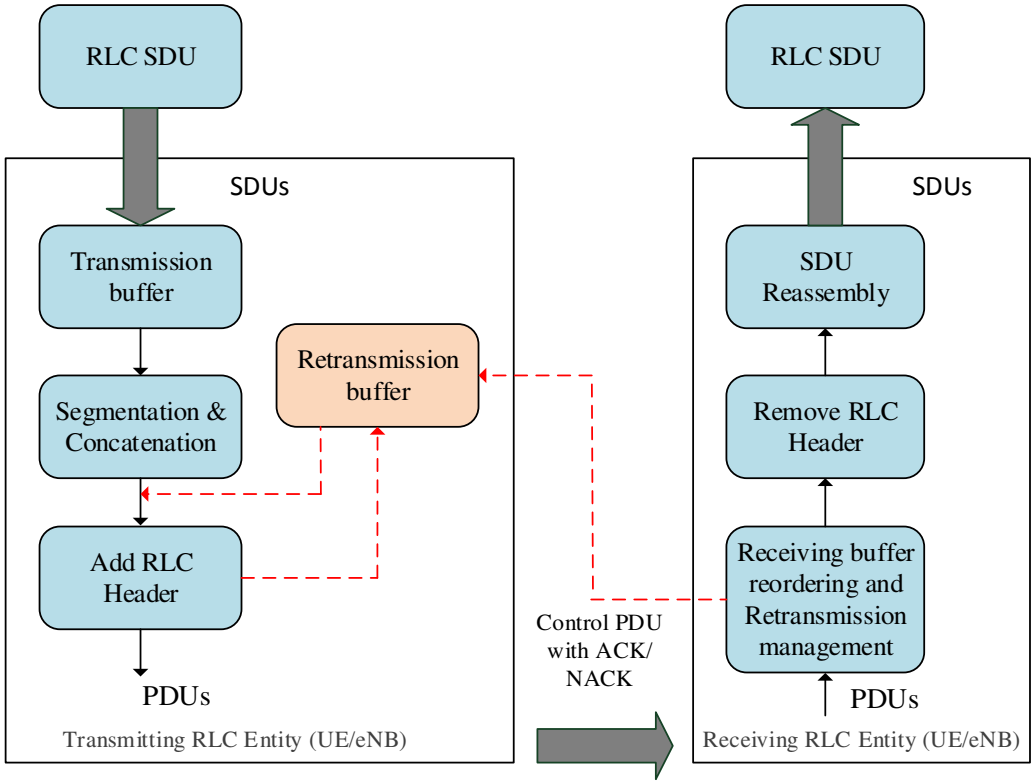



Figure 2.4: RLC AM/UM Operation (AM operational elements shown in red)

### 2.1.3 MAC Layer

The MAC layer takes a central role within the LTE radio protocol stack. Apart from mapping the logical channels, carrying RLC PDUs, to the transport channels and multiplexing multiple radio bearers for delivery by the physical layer [18], the MAC layer implements some of the key radio resource management functions. These include packet scheduling, link adaptation and hybrid ARQ (HARQ).

With reference to Figure 2.5, the packet scheduler exploits the time and frequency selective fading, common in wireless environments, to dynamically distribute the available resources among UEs having data in the eNB RLC buffers and awaiting scheduling. Such scheduling decisions are carried out at every transmission time interval (TTI), which is set to 1ms in LTE, and involves selecting the users to be scheduled together with the amount of resources to be assigned depending upon their QoS requirements.

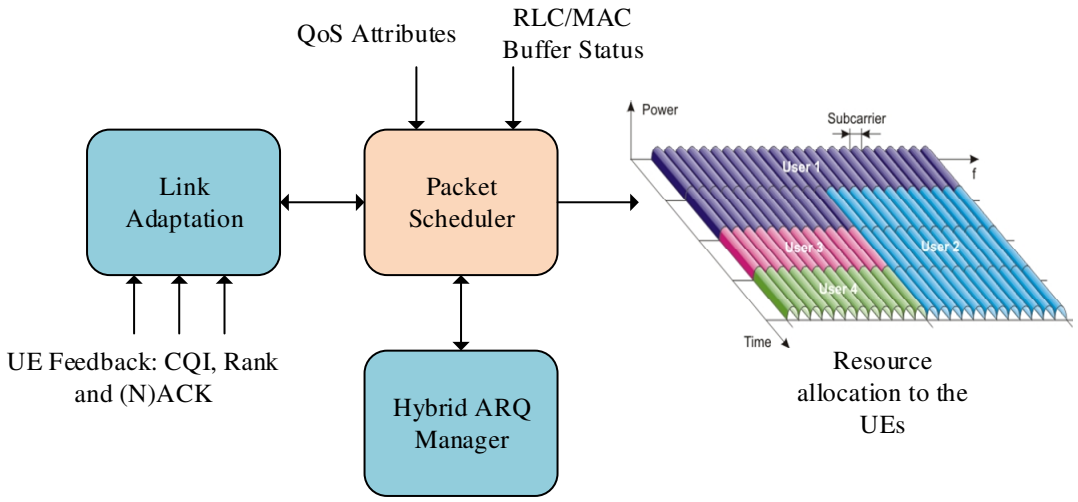



Figure 2.5: MAC Layer Functionality

The received signal level and quality is affected by several impairments inherent to wireless environments including multipath propagation, path loss and interference. In order to counter these affects the link adaptation entity selects the transmission parameters, such as modulation and coding scheme (MCS) and multiple-input multiple output (MIMO) transmission rank and pre-coding, to match the channel conditions on the resources allocated by the scheduler [19]. In order for the packet scheduler and link adaptation entities to gain channel awareness, the UE sends channel feedback information to the eNB including a Channel Quality Indicator (CQI) report coupled with a Pre-coding Matrix Indicator (PMI) report, Pre-coding type indicator (PTI) and Rank Indicator (RI) depending on the MIMO transmission mode and eNB configuration [20].

In order to further improve system performance and increase robustness against link adaptation errors, LTE implements a hybrid ARQ protocol which offers a fast, layer 1 retransmission mechanism. HARQ effectively combines forward error correction (FEC) with an ARQ protocol such that transport blocks received in error are used in FEC decoding rather than simply discarded. The HARQ retransmission protocol implemented in LTE is based on an 8-process Stop-And-Wait (SAW) protocol between the UE and eNB and the receiver uses ACK/NACK messages to inform the transmitter whether the current transport block was received correctly or otherwise. LTE supports two HARQ combining methods, namely [21]:

- Chase Combining where each retransmission is an exact copy of the data first transmitted allowing time diversity and soft combining gain at a low complexity;
- Incremental Redundancy, where additional redundant information is included with each re-transmission allowing the effective coding rate to increase based on the number of retransmissions.

### 2.1.4 PHY Layer

As opposed to previous generations of mobile communication systems, LTE employs Orthogonal Frequency Division Multiple Access (OFDMA) for the downlink whilst the uplink is based on Single Carrier Frequency Division Multiple Access (SC-FDMA). In an OFDMA transmitter the high bit rate data to be transmitted is first split into several parallel low bit rate data streams. Each of these streams is then modulated on a set of narrow, mutually orthogonal, sub-carriers using a conventional modulation scheme [22]. In LTE, the sub-carrier spacing is set to 15 kHz whilst the number of sub-carriers varies depending on the overall channel bandwidth. The process of splitting the high bit rate data into multiple low bit rate data streams results in a larger OFDM symbol duration thereby partially counteracting the effect of inter-symbol interference (ISI), caused by multi-path delay spread, whilst reducing error rates. Nevertheless, LTE further employs the use of a cyclic prefix (CP) to avoid ISI whereby a posterior portion of the symbol is copied and prepended to the symbol.

Apart from the good spectral properties and the handling of multiple bandwidths, the implementation of OFDMA is made relatively efficient through the use of the Fast Fourier Transform (FFT). Moreover, a key attribute of OFDMA is that it allows the scheduler in the base station transmitter to allocate users to different sub-carriers thereby exploiting the time-frequency selective nature of wireless channels and increasing robustness. Nevertheless, OFDMA presents a number of challenges namely [23]:

- Sensitivity to frequency offsets caused by hardware imperfections and Doppler shift in high mobility scenarios;

- High Peak-to-Average Power Ratio (PAPR) of the transmitted signal requiring the use of high linearity power amplifiers operating with a large power back-off leading to poor power efficiency. This is the key reason why 3GPP opted to use OFDMA in the downlink direction whilst using the power efficient SC-FDMA in the uplink direction.

In addition to OFDMA, LTE uses multiple transmit and receive antennas, at both the eNB and UE side, in order to exploit the space-time diversity, present in a multipath rich environment, giving rise to spatial multiplexing. This creates multiple parallel data transmission paths which are orthogonal in the space-time domain and thus cause minimal interference to each other [17].

With reference to Table 2.1, 3GPP LTE specifies a number of downlink MIMO transmission modes (TM) exploiting several MIMO concepts. Each transmission mode defines its own channel state information feedback requirements leading to a complex list of configuration options [20]. Nevertheless, transmission modes 3 and 4 utilising a 2 x 2 MIMO antenna configuration represent the most commonly deployed setup in live commercial LTE networks due to antenna size, cost and performance considerations [17]. Both TM 3 and 4 are usually operated with an adaptive transmission mode algorithm which allows the eNB to reconfigure the MIMO TM during a data session based on the channel feedback reports sent by the UE. Considering TM3, in good radio conditions with a rich multipath environment the eNB uses spatial multiplexing to maximise the user data throughput whilst in cell edge conditions or poor multipath environment, the eNB switches to a transmit diversity (TxD) mode (TM2) to exploit diversity and increase robustness.

| Downlink transmission mode TM | Key Characteristics | UE Feedback Type | 3GPP Release |
|---|---|---|---|
| 1 | • Single antenna port transmission | CQI | 8 |
| 2 | • Transmit Diversity (TxD).<br>• Transmit the same information in different layers through different antennas.<br>• Increases robustness in low SINR conditions. | CQI | 8 |
| 3 | • Open loop Spatial Multiplexing (OLSM).<br>• Transmit different information in different layers through different antennas.<br>• Increases throughput in high SINR conditions. | CQI, RI | 8 |
| 4 | • Closed loop Spatial Multiplexing (CLSM).<br>• Optimises the precoding process by taking into account the current channel conditions for each layer as reported by the UE through the PMI. | CQI, RI, PMI | 8 |
| 5 | • Multi-User MIMO (MU-MIMO)<br>• Increases overall cell capacity rather than the throughput of a single user.<br>• Spatial multiplexing is used to transmit two codewords belonging to different users at the same time. | CQI, PMI | 8 |
| 6 | • Single-User MIMO (SU-MIMO)<br>• Similar to CLSM but uses a single codeword (CLSM Rank-1 scheme) and aimed at enhancing robustness. | CQI, PMI | 8 |
| 7 | • Beamforming (referred to as "virtual" antenna port 5).<br>• Uses UE specific reference signal. | CQI | 8 |
| 8 | • Combined beamforming and MIMO up to 2 antennas.<br>• Combine MIMO and beamforming techniques such that LTE signal is directed towards the addressed UE. | CQI, RI, PMI | 9 |
| 9 | • Combined beamforming and MIMO up to 8 antennas.<br>• Evolution over TM8 adding the support of MIMO 8x8 in combination with beamforming. | CQI, RI, PMI/PTI | 10 |
| 10 | • Combined beamforming and MIMO up to 8 antennas with Coordinated Multi-point Transmission (CoMP).<br>• Similar to TM9 with the addition of CoMP support whereby antennas may be physically in different locations. | CQI, RI, PMI/PTI | 11 |

Table 2.1: Downlink Transmission modes for 3GPP LTE [20]

## 2.2 LTE Radio Frame Structure and Physical Channels

The FDD LTE radio frame structure together with a definition of a physical resource block (PRB) and resource element (RE) are shown in Figure 2.6. The radio frame is composed of 20 slots, each lasting 0.5ms. As discussed in Section 2.1.4, the scheduler allows multiple users to receive data traffic at the same time by allocating users in

different parts of the radio channel. In LTE, this allocation is done on a PRB basis, each consisting of 12 sub-carriers, resulting in a minimum bandwidth allocation of 180 kHz, and lasts for one slot. A PRB has six or seven OFDM symbols, depending on whether an extended or normal cyclic prefix is configured [17] whilst a resource element (RE), defined as one subcarrier lasting one symbol, can carry reference signals (RS), modulated information (user plane or control plane traffic) or nothing (muted).

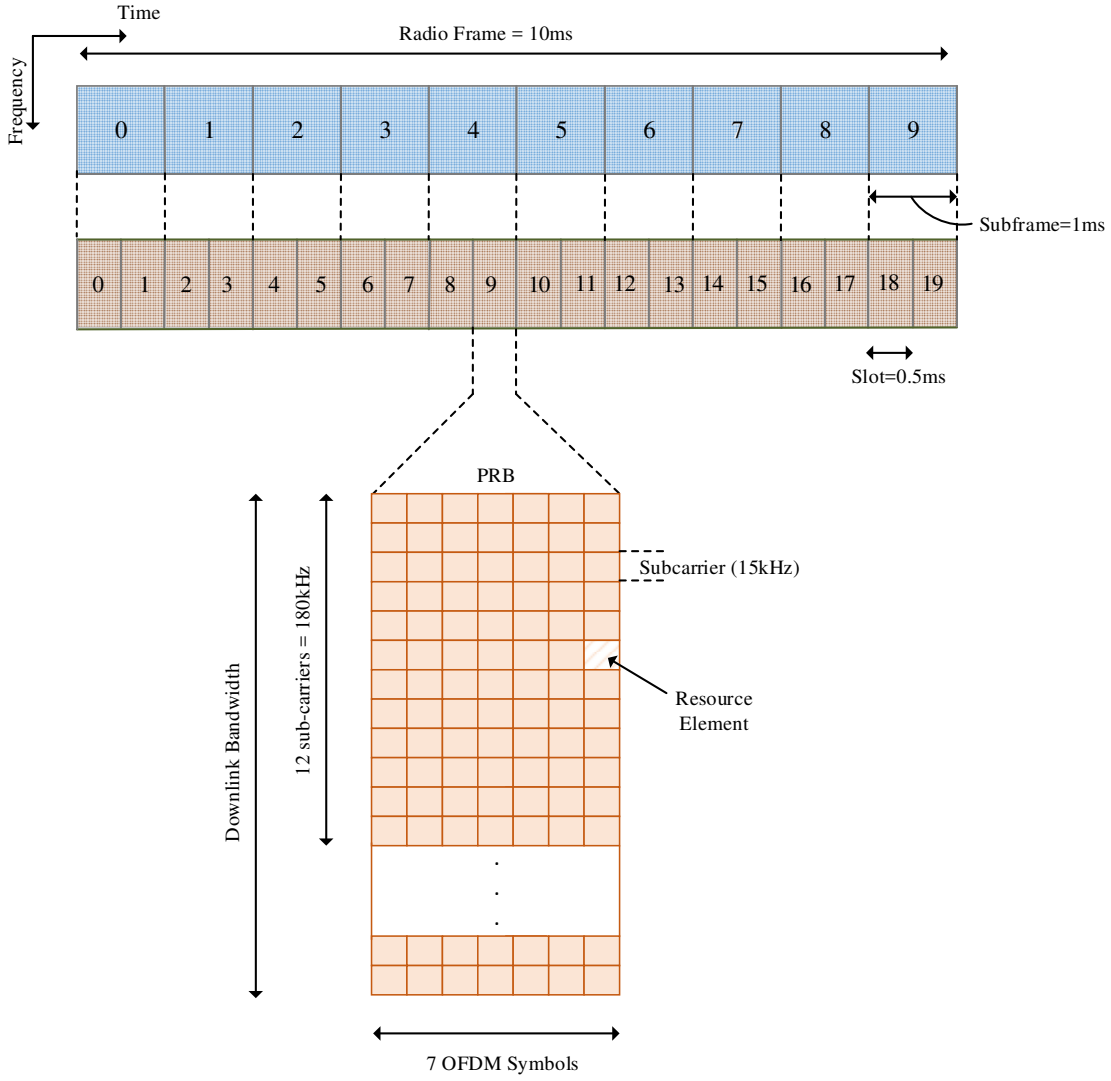



Figure 2.6: LTE Radio frame structure, Physical Resource block and Resource element

The transport channels at the MAC layer are mapped onto several physical channels depending on their transmission direction (uplink (UL) / downlink (DL)) and use as shown in Table 2.2.

| Physical Layer Channel | Direction | Functionality |
|---|---|---|
| Physical Broadcast Channel (PBCH) | DL | Transports RRC broadcast messages |
| Synchronisation Channel (SCH) | DL | Used to identify Cell ID, frame and slot timing |
| Downlink Reference Signal (DL-RS) | DL | Used for cell signal quality estimation |
| Physical Downlink Shared/Control Channel (PDSCH/PDCCH) | DL | Transports the user plane and control plane data to UEs |
| Physical control format indicator channel (PCFICH) | DL | Indicates to the UE the number of OFDM symbols used for the PDCCHs |
| Physical Hybrid-ARQ Indicator Channel (PHICH) | DL | Transports HARQ (N)ACK for uplink data transfers |
| Physical Uplink Shared/Control Channel (PUSCH/PUCCH) | UL | Transports user plane and control plane data to the eNB |
| Demodulation Reference Signal (DM-RS) | UL | Channel estimation for uplink coherent demodulation/detection of the uplink control and data channels |
| Sounding Reference Signal (SRS) | UL | Used to provide uplink channel quality estimation feedback to uplink scheduler for channel aware scheduling in the eNB |
| Physical Random Access Channel (PRACH) | UL | Transports RACH preambles |

Table 2.2: LTE Physical Layer Channels [17]

In order for the UE to estimate the DL RF channel characteristics, 3GPP LTE uses coherent demodulation by transmitting RS in the OFDM time-frequency grid structure. Several RS have been defined in 3GPP LTE however the three main ones are the cell-specific, UE-specific and channel state information reference signals. The latter has been introduced in 3GPP R10 to support 8-layer spatial multiplexing. Whilst focusing on the cell-specific RS and with reference to Figure 2.7, these RS are transmitted within the first and fifth OFDM symbol of each slot and having a frequency domain spacing of six subcarriers. Moreover, there exists a frequency domain staggering of three subcarriers between the first and second reference symbols resulting in four reference symbols within each PRB. The RS mapping depends on the number of antennas and whether normal or extended CP is used, leading to the existence of several different mappings [17].

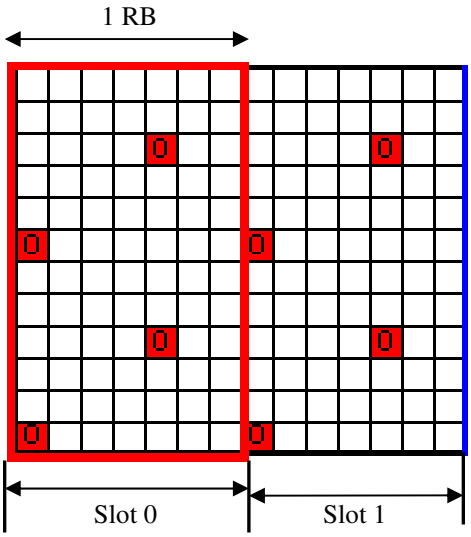



Figure 2.7: Mapping of downlink cell-specific reference signals (1- antenna, normal cyclic prefix)

## 2.3 Channel Feedback Report Types

As discussed in the previous sections, in order for the packet scheduler and link adaptation entities in the eNB to gain channel awareness, the UE can send four types of channel feedback information depending on the MIMO and eNB configuration. These include:

- RI which indicates the UE recommendation for the number of layers to be used in spatial multiplexing and only relevant when the UE is operating in MIMO modes with spatial multiplexing;
- PMI is strictly relevant to closed loop MIMO operation and indicates the pre-coding matrix to be used by the eNB which allows the mapping of each layer onto one or more antennas;
- PTI was introduced in Release 10 to support the extended MIMO transmission modes and is used to distinguish between slow and fast fading environments;
- CQI which is considered a critical element in the channel feedback information sent by the UE and the focus of this study.

With reference to the 3GPP LTE defined CQI index table, shown in Table 2.3, the 4-bit CQI index corresponds to the UE recommendation of the highest MCS and TBS it can decode at the time with a block error rate (BLER) not exceeding 10% [23].

| CQI Index | UE Cat. 1 - 10 | | UE Cat. 11 - 12 | |
|---|---|---|---|---|
| | **Modulation** | **Efficiency** | **Modulation** | **Efficiency** |
| 1 | QPSK | 0.1523 | QPSK | 0.1523 |
| 2 | QPSK | 0.2344 | QPSK | 0.377 |
| 3 | QPSK | 0.377 | QPSK | 0.877 |
| 4 | QPSK | 0.6016 | 16QAM | 1.4766 |
| 5 | QPSK | 0.877 | 16QAM | 1.9141 |
| 6 | QPSK | 1.1758 | 16QAM | 2.4063 |
| 7 | 16QAM | 1.4766 | 64QAM | 2.7305 |
| 8 | 16QAM | 1.9141 | 64QAM | 3.3223 |
| 9 | 16QAM | 2.4063 | 64QAM | 3.9023 |
| 10 | 64QAM | 2.7305 | 64QAM | 4.5234 |
| 11 | 64QAM | 3.3223 | 64QAM | 5.1152 |
| 12 | 64QAM | 3.9023 | 256QAM | 5.5547 |
| 13 | 64QAM | 4.5234 | 256QAM | 6.2266 |
| 14 | 64QAM | 5.1152 | 256QAM | 6.9141 |
| 15 | 64QAM | 5.5547 | 256QAM | 7.4063 |

Table 2.3: CQI Table [20]

The CQI index is calculated by the UE and is based on the DL Signal-to-Interference-and-Noise Ratio (SINR) measured using the DL cell-specific reference signals transmitted by the eNB. For each PRB $k$, the received SINR is given by [23]:

$$\gamma_k = \frac{P_{i,k} G_{i,k}}{\sum_{j \neq i} P_{j,k} G_{j,k} + n_k} \tag{2.1}$$

where $P_i$ and $G_i$ are the transmit power and antenna gain of the serving eNB $i$ whilst $P_j$ and $G_j$ are the transmit power and antenna gain of the interfering eNB $j$, and $n_k$ is the additive Gaussian noise.

The measured SINR is then mapped to a discrete CQI index, using a UE chipset vendor specific mapping table, and thus when reporting the CQI index the UE considers both the downlink channel quality and the capabilities of the UE receiver. This allows a UE with enhanced receiver performance to feedback a higher CQI index for the same DL SINR conditions. In order to allow the eNB gain rich channel awareness and fully exploit the advantages of frequency domain scheduling (FDS), each UE must report detailed channel state information measured across the whole channel bandwidth. However, as the number of UEs increases, the uplink signalling overhead becomes prohibitively high and thus a time domain feedback compression

technique in the form of periodic or aperiodic reporting has been adopted in LTE. Whilst periodic reporting makes extensive use of the capacity constrained PUCCH, which also carries uplink scheduling requests and HARQ signalling, it allows the eNB to gain coarse channel information on a regular basis. Meanwhile, aperiodic reporting allows the UE to send rich channel information over the PUSCH which doesn't have the same capacity limitations as the PUCCH. However, the PUSCH is a dedicated resource and thus only one UE can be scheduled on a single portion of the spectrum at a time [23].

In 3GPP LTE, a connected user is defined as a UE which has at least one established DRB whilst not necessarily transferring data. In general, every connected user in a cell is required to send periodic CQI reports and thus the signalling load on the PUCCH resources increases rapidly with the increasing number of UEs in the cell. With reference to Figure 2.8, the PUCCH resources are located at the edges of the LTE channel such that the PUSCH region is maximised. Thus, the dimensioning of the PUCCH region is critical as the allocation of a wide PUCCH region results in lower resources reserved for the PUSCH leading to lower uplink capacity for data traffic. Meanwhile, a narrow PUCCH region may result in a degradation of the data call setup success rate in case of an increase in the number of connected users in the cell due to the lack of available control channels. Apart from the number of connected users, the PUCCH capacity dimensioning process depends on the amount of CQI feedback signalling. A network operator has little control on the number of connected users in the cell and hence techniques to reduce the CQI feedback signalling are highly desirable especially in high load conditions.

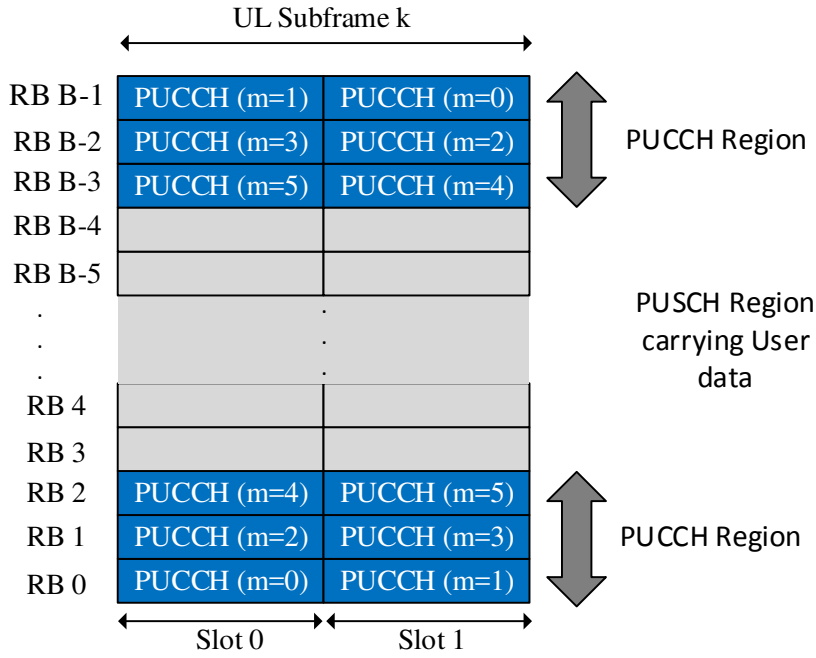



Figure 2.8: PUCCH/PUSCH region

## 2.4 CQI Feedback Signalling Reduction Schemes

The main goal of any CQI feedback signalling reduction scheme is to allow the eNB to gain reasonably accurate channel awareness to maintain an adequate system performance with limited uplink feedback. Several CQI feedback signalling reductions schemes have been proposed however these can be grouped into two types; namely frequency domain and threshold/prediction based techniques**.**

### 2.4.1 Frequency-domain schemes

Frequency domain schemes are typically used in conjunction with the time domain schemes described in section 2.3 and are broadly divided into three; namely full-band, sub-band and wideband compression schemes.

Full-band channel reporting represents the ideal scenario whereby the UE reports the channel state information across all the PRBs. This type of detailed reporting, although optimal, is not feasible and a number of techniques have been proposed to reduce the signalling overhead whilst allowing the eNB to reconstruct a good approximation of the entire system bandwidth. These techniques are generally based on the use of mathematical transforms such as the discrete cosine transform (DCT) [24, 25] and the Haar wavelet transform [26]. Although these techniques allow the

eNB to reconstruct a good approximation of the entire system bandwidth, using a relatively small number of transform coefficients reported by the UE, their performance strongly depends on the delay spread of the channel which varies with UE speed. This makes them highly unreliable.

In sub-band compression, a UE transmits a selected set of CQI values to the eNB thereby allowing the eNB to gain limited channel awareness in the frequency domain and allow FDS. The 3GPP LTE standard defines two sub-band compression techniques namely eNB configured and UE-selected sub-band feedback. With reference to Figure 2.9, in the eNB configured sub-band feedback scheme the system bandwidth is divided into $q$ sub-bands of $N$ consecutive PRBs and the UE reports a 4-bit wideband CQI value together with a 2-bit differentially encoded CQI value for each sub-band [23]. The sub-band size $N$ is dependent on the channel bandwidth $B$ and is given in Table 2.4 [20].

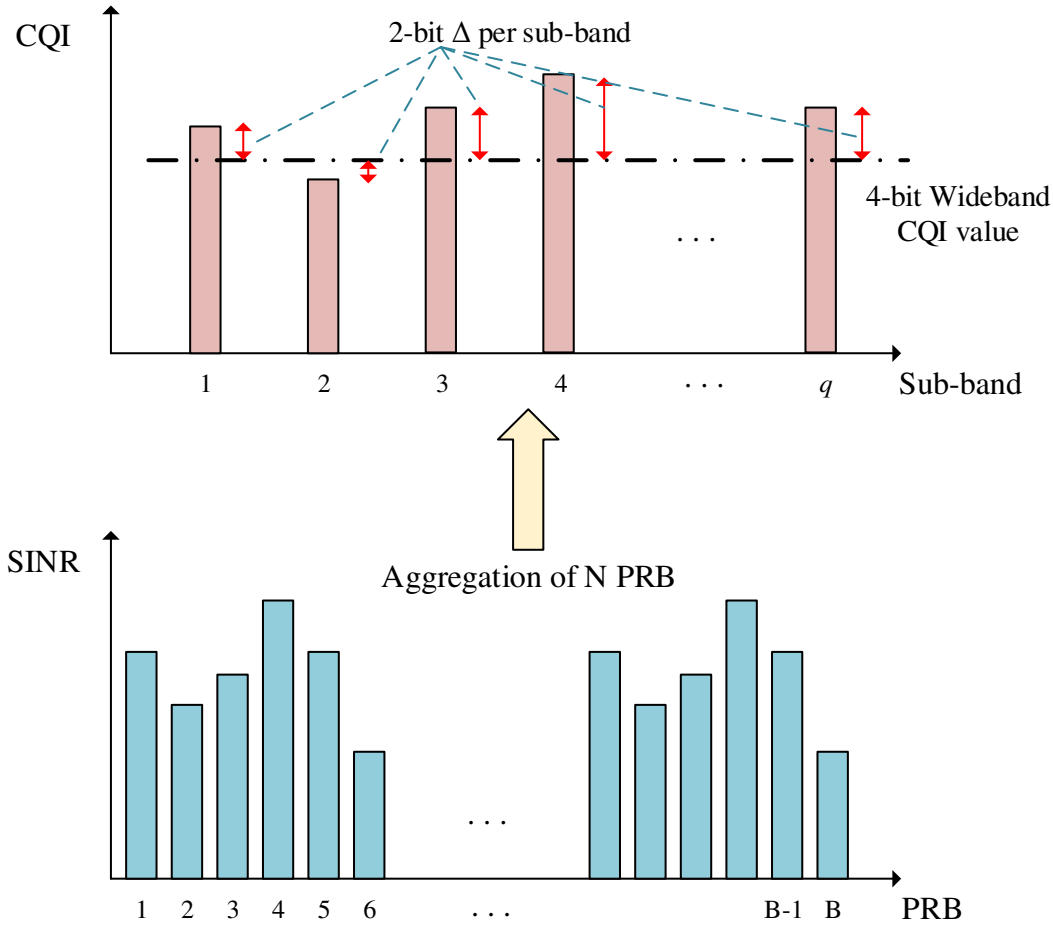



Figure 2.9: eNB-Configured sub-band feedback compression

| System Bandwidth *B* | Sub-band size *N* |
|---|---|
| 6 – 7 | NA |
| 8 – 10 | 4 |
| 11 – 26 | 4 |
| 27 – 63 | 6 |
| 64 – 110 | 8 |

Table 2.4: eNB configured sub-band feedback – sub-band size *N*

The UE-selected sub-band feedback is based on the Best-M techniques presented in [27] and represents an effective balance between the system performance and the uplink feedback signalling load [23,28,29]. As shown in Figure 2.10, the system bandwidth is first divided into *q* sub-bands of *N* consecutive PRBs and each UE selects *J* preferred sub-bands. The UE will then report a 4-bit wideband CQI value coupled with a 2-bit differentially encoded CQI value representing the mean quality of the *J* preferred sub-bands. In addition, the location of the selected *J* preferred sub-bands is also transmitted thereby incurring an incremental increase in feedback load. The value of *J* together with the sub-band size *N* is given in Table 2.5.

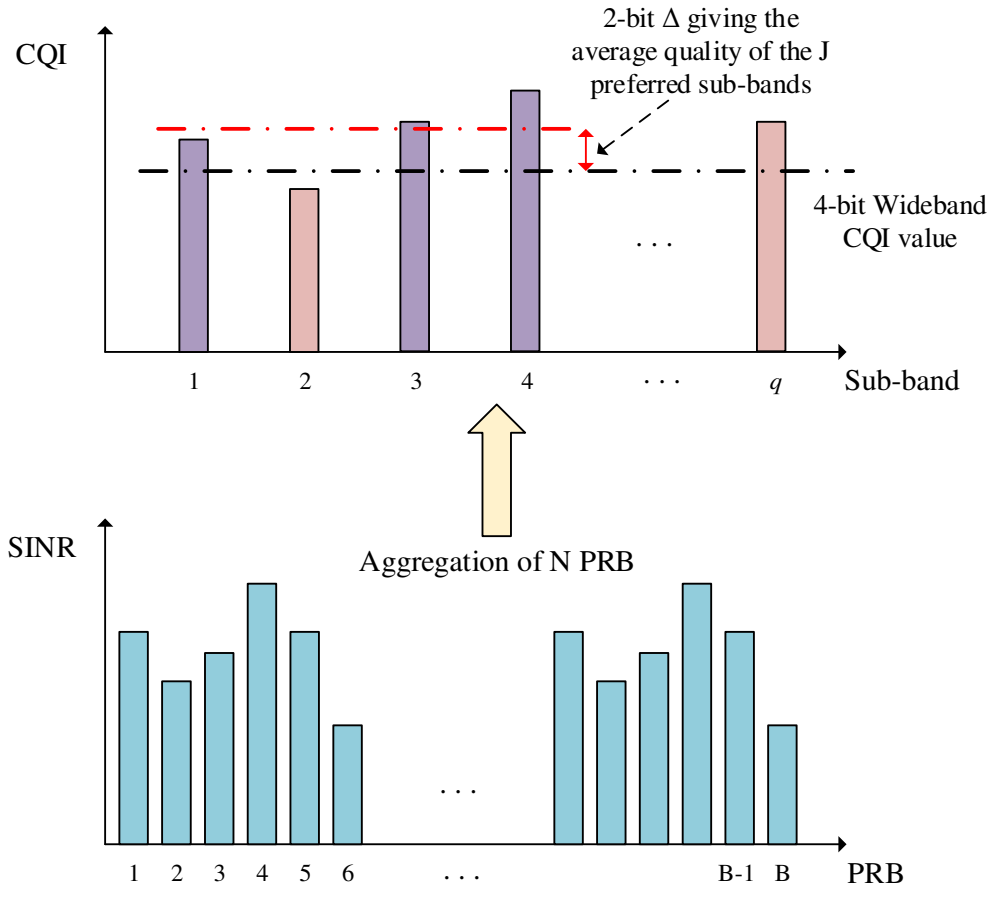



Figure 2.10: UE-selected sub-band feedback compression (preferred sub-bands marked in violet)

| System Bandwidth *B* | Sub-band size *N* | Number of Sub-bands *J* |
|---|---|---|
| 6 – 7 | NA | NA |
| 8 – 10 | 2 | 1 |
| 11 – 26 | 2 | 3 |
| 27 – 63 | 3 | 5 |
| 64 – 110 | 4 | 6 |

Table 2.5: UE-Selected sub-band feedback – sub-band size *N* and number of sub-bands *J*

With reference to Figure 2.11, wideband compression allows the UE to feedback a 4-bit CQI value for all the PRBs in the bandwidth thereby achieving a high compression factor. The lack of channel information in the frequency domain precludes the use of FDS leading to a lower system performance by as much as 40% in the average cell throughput at low mobility [23]. Nevertheless, the 3GPP LTE standard includes wideband compression since Release 8 and is commonly used in live commercial networks especially in highly loaded networks.

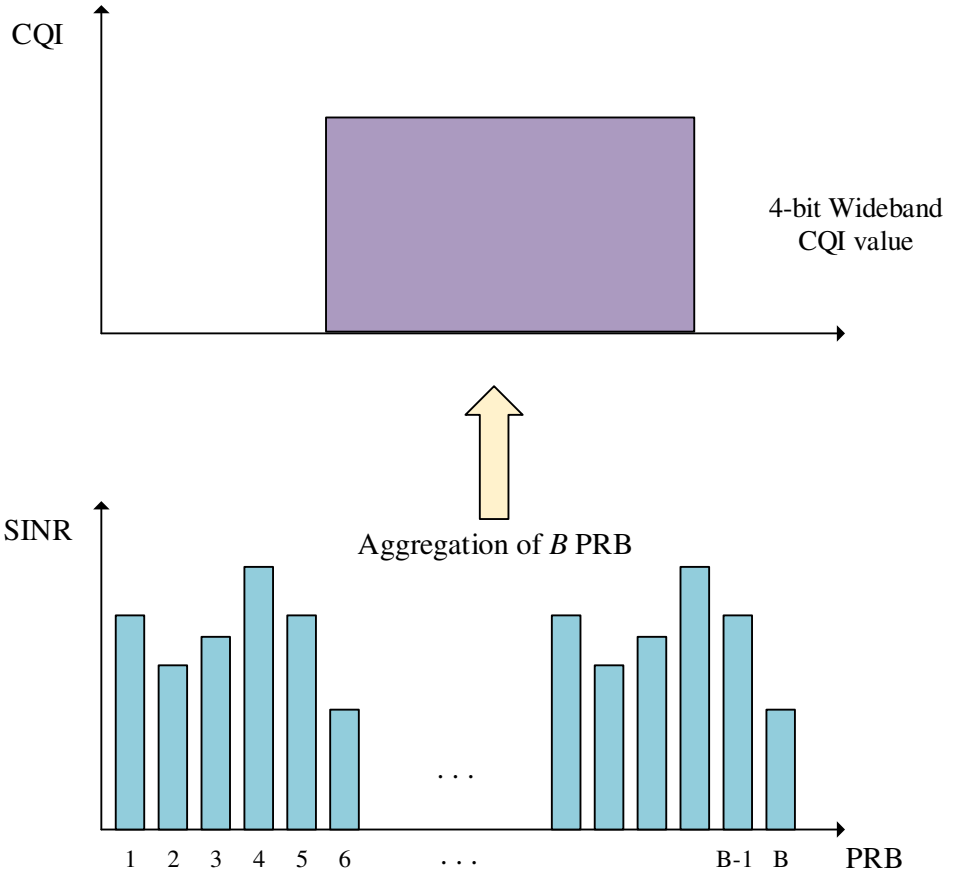



Figure 2.11: Wideband CQI feedback compression

Although the channel frequency response across a 180 kHz PRB is typically flat, the effect of frequency selective fading results in different PRBs experiencing different fading characteristics. Hence, when combining the instantaneous SINR of multiple PRBs, such as in the case of wideband or sub-band feedback compression schemes, the effective SINR, $\gamma_{eff}$, is computed using an effective SINR mapping (ESM) model such as the exponential ESM (EESM) [30,31,32] given by:

$$\gamma_{eff} = -\lambda \log\left( \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{-\frac{\gamma_k}{\lambda}} \right) \tag{2.2}$$

where $\lambda$ is a factor that is MCS dependent and calibrated using link-level simulations [32], $N$ represents the number of PRBs to group and $\gamma_k$ is the instantaneous SINR of the $k^{th}$ PRB found using (2.1).

The bit cost for the full-band, eNB configured sub-band, UE-selected sub-band and wideband feedback schemes for $T_u$ UEs in the sector is shown in Table 2.6. For the case of full-band reporting, a UE reports a 4-bit wideband CQI value together with a 2-bit differentially encoded CQI value for each PRB $B$.

| Feedback scheme | Bit cost |
|---|---|
| Full-band | $(4+2B)T_u$ |
| eNB configured sub-band | $(4+2q)T_u$ |
| UE selected sub-band | $\left(4+2+\left\lceil \log_2\binom{B}{J} \right\rceil\right)T_u$ |
| Wideband | $4T_u$ |

Table 2.6: Bit cost for various feedback schemes

### 2.4.2 Threshold / Prediction based schemes

This category groups together techniques based on either event-based triggering or channel prediction to reduce the uplink signalling feedback. Although several predictive channel quality methods have been proposed, the large majority are designed to limit the degradation in system performance caused by CQI delay rather than addressing the CQI feedback reduction problem. In [33], the authors explore the use of a normalised least mean square (NLMS) adaptive filter to compensate for CQI

delay at the eNB whilst an autoregressive prediction filter is used in [34]. Meanwhile, the authors in [35, 36] investigate the use of various CQI prediction schemes based on a number of techniques varying from cubic spline extrapolation and short-term averaging to the application of Wiener filter. In [37], a cross layer CQI feedback scheme for the IEEE 802.16e standard is explored. In this work, the authors extend the work found in [38-41] and propose a feedback window adaptation algorithm to adapt the CQI feedback rate for each UE. The proposed prediction-based feedback scheme is based on a base station side recursive least square (RLS) algorithm however the authors fail to adequately demonstrate its efficacy as it has been tested in a single sector scenario.

In [42-46] several signalling feedback reduction schemes, based on the use of a fixed threshold set on the users that are most likely to be scheduled, are proposed. In these works, only users which exceed this set threshold are allowed to feedback their channel state information to the base station leading to a questionable fairness situation. In [47], the authors present a contrasting strategy whereby low SINR users are selected as candidates to use full-band feedback whilst the rest use wideband frequency domain compression. Even though this scheme exhibits some gains, it is designed for scenarios where channel conditions vary slowly. Moreover, the impact of the use of the full-band feedback is not appropriately evaluated and the study has been carried out in a single cell scenario ignoring the underlying network dynamics due to interference and varying traffic load. Meanwhile, in [48], the disadvantages of fixed threshold-based schemes are addressed by proposing an adaptive threshold feedback compression scheme based on a particle swarm optimisation (PSO) technique requiring a relatively large number of iterations. In [49], the authors explore the use of a Gaussian Process (GP) regression method to estimate the CQI in the eNB. This technique is combined with a dual-control scheme based on the use of active learning thereby allowing the eNB to estimate for each user the optimal prediction time. Although the proposed scheme is able to achieve a reduction of 77% in the uplink signalling overhead, when compared to the 3GPP standard eNB-configured sub-band feedback scheme, this comes at the expense of high computational cost and memory complexity induced by the GP regression technique coupled with the adaptive dual-control scheme requiring several iterations per measurement for each user

Although not directly related to feedback compression, the same authors of [49] extend their work in [50], by proposing the use of an adaptive sub-band feedback scheme whereby a reinforcement learning framework is used to adapt, on a per user basis, the type of feedback to use. In this work, the authors show that coarse CQI feedback is adequate in sectors with a high number of users. Meanwhile, in [51], the authors explore the use of complimentary PUSCH grants to a select number of users so as to send rich channel quality feedback reports on the PUSCH even though they do not have any pending data for uplink transmission.

A summary of the various feedback reduction schemes discussed in this section is given in Table 2.7.

| Scheme | Category | | | | Key Characteristics |
|---|---|---|---|---|---|
| | Frequency-domain | Threshold-based | Prediction-based | Other | |
| DCT based Full-Band schemes [24,25] | • | | | | • Allows the reconstruction of the entire system bandwidth.<br>• Requires the reporting of a small number of transform coefficients and their respective positions.<br>• Unreliable, as performance depends on delay spread of channel. |
| Haar based Full-Band schemes [26] | • | | | | |
| eNB selected sub-band scheme [23] | • | | | | • 3GPP standardised.<br>• Wideband scheme is widely used in commercial LTE networks especially in crowd event scenarios. |
| UE selected sub-band scheme [23] | • | | | | |
| Wideband scheme [23] | • | | | | |
| Fixed Threshold CQI reporting schemes [42-46] | | • | | | • Fail to address users with low signal quality leading to inefficient link adaptation and unfair scheduling. |
| Fixed Threshold based full-band/wideband reporting scheme [47] | | • | | | • Achieves marginal gain at low mobility.<br>• Study carried out in single cell scenario.<br>• Impact of full-band reporting is not appropriately evaluated. |
| Adaptive Threshold CQI reporting scheme [48] | | • | | | • High complexity due to adopted PSO technique.<br>• Sub-band position reporting is required. |
| Adaptive feedback window CQI reporting scheme [37] | | | • | | • Designed for IEEE 802.16e and considers the use of RLS at base station side for prediction.<br>• Tested in single cell scenario. |
| Gaussian Prediction based CQI feedback reporting prediction time [49] | | | • | | • Exploits active learning to determine optimal CQI prediction time for each user.<br>• Achieves up to 77% signalling overhead reduction with a packet loss rate of 10%. |
| Use of complimentary PUSCH grants [51] | | | | • | • Studies the use of TCP ACKs to carry CQI feedback bits.<br>• Use of complimentary PUSCH grants to select number of users to send rich channel quality feedback. |

Table 2.7: Summary of CQI feedback reduction schemes

# Chapter 3 CQI Feedback Reduction Schemes

This chapter presents a detailed description of the proposed CQI feedback reduction schemes and is organised as shown in Figure 3.1. Two main scenarios are considered namely a standard mobility scenario, modelling multiple sectors with a number of UEs moving at various speeds, and a crowd event scenario, involving a relatively large number of stationary UEs in a confined area.

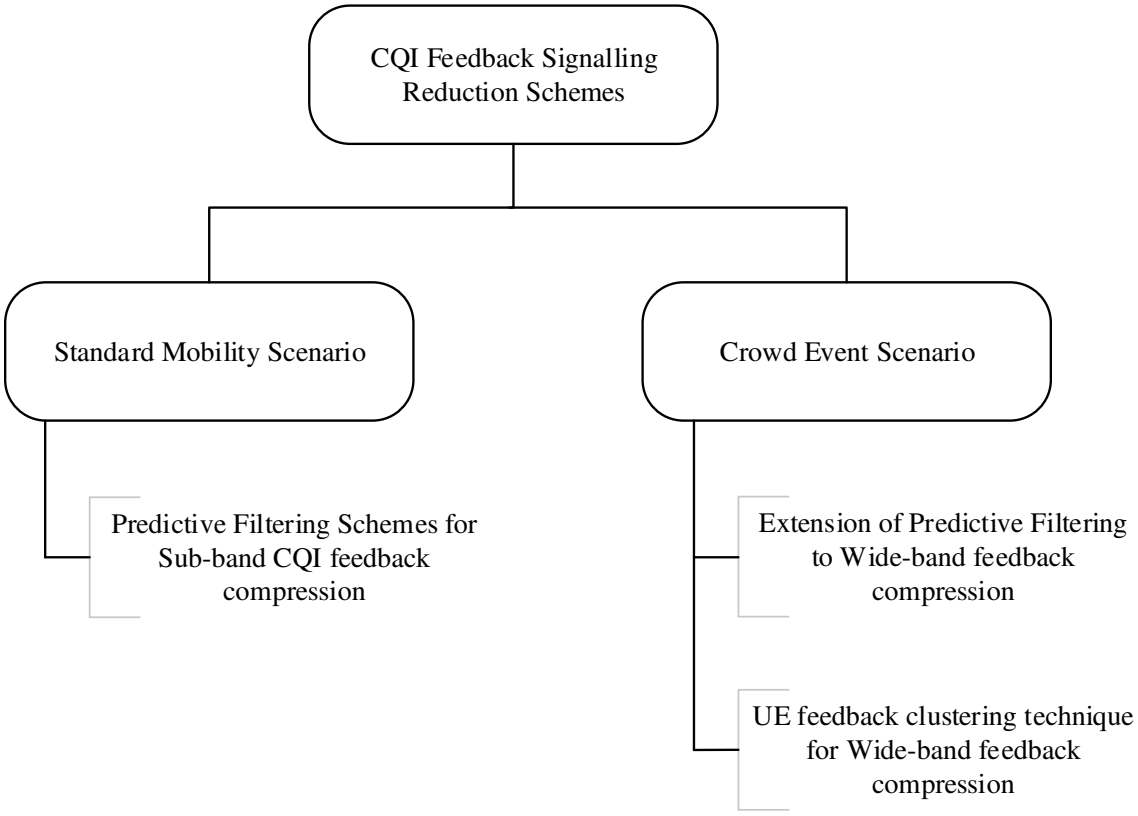



Figure 3.1: Feedback Reduction Scheme Organisation

## 3.1 Predictive Filtering Schemes for Sub-band CQI feedback compression

The proposed sub-band CQI feedback reduction technique exploits short-term channel prediction through the use of a UE-assisted predictive filtering algorithm. This is realised by using low computational complexity adaptive filters in both the UE and eNB.

### 3.1.1 System model

With reference to the system model shown in Figure 3.2, consider that the channel bandwidth composed of $L$ PRBs, where $L$ is defined by the 3GPP Release 8 standard [20], is divided into $q$ sub-bands of $N$ consecutive PRBs. As discussed in section 2.3, a UE uses the cell-specific reference signals transmitted by the eNB to measure the SINR on each PRB, denoted by $\gamma_L$, and calculates the effective SINR of each sub-band, denoted by $\gamma_q^{sub}$, using EESM given by (2.2). The effective SINR is then mapped to a CQI value using the SINR to CQI mapping table, shown in Table 3.1, derived from the BLER-SINR curves obtained through LTE link level simulations [52, 53].

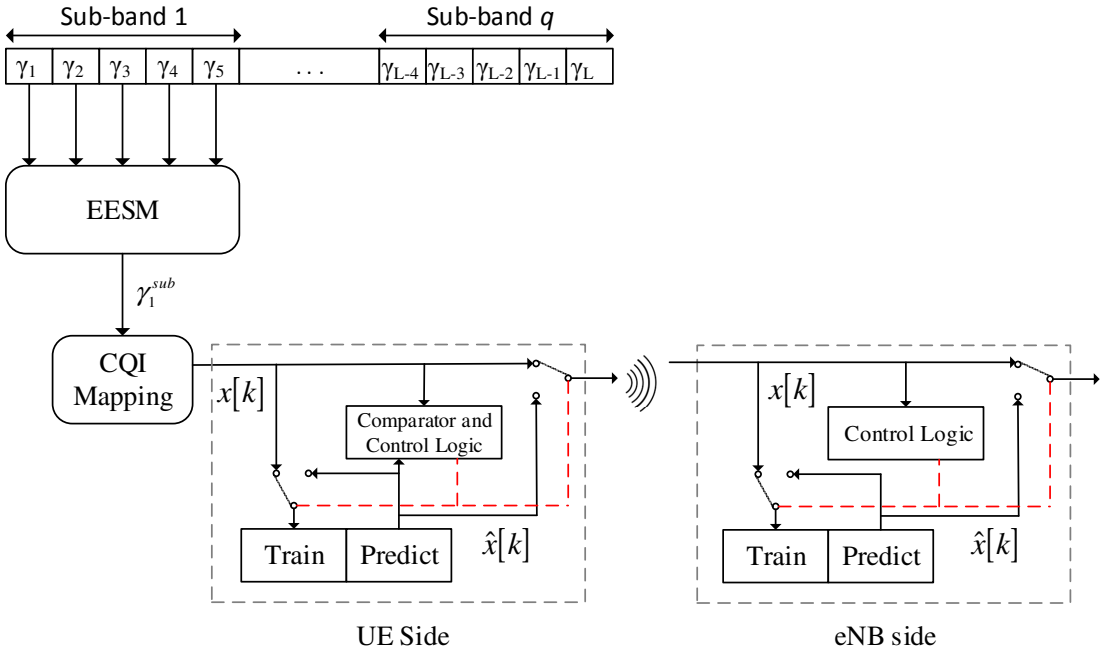



Figure 3.2: Prediction filtering technique ($N$=5) showing the report for sub-band 1

Now consider a series of CQI measurements for the $q^{th}$ sub-band, $x[k]$, that has to be reported by the UE to the eNB at every time step $k$. Moreover, consider that the required accuracy of the measurements at the eNB is within $\pm E_{max\text{-}cqi}$. Then, using the proposed predictive filtering algorithm, instead of reporting every CQI measurement in $x[k]$, we selectively report some elements of the measurement stream such that the eNB is able to reproduce the complete measurement stream for the $q^{th}$ sub-band within the given accuracy. This is achieved by maintaining a set of predictors, one for each sub-band, in the UE and eNB which are an exact replica of each other. These

predictors are based on adaptive filter theory and compute an estimate of the next element in the measurement stream $\hat{x}[k]$, given some previous elements.

| SINR | CQI |
|---|---|
| -6.15 | 1 |
| -4.37 | 2 |
| -2.37 | 3 |
| -0.42 | 4 |
| 1.53 | 5 |
| 3.43 | 6 |
| 5.46 | 7 |
| 7.25 | 8 |
| 9.28 | 9 |
| 11.11 | 10 |
| 13 | 11 |
| 14.9 | 12 |
| 16.64 | 13 |
| 18.41 | 14 |
| 20.54 | 15 |

Table 3.1: SINR-CQI Mapping

With reference to Figure 3.3, when the average error between the predicted value $\hat{x}[k]$ and the actual value $x[k]$ increases beyond the allowed error budget, $\pm E_{max\text{-}cqi}$, then $x[k]$ is transmitted to the eNB; otherwise the actual measurement is not transmitted and the eNB uses its own prediction, $\hat{x}[k]$, resulting in feedback reduction.

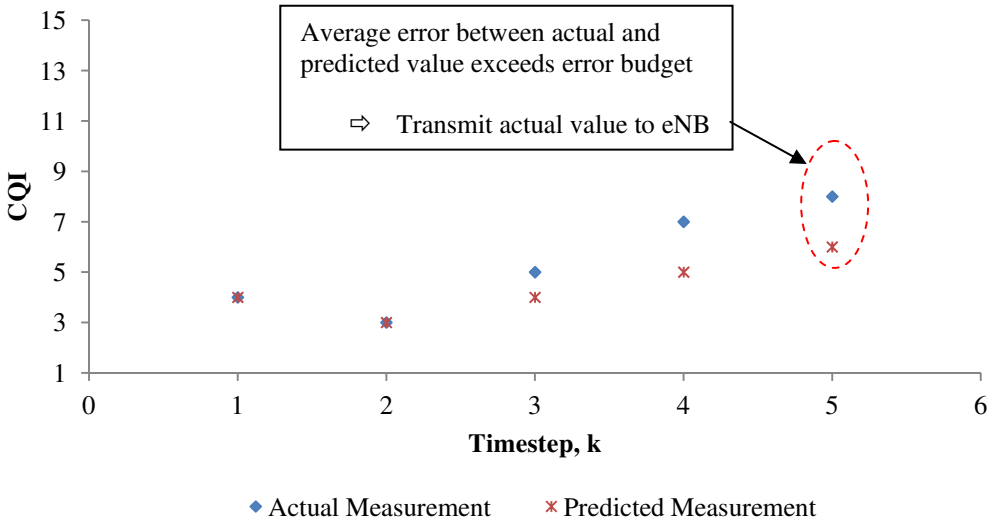



Figure 3.3: CQI feedback reduction

### 3.1.2 Adaptive Filters

Several adaptive filters can be found in literature and the selection of an adaptive filtering algorithm over another depends on the trade-off of a number of factors such as the computational complexity, robustness and rate of convergence. In this work, the least mean square (LMS) and the normalised LMS (NLMS) algorithms were selected in view of their simplicity, and relatively good performance in a wide variety of applications within the wireless communication sphere. The operation of the LMS and NLMS adaptive filters is fundamentally governed by equations (3.1), (3.2) and (3.3) which allow the computation of the output, error and LMS weight updating respectively.

$$y[k] = \underline{w}^T[k]\underline{x}[k] \tag{3.1}$$

$$e[k] = d[k] - y[k] \tag{3.2}$$

$$\underline{w}[k+1] = \underline{w}[k] + \mu \underline{x}[k]e[k] \tag{3.3}$$

where $y[k]$ is the output of the adaptive filter, $e[k]$ is the prediction error, $d[k]$ is the desired signal, $\mu$ is the step size parameter, $M$ is the filter order whilst $\underline{w}[k]$ and $\underline{x}[k]$ denote the $M \times 1$ column vectors:

$$\underline{w}[k] = \left[w_1[k], w_2[k], w_3[k], ..., w_M[k]\right]^T \tag{3.4}$$

$$\underline{x}[k] = \left[x[k-1], x[k-2], x[k-3], ..., x[k-M]\right]^T \tag{3.5}$$

The LMS adaptive filtering algorithm is notoriously sensitive to the scaling of its input $x[k]$ thereby making it challenging to select a step size that ensures convergence [54]. Instead, the NLMS adaptive filtering algorithm uses a modified weight updating equation, given by (3.6), such that the step size parameter is normalised with the input signal power. A detailed description of the LMS/NLMS algorithm can be found in [54].

$$\underline{w}[k+1] = \underline{w}[k] + \frac{\mu_T}{\alpha + \left\|x[k]\right\|^2}\underline{x}[k]e[k] \tag{3.6}$$

where $\mu_T$ is the step size parameter and $\alpha$ is a constant to ensure that $\underline{w}[k+1]$ is bounded even when the input is very small.

### 3.1.3 Predictive Filtering Algorithm

With reference to Algorithm 1 (see Figure 3.4), the operation of the predictive filtering algorithm, which applies to both the LMS and NLMS adaptive filter case, is divided in two modes of operation namely a training mode and a prediction mode. In training mode, the UE reports CQI measurements $x[k]$ to the eNB using the standard CQI reporting process described in section 2.4. Meanwhile, in order to ensure synchronisation of the prediction filters, the UE and eNB compute $\hat{x}[k]$, for the upcoming measurement, using the last $M$ measurements. Moreover, the filter coefficients $w$ are updated based on the actual prediction error, using (3.3) or (3.6) depending on whether an LMS or NLMS adaptive filter is used. The UE remains in training mode until the eNB receives $T_{max}$ measurements from the UE and the average absolute prediction error over the past $M$ measurements is below a preconfigured error budget, $|E_{max\text{-}cqi}|$, for $\beta$ consecutive times, thereby providing protection against oscillations between the training and prediction modes of operation. Otherwise, the UE will switch to prediction mode.

In prediction mode the UE continues to perform CQI measurements however it suppresses the transmission of these reports to the eNB. Instead, the UE computes $\hat{x}[k]$ at each time step $k$, compares it to the actual measurement and calculates the prediction error. As long as the prediction error remains within an error budget, the UE uses the predicted value to train the predictive filter, instead of using the actual measurement, such that the predictive filters in the UE and eNB remain synchronised. In this mode, $e[k]$ is forced to zero thereby avoiding the adaptation of the filter weights and saving computational overhead. Similarly, the eNB computes $\hat{x}[k]$ at each time step $k$ and feeds this value to its own prediction filter. In case, the UE observes that the average prediction error over the past $M$ measurements is deviating from the maximum allowed error for $\beta$ consecutive times, it will resume the transmission of the measurements $x[k]$ to the eNB and switch back to training mode.

**Algorithm 1: Prediction Filtering Algorithm**

1: Initialisation:
2: k $\leftarrow$ 0; mode $\leftarrow$ training; txCounter $\leftarrow$ 0
5: for k do
6: if mode = training then
7: UE transmits measurement to eNB
8: txCounter $\leftarrow$ txCounter + 1
9: if txCounter $\geq$ M then
10: UE and eNB uses last M measurements to compute:
11: Prediction, $\hat{x}[k]$
12: Prediction error, e[k]
13: Update filter weights, w
14: if txCounter $\geq$ $T_{max}$ & e $\leq$ error budget for $\beta$ times then
15: errCounter $\leftarrow$ 0
16: mode $\leftarrow$ prediction
17: end if
18: end if
19: else
20: UE and eNB uses last M measurements to compute prediction, $\hat{x}[k]$
21: UE and eNB shift $\hat{x}[k]$, into prediction filter
22: UE uses $\hat{x}[k]$ and x[k] to compute prediction error, e[k]
23: if e $\geq$ error budget then
24: errCounter $\leftarrow$ errCounter + 1
25: if errCounter = $\beta$ then
26: mode $\leftarrow$ training
27: end if
28: else
29: errCounter $\leftarrow$ errCounter $-1$
30: if errCounter < 0 then
31: errCounter $\leftarrow$ 0
32: end if
33: end if
34: end if
35: k $\leftarrow$ k + 1
36: end for

Figure 3.4: Algorithm 1 - Predictive filtering algorithm

### 3.1.4 Parameter tuning

Several parameters of the predictive filtering algorithm need to be tuned prior to evaluating the effectiveness of the proposed technique. These include:

- Step size $\mu$ or $\mu_T$: This is a fundamental parameter as it determines the rate of convergence of the adaptive filter;
- Filter length $M$: This impacts the computational load and memory footprint of the filters. The LMS algorithm requires $2M+1$ multiplications and $2M$ additions per prediction whereas the NLMS algorithm requires $3M+1$ multiplications and $2M$ additions per prediction;
- Hysteresis parameter $\beta$;
- Maximum allowed prediction error $|E_{max\text{-}cqi}|$.

The above mentioned parameters were tuned by implementing and testing the predictive filtering algorithm, described in section 3.1.3, in MATLAB® using CQI measurements extracted from several UEs simulated in an LTE system level simulator [55]. The LTE simulation parameters used for this study are shown in Table 3.2 whilst the optimal parameter values for both the LMS and NLMS prediction filtering scheme (PFS) are given in Table 3.3.

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Reporting Type | eNB configured with sub-band size = 5 |
| CQI Reporting Period | 20ms |
| Traffic Model | Best Effort (Infinite Buffer) |
| Frequency / Re-use | 2GHz / 1 |
| Cellular Layout | Hexagonal grid, 19 cells |
| Inter-site distance (ISD) | 500m |
| UE speed | 3km/hr, 120km/hr |
| Mobility model | Random Direction |
| UE per Cell | 10-40 UEs |
| $T_{max}$ | 100 |
| Propagation Model | 3GPP Typical Urban<br>$L$= 128.1+37.6log($d$) @ 2GHz [55] |
| Fast Fading Realisation | Jakes Model [55] |

Table 3.2: LTE Simulation Parameters

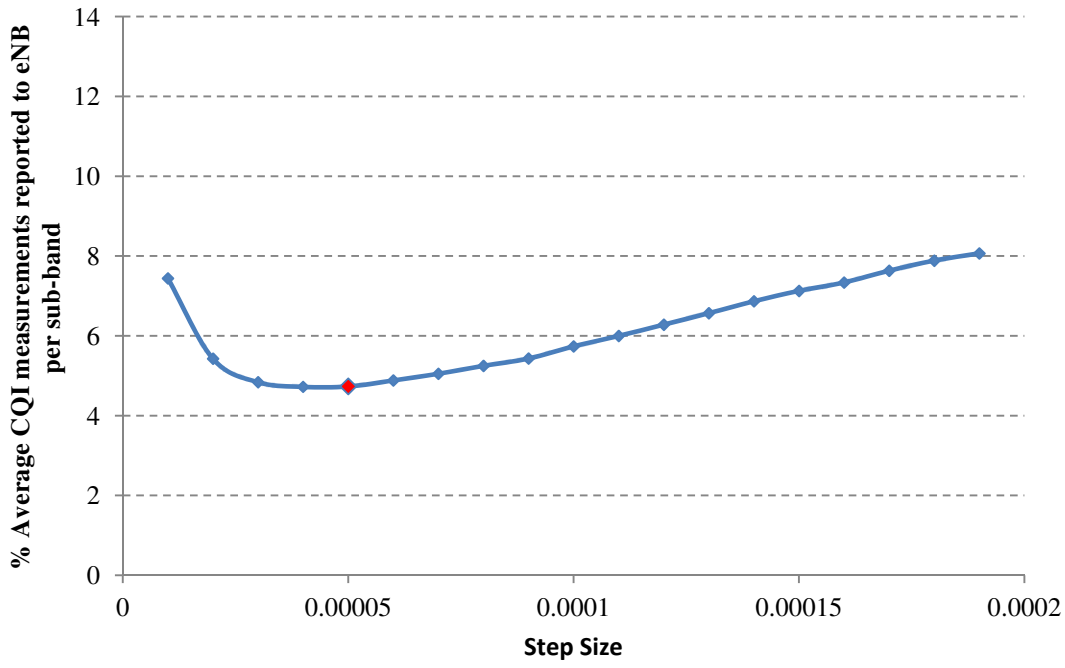
| Parameter | CQI-LMS | CQI-NLMS |
|---|---|---|
| Filter order $M$ | 4 | 4 |
| Step size ($\mu$ or $\mu_T$) | 0.00005 | 0.01 |
| Hysteresis $\beta$ | 3 | 3 |
| $E_{max}$ | 1.5 | 1.5 |

Table 3.3: Optimal parameter values for LMS and NLMS PFS

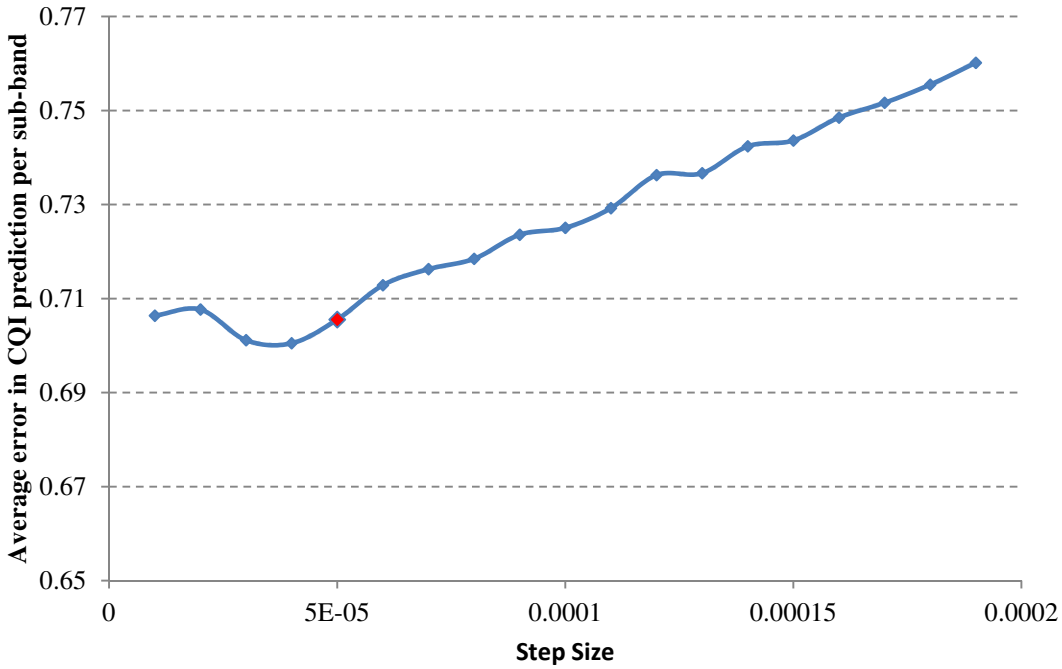
The values shown in Table 3.3 were empirically determined by considering trade-offs between:

- The proportion of CQI measurement reports sent by the UE from the total measurements;
- The average error between the actual and predicted CQI measurement;
- The computational load in terms of arithmetic operations per measurement.

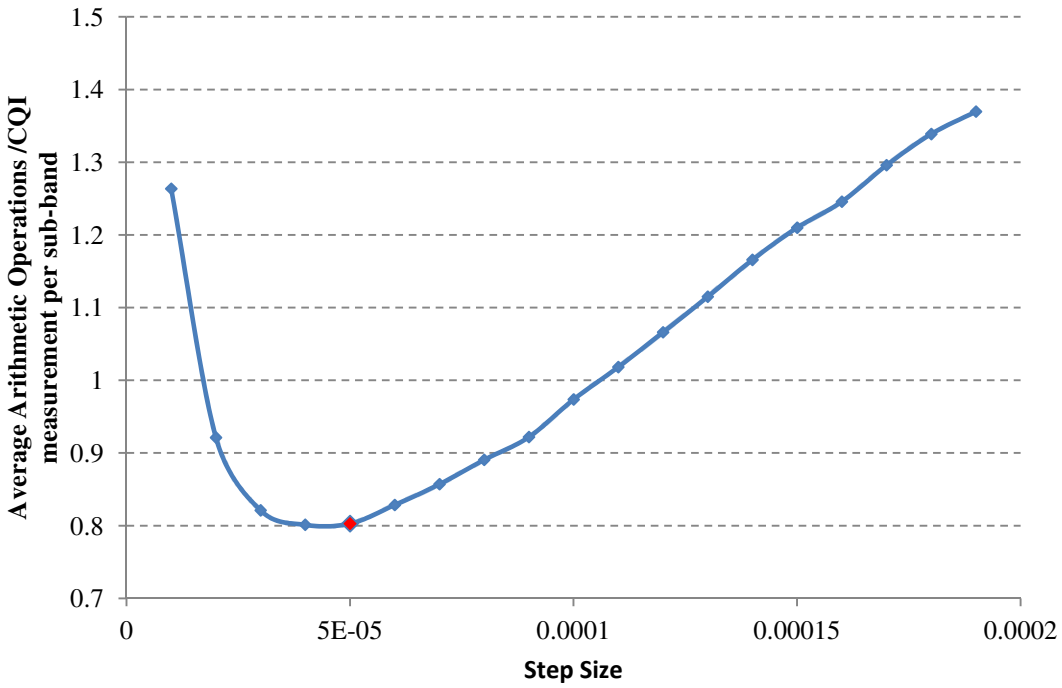
Considering the LMS-based PFS, the effect of varying the step size $\mu$ and filter length $M$ on the abovementioned trade-offs are shown in Figure 3.5(a)-(c) and Figure 3.6(a)-(c) respectively. With reference to Figure 3.5 and 3.6, selecting a $\mu$ of 0.00005 and filter length $M$ of 4 yields the best trade-off between the selected metrics. The rest of the parameters for the LMS and NLMS schemes were determined using a similar methodology.




(a)

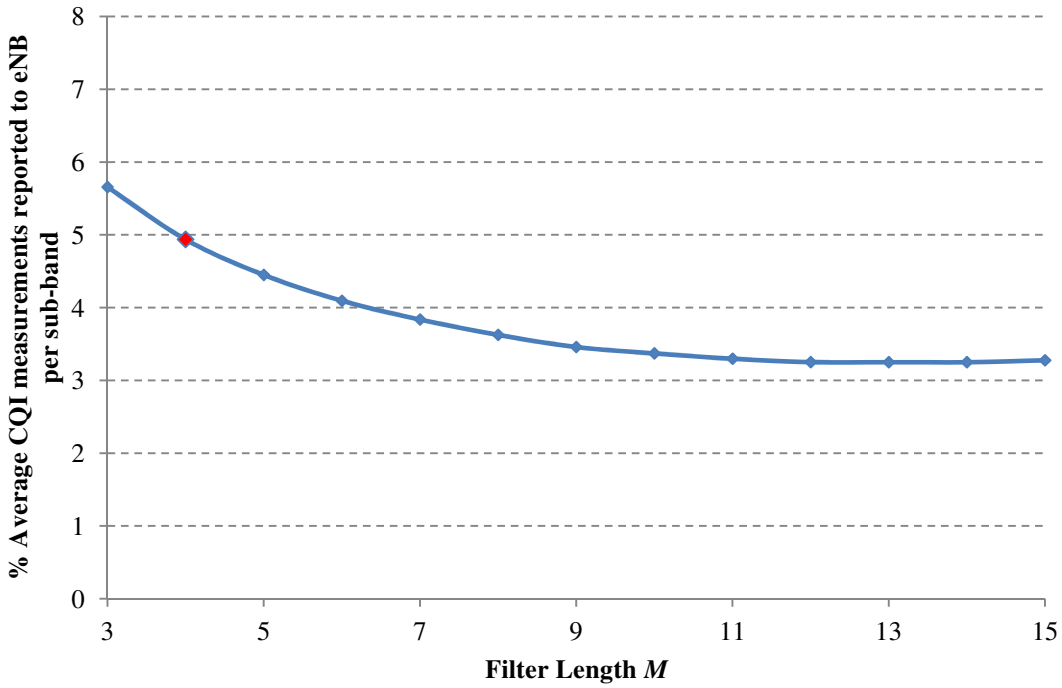



(b)

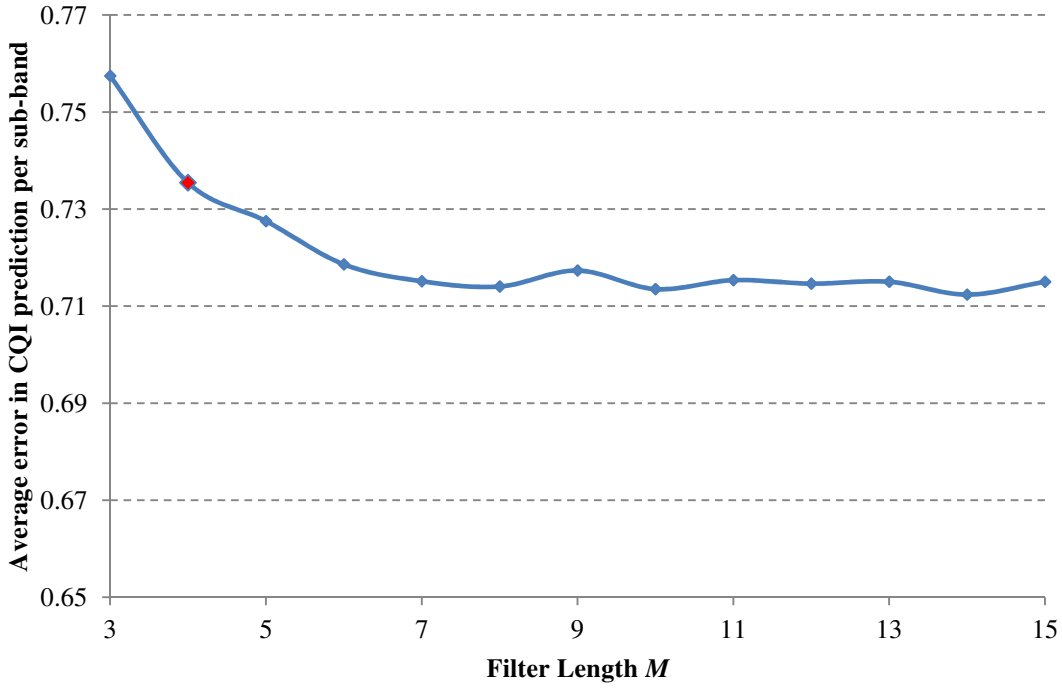



(c)

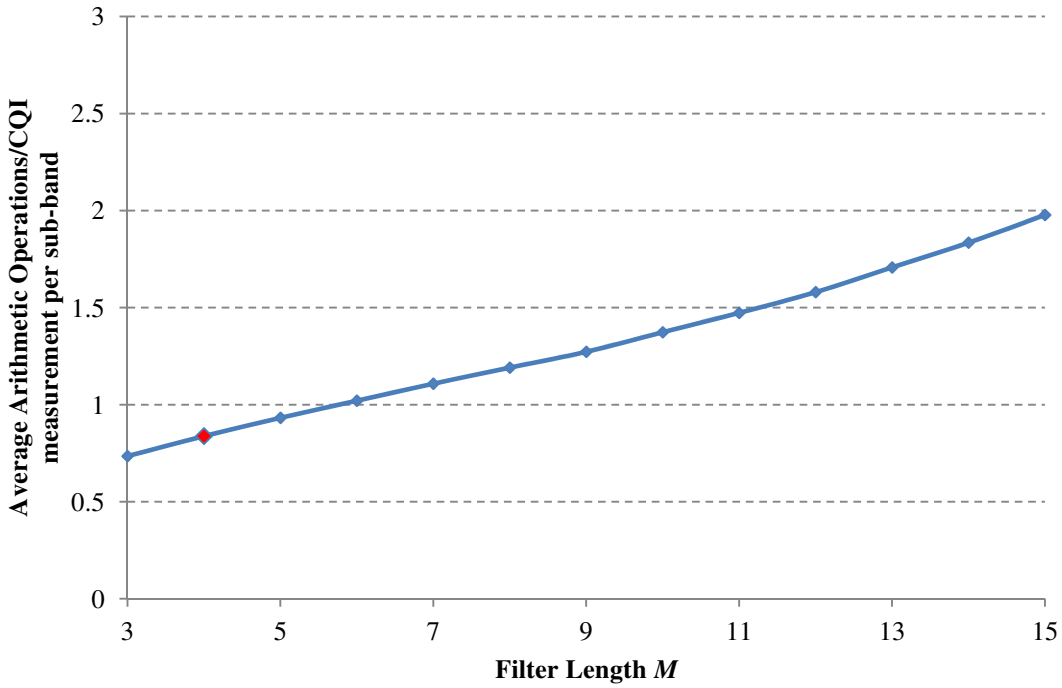
Figure 3.5: LMS adaptive filter parameter tuning – Step size $\mu$. (a) Average proportion of CQI measurements reported to eNB from total per sub-band, (b) Average error of the predicted CQI value per sub-band, and (c) Average arithmetic operation/CQI measurement per sub-band


% Average CQI measurements reported to eNB per sub-band
8
7
6
5
4
3
2
1
0
3
5
7
9
11
13
15
Filter Length M

(a)


Average error in CQI prediction per sub-band
0.77
0.75
0.73
0.71
0.69
0.67
0.65
3
5
7
9
11
13
15
Filter Length M

(b)

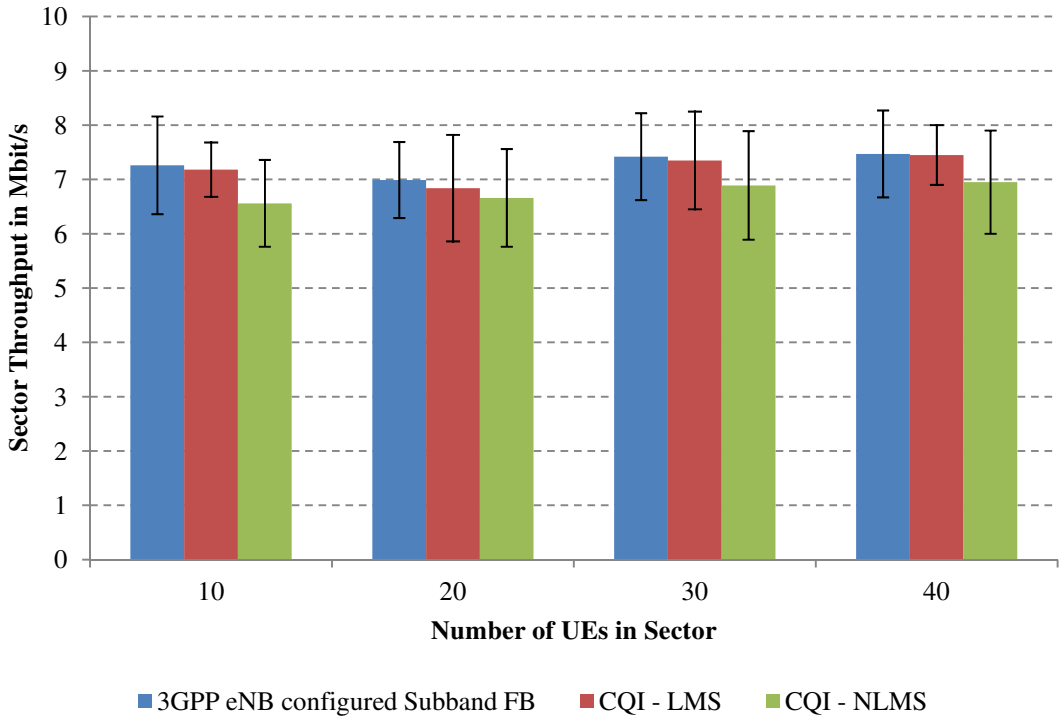



(c)

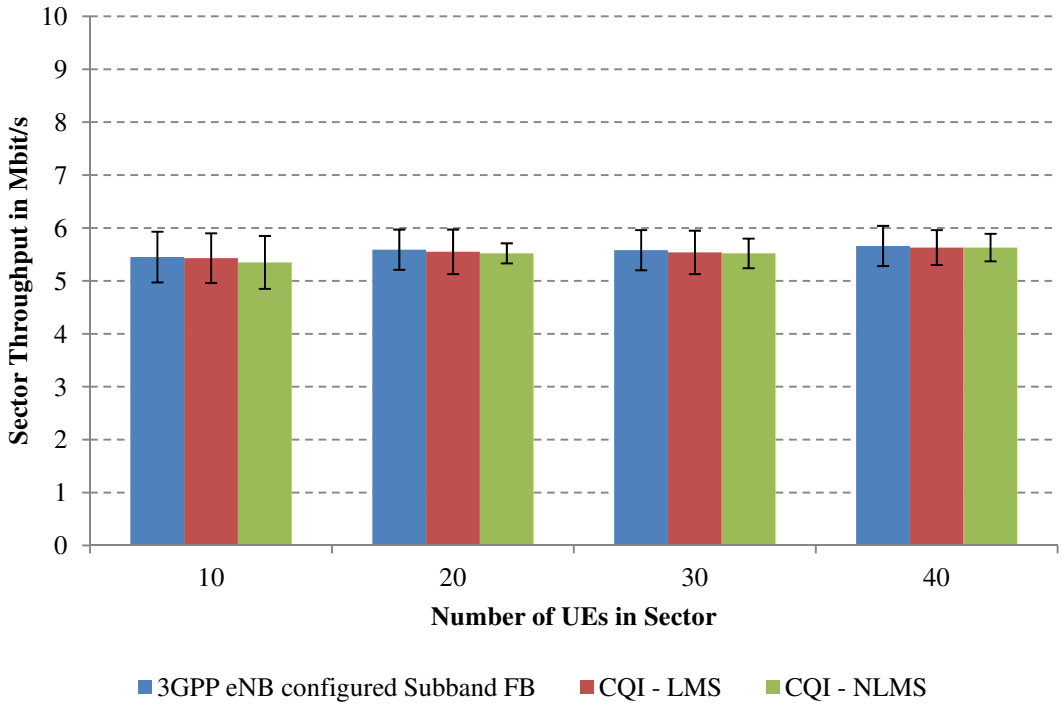
Figure 3.6: LMS adaptive filter parameter tuning – Filter Length *M*. (a) Average proportion of CQI measurements reported to eNB from total per sub-band, (b) Average error of the predicted CQI value per sub-band, and (c) Average arithmetic operation/CQI measurement per sub-band

### 3.1.5 Simulation Results

The predictive filtering algorithm described in section 3.1.3 was implemented in the system level simulator [55] in order to verify its effectiveness in terms of average sector throughput, uplink signalling reduction and packet loss rate. Simulations using the parameters given in Tables 3.2 and 3.3 were carried out over 50 random seeds in order to achieve statistical relevant results. The mean and standard deviation (shown in parenthesis) of the sector throughput for different UE speeds and sector loading is shown in Figure 3.7(a)-(b) whilst the CQI feedback reduction for the LMS and NLMS PFS is shown in Table 3.4-3.5 respectively.




(a)




(b)

Figure 3.7: Sector Throughput (standard deviation is indicated by vertical bars) for (a) UE speed of 3km/hr, and (b) UE speed of 120km/hr

| Number of UEs in Sector | UE Speed | |
|---|---|---|
| | 3km/hr (low mobility) | 120km/hr (high mobility) |
| 10 | 91.8% (±0.4) | 92.6% (±0.2) |
| 20 | 92.4% (±0.5) | 92.4% (±0.2) |
| 30 | 92.6% (±0.2) | 92.5% (±0.2) |
| 40 | 92.6% (±0.2) | 92.4% (±0.2) |
| **Average** | **92.3% (±0.3)** | **92.4% (±0.1)** |

Table 3.4: Feedback Reduction – LMS Predictive Filtering Algorithm

| Number of UEs in Sector | UE Speed | |
|---|---|---|
| | 3km/hr (low mobility) | 120km/hr (high mobility) |
| 10 | 92.6% (±1.2) | 93.9% (±0.1) |
| 20 | 93.1% (±0.4) | 93.6% (±0.1) |
| 30 | 93.1% (±0.4) | 93.6% (±0.1) |
| 40 | 93.1% (±0.6) | 93.9% (±0.1) |
| **Average** | **92.9% (±0.2)** | **93.8% (±0.2)** |

Table 3.5: Feedback Reduction – NLMS Predictive Filtering Algorithm

Whilst considering a marginal loss in performance, when compared to the baseline configuration using the 3GPP standardised eNB configured sub-band scheme, both schemes are able to significantly reduce the CQI signalling overhead. With reference to Table 3.4 and 3.5, the NLMS PFS is marginally better than the LMS counterpart, in terms of feedback reduction, particularly in high mobility scenarios due to its faster convergence speed. Nevertheless, in such scenarios, both schemes achieve a similar performance in terms of sector throughput, whilst in low mobility scenarios, the performance of the LMS-based scheme ranks marginally better when compared to the NLMS-based scheme. The mean and standard deviation (shown in parenthesis) of the packet loss rate (PLR) for both schemes are shown in Table 3.6. Although both schemes exhibit degradation in PLR, the CQI-LMS based PFS ranks better and is within 5% of the baseline configuration whilst achieving a similar performance in high mobility scenarios. In view of the performance exhibited by the CQI-LMS based PFS together with the computational efficiency of the adaptive filters, this scheme offers the best compromise between the selected metrics.

| Scheme | UE Speed | |
|---|---|---|
| | 3km/hr (low mobility) | 120km/hr (high mobility) |
| 3GPP eNB configured sub-band | 2.5% (±0.3) | 17.4% (±1.7) |
| CQI - LMS | 6.7% (±0.7) | 17% (±2.3) |
| CQI - NLMS | 11.6% (±1.4) | 19.9% (±1.7) |

Table 3.6: Packet Loss Rate comparison of the tested schemes

### 3.1.6 Efficiency analysis of the CQI-LMS PFS

The proposed scheme requires the implementation of an LMS adaptive filter for each connected user in the sector per sub-band and thus it is beneficial to analyse the additional computational load and memory requirements. Considering the computational requirements of the LMS adaptive filter algorithm when operating in training and prediction phase, shown in Table 3.7, the additional computational load in the eNB, $C_{enb}$, and UE, $C_{ue}$, is given by:

$$C_{enb} = \alpha\left[\frac{1000}{\tau_p}(4M+1)N_{sub}T_u\right] + (1-\alpha)\left[\frac{1000}{\tau_p}(2M-1)N_{sub}T_u\right] \tag{3.7}$$

$$C_{ue} = \alpha\left[\frac{1000}{\tau_p}(4M+1)N_{sub}\right] + (1-\alpha)\left[\frac{1000}{\tau_p}(2M)N_{sub}\right] \tag{3.8}$$

where $M$ is the filter order, $N_{sub}$ is the number of sub-bands, $T_u$ is the number of connected users in the cell, $\alpha$ is the proportion of time spent in training phase and $\tau_p$ is the CQI reporting period which is commonly set to 20ms in commercial LTE networks.

| | Training Mode | Prediction Mode |
|---|---|---|
| Multiplications | $2M + 1$ | $M$ |
| Additions | $2M$ | $(M - 1)$ for eNB, $M$ for UE |

Table 3.7: Additions and Multiplications for the LMS adaptive filter

Based on the parameters listed in Tables 3.2 and 3.3 whilst assuming $T_u$ to be 200 and $\alpha$ is set to 20%, the overall computational complexity of the proposed scheme is

0.9Mflops and 4Kflops in the eNB and UE respectively. Moreover, by considering the use of the IEEE 754-20008 single precision floating point format [56] for all stored variables required in Algorithm 1, the memory footprint necessary for both the UE and eNB is very low as shown in Table 3.8. Although the variable *mode* is a binary flag to indicate whether the algorithm is operating in training or prediction mode, the use of such bit flag requires additional masking operations to read/write from/to memory. For this reason, an optimised method was chosen to represent this variable by using a byte representation.

| **Variable** | **eNB memory footprint (bytes)** | **UE memory footprint (bytes)** |
|---|---|---|
| $\boldsymbol{k}$ | 4 | 4 |
| mode | 1 | 1 |
| txCounter | 4 | 4 |
| errorCounter | 4 | 4 |
| $\boldsymbol{x}[\boldsymbol{k}]$ | 4 | 4 |
| $\boldsymbol{e}[\boldsymbol{k}]$ | 4 | 4 |
| Filter weights Size $\boldsymbol{M}$ | 4M | 4M |
| Sub-bands | 10 | 10 |
| Users | 200 | 1 |
| **Total Memory Footprint** | **74kbytes** | **0.37kbytes** |

Table 3.8: eNB and UE memory footprint

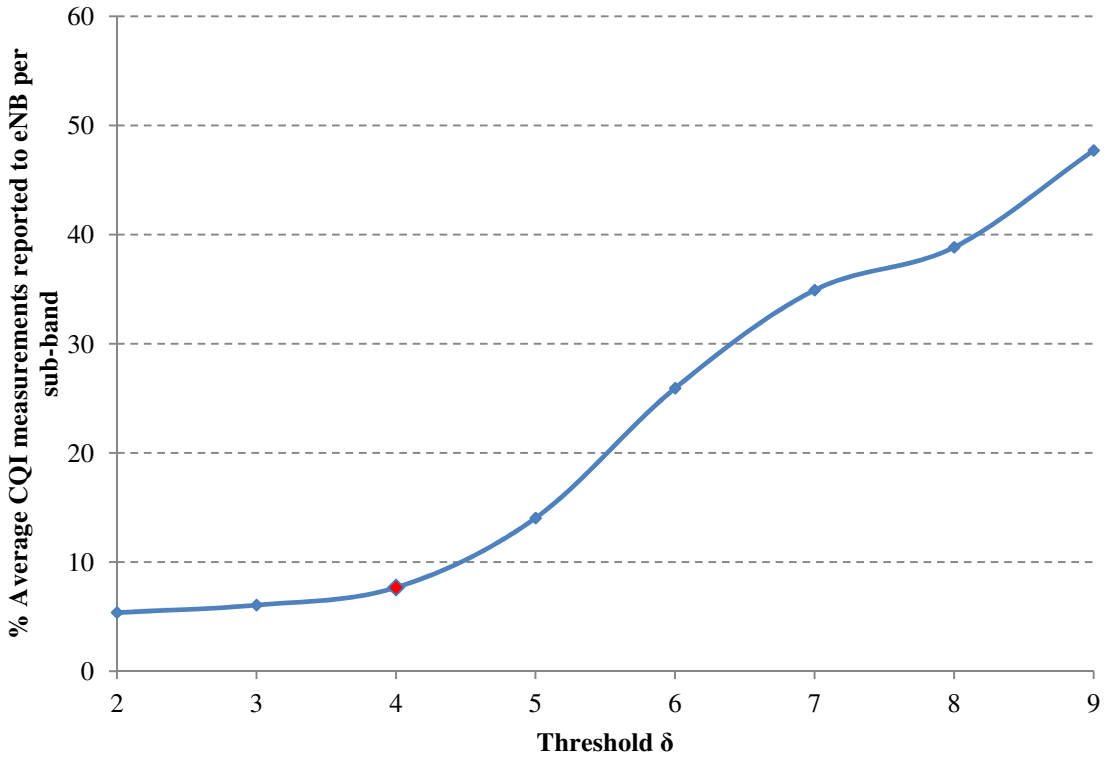
### 3.1.7 An efficient CQI-LMS based PFS

As discussed in the previous section, the computational load and memory footprint of the proposed scheme is low when compared to other operations carried out by the eNB baseband processor. Nevertheless, since the majority of the computations are carried out in the weight update process, a reduced complexity LMS adaptive filter, based on the use of the sign function defined by (3.9), was also investigated [57]. With reference to the new weight update function (3.10), when the tap input vector for time sample $k$ is less than the specified threshold $\delta$, $\underline{x}[k]$ will be equal to zero and thus the corresponding weights does not need to be adapted thereby yielding a reduction in the computational load.

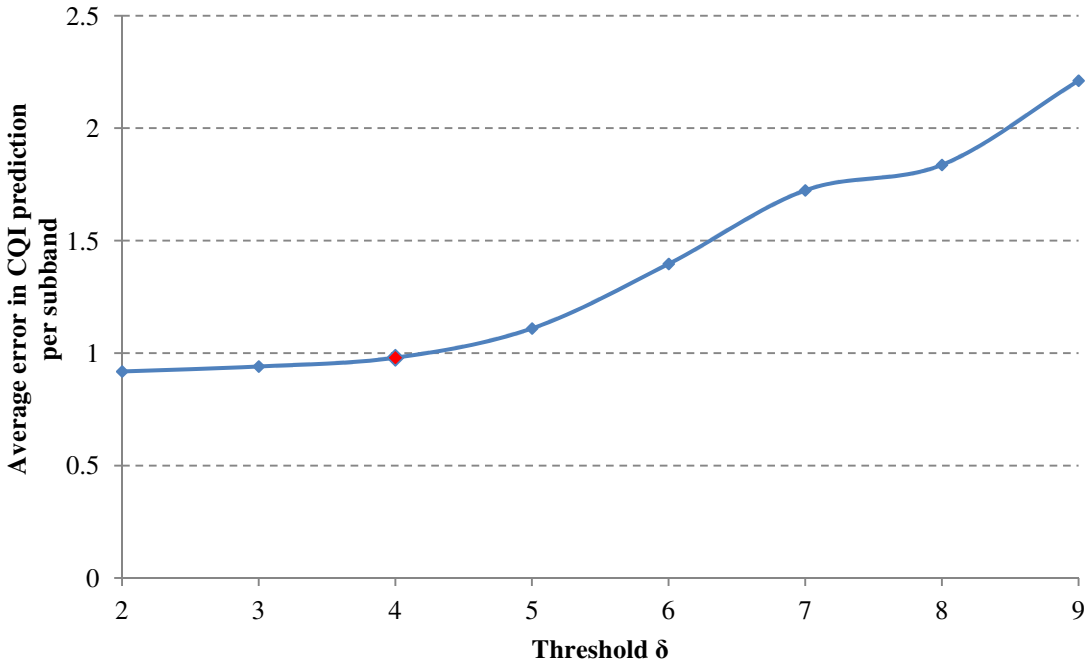
$$sign(\underline{x}[k],\delta) = \begin{cases} \underline{x}[k] & \text{if } x[k] \geq \delta \\ 0 & \text{if } x[k] < \delta \end{cases} \tag{3.9}$$

$$\underline{w}[k+1] = \underline{w}[k] + \mu e[k] sign(\underline{x}[k], \delta) \tag{3.10}$$

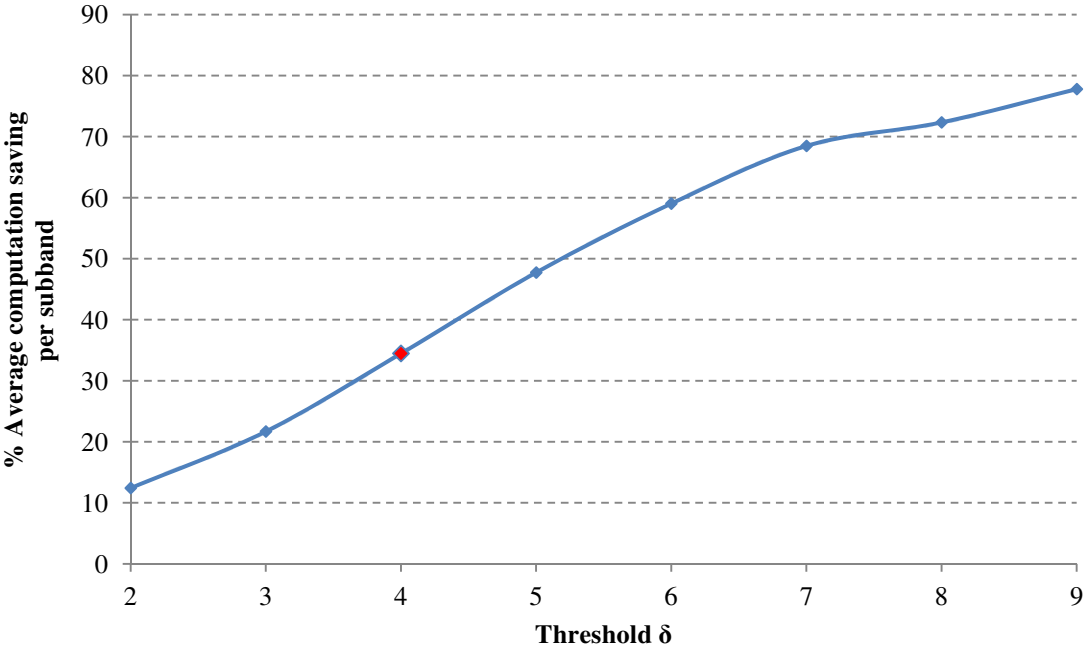
In order to study the effect of $\delta$ on the (i) ratio of CQI measurement reports transmitted to the eNB from the total measurements, (ii) the average error between the actual and predicted CQI values, and (iii) the computational load reduction when compared with the LMS-based PFS, simulations were conducted in MATLAB® based on the parameter set for the CQI-LMS based PFS given in Table 3.3.




(a)




(b)




(c)

Figure 3.8: Efficient CQI-LMS based PFS parameter tuning – Threshold δ. (a) Average proportion of CQI measurements reported to eNB from total per sub-band, (b) Average error of the predicted CQI value per sub-band, and (c) Average computational load saving per sub-band when compared to the LMS PFS

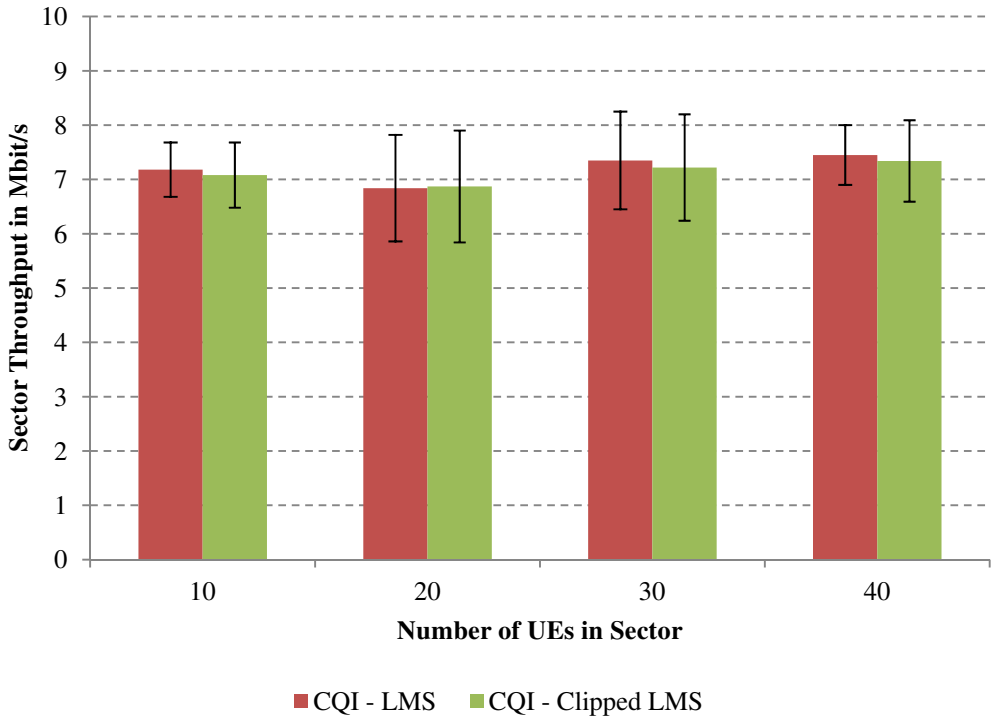
As shown in Figure 3.8(a)-(c), the computation load saving increases with the threshold $\delta$ as an increasing proportion of the input vector $\underline{x}[k]$ will be equal to zero. Nevertheless, this results in degradation in the CQI prediction error caused by the adaptive filter not being updated regularly. Setting $\delta$ to 4 yields the best trade-off between the selected metrics. In order to study the impact of the marginal increase in prediction error on the network performance, the sign-clipped LMS-based PFS was implemented and tested in an LTE system level simulator [55], over 50 random seeds, using the parameters listed in Table 3.2 and 3.3. The mean and standard deviation (shown in parenthesis) of the signalling overhead reduction, PLR, sector throughput and computational load saving for various UE speeds and cell loading with the sign-clipped CQI-LMS based PFS are shown in Table 3.9, 3.10 and Figure 3.9 respectively.

| **UE Speed** | **Prediction Filtering Scheme** | **Number of UEs in Sector** | | | | **Average** |
|---|---|---|---|---|---|---|
| | | **10** | **20** | **30** | **40** | |
| 3km/hr (low mobility) | CQI - LMS | 91.8% (±0.4) | 92.4% (±0.5) | 92.6% (±0.2) | 92.6% (±0.2) | **92.40% (±0.38)** |
| | Sign clipped CQI - LMS | 89.4% (±0.3) | 90.1% (±0.5) | 90.2% (±0.3) | 89.9% (±0.3) | **89.90% (±0.36)** |
| 120km/hr (high mobility) | CQI - LMS | 92.6% (±0.2) | 92.4% (±0.2) | 92.5% (±0.2) | 92.4% (±0.2) | **92.50% (±0.1)** |
| | Sign clipped CQI - LMS | 88.5% (±0.3) | 88.9% (±0.3) | 89.3% (±0.2) | 89.2% (±0.1) | **89.00% (±0.36)** |

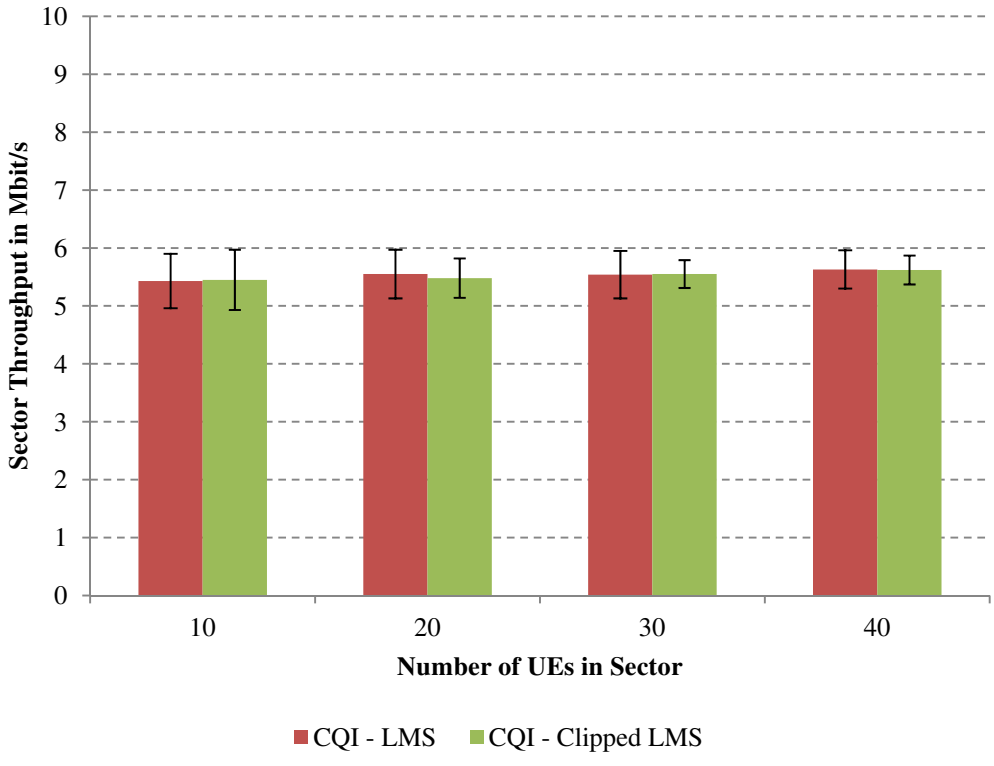
Table 3.9: CQI feedback reduction – Efficient CQI-LMS based scheme

| **Prediction Filtering Scheme** | **UE Speed** | |
|---|---|---|
| | **3km/hr (low mobility)** | **120km/hr (high mobility)** |
| CQI - LMS | 6.7% (±0.7) | 17% (±2.3) |
| Sign clipped CQI - LMS | 8.6% (±1.3) | 17.2% (±1.3) |

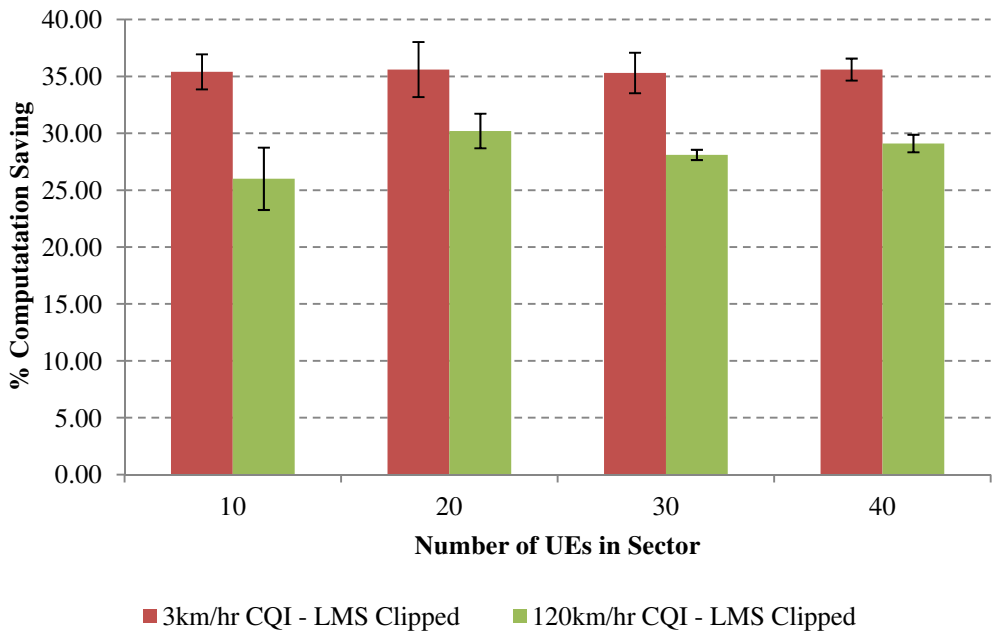
Table 3.10: Packet loss rate – Efficient CQI-LMS based scheme


10
9
8
7
6
5
4
3
2
1
0
Sector Throughput in Mbit/s
10
20
30
40
Number of UEs in Sector
CQI - LMS
CQI - Clipped LMS

(a)


10
9
8
7
6
5
4
3
2
1
0
Sector Throughput in Mbit/s
10
20
30
40
Number of UEs in Sector
CQI - LMS
CQI - Clipped LMS

(b)




(c)

Figure 3.9: Sector throughput and computational load comparison (standard deviation is indicated by vertical bars). (a) Sector throughput @3km/hr, (b) Sector throughput @120km/hr, and (c) % computational load reduction

Whilst considering a marginal loss in performance in terms of packet loss and uplink signalling reduction, the sign-clipped LMS PFS achieves a mean computation load saving of 35% and 26% at 3km/hr and 120km/hr respectively when compared to the LMS predictive filtering scheme.

### 3.1.8 Evaluating the LMS-CQI PFS using field test measurements

In order to further evaluate the performance of the proposed LMS-CQI PFS, radio measurements were collected from a live commercial LTE network, having the relevant radio parameters shown in Table 3.11. The measurements were collected using industry standard test equipment [58] installed in a purposely setup drive test vehicle. With reference to Figure 3.10 and in accordance with [59], the UE is connected to a pair of omni-directional antennas mounted on top of the drive test vehicle through a combining network exhibiting 12dB attenuation thereby compensating for the antenna gain and vehicle penetration loss. The drive test route, shown in Figure 3.11 exhibits a rich multi-path channel leading to highly varying radio conditions.

| Parameter | Value |
|---|---|
| Frequency Band | 1.8GHz |
| Bandwidth | 20MHz |
| CQI reporting period | 20ms |
| CQI reporting type | Wideband |
| MIMO Type | 2x2 TM4 |
| Transmit Power | 2 x 40W |

Table 3.11: Relevant commercial LTE network radio parameters

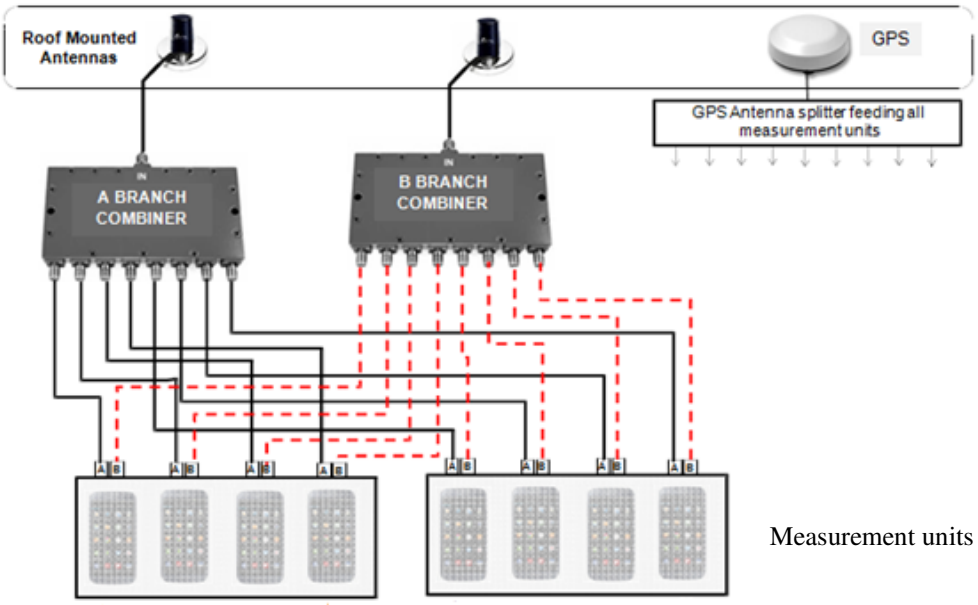



Figure 3.10: Test vehicle setup showing the antenna combining network and connection with UE

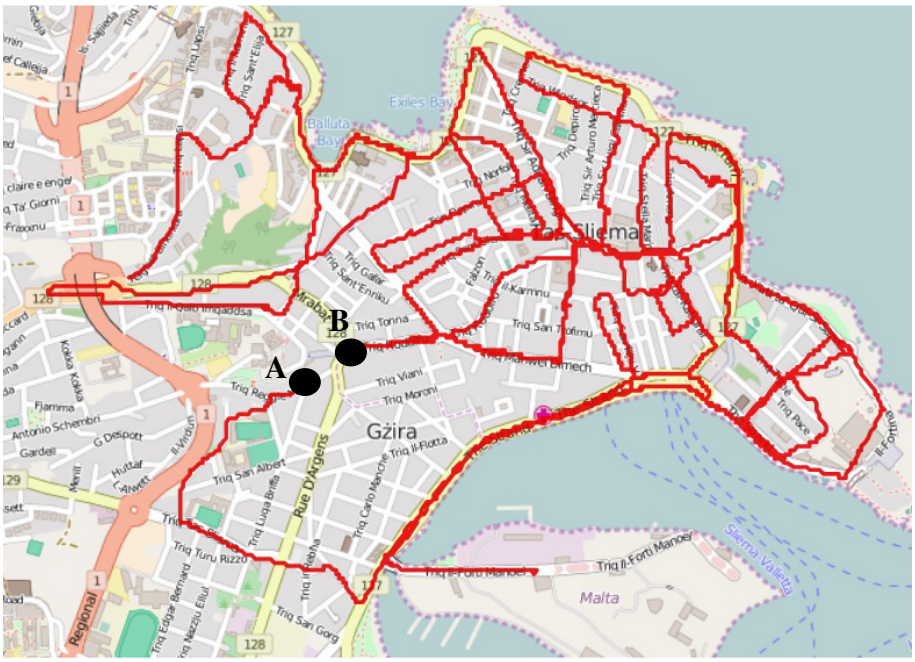



Figure 3.11: Test route – Starting at point A and stopping at point B [60]

The CQI measurements conducted by the UE and reported to the network throughout the drive test were collected and fed to the proposed LMS-based prediction filtering scheme described in Algorithm 1. In contrast with the simulated environment, the tested commercial network was configured to use wideband CQI reporting and thus the proposed scheme was adapted to use a single LMS adaptive filter covering the whole bandwidth in both the UE and eNB. Out of the 13,944 CQI measurements reported, only 2,383 measurements were transmitted representing a CQI feedback signalling reduction of approximately 83% whilst achieving an absolute CQI prediction error of 1.2. The distribution of the CQI prediction error is shown in Figure 3.12 whilst a sample of the actual and predicted measurements, indicating the two modes of operation described in section 3.1.3 is shown in Figure 3.13. As shown in Figure 3.12, the predictive filtering scheme tends to overestimate the CQI value which might lead to an increased propensity to physical layer errors. This might not be significantly detrimental for the case of non-real time traffic, which make use of AM RLC and upper layer retransmission protocols, however such link adaptation mismatch would lead to a lower quality of service for real time traffic such as low latency video.

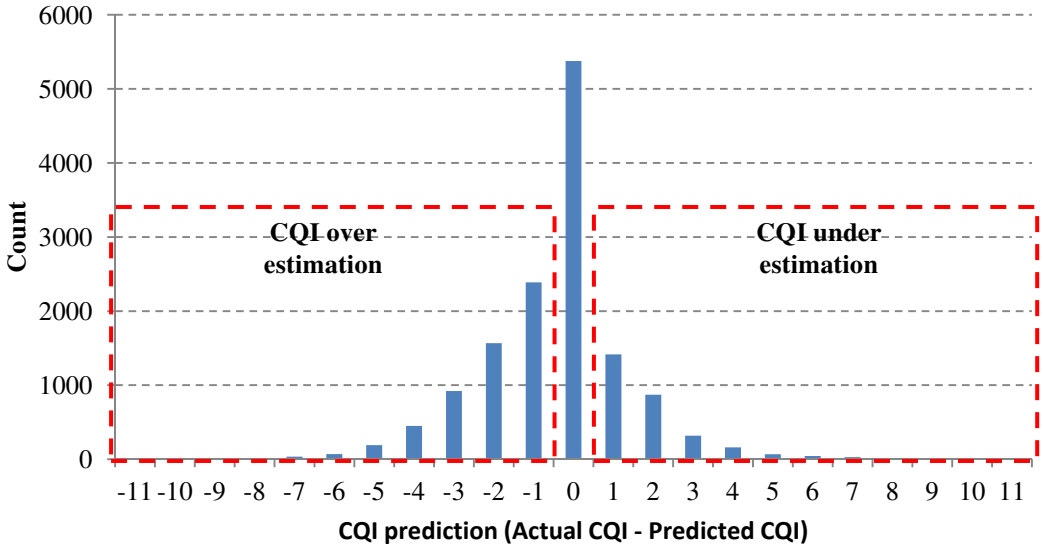



Figure 3.12: Distribution of CQI prediction error

With reference to Figure 3.13, when the predictive filtering algorithm is in prediction mode, the CQI measurements collected by the UE are not transmitted to the eNB. Meanwhile, the predictors attempt to follow the variations in the channel conditions while keeping track of the prediction error. As the prediction error degrades, the algorithm switches back to the training mode and CQI feedback measurement reports

are transmitted to the eNB thereby allowing the re-training of the adaptive filters. Once the prediction errors falls below the error budget threshold the algorithm switches into the prediction mode thereby ceasing the transmission of CQI feedback measurements.

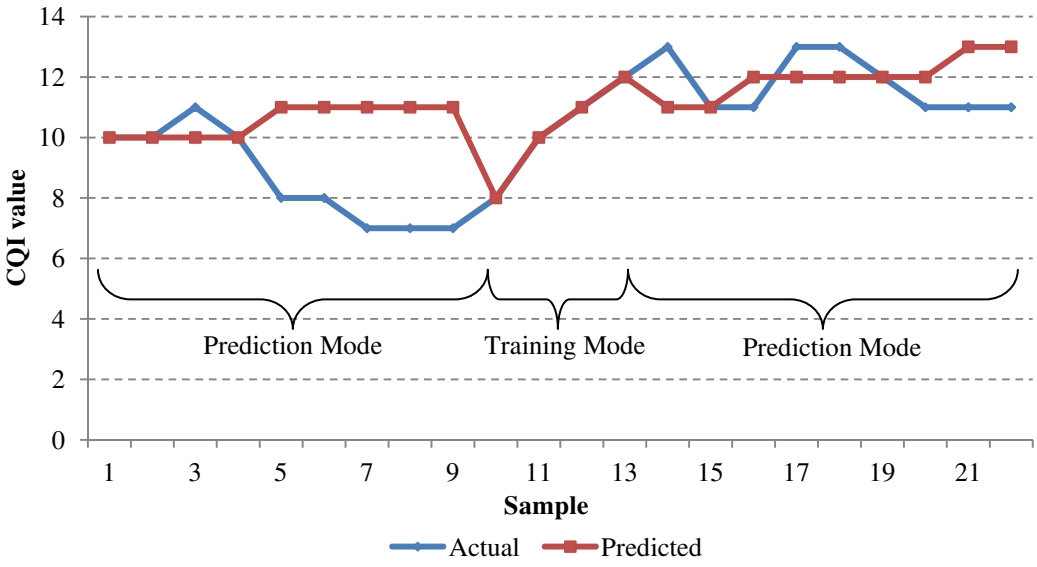
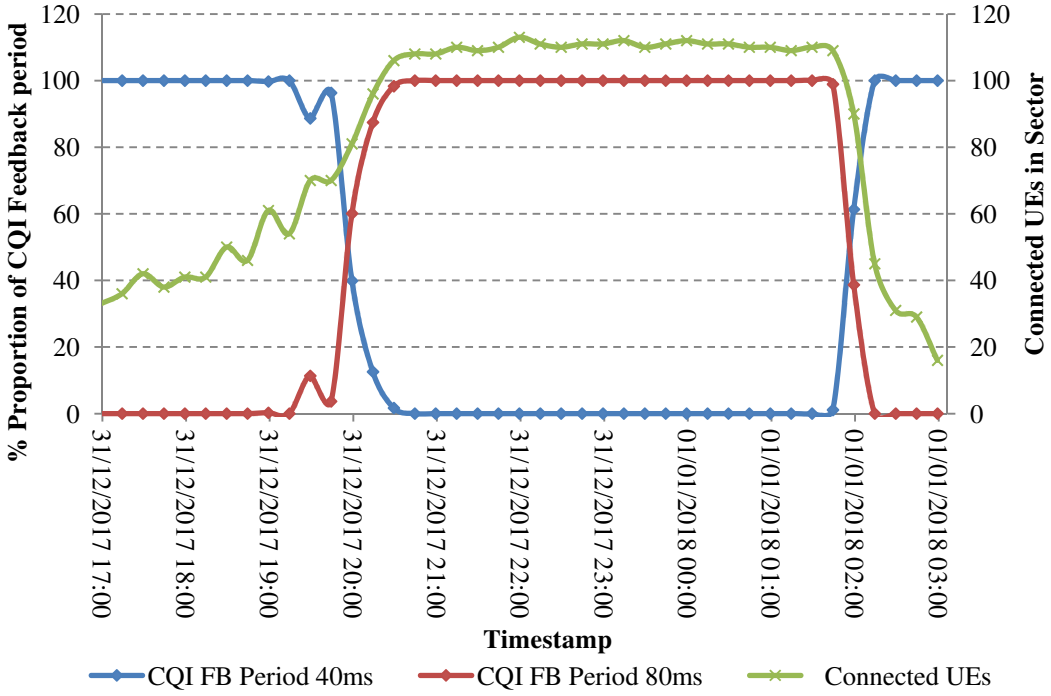



Figure 3.13: Actual and predicted CQI field test measurements [60]

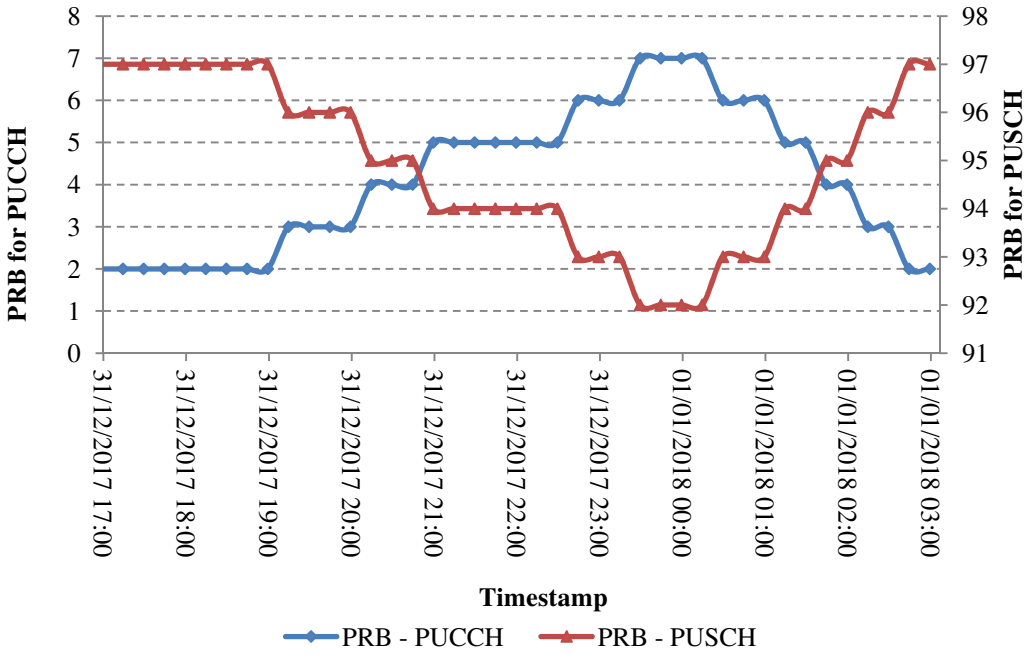
## 3.2 LMS-based predictive filtering scheme in crowd event scenarios

A major problem associated with crowd event scenarios is the uplink signalling feedback congestion. In order to control uplink signalling feedback in such scenarios, network operators typically resort to wideband feedback compression and feedback time adaptation coupled with dynamic PUCCH resource allocation. Nevertheless, as evidenced by Figure 3.14 that shows measurements collected from an eNB in a live commercial network covering a typical crowd event scenario, these actions are not enough. With reference to Figure 3.14 (a), the increase in the number of connected users in the sector leads to the dynamic PUCCH algorithm in the eNB to increase the feedback time interval from 40ms to 80ms. Moreover, as shown in Figure 3.14 (b), as the number of connected users continues to increase, the PUCCH region is incrementally expanded, thereby reducing the number of available PRB for the PUSCH, until it reaches a pre-configured limit. Beyond this point the eNB starts rejecting users due to lack of PUCCH resources. Thus, the use of effective feedback reduction schemes would benefit the end user experience. With reference to the LMS-

based PFS proposed in section 3.1 and the field tests results presented in section 3.1.8, this scheme could easily be adapted to the wideband CQI feedback case and can offer significant benefits especially in crowd event scenarios.




(a)




(b)

Figure 3.14: PUCCH measurements collected from a live eNB in a commercial network (a) CQI Feedback period adaptation, and (b) PUCCH and PUSCH dynamic allocation

### 3.2.1 Crowd event scenario simulation

A common network design strategy implemented by mobile network operators to tackle crowd events is to partition the crowd into smaller areas by the use of additional radio sectors. In this study, a typical sport event scenario is considered whereby users are seated in a three-tier football stadium, shown in Figure 3.15, which has been partitioned into several areas with a number of radio sectors.


Figure 3.15: Three-tier football stadium and radio sectors under study [61]

With reference to Figure 3.15, sector B in the south stand is the sector under study, whilst the adjacent sector A and C contribute to interference. Considering the seating configuration parameters which are aligned with the Fédération Internationale de Football Association (FIFA) football stadium technical guidelines [62], shown in Table 3.12, each area has a seating capacity of 1540 spectators. Moreover, considering the activity parameters shown in Table 3.13, the maximum number of simultaneous active UEs, at any one time, in the sector is approximately 130.

| Parameter | Value |
|---|---|
| Area dimension | 30m x 30m |
| Front/Back passage ways | 1m |
| Lateral passage ways | 2m |
| Inter-sear spacing in x-direction | 0.5m |
| Inter-sear spacing in y-direction | 1m |

Table 3.12: Seating configuration parameters

| Parameter | Value |
|---|---|
| Operator market share | 40% [63] |
| Spectators sharing content in crowd events | 54% [64] |
| Simultaneous user ratio | 40% |

Table 3.13: Activity parameters

### 3.2.2 Radio Design

The use of multiple radio sectors in a confined area such as a football stadium requires strict control of the radio network design in order to minimise co-channel interference. This is achieved through the use of sculpted radiation pattern antennas, shown in Figure 3.16, having a sharp -3dB cut-off point thereby minimising the cell overlap. As shown in Figure 3.17, these antennas are typically fixed to the underside of the stadium roof thereby allowing the signal to be focused in a specific area of the crowd whilst minimising inter-cell interference.

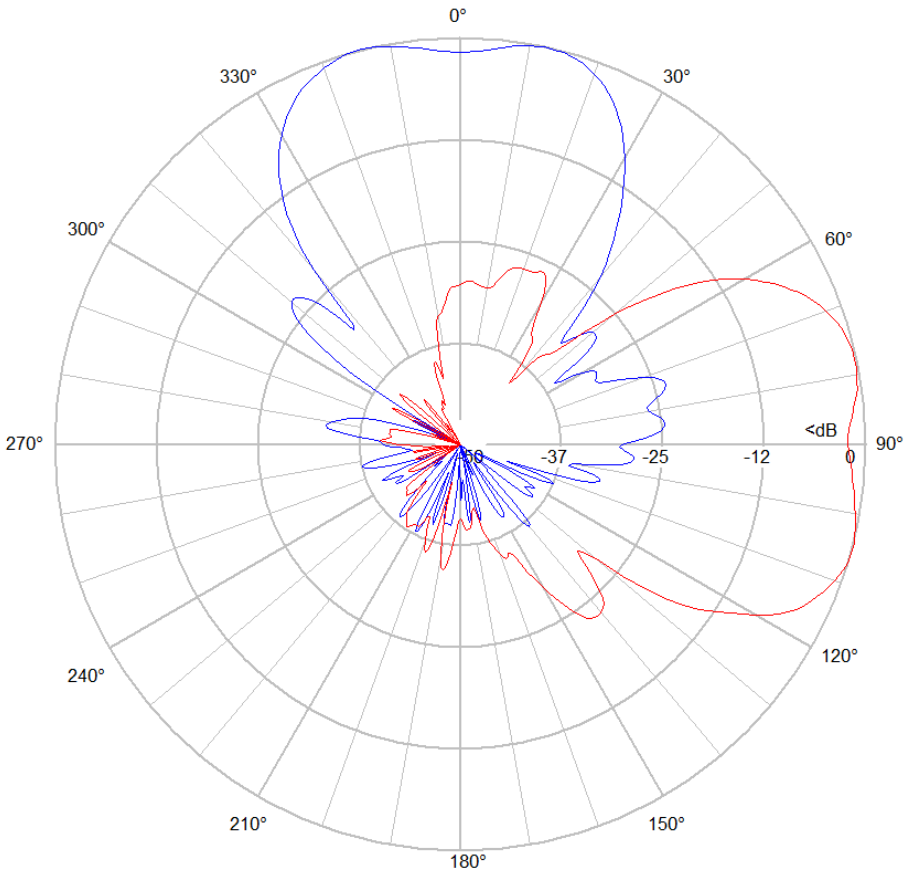



Figure 3.16: Antenna radiation pattern (Horizontal plane – red, Vertical plane – blue) [65]

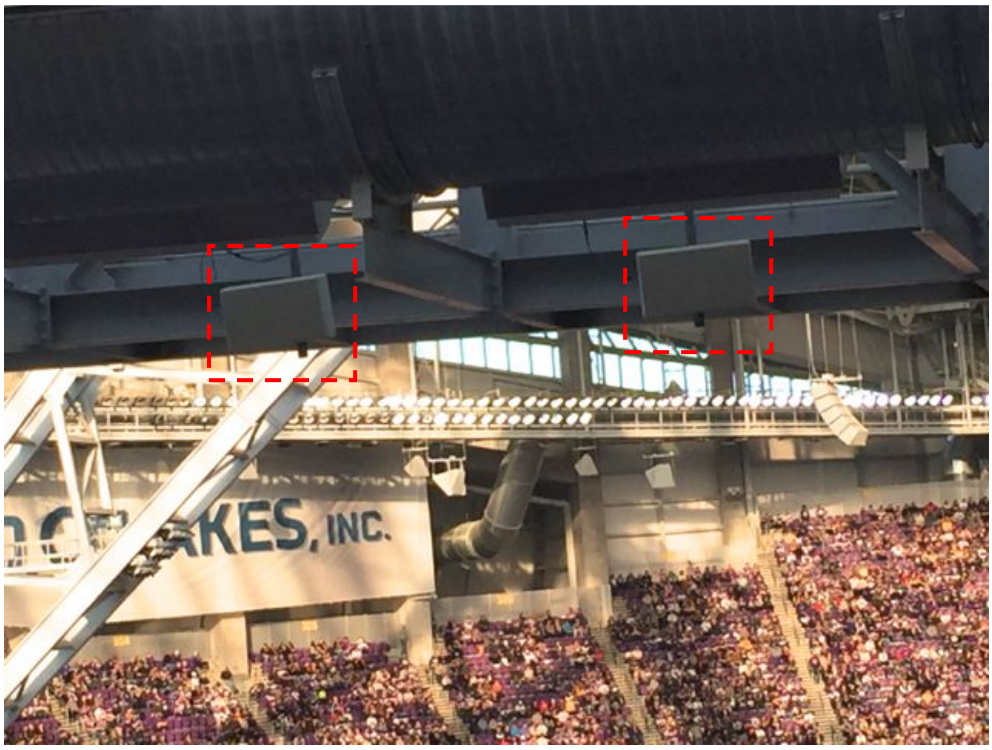



Figure 3.17: Antenna installation in stadium [66]

A radio coverage simulation of the area showing the reference signal received power (RSRP) in dBm provided by sector B at 0.5m above ground level (AGL) is shown in Figure 3.18 whilst the relevant simulation parameters are provided in Table 3.14.

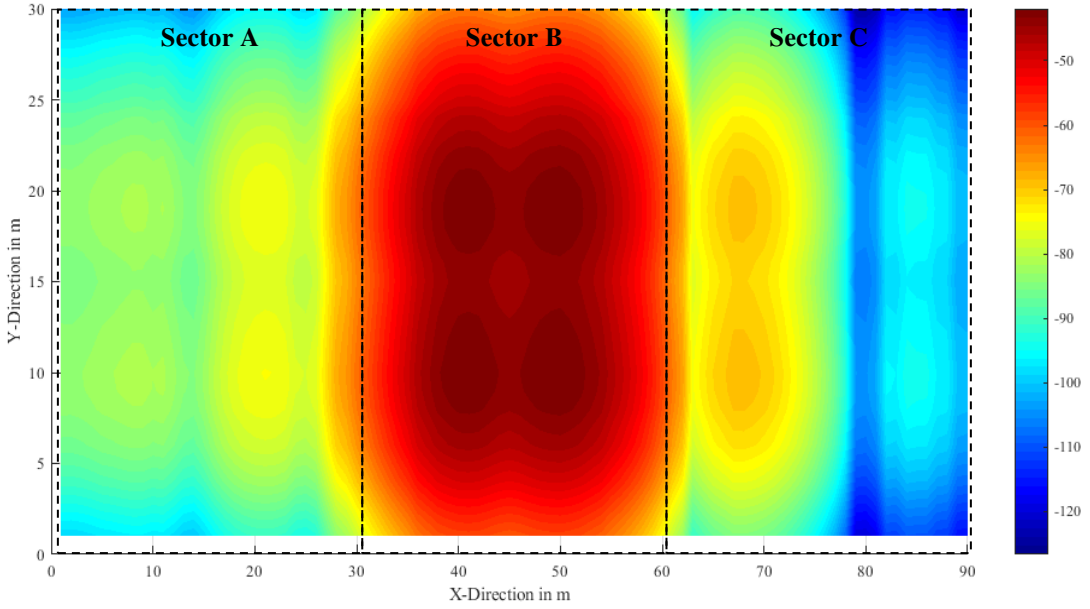



Figure 3.18: RSRP in dBm of Sector B at 0.5m AGL (Elevation)

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| CQI Feedback Reduction | LMS CQI Predictive Filtering (see Table 3.3) |
| Traffic Model | Full Buffer |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log($d$+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | {50, 70, 90, 110, 130} |

Table 3.14: Simulation parameters – wideband LMS-based PFS

### 3.2.3 Simulation Results

In order to verify the efficacy of the wideband LMS-based PFS, when applied to a crowd event scenario, the setup described in Section 3.2.1 and 3.2.2 was implemented and tested in an LTE system level simulator [55], over 50 random seeds, using the parameters shown in Table 3.14. The feedback reduction and average sector throughput, when compared to the 3GPP wideband scheme, for various users in the sector are shown in Table 3.15 and Figure 3.19 respectively whilst the average packet loss rate for the 3GPP wideband scheme and the wideband LMS-based PFS is shown in Table 3.16.

| UE in Sector | % Feedback reduction |
|---|---|
| 50 | 86.5 (±0.3) |
| 70 | 87.5 (±0.6) |
| 90 | 86.7 (±0.4) |
| 110 | 86.9 (±0.6) |
| 130 | 87.1 (±0.5) |

Table 3.15: Feedback reduction (standard deviation is shown in parenthesis)

| 3GPP Wideband | LMS WB |
|---|---|
| 0.9 (±0.1) | 3.3 (±0.16) |

Table 3.16: Average Packet Loss Rate (standard deviation is shown in parenthesis)

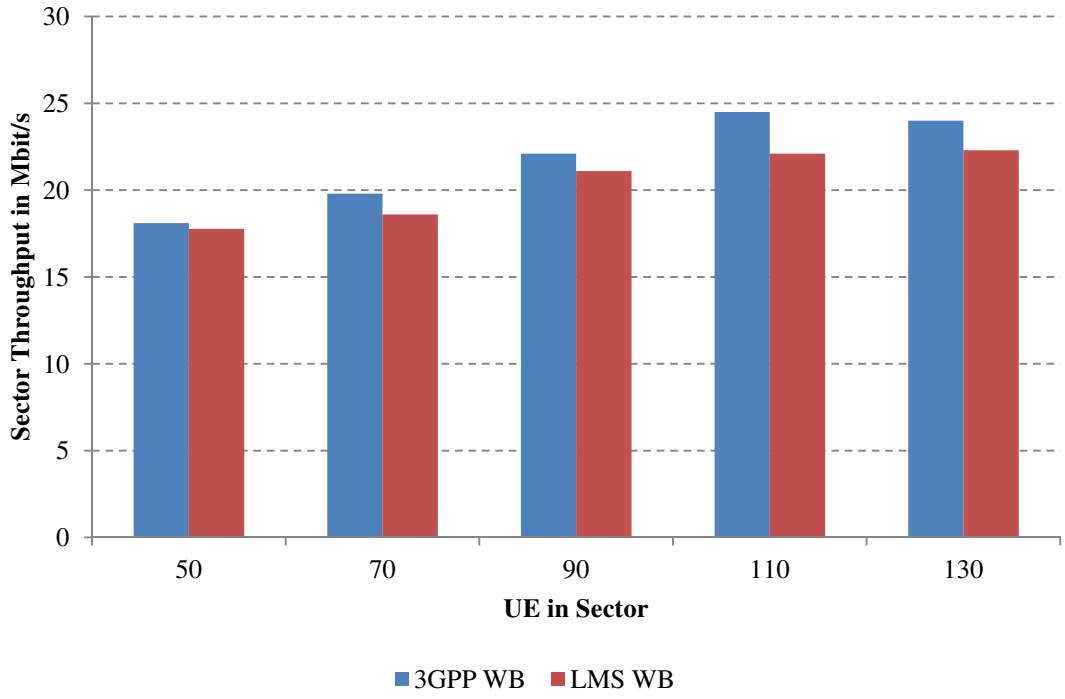



Figure 3.19: Sector throughput comparison – 3GPP wideband and wideband LMS PFS

Although the deployment of the LMS-based PFS in a crowd event scenario results in a marginal loss in performance when compared to the 3GPP standard wideband scheme, it manages to achieve a respectable average feedback signalling reduction of 87%.

## 3.3 Wideband CQI feedback clustering technique

An alternative to the predictive filtering scheme presented in section 3.1 and its application to a crowd event scenario in section 3.2 is the use of a CQI clustering scheme which exploits the slow varying radio channel conditions of static users in the same radio sector. The operation of the wideband CQI feedback clustering technique is shown in Figure 3.20. During the initialisation phase, the UEs are operating in *legacy_mode* whereby periodic wideband CQI measurements conducted by the UEs are reported in order to allow the eNB to build a CQI report history for each connected UE. The eNB keeps track of the number of CQI reports received per UE and when a UE reaches $\tau_{UE}$ reports, the eNB updates the status of the UE to *cluster_wait* mode. During this mode, the UE continues to report the CQI measurements however the eNB will overwrite old CQI measurement values thereby ensuring that the eNB retains the

latest $\tau_{UE}$ reports for each UE. When all connected UEs are in *cluster_wait* mode, the eNB initiates the cluster formation phase composed of three main operations namely:

1. **Compute CQI statistics:** For each UE, the eNB computes the mean $\mu_i$ and variance $\sigma_i^2$ using the last $\tau_{UE}$ CQI measurement reports and normalises the data.

2. **Clustering:** eNB applies a data clustering algorithm on the two-dimensional data vector, composed of the computed normalised statistics in step 1, to cluster users into *k* clusters and assigns a cluster ID to each UE. In this work, the *k*-means algorithm was chosen in view of its simplicity, robustness and effectiveness.

3. **Elect cluster leaders:** For each cluster; calculate the centrality of each UE within the cluster using (3.11). The centrality of a UE, $C_i$, defines how close a UE, *i*, is to the other UEs, *j*, in the same cluster, based on the computed normalised statistics. The UE with the lowest centrality value (i.e. highest centrality) is then elected as cluster leader, $UE_{CL}$. This process ensures that the cluster leader CQI reports are representative of the whole cluster.

$$C_i = \sum_{j \neq i} \sqrt{\left(\mu_i - \mu_j\right)^2 + \left(\sigma_i^2 - \sigma_j^2\right)^2} \tag{3.11}$$

After the cluster formation phase is complete, the eNB moves to the clustered phase during which the elected cluster leaders will continue sending CQI reports to the eNB. The eNB uses these reports as representative CQI reports for the other UEs in the cluster whilst the other UEs in the cluster are set to *CQI_Suppressed* mode indicating that these UEs will not transmit CQI measurements reports. The eNB remains in the clustered phase until it receives $\gamma_{CL}$ CQI measurement reports from all the cluster leaders, after which, the eNB moves back to the initialisation phase and restarts the process.

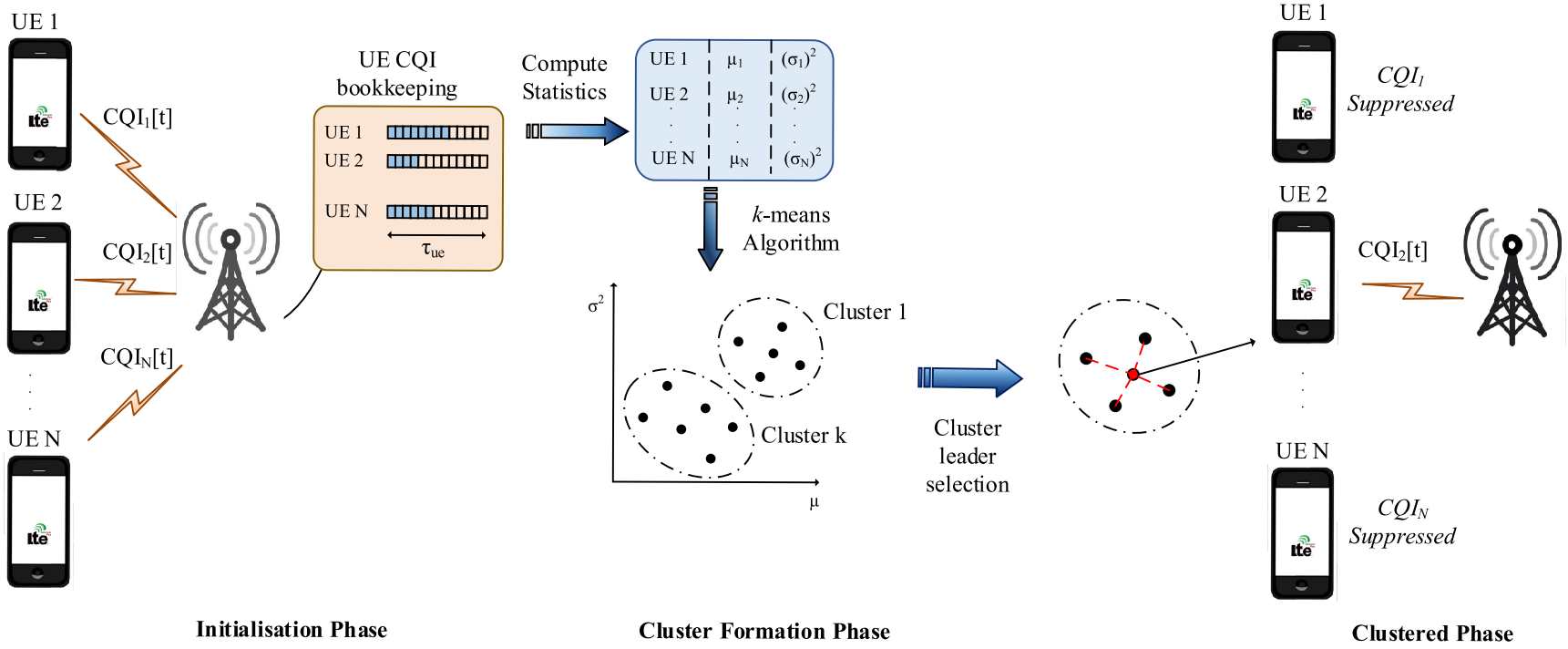



Figure 3.20: Wideband CQI feedback clustering scheme [67]

### 3.3.1 Statistical Model Formulation for the number of feedback clusters

The required number of clusters, *k*, is a critical input parameter for the majority of data clustering techniques. Nevertheless, choosing *k* automatically is notoriously difficult and is usually either fixed or else an iterative technique to merge/split clusters is adopted [68] at the expense of increased computational complexity. In this work, we formulate a statistical model able to estimate the number of clusters based on the number of active users in the radio sector using domain-specific knowledge. In order to collect data for the proposed statistical model, the cluster formation algorithm, described earlier, was implemented and tested in MATLAB® using CQI measurements generated from several UEs, simulated in a system level simulator [55], for various values of *k*. The parameters used in these simulations are provided in Table 3.17.

With reference to Figure 3.21, it is evident that increasing the number of clusters, *k*, results in more compact clusters as the sum of within-cluster point-to-cluster centroid distance, *D*, decreases. Nevertheless, the increase in *k* implies an increased number of cluster leaders transmitting CQI reports to the eNB. In order to strike a balance between these two opposing factors, a cost function relating the normalised CQI report transmission cost, $\eta$, and cluster compactness metric, *D*, is used to understand the variation of the total cost, $\phi$, with the number of clusters *k*. Considering the reporting period and CQI message length shown in Table 3.17, the cost function is given by:

$$\begin{aligned} \phi(k) &= \alpha \cdot D + \beta \cdot \eta \\ \eta &= 200k \quad \text{(bits/sec)} \end{aligned} \tag{3.12}$$

where $\alpha$ and $\beta$ are weighting constants empirically set to 15 and 1.2 respectively.

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| Traffic Model | Full Buffer |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log($d$+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | {50, 70, 90, 110, 130} |
| $\tau_{UE}$ and $\gamma_{CL}$ | 100 and {3000, 5000, 8000} |

Table 3.17: Simulation parameters – CQI clustering scheme

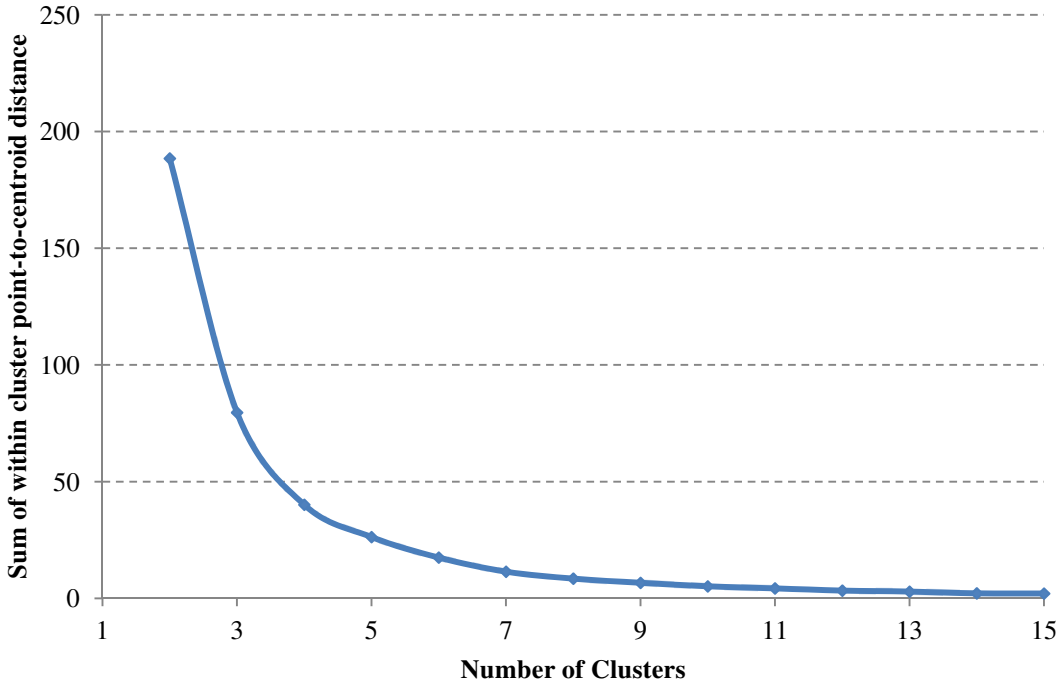



Figure 3.21: Variation in cluster compactness with number of clusters

(number of UEs in sector = 130)

Plotting the cost function on a logarithmic scale for a different number of connected UEs, $T_u$, yields a set of upward concave parabolas shown in Figure 3.22. Further inspection shows that an optimal number of clusters, $k_{opt}$, exist for different values of $T_u$.

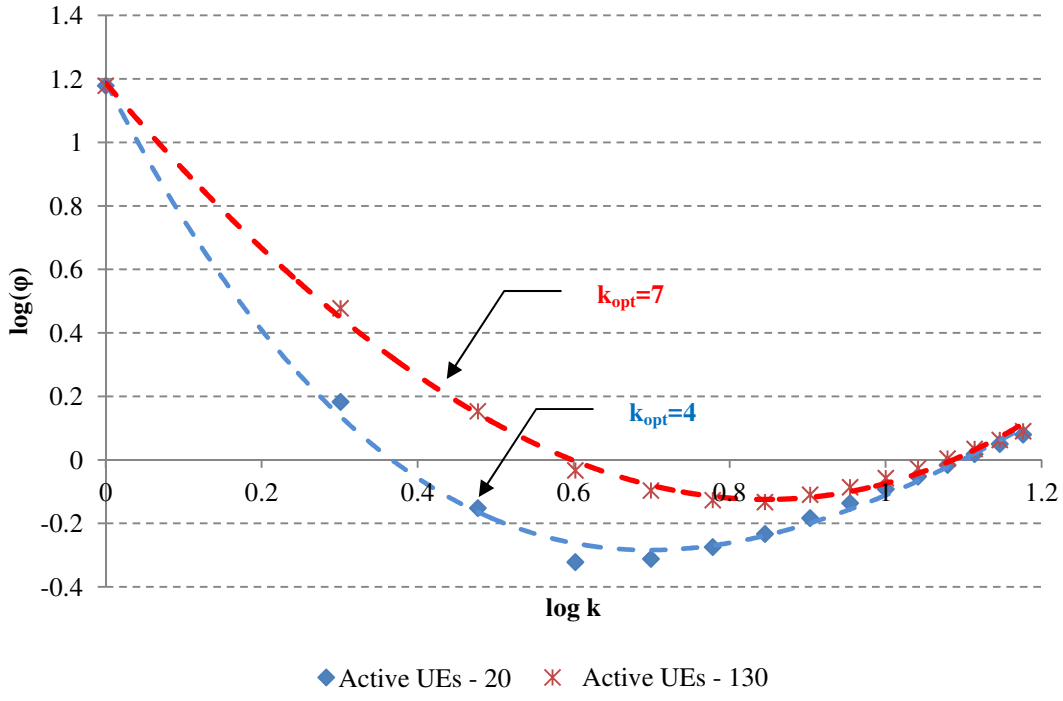



Figure 3.22: Cost function dynamics

Repeating this analysis for various values of $T_u$ and finding the optimal number of clusters for each case reveals the inverse relationship shown in Figure 3.23. As the number of connected UEs increases, the optimal number of clusters rises until it reaches an asymptotic value. This is expected as the CQI transmission cost increases linearly with $k$ thereby enforcing a hard limit on the optimal number of clusters. Meanwhile, at low values of $k$, the inverse relationship will return a progressively lower number of clusters. Nevertheless, experimental results show that below 15 active users (which does not constitute a crowd event in any case), the proposed clustering scheme benefits diminish and hence it would be more appropriate to switch back to the legacy CQI reporting scheme, whereby UEs report the CQI directly to the eNB. The data was further analysed in IBM SPSS© statistical package and an inverse model, given by (3.13) and having a coefficient of determination, $R^2$, of 0.966 was derived using the least square estimation technique [69].

$$\log(k) = -2.61 \cdot \left( \frac{1}{T_u} \right) + 0.867 \tag{3.13}$$

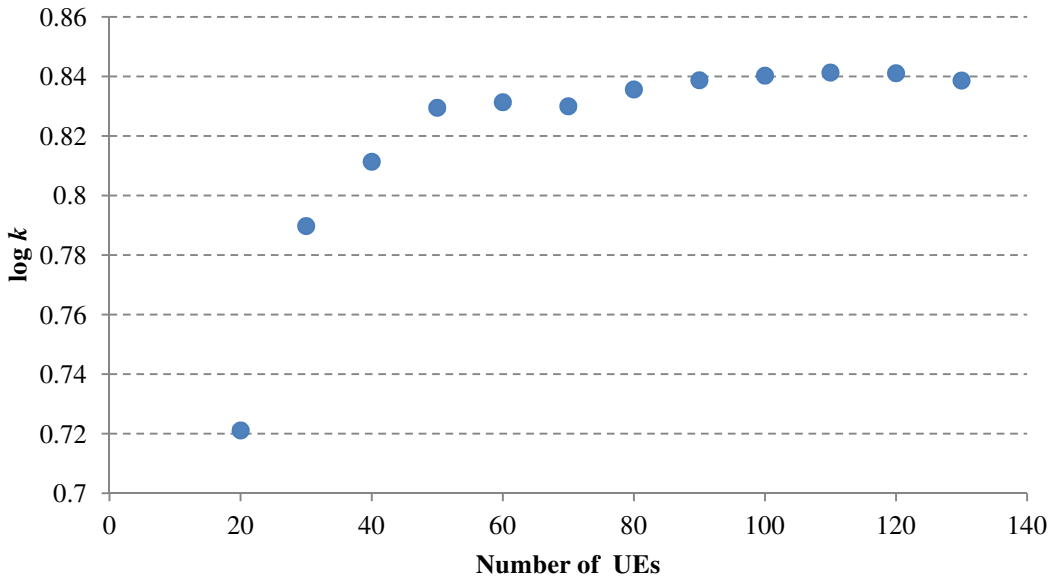



Figure 3.23: Inverse relationship between log($k$) and $T_u$

### 3.3.2 Simulation Results

The proposed CQI clustering together with the derived statistical model given by (3.13) were implemented in a system level simulator [55] to verify its effectiveness in terms of average cell throughput and feedback reduction. Simulations, using the parameters given in Table 3.17 were carried out over 50 random seeds in order to achieve statistical relevant results. The average sector throughput and feedback reduction, when compared to the 3GPP standardised wideband feedback scheme, for various values of $T_u$ and $\gamma_{CL}$ set to 5000, are shown in Figure 3.24 and Table 3.18 respectively.

The proposed CQI feedback clustering scheme manages to achieve an average CQI feedback reduction of 88.2% when compared to the 3GPP wideband scheme. This level of feedback reduction was achieved whilst maintaining approximately the same level of performance, in terms of sector throughput and PLR, as evidenced in Figure 3.24 and Table 3.19. Moreover, with reference to Table 3.18, the feedback reduction gradually increases as the clustering becomes more effective with increasing number of connected UEs.

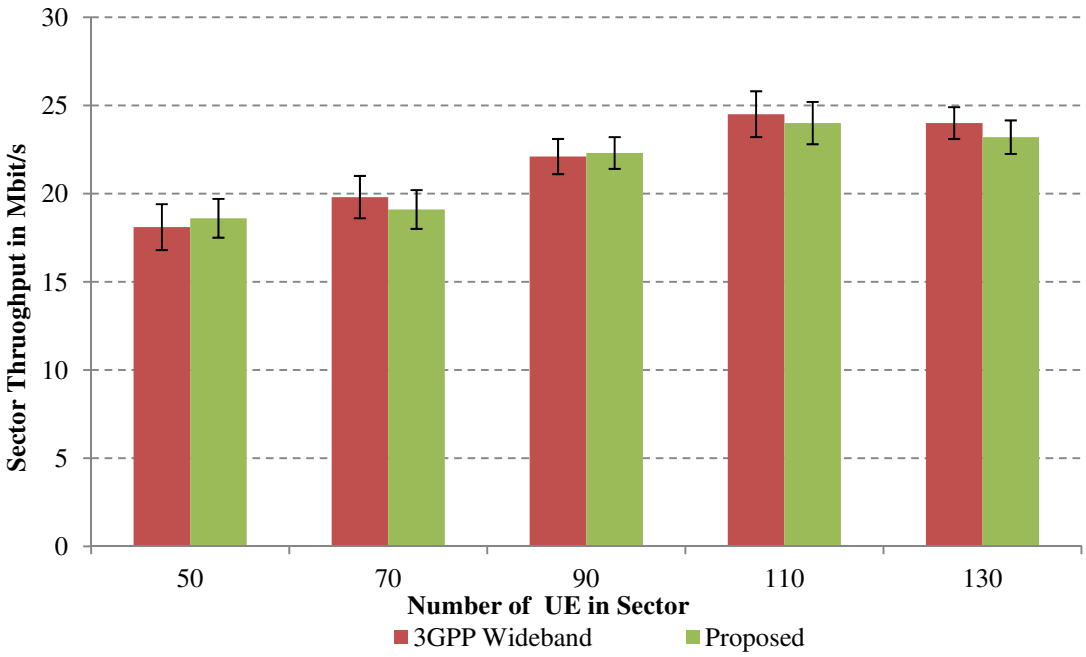



Figure 3.24: Sector throughput comparison for various numbers of connected UEs in sector (standard deviation is indicated by vertical bars)

| Number of UEs | % Feedback Reduction |
|---|---|
| 50 | 83.1 (±0.3) |
| 70 | 86.9 (±0.5) |
| 90 | 89.1 (±0.4) |
| 110 | 90.4 (±0.6) |
| 130 | 91.4 (±0.3) |

Table 3.18: Feedback reduction with respect to 3GPP wideband scheme (standard deviation is shown in parenthesis)

| 3GPP Wideband | CQI Feedback Clustering |
|---|---|
| 0.9 (±0.1) | 1.5 (±0.19) |

Table 3.19: Average Packet Loss Rate (standard deviation is shown in parenthesis)

The effect of the parameter $\gamma_{CL}$ on the performance of the proposed scheme was also investigated as shown in Table 3.20. Intuitively, as $\gamma_{CL}$ increases, we prolong the clustered phase thereby achieving a marginally improved feedback reduction. However, this comes at the expense of degradation in sector performance due to the user clustering and cluster leader selection becoming increasingly sub-optimal.

| $\gamma_{CL}$ | **3000** | **5000** | **8000** |
|---|---|---|---|
| Sector Throughput in Mbit/s | 22.3 | 22.2 | 21.1 |
| % CQI Feedback Saving | 89.1 | 90.3 | 90.9 |

Table 3.20: Effect of parameter $\gamma_{CL}$ on the performance of the proposed CQI feedback clustering scheme (Number of UE's = 90)

## 3.4 Summary

This chapter has presented a thorough study on the use of a UE-assisted PFS for sub-band CQI feedback compression. The proposed technique is based on short-term channel prediction and is able to lower the CQI signalling overhead through the use of adaptive filters at the UE and eNB side. Two schemes were implemented and tested with the CQI-LMS based PFS ranking better than the NLMS based scheme with up to 92.4% CQI signalling overhead reduction. This level of feedback signalling reduction was achieved whilst maintaining the same level of sector throughput and keeping the PLR within 5% when compared to the standardised 3GPP eNB configured sub-band feedback scheme.

Although the adaptive filters used in the proposed scheme exhibit a low computational load and memory complexity, an efficient CQI-LMS PFS able to reduce the average computational load by 35% in low mobility has also been studied. This was achieved in lieu of a marginal loss in performance compared to the CQI-LMS based scheme. A comparison of the proposed techniques, showing the improvement when compared to [49] and the 3GPP standardised scheme [23], is shown in Table 3.21. Even though the proposed scheme has been designed for sub-band CQI feedback compression, it is easily adapted to the wideband CQI feedback reporting scenario by simply using a single LMS adaptive filter in both the UE and eNB. This was tested using field test measurements collected from a live commercial LTE network configured to use wideband CQI reporting yielding a CQI feedback signalling reduction of 83% and an absolute CQI prediction error of 1.2.

| Scheme | Mobility Characteristics | Feedback Reduction | PLR | Throughput degradation | Complexity |
|---|---|---|---|---|---|
| 3GPP eNB configured sub-band feedback [23] | Low Mobility | - | 2.5% (±0.3) | - | - |
| | High Mobility | - | 17.4% (±1.7) | - | - |
| Chimento *et al* [49] | Mixed - generally low mobility | 77% | 10% | Focus of study is on PLR | High. Refer to section 2.4.2 |
| LMS-CQI prediction filtering scheme | Low Mobility | 92.4% (±0.4) | 6.7% (±0.7) | 1.1% (±0.7) | 0.9Mflops and 4Kflops in the eNB and UE respectively |
| | High Mobility | 92.5% (±0.1) | 17% (±2.3) | 0.6% (±0.2) | |
| Sign Clipped LMS-CQI prediction filtering scheme | Low Mobility | 89.9% (±0.4) | 8.6% (±1.3) | 2.2% (±0.5) | -35.5% |
| | High Mobility | 89% (±0.4) | 17.2% (±1.3) | 0.8% (±0.8) | -28.4% |

Table 3.21: Comparison of the key performance indicators for the adaptive filtering schemes

The second part of this chapter dealt with a study on the use of the wideband UE-assisted LMS based predictive filtering scheme and a UE clustering scheme in a crowd event scenario having a relatively high number of stationary users in the sector. As evidenced by the simulation results, summarised in Table 3.22, both schemes yield significant feedback signalling reduction with the UE clustering scheme achieving better overall performance in terms of average sector throughput and packet loss rate. Nevertheless, both schemes could be considered as potential candidates for use in crowd event scenarios.

| Scheme | % Feedback Reduction | Throughput in Mbit/s | % PLR |
|---|---|---|---|
| 3GPP Wideband Scheme [23] | - | 21.7 | 0.9 |
| Wideband UE assisted LMS based prediction filtering scheme | 87 | 20.4 | 3.3 |
| UE Clustering scheme | 88.2 | 21.4 | 1.5 |

Table 3.22: Comparison of the key performance indicators for the wideband schemes

# Chapter 4 High Efficiency Video Coding

This chapter starts with an introduction to the latest standard video encoder, namely the High Efficiency Video Coding (HEVC) or H.265, together with its Multi-View (MV)/3D extensions and the concept behind view synthesis, which enables the rendering of novel views and forms the foundation of FVV. The next section introduces the depth-texture view bit rate allocation problem together with a review of works aimed at tackling this challenge. Finally, the last section focuses on the need of rate control coupled with a review of works on this key video encoder element.

## 4.1 Introduction to High Efficiency Video Coding (HEVC)

The preceding standard, H.264/AVC, garnered significant attention within the research community and enjoyed widespread commercial use ranging from broadcast of high definition (HD) TV signals over satellite, cable and terrestrial transmission systems to security applications, internet and mobile network video [70]. Nevertheless, the new generation of higher resolution and quality content coupled with stereo or multi-view display mandated the need of a video encoder with a superior coding efficiency than that offered by the H.264/AVC capabilities. This need was further accentuated with the rapid increase in mobile video traffic imposing severe challenges to mobile network operators. For these reasons, the Joint Collaborative Team on Video Coding (JCT-VC) formed between the ITU-T VCEG and the ISO/IEC MPEG standardisation bodies released the first edition of the H.265/HEVC standard in January 2013 [70]. The key principles governing the design of the new standard include [70]:

- Handling of higher video resolution beyond HD video format;
- Increased use of parallel processing architectures;
- Significant bit-rate savings for the same quality relative to the performance of H.264/AVC.

### 4.1.1 HEVC video coding

With reference to Figure 4.1, HEVC adopts the same block-based hybrid approach used in legacy video compression standards. Each picture is first split into blocks and a block-wise prediction residual is computed between the original block and its prediction obtained using either intra-picture prediction or inter-picture prediction. Intra-picture prediction, used for the first picture of a video sequence and subsequent pictures at each clean random access point, makes use of previously decoded neighbouring samples from the same picture whilst inter-picture prediction uses corresponding regions of previously decoded pictures and is applied for all the remaining pictures of a sequence or between random access points. In inter-picture prediction the encoding process involves the selection of motion data composed of a motion vector (MV) and associated reference picture, to be used by the decoder in order to predict the samples of each block. The computed residual signal of the intra- or inter-picture prediction for each block is transformed using a linear spatial transform and the transform coefficients are scaled, quantised and entropy coded prior to being transmitted together with the prediction information [70].

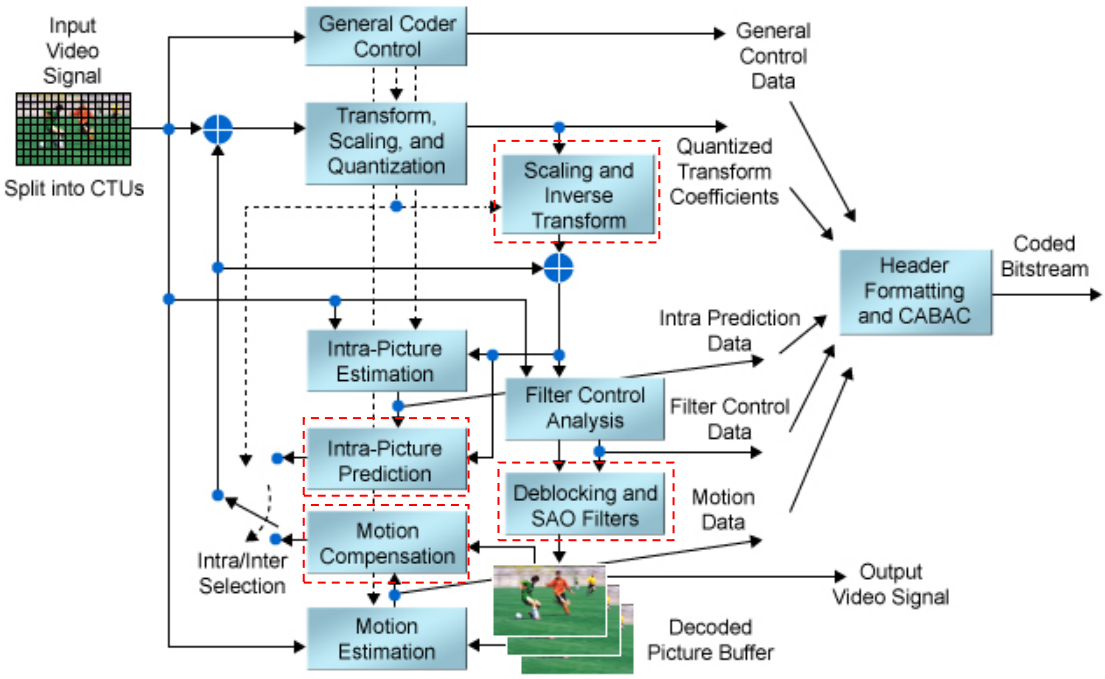



Figure 4.1: HEVC video encoder (decoder modelling blocks shown in red dashed border) [70]

To ensure that the video encoder and decoder generate the same predictions for subsequent data, the encoder implements the decoder processing loop as shown in

Figure 4.1. Thus, the quantised transform coefficients are reconstructed by first applying an inverse scaling function followed by an inverse transform, yielding a decoded approximation of the residual signal. This signal is added to the prediction and processed further to smooth out artefacts induced by block-wise processing and quantisation [70]. The resulting reconstructed picture is then stored in a decoded picture buffer to be used for the prediction of subsequent pictures.

In order for HEVC to achieve significantly improved coding efficiency, a number of features were introduced which sets it apart from its predecessor. These include [70]:

- Code Tree Units (CTU) and Coding Tree Block (CTB) structure: Legacy video standards define the macro-block as the core element of the video coding layer composed of a 16×16 block of luma samples and, considering a 4:2:0 colour sampling, two corresponding 8×8 blocks of chroma samples. In HEVC, the analogous structure to the macro-block is the CTU, also referred to as largest coding unit (LCU), which can be larger than the legacy macro-block. The CTU consists of a luma CTB and the corresponding chroma CTBs where the size of the luma CTB can be 16×16, 32×32 or 64×64 samples with the larger sizes enabling better compression. Moreover, HEVC supports partitioning of the CTBs into smaller blocks called coding units (CU) using a tree structure and quad tree-like signalling.
- Coding Units (CUs), Prediction Units (PUs) and Transform Units (TUs): A CU is further partitioned into PU, each of which can be predicted using either intra or inter prediction. This partitioning enables better matching to the boundaries present in the picture [71]. The PUs are in turn subdivided into TUs which is the basic unit for the transform and quantisation process.
- Intra-frame prediction: Similar to H.264/AVC, the decoded boundary samples of adjacent blocks are used as reference data for spatial prediction however in contrast with AVC, intra-frame prediction in HEVC supports 33 directional modes in addition to planar (surface fitting) and DC (flat) prediction modes.
- Inter-frame prediction and motion compensation: The functionality of inter-frame prediction is essentially the same as H.264/AVC however motion compensation is enriched whereby quarter-sample precision is used for the

motion vectors and 7-tap or 8-tap filters are used for interpolation of fractional-sample positions.

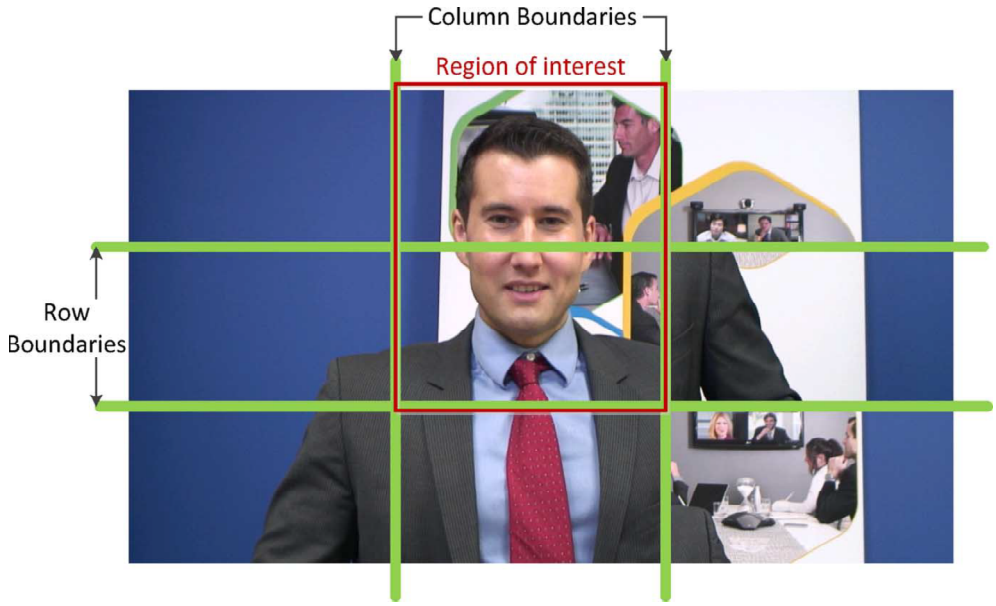
- Quantisation control: HEVC uses uniform reconstruction quantisation coupled with quantisation scaling matrices supported for the various transform block sizes.
- Entropy coding: Similar to its predecessor, context adaptive binary arithmetic coding (CABAC) is used for entropy coding however it has been enhanced to improve its throughput speed and compression performance whilst reducing its context memory requirements. Opposed to H.264/AVC, CAVLC is not supported in HEVC.

### 4.1.2 Slices and Tiles

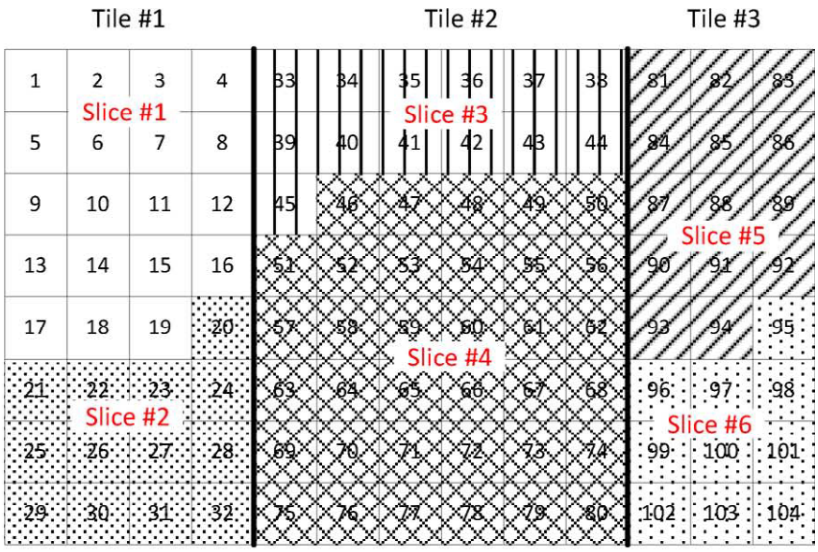
The concept of slices has been introduced in earlier video coding standards [72] to organise the generated video bit-stream into self-contained units which can be independently decoded. This allows resynchronisation in case of data loss in part of the bit stream, thereby limiting error propagation, at the expense of the loss of prediction across slice boundaries. Moreover, slices are commonly used to limit the maximum number of bits thereby aiding the transport of the encoded bit stream over a packet-switched network. With reference to Figure 4.2, slices in HEVC define groups of independently decoded CTB which follow raster scan order within a picture. The number of CTU in a slice varies depending on the video scene activity.


Figure 4.2: HEVC Slices

In addition to slices, H.264/AVC introduced flexible macro-block ordering (FMO) [73], to organise macro-blocks into slices in a highly flexible manner. Although FMO improves error resilience and allows the use of region of interest (ROI) awareness in high packet loss scenarios, the increased complexity coupled with lower coding efficiency, caused by the lack of prediction across slice boundaries, made FMO an unpopular tool. Alternatively, HEVC introduced the concept of tiles in order to aid parallelism, improve maximum transmission unit (MTU) size matching and offer additional ROI functionality [74]. As shown in Figure 4.3, tiles are defined by vertical and horizontal boundaries with intersections that partition a picture into rectangular regions [75]. Both the tiles and the LCUs in each tile are processed in raster scan order within a picture. Moreover, the column and row boundaries of tiles need not be uniform thereby offering greater flexibility and thus could be used to offer error resiliency of specific ROIs. The combined use of slices and tiles in HEVC offer various possibilities to control the ROI whilst improving MTU size matching.




(a)




(b)

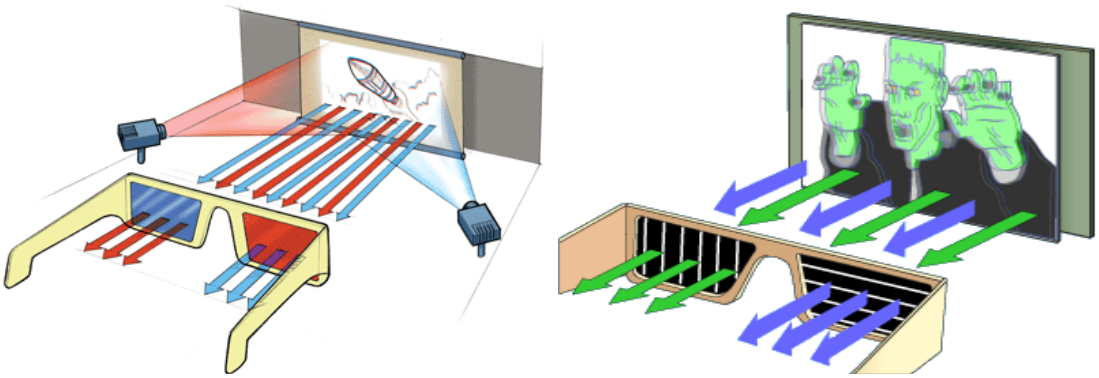
Figure 4.3: Slices and Tiles [74]. (a) Tile definition, and (b) Combined Slices and Tiles

### 4.1.3 3D video fundamentals and video formats

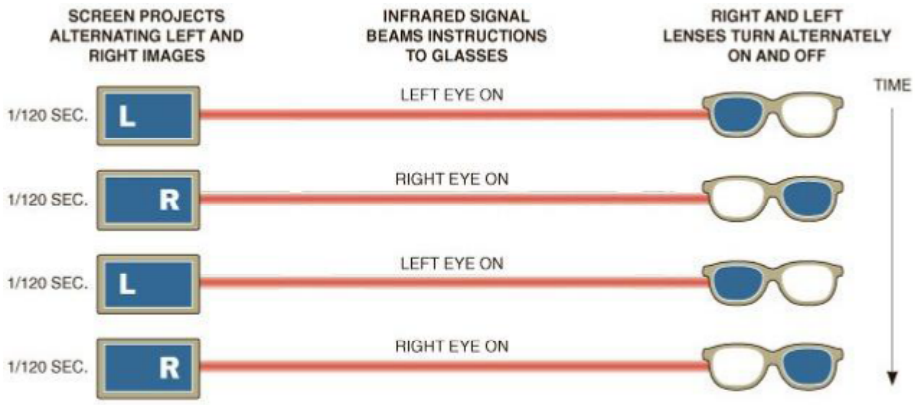
The human visual system allows the perception of 3D by using two main mechanisms namely binocular and motion parallax. Binocular parallax allows viewers to perceive depth by viewing a different angle of the same object in space through each eye whilst motion parallax relies on the relative motion of the object to the human eyes. Based on these concepts, one of the simplest and most cost effective 3D video formats is the conventional stereo video (CSV). In this 3D video format two views, one for each human eye, are transmitted and the viewer makes use of a headgear (glasses) to direct each view to the respective eye thereby allowing 3D perception.

With reference to Figure 4.4, headgear typically uses one of three common methods to separate the left and right views; namely anaglyph filtering, polarised filtering and time division. In the latter, frame sequencing is employed to project the left and right views to the respective eye, one at a time, using an active shutter glass system synchronised to the display. Nevertheless, the use of headgear to perceive 3D presents several drawbacks ranging from eye-fatigue, cross-talk and the lack of any motion parallax. For this reason, auto-stereoscopic displays utilizing lenticular lenses or parallax barriers have been developed in order to direct the left and right views to the respective eye without the need of headgear. Generally, such displays offer stereoscopic views at

multiple viewpoints, thereby allowing the viewer to experience motion parallax as the viewer moves from one viewpoint to another and enhancing the 3D perception.



(a) (b)




(c)

Figure 4.4: Headgear techniques [76]; (a) Anaglyph filtering, (b) Polarised filtering, and (c) Time-division active shutter system

In order to transmit stereoscopic 3D video, the sequence is either coded using simulcast, whereby the left and right views are individually compressed using a conventional video compression scheme such as H.265/HEVC, or using one of the frame compatible video formats shown in Figure 4.5. In the latter, the left and right views are first sub-sampled and interleaved into a single high definition frame prior to compression. Alternatively, a time multiplexing technique is used whereby the left and right views are interleaved as alternating frames or fields [77]. Although such frame compatible formats are attractive, as they facilitate the deployment of stereoscopic video to capable displays using conventional, single-view, video compression schemes, the spatial or temporal resolution is degraded.

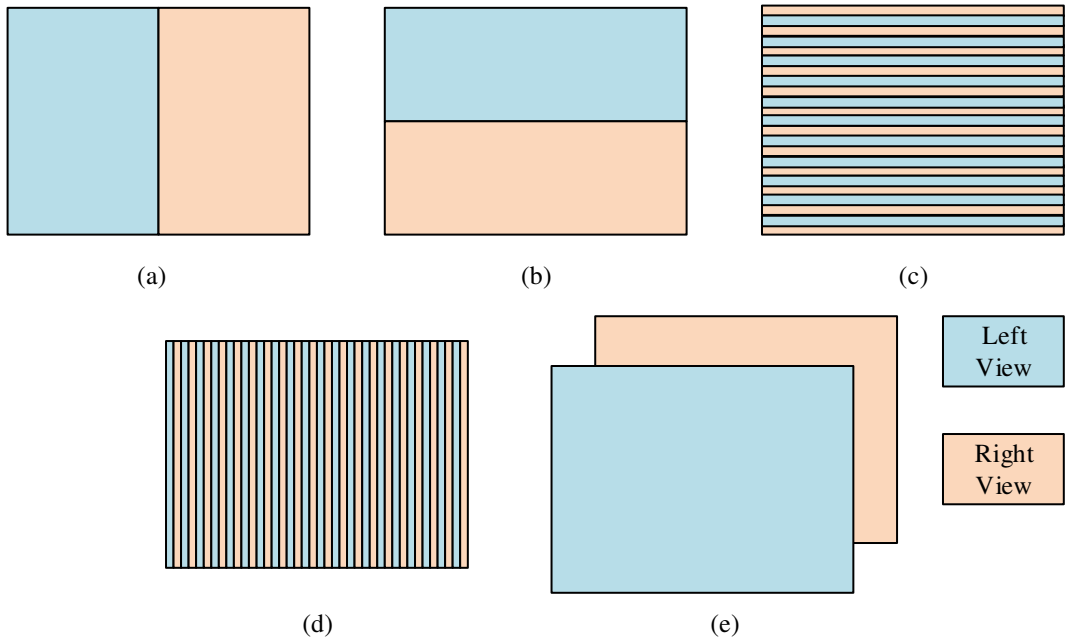



Figure 4.5: Frame compatible video format; (a) Left-Right, (b) Top-Bottom, (c) Horizontal interlacing, (d) Vertical interlacing, and (e) Temporal

Even though this representation is effective, the main disadvantage of such conventional stereo video format is its restriction to a pair of views at fixed spatial positions. Other alternative 3D video formats have been proposed, including model-based [78] and point-sample representation [79], however these are either restricted to video sequences having known objects or require the transmission of camera parameters in addition to other meta-data. The introduction of the multi-view video (MVV) [80] representation coupled with auto-stereoscopic displays led to significant developments in 3D video in recent years. With reference to Figure 4.6, in the MVV representation a visual scene is captured using multiple cameras placed at different angles thereby allowing viewers to experience depth perception of the visual scene, through the use of appropriate 3D displays, from multiple viewing locations [81, 82]. The generated video data increases proportionally with the number of cameras capturing the scene and thus multi-view video coding schemes are typically used to transmit multiple views efficiently. Nevertheless, even though high compression rates can be achieved by exploiting redundancies in both the spatial and temporal domain and in between cameras, the need of transmitting a large number of views to generate an immersive 3D experience led to the development of the multi-view video plus depth (MVD) [83] representation.

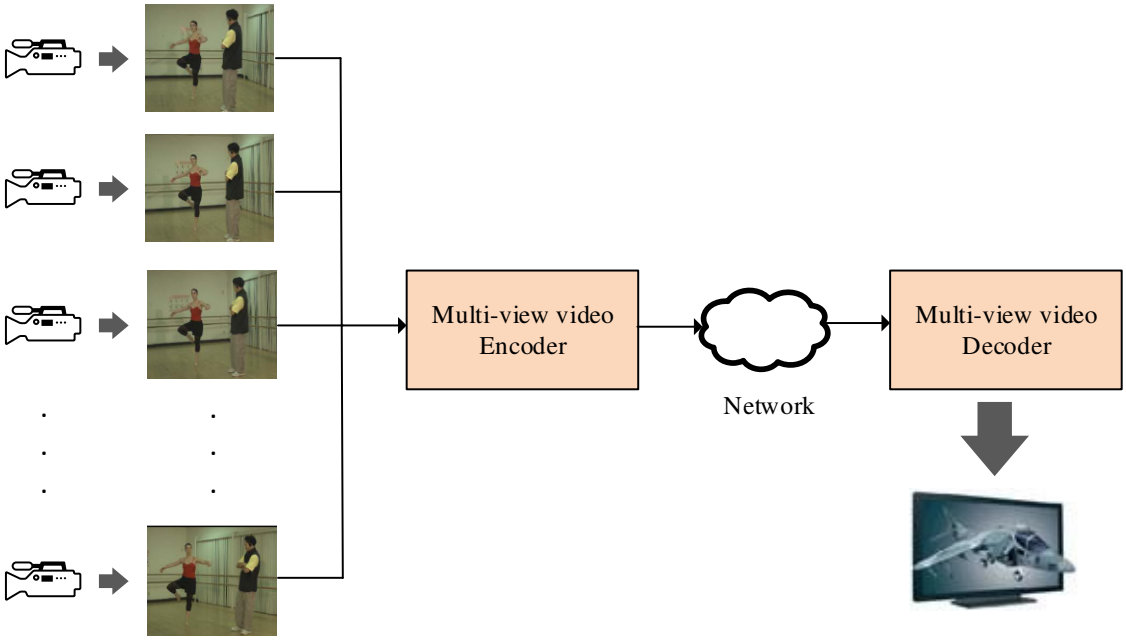



Figure 4.6: Multi-view video representation

With reference to Figure 4.7, the transmission of a small number of texture and depth map views allows the receiver to generate any number of views in-between the transmitted views by using the geometry data found in the depth maps and a view synthesis technique, such as depth-image-based rendering (DIBR). The ability to synthesis novel views using a limited set of texture and depth map views does not only allow the efficient delivery of enhanced 3D video streams, which gives a depth impression of the observed scene, but also enables free-viewpoint video (FVV). The latter is a complementary application of the MVD 3D video representation and has gained significant commercial interest, specifically for sporting events [8], as it allows viewers to interactively navigate a scene through different viewpoints.

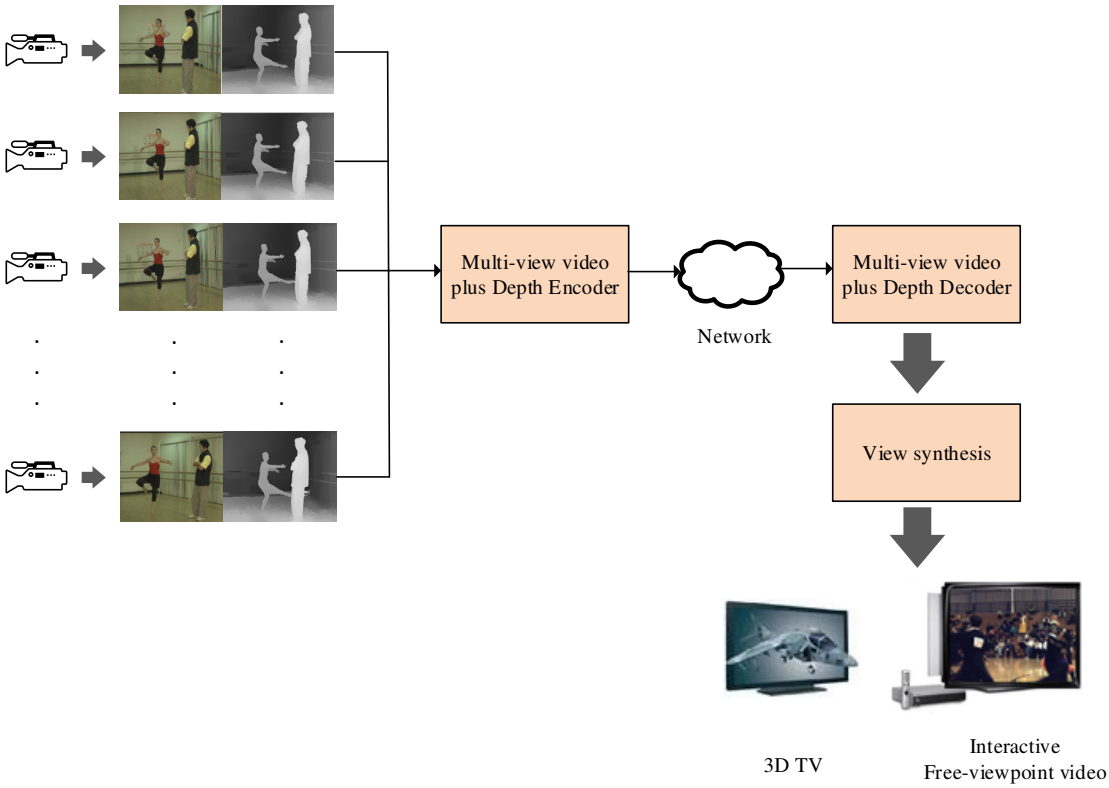



Figure 4.7: Multi-view video plus Depth (MVD) representation

### 4.1.4 HEVC – Multi-view and 3D extensions

In order to support these applications, a new Joint Collaborative Team on 3D Video Coding Extensions Development (JCT-3V) was formed between the ISO/IEC and ITU-T to develop extensions on top of HEVC including 3D and multi-view standards [81, 82]. The development of 3D and multi-view extensions of HEVC is essentially split into two; namely Multi-view HEVC (MV-HEVC) and 3D-HEVC. MV-HEVC uses the same design principles of MVC in the H.264-AVC framework [84] whilst retaining the same block-level decoding process. The main advantage of this scheme is that it provides backwards compatibility for monoscopic video decoding whilst allowing the use of inter-view prediction thereby exploiting both inter-view and temporal redundancy yielding high compression rates. Although the primary application of this coding architecture was intended to efficiently compress multi-view texture streams, it can also be applied to compress multi-view depth map streams thereby providing a simple and relatively low complexity technique to efficiently compress MVD streams as shown in Figure 4.8.

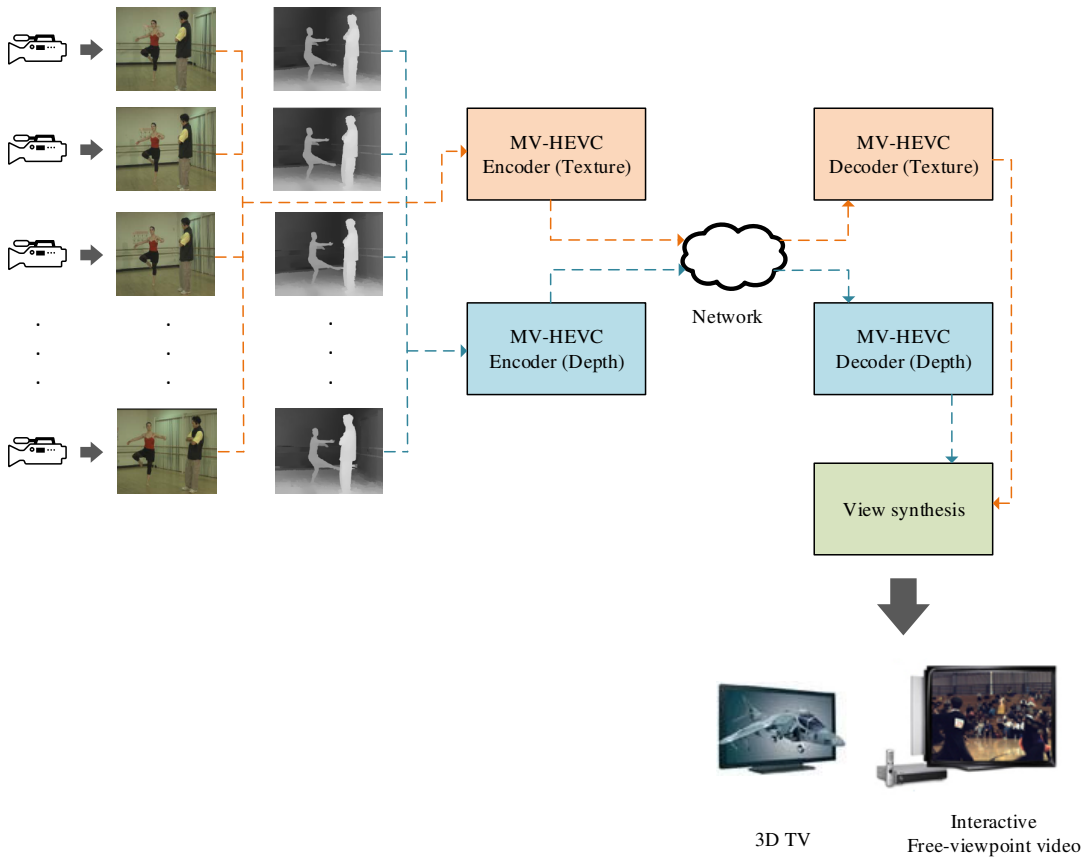



Figure 4.8: MVD using dual MV-HEVC streams, one for texture and the other for depth

To achieve high compression rates, 3D-HEVC makes use of additional coding tools for the dependent views to exploit the correlation of motion and residual data between views [82]. The salient tools in this release include Neighbouring Block-Based Disparity Vector Derivation [85], illumination compensation to enhance coding efficiency for blocks predicted from inter-view reference pictures [86] and inter-view motion and residual prediction [87,88]. In addition to these tools, 3D-HEVC introduces a set of techniques aimed at video-plus-depth video format compression. Unlike MV-HEVC the depth maps, which are characterised by large homogenous areas and sharp edges [82], are coded using techniques designed to cater for such unique characteristics. Depth maps are not intended to be viewed by end users but instead used for the generation of synthesised views. For this reason, it is imperative that the depth map edges are preserved since inaccurate edge reconstruction has a significant impact on synthesised views. To better represent the depth information, partition-based depth intra coding modes such as depth modelling modes (DMM) [89], are used. In addition, in order to reduce inter-view redundancy, the depth information

is used in view synthesis prediction (VSP) to warp the texture data from a reference view to the current view so that a predictor for the current view can be generated [82].

Both the MV-HEVC and 3D-HEVC extensions yield significant gains when compared to simulcast. The use of inter-view prediction in MV-HEVC results in an average bit rate saving of 40% for the three-view scenario when compared to simulcast [90]. Meanwhile, the performance improvement of 3D-HEVC relative to MV-HEVC is approximately 10% but increases to 21% when considering the depth map views and the synthesised view quality thereby showing the effectiveness of the coding tools introduced in 3D-HEVC [90]. Nevertheless, the latter comes at the expense of increased computational complexity [81]. Moreover, the current 3D-HEVC standard does not allow the texture and depth map view to have differing resolutions thereby precluding the re-sampling of the texture or depth map views prior to encoding and transmission. In contrast, MV-HEVC encodes the texture and depth views independently giving the possibility of higher bit rate savings through re-sampling techniques.

## 4.2 Depth-Texture Map View Bit Allocation

The compressed MVD stream is usually transported over a bandwidth limited network, such as a wireless cellular network, and thus a depth-texture bit allocation scheme coupled with appropriate rate control for the texture and depth-map views is necessary. Even though the depth map is treated as side-information and used only for novel view rendering, studies have shown that coding the depth map views at a low bit rate introduces artefacts which can have a significant impact on the quality of the rendered novel views [91]. Therefore, under the constraint of the total available bandwidth provided by the underlying network, the correct balance between texture and depth rate allocation is critical. Several proposals can be found in literature however these are either based on the use of a fixed depth-texture allocation or use a pre-encoding technique to determine model parameters offline.

The use of a fixed depth-texture rate allocation, although simple and suitable for real-time applications, does not guarantee an optimal virtual view quality. Various fixed ratios, varying from 5-30% [92, 93], have been proposed in literature with 20% being commonly used. In order to address the limitation imposed by this fixed depth-texture allocation, Morvan *et al.* proposed a scheme based on the combination of the texture and depth rate-distortion (R-D) curves yielding a single R-D surface [94]. In order to generate the R-D curves, the texture and depth map views are pre-encoded using various quantisation parameter (QP) values thereby incurring high computational cost. The R-D surface is then used with a hierarchical optimisation algorithm, based on an orthogonal search pattern which exploits the smooth monotonic properties of the R-D surface, to find the optimal depth-texture rate allocation ratio. Apart from the high computational cost and off-line analysis required to generate the R-D surface, the authors assume that a full reference assessment of the synthesised view is always possible.

In an attempt to reduce the high computational cost and the full reference assessment constraint, several works based on model-based bit allocation schemes have been proposed. In [95], the authors propose a complex distortion model incorporating the video coding, depth quantisation and geometry induced distortion. This model is then used in conjunction with an off-line joint rate allocation method to find the correct balance between the texture and depth bit rate which maximises the novel view synthesis quality. A cubic distortion model, whose parameters are established using an off-line processing technique, is proposed in [96]. Here, the optimal quantisation parameters for the texture and depth map views are modelled by the shortest path in a custom constructed 3-D trellis. In [97], the authors formulate a simple distortion model for the synthesised view based on the average distortion of the left and the right view and the average distortion of the compressed depth maps. Based on this model together with the fractional rate-quantisation model [98], a joint bit allocation problem is formulated as a constrained optimisation problem and resolved using the Lagrangian multiplier method. Similarly, the solutions proposed in [99, 100, 101, 102], follow the same principles and require the pre-encoding of a number of frames to derive the model parameters from which the optimal QP pair for the texture and depth map views is calculated. Thus, in these schemes, although the computational complexity is

reduced, they require pre-encoding of a number of frames thereby limiting these schemes to non-real time applications. As opposed to previous studies, the authors in [103] propose a bit allocation algorithm based on the estimated synthesised view distortion. This is based on a simplified virtual view distortion model and is able to adjust the bit allocation based on the video characteristics. Although this method does not require any pre-encoding of frames, the authors still use a set of pre-configured parameters to ensure a fast convergence rate. Moreover, the algorithm assumes a fixed total available bit rate and its performance was not tested in a varying bit rate scenario.

In our early work within this study [104], we explored the use of the macro blocks (MBs) prediction mode distribution at the discontinuity regions of the depth map video to estimate the texture-depth bit rate allocation in H.264/MVC, this being the standard codec at that time. Although the proposed scheme exhibits a low estimation error and low complexity, its formulation is intrinsically tied with the H.264/MVC encoder and does not provide any input related to the available bandwidth. Moreover, to establish the prediction mode distribution of the MBs at the discontinuity region, one must start encoding the video sequence at some arbitrary quantisation parameter (QP) value. In [105], the authors propose a simplistic model to directly select the QP of the depth map view given the quantisation parameter of the texture video. Similar to [104], the authors focus on H.264/MVC and the available bandwidth does not feature as one of the control inputs. A summary of the depth-texture bit allocation schemes reviewed in this section is shown in Table 4.1.

| **Scheme** | **Depth-Texture bit allocation scheme** | | | **Computational Complexity** | **Pre-encoding of frames** |
|---|---|---|---|---|---|
| | **Fixed** | **Model Based** | **Other** | | |
| Fehn [92], Klimaszewski *et al* [93] | Various 5-30% (20%) | | | Low | No |
| Morvan *et al* [94] | | • | | High | Yes |
| Liu *et al* [95] | | • | | Medium | Yes |
| Cheung *et al* [96] | | • | | Low | Yes |
| *Various* [97, 99-102] | | • | | Medium | Yes |
| Yang *et al* [103] | | • | | Medium | No - but requires a set of pre-configured parameters |
| Cordina *et al* [104] | | | • | Low | No |
| Klimaszewski *et al* [105] | | | • | Low | No |

Table 4.1: Summary of Depth-Texture bit allocation schemes

## 4.3 Rate Control

The overarching goal of the rate control entity in an encoder is to efficiently distribute the bit budget among frames whilst striking a balance between the image quality and smoothness under a given bandwidth constraint. Rate control is a very active research topic and several schemes have been proposed for both legacy and contemporary video coding standards [106]. In HEVC, two main rate control schemes have been adopted; namely the unified rate quantisation (URQ) algorithm and the R-λ model based rate control, defined in JCTVC-H0213 [107] and JCTVC-K0103 [108] respectively. The URQ rate control scheme makes use of the conventional quadratic model used in earlier video coding standards and has been further improved through the enhancement of the model accuracy. In [109], the authors adopt the sum of absolute transformed differences (SATD) whilst in [110] a per pixel gradient value is used in the R-Q model. Meanwhile, the work carried out by Li *et al* in [111] shows that the bit rate $R$ has a stronger relationship with the Lagrange multiplier $\lambda$ than the quantisation parameter QP, thereby leading to the development of the R-λ model. Moreover, the authors in [108] show that the rate-distortion (R-D) model can be modelled using a hyperbolic function given by:

$$D(R) = CR^{-K} \tag{4.1}$$

where $C$ and $K$ are model parameters related to the source characteristics.

Using (4.1), the Lagrange multiplier $\lambda$ is defined as the slope of the R-D curve and thus:

$$\lambda = -\frac{\partial D}{\partial R} = CKR^{-K-1} = \alpha R^{\beta} \Rightarrow R = \left(\frac{\lambda}{\alpha}\right)^{-\beta} = \alpha_1 \lambda^{\beta_1} \tag{4.2}$$

where $\alpha_1$ and $\beta_1$ are model parameters related to the source characteristics.

The R-λ model has been adopted as the reference algorithm for HEVC and has since been further improved through the use of better intra-frame rate control [112, 113]. Moreover, an efficient bit allocation scheme based on the HEVC hierarchical coding structure is proposed in [114] whilst a Laplace distribution based CTU rate control algorithm is presented in [115]. Meanwhile, Lee *et al.* investigated the use of separate

R-Q models for the texture and non-texture areas in videos and proposed a frame level rate control scheme for HEVC [116] whilst a region of interest (ROI) based rate control algorithm was proposed in [117]. In MV-HEVC, the same reference rate control algorithm found in HEVC is typically adopted and this is also implemented in the reference encoder software model [118]. Meanwhile, in [119], the authors propose an enhancement over the URQ and R-λ rate control schemes which consider the inter-view prediction present in MVV encoding whilst in [120] a convex optimisation algorithm is used to solve a recursive rate-distortion model which takes into account these inter-view dependencies. Similarly, in MV-HEVC for the MVD representation, the same R-λ rate control algorithm used in HEVC is adopted across all the texture and depth map views regardless of the fact that the texture and depth map videos have strikingly different characteristics. Moreover, in 3D-HEVC, the rate control algorithms developed for HEVC were extended and applied for both texture and depth map views. The authors in [121], exploit the depth map to enhance the accuracy of the mean absolute difference (MAD) by considering the inter-view disparity in the texture view MAD estimation whilst in [122], the authors propose an adaptive frame level rate control algorithm for 3D-HEVC based on a new initial quantisation parameter decision and bit allocation scheme. Although the results look promising, the authors do not indicate any improvements in the view quality.

Considering that the depth map views are significantly important for the rendering of synthesised views, a custom designed rate control algorithm for such views, based on their characteristics, is beneficial. Nevertheless, work on such rate control schemes is limited. In [124], the authors propose a view synthesis distortion model based frame level rate control optimisation scheme for multi-view depth video coding based on a Nash bargaining solution. Nevertheless, this scheme requires the pre-encoding of frames to determine model parameters. Meanwhile, in [125], Lei *et al* propose a region adaptive R-λ model based rate control scheme for depth map views. In this work, the authors first propose a modified frame level bit allocation method based on the depth map view coding bits distribution. The depth map views are then decomposed into two types, each having its own R-λ model, depending on their importance for virtual view rendering. The proposed scheme is able to achieve an average improvement of 0.3dB in Bjøntegaard Delta PSNR (BDPSNR) however; similar to [124], this scheme requires the pre-estimation of model parameters.

# Chapter 5 Depth-Texture Rate Allocation Estimation Technique and Depth Map Rate Control Scheme

As discussed in chapter 4, the depth-texture rate allocation problem has attracted a number of solutions however these typically require pre-encoding to fit model parameters aiding the optimal estimation of the depth-texture rate allocation. Moreover, rate control schemes are generally designed for texture views and are re-used for the depth map views without any consideration of the depth map characteristics. This chapter starts with a detailed account of the proposed cross-layer MV-HEVC depth-texture rate allocation estimation scheme and is followed by an enhanced depth map rate control scheme to improve the synthesised view quality. As opposed to previous works, the proposed depth-texture rate allocation scheme does not require the pre-encoding of frames and considers both the image characteristics and the total available bit rate for the MVD stream thereby allowing online adaptation and making it suitable for real-time applications.

## 5.1 Cross-layer MV-HEVC Depth-Texture Rate Allocation Estimation Technique

### 5.1.1 System model and problem formulation

With reference to the radio network sub-system of an MV-HEVC based MVD video delivery system for an FVV application, shown in Figure 5.1, consider that the dual MV-HEVC video encoder stack is able to simultaneously encode multiple texture and depth map views. Moreover, consider an arbitrary case of encoding two texture views, view 0 and view 2, together with their corresponding depth map views and assume that the mobile terminal (UE) is able to synthesise view 1 using DIBR. As shown in Figure 5.1, the encoded texture and depth map views are combined into an MVD bit stream and transmitted to the UE over a standard 3GPP LTE Release 8 compliant eNB whilst the UE periodically reports CQI feedback reports to the eNB. The latter is used by the eNB scheduler to determine the users to be scheduled in the current scheduling interval together with the number of resource blocks, MIMO mode of operation, MCS and TBS to be assigned to the scheduled UE. These parameters effectively determine the maximum scheduled physical layer bit rate (channel bandwidth) supported by the UE,

denoted by $R_p$, at the current radio and loading conditions. With reference to chapter 2, this physical layer bit rate includes several overhead factors stemming from the various protocol layers between the application and physical layer. Nevertheless, an eNB is able to measure the scheduled IP throughput at the PDCP layer [14], denoted by $R_a$, which is closely related to the physical layer throughput. These measurements are made available to the upper layers thereby introducing a cross-layer functionality which will be discussed in more detail in the next sections.

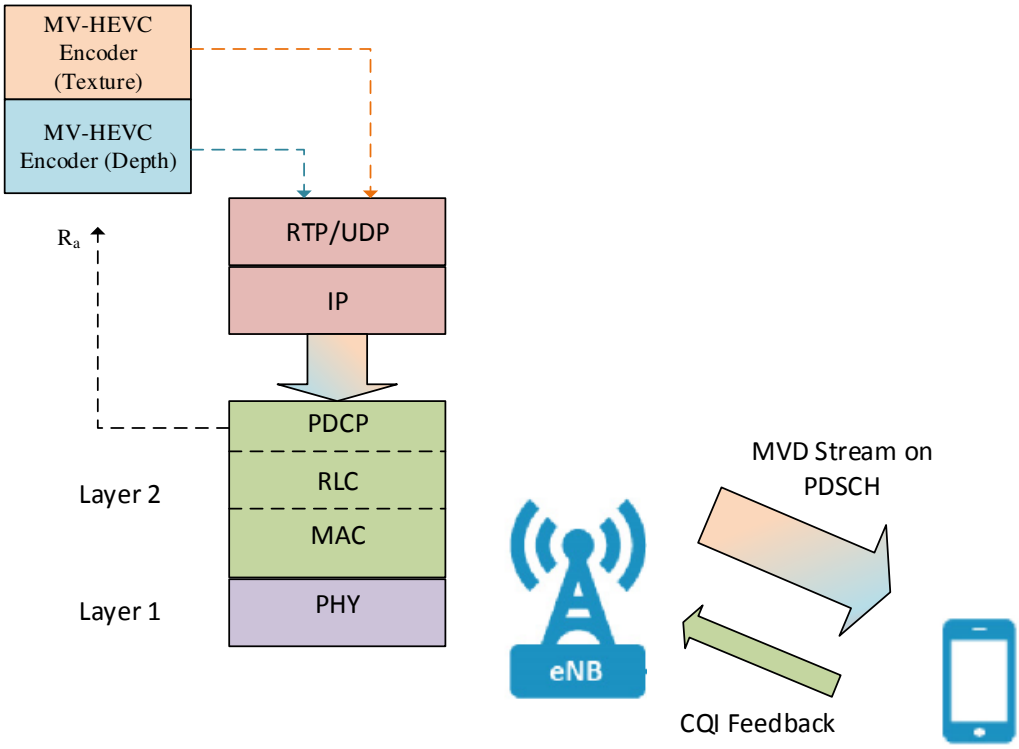



Figure 5.1: System Model

As discussed in section 4.2, under the constraint of the channel bandwidth, the synthesised view quality depends on the optimal selection of the depth-texture bit rate allocation. Moreover, as evidenced from previous studies, this depends on the sequence characteristics. Hence, the objective of this work is to formulate a statistical model, given by (5.1), able to estimate, online, the optimal depth-texture rate allocation ratio, $\beta_{opt}$, based on the application layer bit rate and image statistics.

$$\beta_{opt} = f\left(\Phi_T, \Phi_D, R_a\right) \tag{5.1}$$

where $\Phi_T$ and $\Phi_D$ represent the texture and depth map views image statistics respectively whilst $R_a$ is the application layer bit rate available for the MVD stream.

Based on the model defined by (5.1), the target bit rates for the texture and depth map views to be used by the respective encoder rate control algorithm are given by:

$$R_d = \beta_{opt} R_a$$
$$R_t = (1 - \beta_{opt}) R_a \tag{5.2}$$

where $R_d$ and $R_t$ represent the depth map and texture views bit rate allocation respectively.

### 5.1.2 Experimental setup and initial observations

Inspired by the works in [91, 104], the variation of the optimal depth-texture ratio, $\beta_{opt}$, for various video tests sequences exhibiting different video characteristics and total bit rate was studied by conducting a series of experiments using the test setup shown in Figure 5.2. In this work, the total bit rate refers to the sum of the encoded bit rate of the texture and depth map views which is usually bounded by the available bandwidth at the application layer, $R_a$. As the main interest of this study focuses on real-time video, a low latency prediction structure, shown in Figure 5.3, is used in these experiments. Moreover, the encoding parameters together with the test sequences, defined in the Common Test Conditions (CTC) [125], are shown in Table 5.1 and Table 5.2 respectively, where *L*, *C* and *R* represent the left, rendered and right views respectively.

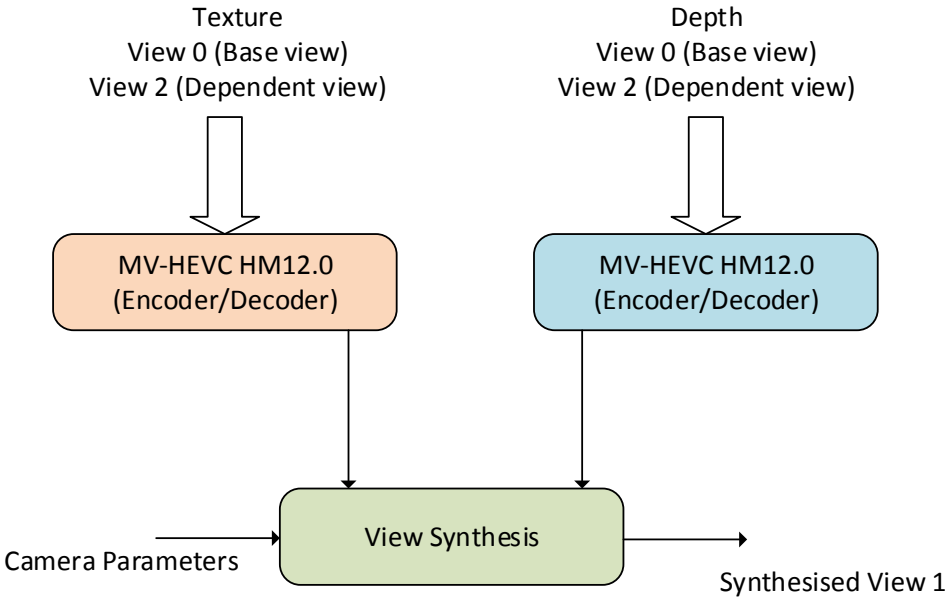



Figure 5.2: Test setup

| Parameter | Value |
|---|---|
| QP | 20 - 44 |
| GOP Size | 4 |
| Intra Period | 12 |
| LCU Size | 64 x 64 |
| Partition Depth | 4 |
| Fast Search | 1 |
| Search Range | 64 |
| RDOQ | 1 |

Table 5.1: Encoding Parameters

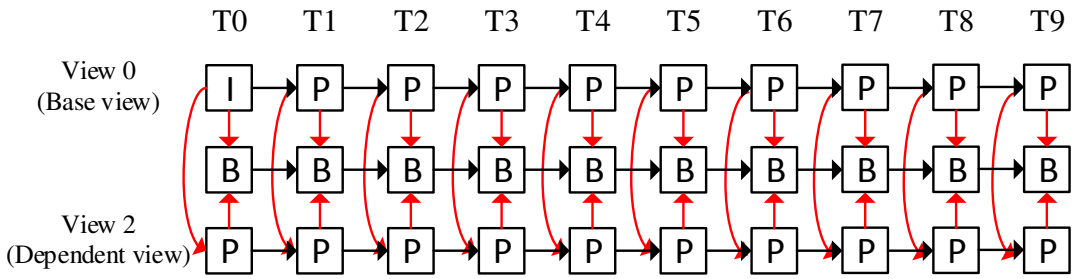



Figure 5.3: Low Latency Prediction Structure [40]

| Sequence | Image Size | Cameras (L-C-R) | Characteristics (frame rate, baseline) |
|---|---|---|---|
| *Balloons* | 1024 x 768 | 1-3-5 | 30 fps; 5cm in 1D plane |
| *Book Arrival (Book)* | 1024 x 768 | 6-8-10 | 15 fps; 6.5cm in 1D plane |
| *Kendo* | 1024 x 768 | 1-3-5 | 30 fps; 5cm in 1D plane |
| *Newspaper (News)* $^{V}$ | 1024 x 768 | 2-4-6 | 30 fps; 5cm in 1D plane |
| *Poznan Hall*$^{V}$ | 1920 x 1088 | 5-6-7 | 25 fps; 13.75cm in 1D plane |
| *Poznan Street*$^{V}$ | 1920 x 1088 | 3-4-5 | 25 fps; 13.75cm in 1D plane |

Table 5.2: Test sequence parameters and characteristics

(Sequences marked with $^{V}$ were used for model verification only)

With reference to Figure 5.2, in order to increase the number of data samples for the statistical model formulation, each test sequence was first split into chunks and each chunk is encoded using MV-HEVC reference software HM12.0 [118]. The size of each chunk was set equal to the intra frame period whilst the total bit rate and depth-texture ratio were varied by adjusting the QP used for encoding the texture and depth map views between 20 and 44 in steps of 4. A plot of the synthesised view Peak Signal-to-Noise Ratio (PSNR) with depth-texture rate allocation ratio $\beta$, given by (5.3), for the *Balloons* video sequence at a given total bit rate is shown in Figure 5.4.

$$\beta = \frac{R_d}{R_d + R_t} \tag{5.3}$$

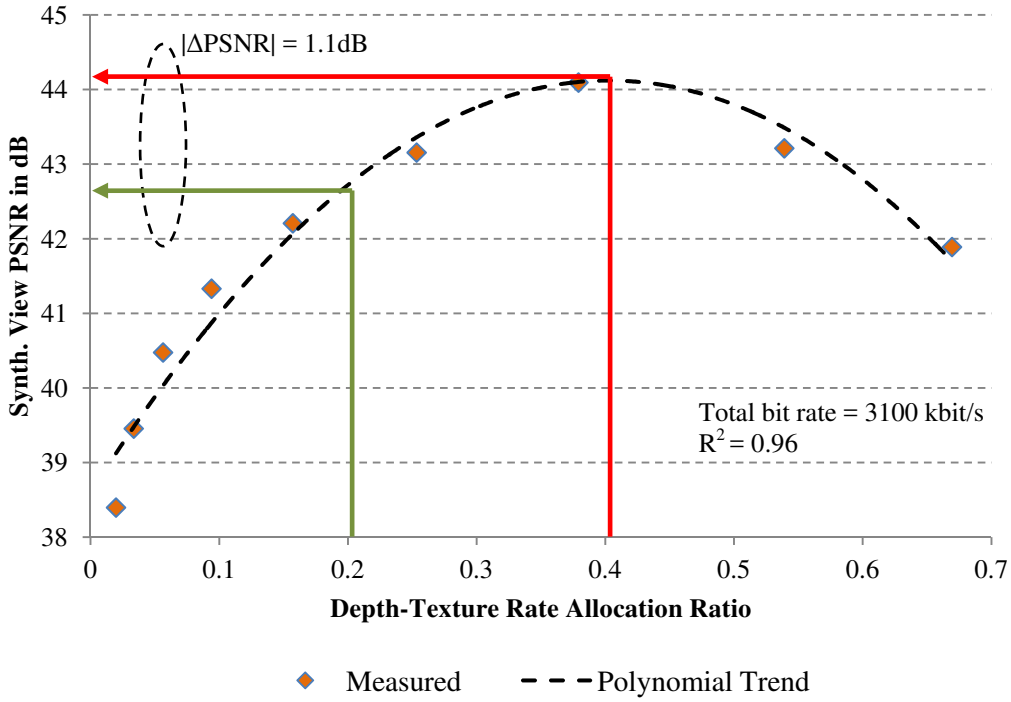



Figure 5.4: PSNR vs depth-texture ratio for the *Balloons* sequence
($R^2$ represents the coefficient of determination)

With reference to Figure 5.4, the synthesised view PSNR follows a downward concave parabola yielding an optimal depth-texture ratio (shown in red) of 40.3%. Moreover, the importance of estimating accurately the optimal depth-texture ratio is made evident since even though the synthesised view quality is fairly constant over a narrow range of depth-texture ratios, any large deviations from this optimal value (shown in green) will result in a relatively large degradation in the synthesised view's quality. With reference to Figure 5.5, the variation of the optimal depth-texture ratio is dependent on the sequence characteristics since different sequences impart contrasting trends. Moreover, as the total bit rate increases, the optimal depth-texture ratio approaches an asymptotic value which never surpasses 60%. This observation is not surprising since increasing the depth map view bit rate beyond a specific value would not yield an improvement in the synthesised view's PSNR. Meanwhile, at low channel bit rates, the bit rate available for the depth map views is very low and thus it would be beneficial to refrain from transmitting the depth maps views and instead allocate the available bandwidth to the texture views only. This would effectively preclude the

UEs to synthesize views at intermediate viewpoints between view 0 and view 2 thereby limiting the number of possible viewing angles.

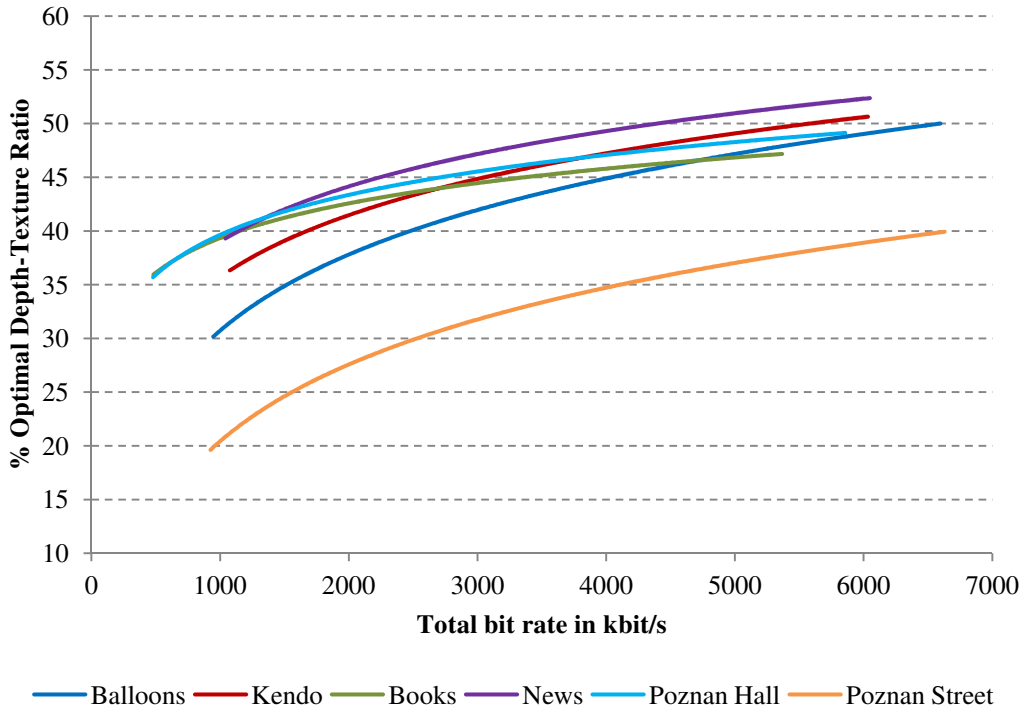



Figure 5.5: Optimal depth-texture ratio with total bit rate

Thus, considering these constraints in (5.2), the estimated optimal depth-texture rate allocation ratio calculated by the statistical model is bounded as shown in (5.4).

$$\begin{aligned} &if\ \beta_{opt} > 0.6 \\ &\quad \beta_{opt} = 0.6 \\ &elseif\ \beta_{opt} < 0.1 \\ &\quad \mathrm{TX}_{T+D} \rightarrow \mathrm{TX}_{T}\ \text{switch} \\ &\text{otherwise use } \beta_{opt} \end{aligned} \tag{5.4}$$

In order to model the sequence dependent characteristics, a number of texture and depth map image metrics were considered. One of the key metrics used is the depth map entropy ratio, $E_{DT}$, which has been shown to influence the optimal depth-texture ratio [91] and is given by:

$$E_{DT} = \frac{E_D}{E_D + E_T} \tag{5.5}$$

where $E_D$ and $E_T$ are the average entropy [91] of the depth map and texture views respectively. Other metrics used in this study includes the gradient magnitude [91] of the texture view, $G_T$, and depth map view, $G_D$, respectively. All image statistics are aggregated across view 0 and view 2.

### 5.1.3 Statistical model formulation

The average chunk values of the image statistics coupled with their associated total bit rate and optimal depth-texture ratio for all test sequence used during model formulation, shown in Table 5.2, were collected and imported into IBM SPSS statistical package for analysis. As the statistical distribution of the dependent variable would impact the selection of the statistical regression model, the optimal depth-texture ratio was tested for normality using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.57. As the p-value is greater than 0.05, the normality criteria is validated and thus a general linear model in the form of a multiple regression model is investigated in this study. A histogram plot of the dependent variable against a normal distribution curve is shown in Figure 5.6.

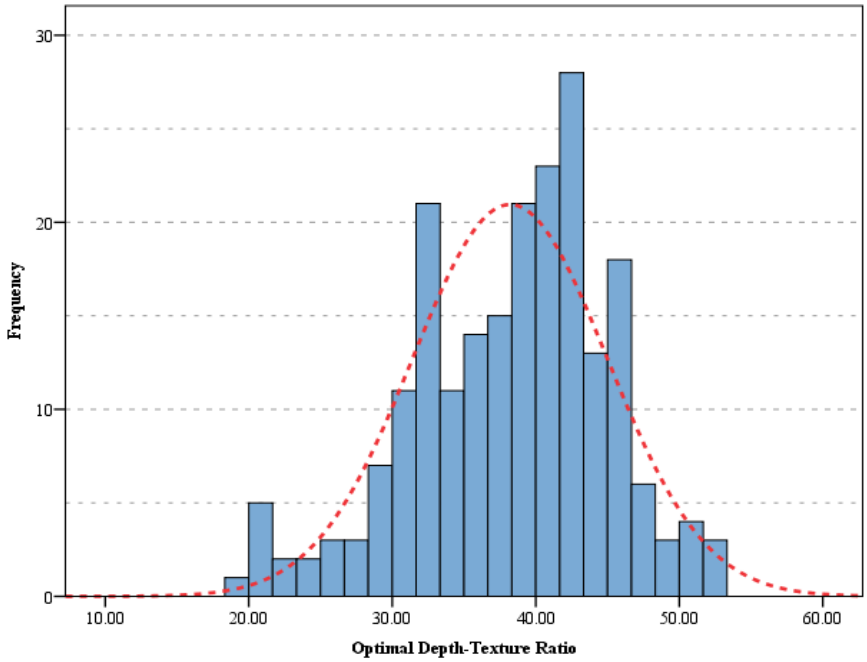



Figure 5.6: Histogram of the optimal depth-texture ratio and normal distribution curve

Multi-regression models have been used extensively in various fields due to their well known statistical properties. For $n$ observed cases, a multi-regression model expresses the dependent variable $y_n$ as a function of $p$ predictors given by [69]:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{5.6}$$

$$\mathbf{y} = \mathbf{x}\beta + \varepsilon \tag{5.7}$$

where **y** is an *n*-vector of responses which exhibit a normal distribution and is linearly related and correlated with each quantitative predictor, **x** is an $n \times p$ matrix whose elements represent the values of each predictor *p* for all observed cases, $\beta$ is a *p*-vector of unknown regression parameters estimated using least squares estimation, and $\varepsilon$ is an *n*-vector of unknown random error terms. Considering that $\varepsilon$ has a normal distribution, the least square estimate, which is equivalent to the maximum likelihood estimator *b*, is given by:

$$b = \left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y} \tag{5.8}$$

Based on the collected data, a multi-regression model was formulated using SPSS© which makes use of the least square estimation method to estimate the parameter values. Although various predictors have been evaluated in SPSS, some of these predictors were found to be either correlated with each other or do not exhibit a strong correlation with the variation in the optimal depth-texture ratio. Thus, in order to reduce multi-collinearity effects, which might affect the robustness of the formulated model, these predictors were removed from the final statistical model yielding the parsimonious multiple regression model given by (5.9) having seven significant predictors and a coefficient of determination, $R^2$, of 0.86.

$$\begin{aligned} \beta_{opt} = {} & 34.516 \\ & + G_D\left[25.714 + 0.001 R_a G_D^{-2}\right] \\ & - G_T[3.899 - 0.0046 R_a - 0.0855 R_a G_T^{-2}] \\ & - R_a E_{DT}[0.0062 E_{DT}^{-2} - 0.0623] \end{aligned} \tag{5.9}$$

where $\beta_{opt}$ is the estimated optimal depth-texture ratio, $G_D$ and $G_T$ are the average chunk gradient of the depth map views and texture views respectively, $E_{DT}$ is the depth map entropy ratio given by (5.5) whilst $R_a$ is the total bit rate available at the application layer.

### 5.1.4 Statistical model diagnostics tests

Apart from the model's coefficient of determination, $R^2$, a number of statistical diagnostics tests were carried out and analysed in further detail. These tests aid the detection of any model uncertainty, stemming from a violation of the underlying assumptions, and overly influential points in the data. One key assumption of multi-regression models is that the residuals should have a normal distribution [69]. This was verified by checking the standardised residuals using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.805 thereby validating the normality criteria. Moreover, in order to assess the presence of any influential points in the data which might degrade the model uncertainty, the Cook's distance, $D_n$, given by (5.10), is computed for all observations $n$, and plotted with the unstandardised predicted optimal depth-texture ratio value. With reference to Figure 5.7, the data contains a relatively low number of influential points (points having relatively high Cook's distance) thereby reducing the formulated statistical model uncertainty [69].

$$D_i = \frac{\sum_{j=1}^{n} \left( \hat{y}_j - \hat{y}_{j(i)} \right)^2}{(p+1)\hat{\sigma}^2} \tag{5.10}$$

where $\hat{y}_j$ and $\hat{y}_{j(i)}$ is the response value of the $j^{th}$ observation with and without the $i^{th}$ observation respectively, $p$ is the number of predictors in the model and $\hat{\sigma}^2$ is the mean squared error.

Finally, a plot of the studentised residual against the unstandardised predicted optimal depth-texture ratio is shown in Figure 5.8. This plot is commonly used to identify any patterns in the residual of the fitted model which might indicate a violation of the underlying assumptions. As the majority of the points are distributed across a horizontal band, indicating homoscedasticity, and lie within the standard ±2 limit, the model is adequate for such data fit with a very small number of outliers (shown in red).

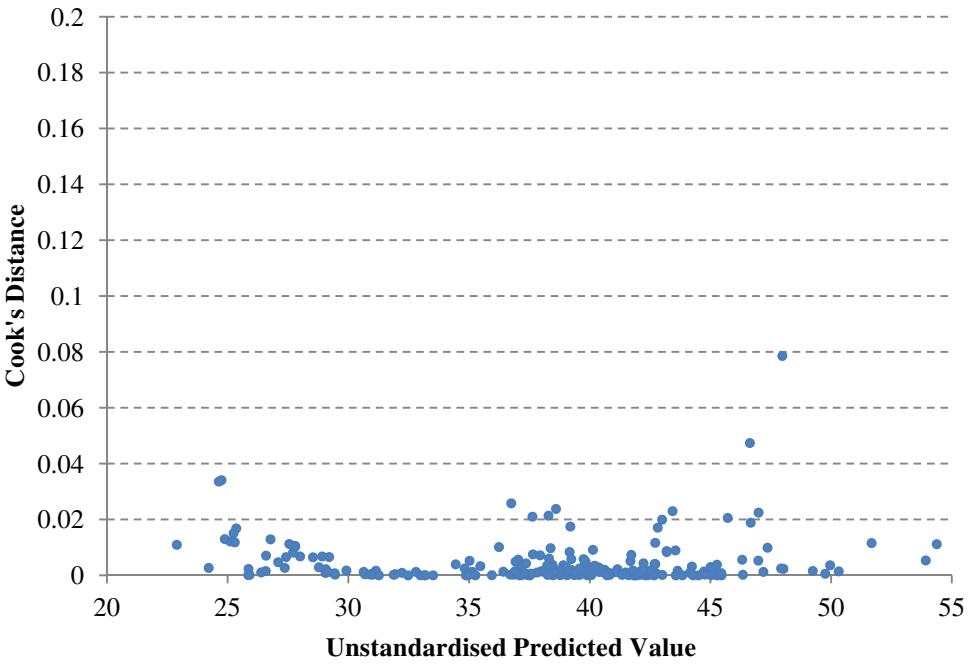



Figure 5.7: Cook's distance

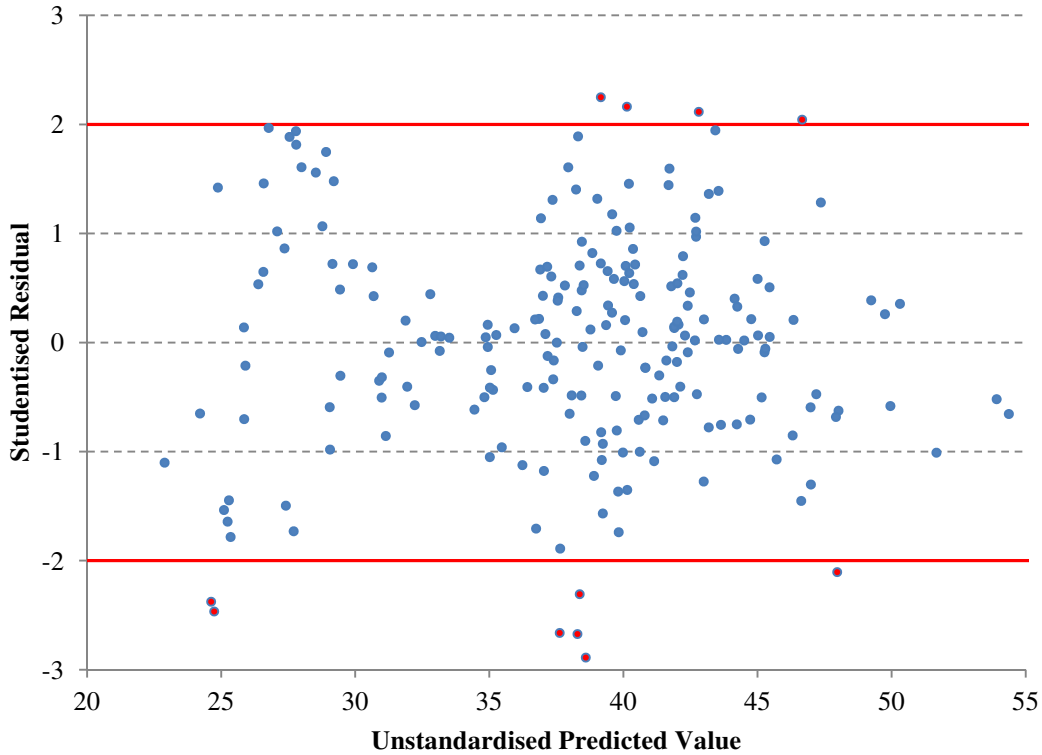



Figure 5.8: Studentised residual plot

### 5.1.5 Model Verification Results

With reference to Figure 5.9, in order to verify the efficacy of the proposed solution the MV-HEVC HM12.0 reference encoder software was modified such that the texture and depth map view image statistics, discussed in section 5.1.3, are computed online for every frame. The computed statistics together with the application layer bit rate, $R_a$, are then fed to the depth-texture rate allocation controller for smoothing and computation of the estimated optimal depth-texture rate allocation ratio, $\beta_{opt}$, using (5.9). Finally, the estimated optimal depth-texture rate allocation ratio is bounded, using (5.4), prior to calculating the target bit rates for the texture and depth map views using (5.2).

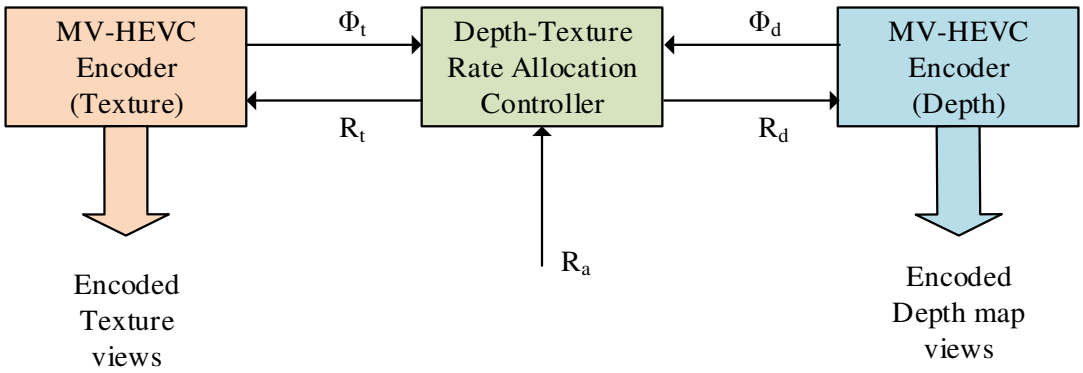



Figure 5.9: Proposed cross-layer depth-texture rate allocation scheme

As discussed in section 5.1.1, the application layer bit rate, $R_a$, depends on the underlying network physical layer bit rate, $R_p$, however for these verification tests, the application layer bit rate is arbitrary set such that the optimal depth-texture ratio can be estimated for various average total bit rates. Using the setup shown in Figure 5.9, the actual and estimated optimal depth-texture rate allocation ratio for the test sequences used during model formulation and verification at various average total bit rates are shown in Figure 5.10 and 5.11 respectively. Meanwhile, the plots for the *Book Arrival* and *Newspaper* test sequences, averaged over a chunk period (which is set equal to the intra-frame period) at a low and high total bit rate, are shown in Figure 5.12 and 5.13 respectively.


Figure 5.10: Optimal depth-texture ratio – model formulation

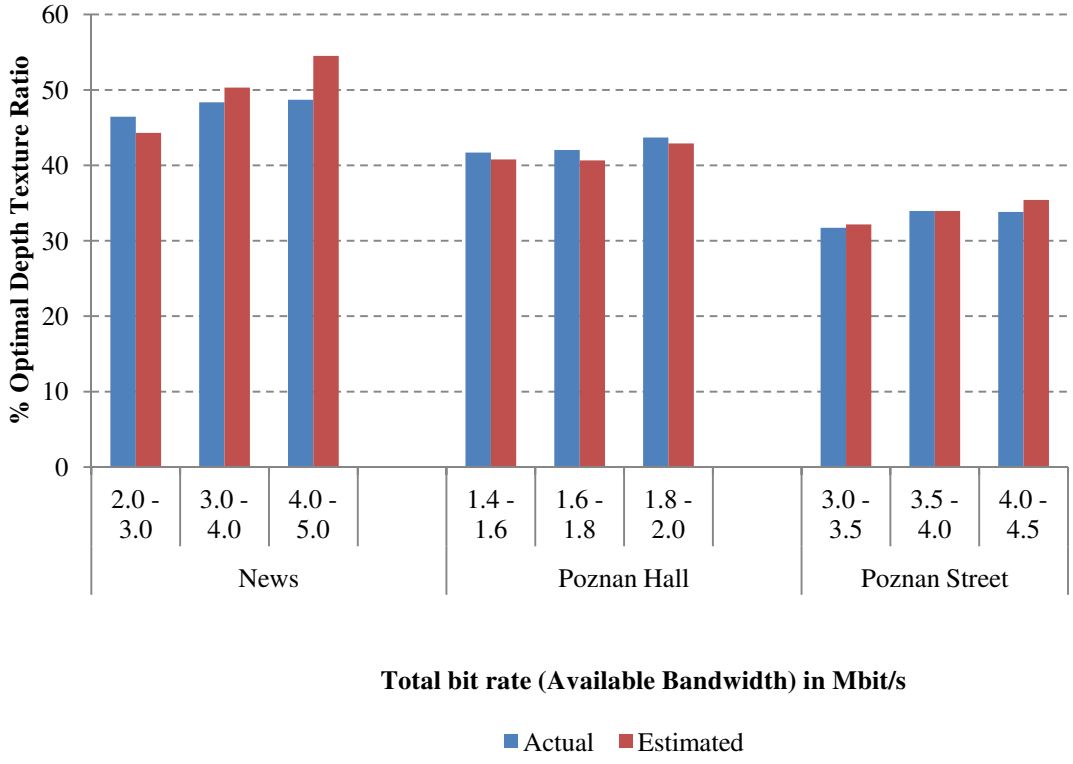
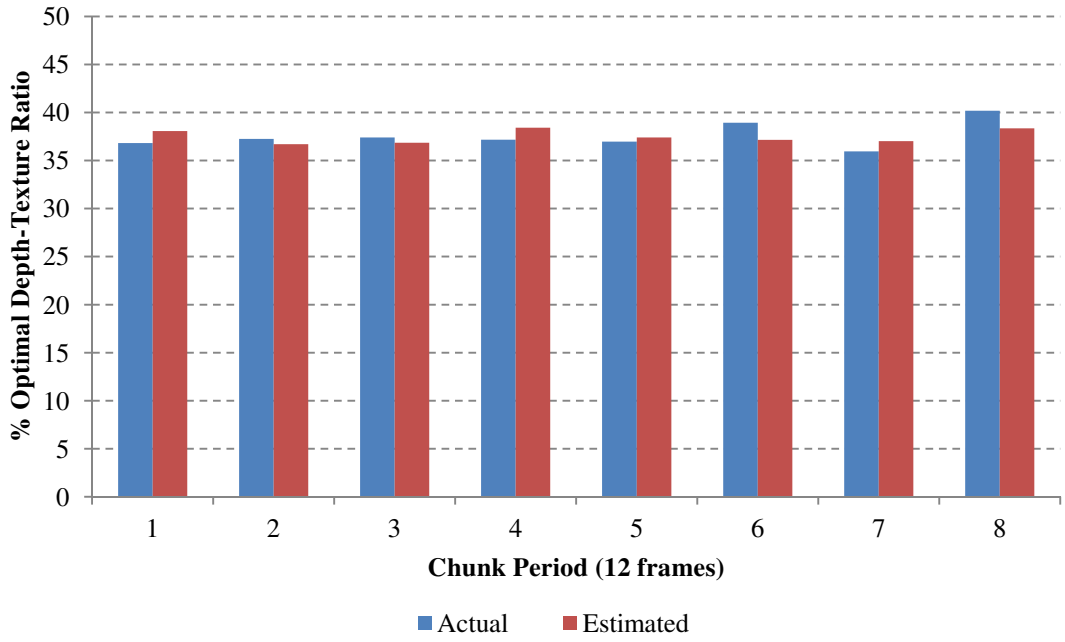
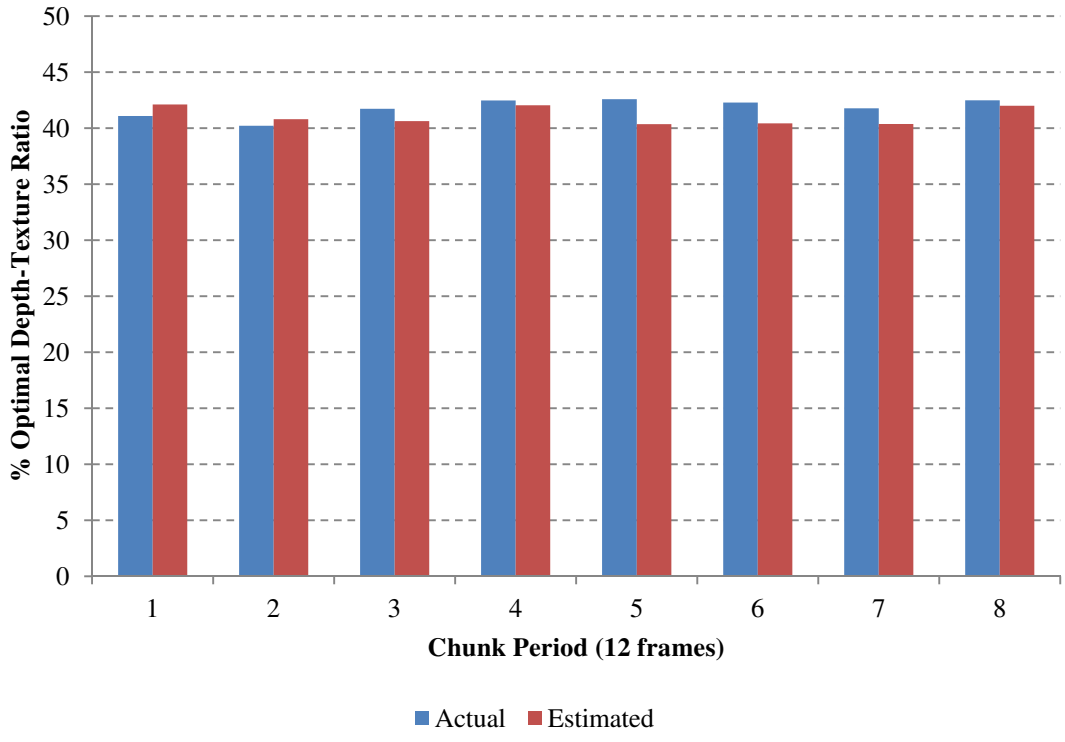



Figure 5.11: Optimal depth-texture ratio – model verification

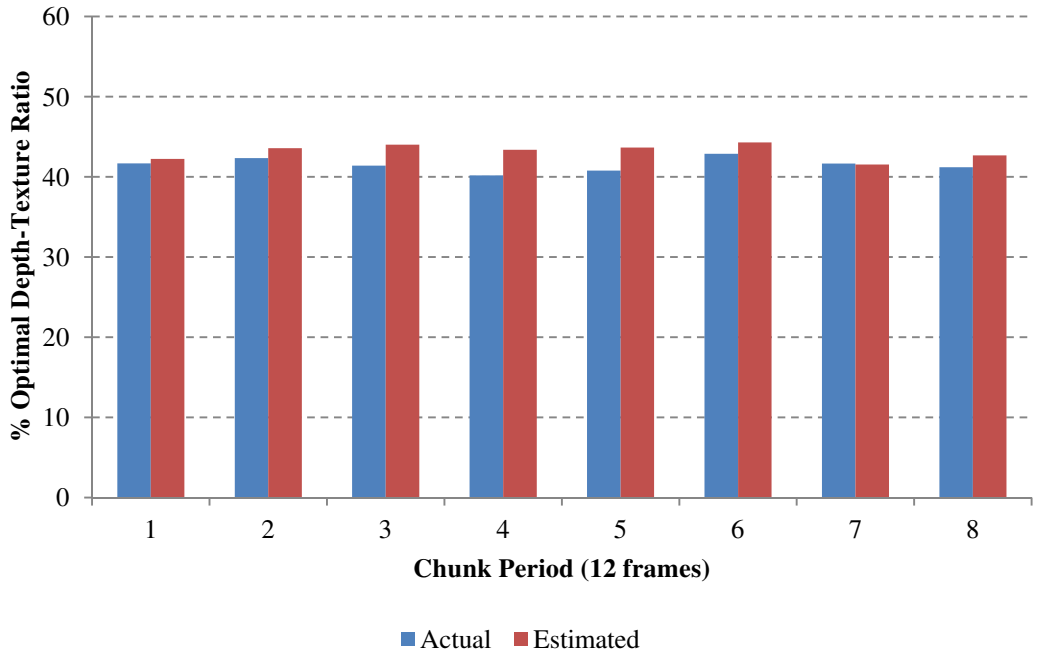



(a)

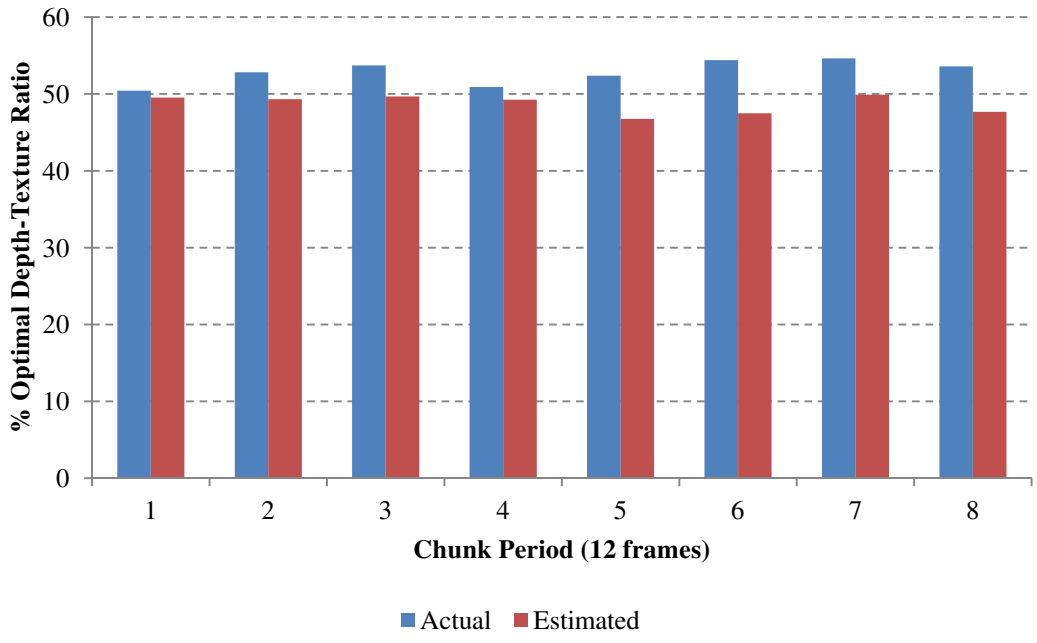



(b)

Figure 5.12: Optimal depth-texture ratio – *Book Arrival* test sequence (a) Low bit rate (720kbit/s), (b) High bit rate (1.56Mbit/s)




(a)




(b)

Figure 5.13: Optimal depth-texture ratio – *Newspaper* test sequence (a) Low bit rate (2.0Mbit/s), (b) High bit rate (3.9 Mbit/s)

As shown in Table 5.3, the proposed solution is able to estimate the optimal depth-texture rate allocation ratio for the test sequences used during model formulation with an absolute mean estimation error of 2.2% and standard deviation of 2%. As expected,

these metrics degrade slightly for the *Newspaper*, *Poznan Hall* and *Poznan Street* verification test sequences with an absolute mean estimation error of 3.3% and standard deviation of 2.3%. Even though the estimation error is relatively low, the effect of the estimation error coupled with the standard deviation on the synthesised view PSNR was evaluated for various test sequences at different total bit rates and this was found to be less than 0.1dB. This is in accord with the observation noted earlier in that the synthesised view PSNR is fairly constant over a relatively narrow range around the optimal value. A series of experiments were conducted on the *Newspaper* test sequence to compare the proposed cross-layer scheme, which adapts the depth-texture rate allocation to the sequence characteristics and available total bit rate, and the fixed depth-texture rate allocation scheme. As expected, with reference to Figure 5.14, the proposed scheme is able to improve the synthesised view PSNR by an average of 1.2dB when compared to a fixed depth-texture rate allocation of 20% [92].

| **Test Sequence** | **% Absolute Mean Estimation error** | **% Standard Deviation of the Estimation error** |
|---|---|---|
| Model Formulation | 2.2 | 2.0 |
| Model Verification | 3.3 | 2.3 |

Table 5.3: Statistical model estimation error

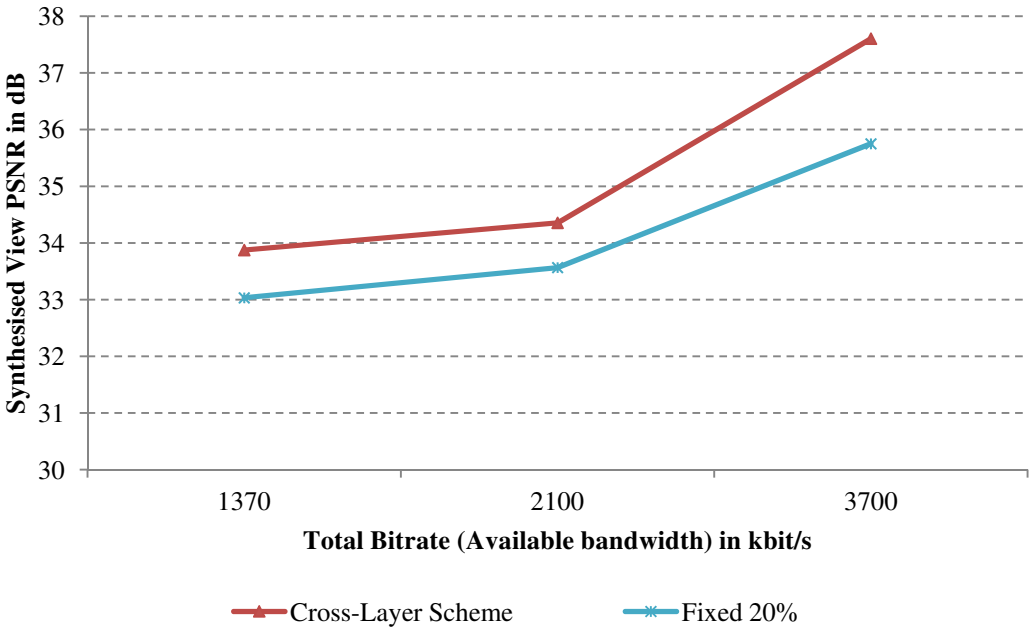



Figure 5.14: Synthesised view PSNR – proposed cross-layer scheme vs. fixed scheme at 20%

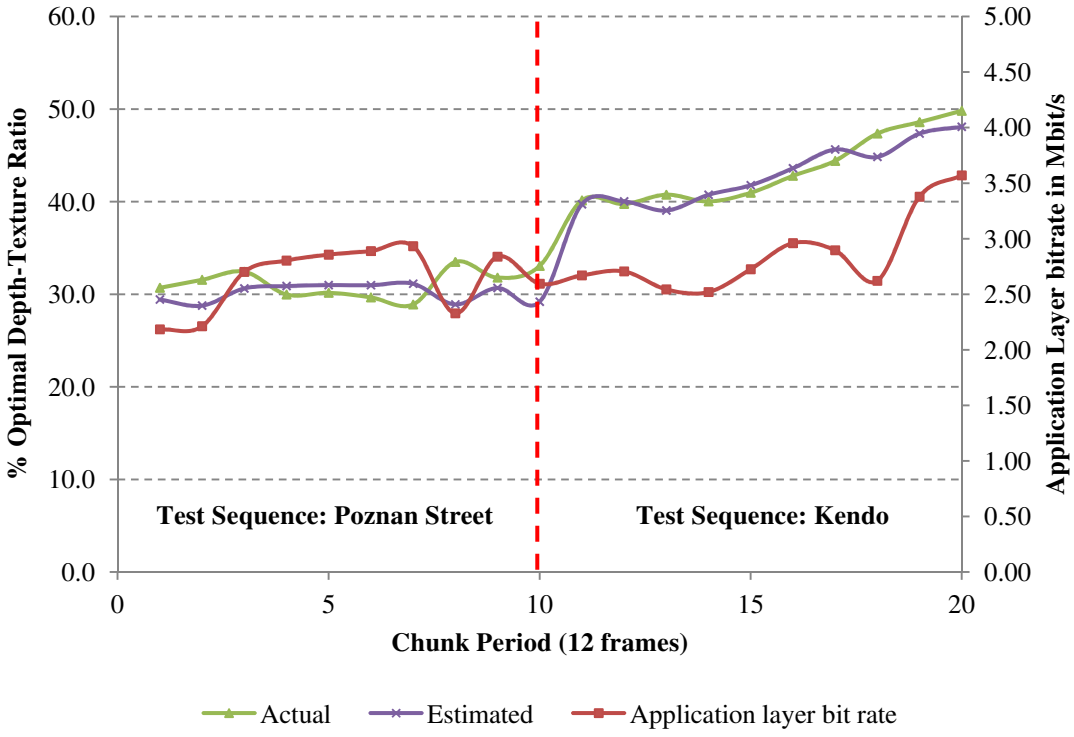
### 5.1.6 Depth-Texture rate allocation estimation in a crowd event scenario

In order to evaluate the adaptation of the depth-texture rate allocation ratio using the proposed scheme, a series of 120 frames (10 chunks) from 3 different test sequences were combined and encoded using the setup shown in Figure 5.9. As opposed to the verification tests in the previous section, the application layer bit rate, $R_a$, was extracted from LTE simulations conducted using the crowd event scenario described in section 3.2.1 whilst the relevant simulation parameters are provided in Table 5.4. In these tests, the LTE simulations were only used to obtain the application layer throughput, per radio frame, experienced by a UE in mid-sector conditions and establish the impact on the application layer throughput caused by a change in the number of users in the cell. Based on this testing, the adaptation of the proposed scheme is assessed for both sequence characteristics and application layer bit rate changes.

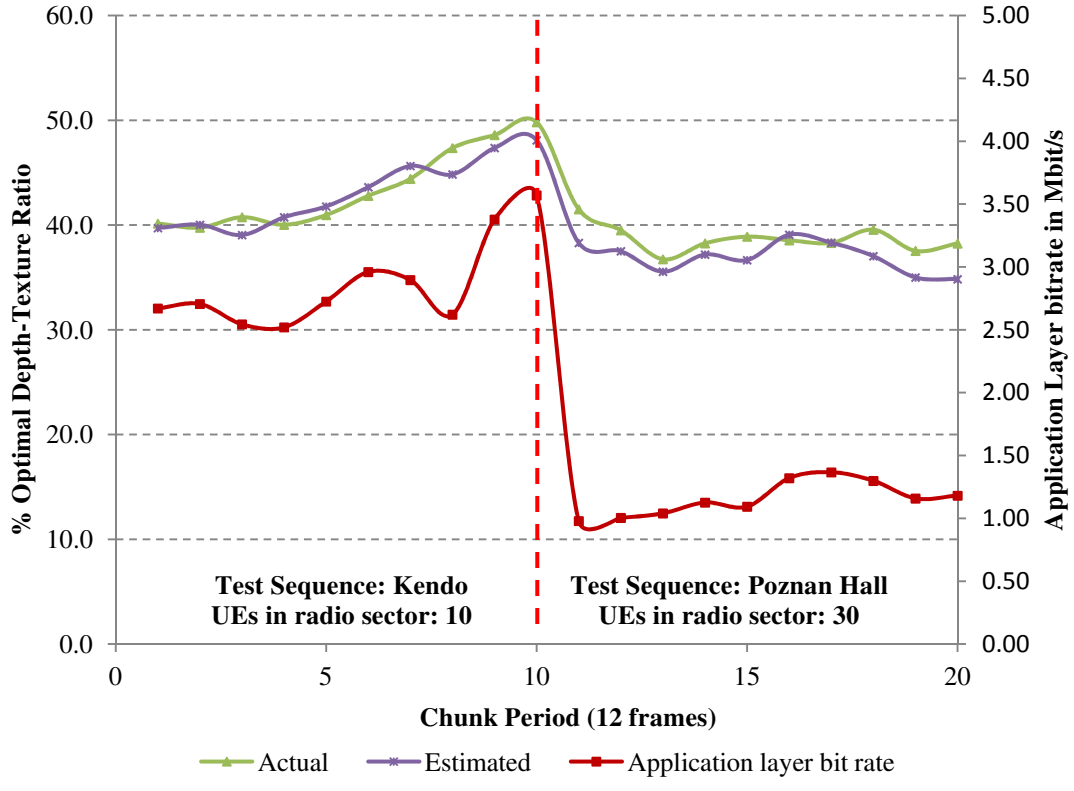
| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| Traffic Model | Full Buffer |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | {10,30} |

Table: 5.4: Simulation parameters used in rate allocation ratio testing

As evidenced in Figure 5.15, the optimal depth-texture rate allocation ratio, aggregated on a chunk level, indicates that the proposed scheme is able to adapt to varying video characteristics, caused by abrupt test sequence change and intra sequence changes, together with channel bandwidth variations caused by an increase in the number of UEs in the radio sector.




(a)




(b)

Figure 5.15: Adaptation of the optimal depth-texture rate allocation ratio caused by a variation in (a) Test sequence characteristics, (b) Application layer bandwidth

## 5.2 Enhanced Rate Control Scheme for depth map views

Following the texture and depth map view bit rate allocation, the estimated texture and depth map view bit rates are assigned as target values to the rate control algorithm in the respective MV-HEVC encoder. The proposed depth map rate control algorithm, shown in Figure 5.16, is based on the R-λ model [108] and operates on each depth map view independently at the group of pictures (GOP), frame, and coding tree unit level.

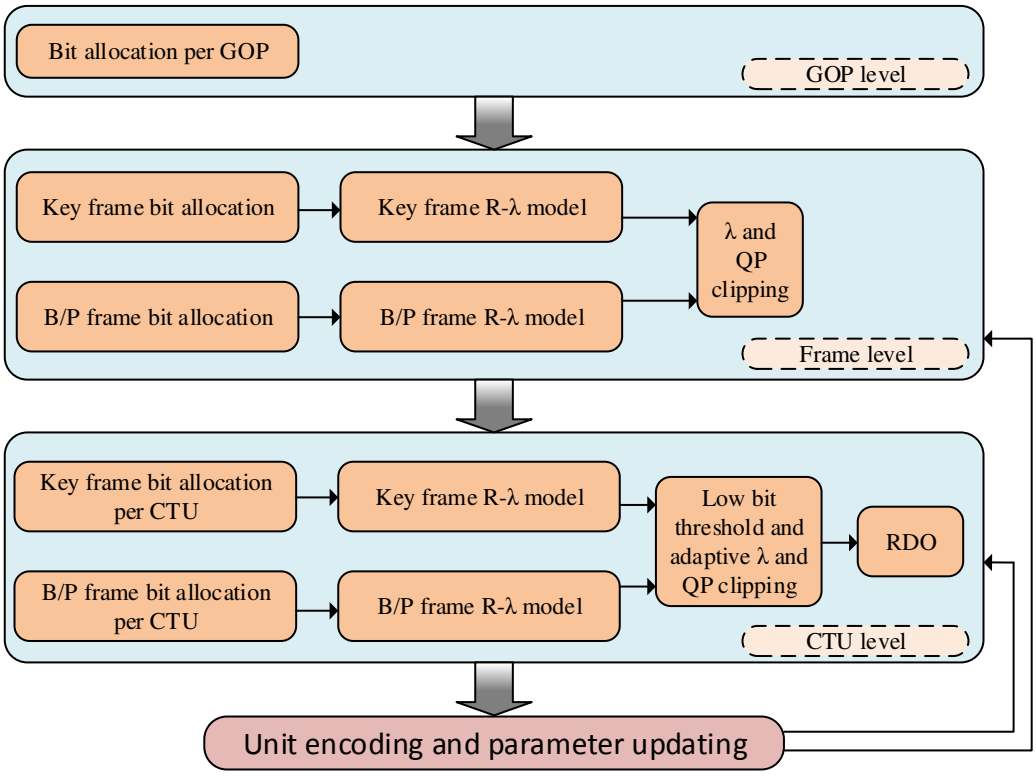



Figure 5.16: Proposed depth map rate control algorithm

### 5.2.1 GOP Level Bit allocation

At the GOP level, the bit allocation is similar to that in [108] and is given by:

$$T_{GOP} = N_{GOP}\left(\frac{\left(\frac{R_{target}}{f}\right)\left(N_{S,CODED}+S_W\right)-T_{S,CODED}}{S_W}\right) \tag{5.11}$$

where $N_{GOP}$ is the number of frames in the GOP, $T_{S,CODED}$ is the bit cost of the coded pictures in the sequence, $N_{S,CODED}$ is the number of pictures in sequence already coded, $R_{target}$ is the target bit rate, $f$ is the frame rate whilst $S_W$ is the smoothing window size

which is set to 40 and is used to allow the controller to adapt the bit rate smoothly and achieve $R_{target}$ after $S_W$ frames.

### 5.2.2 Frame Level Bit allocation

In the proposed scheme, the inter-B/P frame bit allocation and R-λ model is analogous to the scheme proposed in [108] and is given by (5.12) and (5.13) respectively.

$$T_{avgpic} = \frac{T_{GOP} - T_{G,CODED}}{\sum_{NotCodedPics} \omega_i} \cdot \omega_{CurPic} \tag{5.12}$$

$$\lambda_{pic} = \alpha_{B,pic} \left( \frac{T_{avgpic}}{P_x} \right)^{\beta_{B,pic}} \tag{5.13}$$

where $T_{avgpic}$ is the initial bit budget allocated per frame, $T_{GOP}$ is the target number of bits in a GOP and is computed using (5.11), $T_{G,CODED}$ is the bit cost of already coded frames in the current GOP, $\omega$ is the bit allocation weighting factors for the different hierarchical levels as defined in [108], $\alpha_{B,pic}$ and $\beta_{B,pic}$ are the R-λ model parameters initially set to 3.2003 and -1.367 respectively and adaptively tuned during the parameter update procedure described in [108], whilst $P_x$ is the number of pixels in a frame.

As opposed to the standard rate control scheme, in order to signify the importance of the key frames in relation to the other frames in the GOP, the bit allocation of the key frames is refined and is given by:

$$T_{key} = 0.25 * \left( \frac{C}{T_{avgpic}} \right)^{\gamma} \tag{5.14}$$

where $C$ is the complexity measure based on the sum of absolute transformed difference (SATD) [112] and $\gamma$ is a tuning parameter controlling the amount of bit budget allocated to the key frames from the total.

Whilst $\gamma$ is typically set to 0.5582 [112] for the intra frames, this parameter was empirically tuned for the P-key frames by first encoding several depth map test sequences at various target bit rates and evaluating the average and standard deviation of the synthesised view PSNR together with the average PSNR of the posterior frames in the GOP. With reference to Figure 5.17, the average PSNR of the synthesised view

for the *Balloons* test sequence encoded at a target bit rate of 110kbps increases with $\gamma$. Moreover, as expected, the average PSNR of the posterior frames degrades rapidly as $\gamma$ exceeds 0.5 since the number of bits available for the posterior frames decreases. Based on these observations and after studying the variation of $\gamma$ for various test sequences, the value of $\gamma$ was empirically set to 0.41 as this strikes a balance between the identified quality metrics.

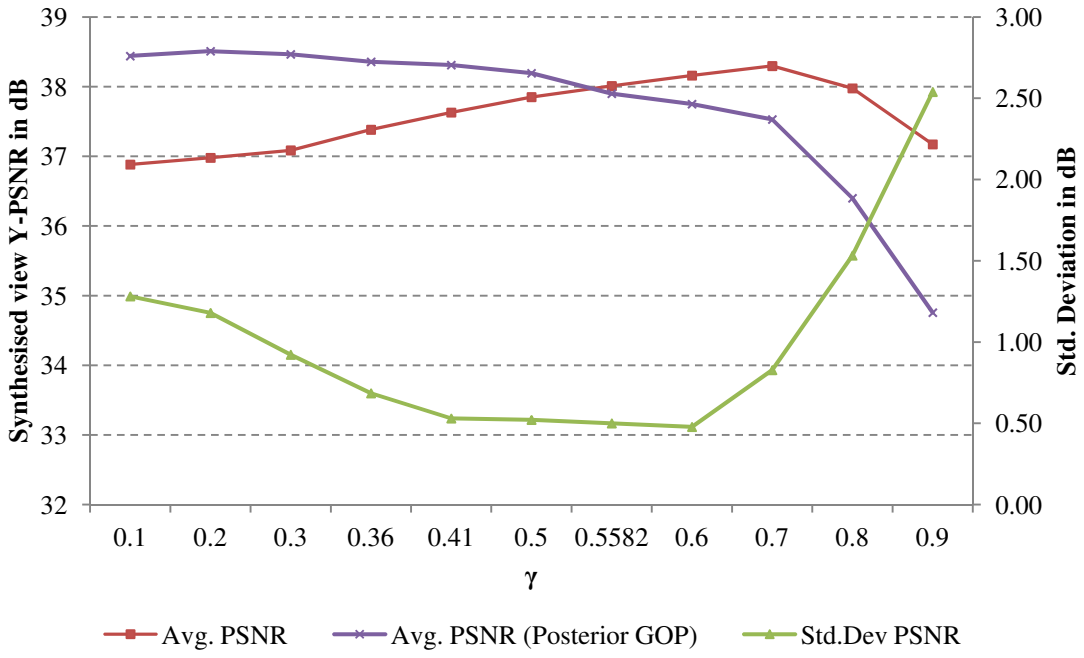



Figure 5.17: Variation of tuning parameter for the *Balloons* test sequence

### 5.2.3 Key frame R-λ model

The R-λ model used in the standard rate control schemes is tuned for the texture views and does not consider the characteristics of depth map views. In fact, tests conducted on texture and depth map views of the test sequences shown in Table 5.2 reveal a significant disparity in zero complexity cost CTUs whereby on average, depth maps have 26% of CTUs with zero complexity cost whilst, in general, texture views do not exhibit any CTUs with zero complexity cost. This is expected as depth maps have a large number of flat areas when compared to texture views and thus the key-frame R-λ model, given by (5.13), was adapted to:

$$\lambda_{pic} = \alpha_{K,pic} \left( \frac{C}{\left( T_{key} / P_x \right)} \right)^{\beta_{K,pic}} \tag{5.15}$$

where $C$ is the complexity measure based on SATD [112], $T_{key}$ is the refined bit budget of the key frame, $P_x$ is the number of pixels in a frame, whilst $\alpha_{K,pic}$ and $\beta_{K,pic}$ are model parameters which are empirically set to fit the hyperbolic model.

In order to fit the parameters of the hyperbolic model, experiments were carried out using the MV-HEVC encoder configured such that the depth map frames across the three views are encoded as key frames. The test sequences, shown in Table 5.2, were encoded at four QP levels, namely 34, 39, 42 and 45 as defined in the Common Test Conditions (CTC) [125]. Meanwhile, the average bits per pixel (bpp), frame complexity $C$, and $\lambda_{pic}$ were noted in order to model the relationship between $\lambda_{pic}$ and ($C/bpp$) for the depth map key frames across the 3 views. The relationship between these two variables and model parameter estimates for the *Balloons* test sequence across the 3 views is shown in Figure 5.18. With reference to Figure 5.18, the model parameters for the reference and non-reference views impart different hyperbolic models and thus different R-$\lambda$ models are considered depending on the view type. Considering the average values of the model parameters estimates for all test sequences yields the R-$\lambda$ parameters shown in Table 5.5.

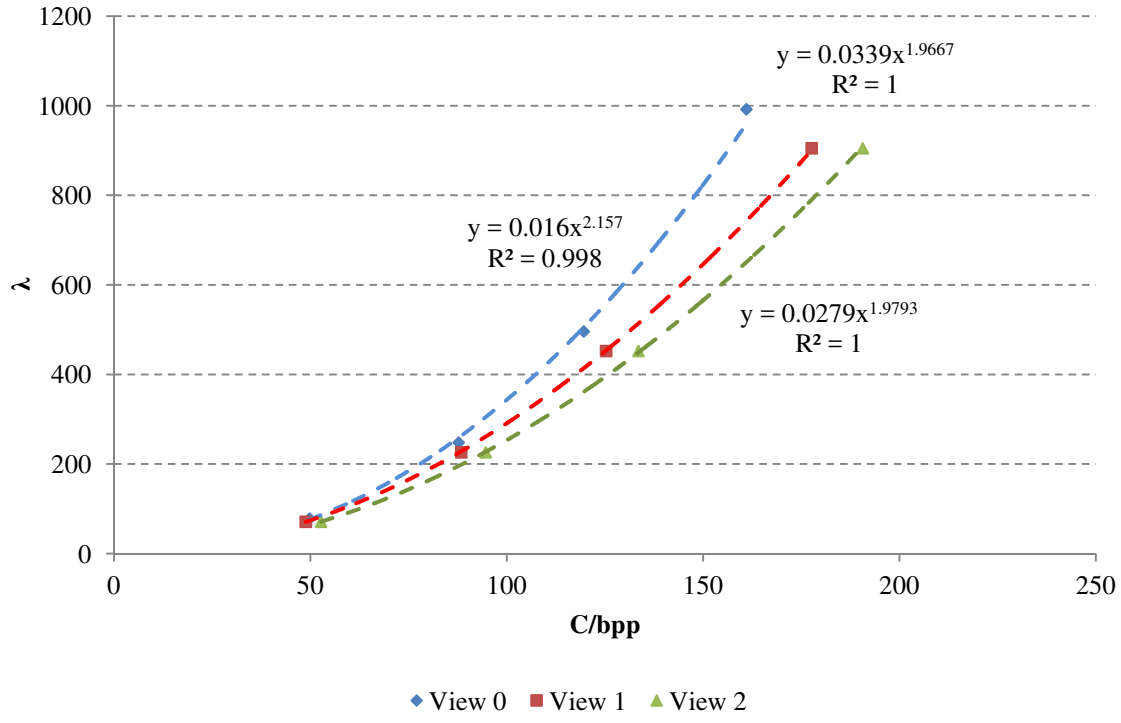



Figure 5.18: R-λ modelling for the *Balloons* test sequence

| View Type | $\alpha_{K,pic}$ | $\beta_{K,pic}$ |
|---|---|---|
| Reference View | 0.018 | 2.104 |
| Non-Reference View | 0.028 | 1.9495 |

Table 5.5: Key Frame R-λ model parameters

Using the estimated $\lambda_{pic}$, the QP for the frame is then calculated using [111]:

$$QP = 4.2005 \ln(\lambda) + 13.7122 \tag{5.16}$$

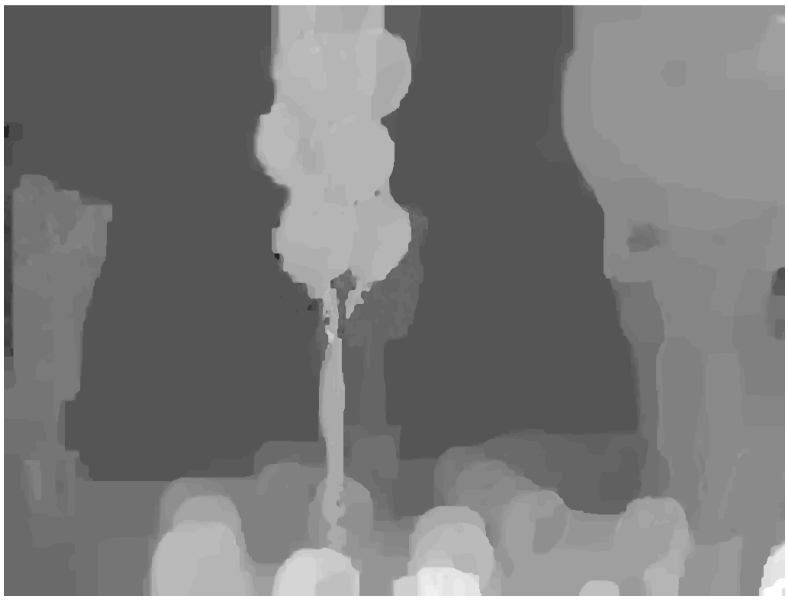
### 5.2.4 CTU Level Bit allocation

The bit budget allocated to the frame is distributed across the CTUs in the frame based on their cost complexity. Although the R-λ models used at the frame level, given by (5.13) and (5.15), are re-applied at the CTU level, the model parameter estimates of the CTUs in a frame are initially cascaded from the frame level R-λ model and then adaptively tuned, on a CTU level basis, during the CTU parameter update procedure described in [108]. Based on the CTU level bit allocation and associated R-λ model, the $\lambda_{CTU}$ is estimated, clipped and then used to calculate $QP_{CTU}$ using (5.16), which is subsequently clipped prior to being used for unit encoding. In the proposed rate control scheme, the clipping range of $\lambda_{CTU}$ and $QP_{CTU}$ is adaptively tuned based on the type of CTU being processed so as to allow a wider clipping range for the CTUs in the depth map view located in the foreground and edges. In order to adapt the clipping range, the CTUs in a frame are classified into three categories namely edge, background and foreground type CTU depending on their characteristics using the algorithm shown in Figure 5.19. This algorithm is based on a binary edge map generated by an edge detection technique and even though one finds a number of edge detection schemes in literature, a Canny edge filter [126] is used in this work in view of its low complexity. Considering the original video depth map shown in Figure 5.20 (a), the Canny edge filter generates the binary edge map shown in Figure 5.20 (b) which is then processed, in conjunction with the original video depth map, using the algorithm shown in Figure 5.19 to generate the CTU mask shown in Figure 5.20 (c). With reference to Figure 5.20 (c), CTUs marked in red are treated as foreground while those marked in green are treated as edges. The rest are treated as background CTUs.

**Algorithm 2: CTU Mask Algorithm**

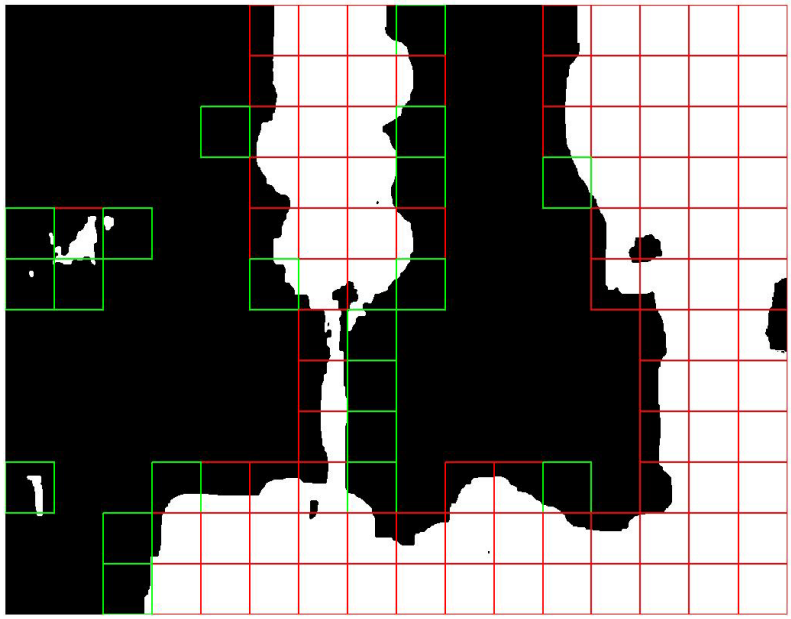
```
1: for all binary edge map frames do
2:     for Each LCU in binary edge map do
3:         Split LCU into 16 16×16 sub-blocks
4:         for each sub-block do
5:             check whether sub-block contains an edge
6:             if sub-block contains an edge then
7:                 mark sub-block as containing an edge
8:             end if
9:         end for
11:        for each sub-block containing an edge in LCU do
12:            check whether neighbouring sub-blocks contain an edge
13:            if neighbouring sub-block containing an edge is 3 or more then
14:                Flag LCU as an Edge
15:            end if
16:        end for
17:    end for
18: end for
20: for all depth map frames do
21:     extract foreground mask by depth map thresholding
22:     for each LCU in foreground mask do
23:         if LCU in foreground mask is also marked as edge LCU then
24:             Mark LCU as Edge
25:         end if
26:         if LCU in foreground mask is not marked as edge LCU then
27:             Mark LCU as Foreground
28:         end if
29:     end for
31:     for all remaining LCU not in foreground mask do
32:         Mark LCU as Background
33:     end for
34: end for
```

Figure 5.19: Algorithm 2 - CTU mask algorithm



(a)

(b)



(c)

Figure 5.20: CTU mask generation for *Balloons* [127] – Camera 1 (red – foreground, green – edge) (a) Original video depth map, (b) binary edge map, (c) CTU mask

Based on the generated CTU mask, the clipping range of $\lambda_{CTU}$ and $QP_{CTU}$ is then adapted based on the current and previous CTU type as shown in (5.17).

If $T_{remain} < T_{thresh}$ then:

$$\lambda_{CTU\_prev} \cdot 2^{-\frac{1}{3}} \le \lambda_{CTU\_cur} \le \lambda_{CTU\_prev} \cdot 2^{\frac{1}{3}}$$

$$\lambda_{pic} \cdot 2^{-\frac{2}{3}} \le \lambda_{CTU\_cur} \le \lambda_{pic} \cdot 2^{\frac{2}{3}}$$

$$QP_{CTU\_prev} - 1 \le QP_{CTU\_cur} \le QP_{CTU\_prev} + 1$$

$$QP_{pic} - 2 \le QP_{CTU\_cur} \le QP_{pic} + 2$$

Else

$$\lambda_{CTU\_prev} \cdot 2^{-\left(\frac{A}{3}\right)} \le \lambda_{CTU\_cur} \le \lambda_{CTU\_prev} \cdot 2^{\left(\frac{B}{3}\right)}$$

$$\lambda_{pic} \cdot 2^{-\frac{2}{3}} \le \lambda_{CTU\_cur} \le \lambda_{pic} \cdot 2^{\frac{2}{3}}$$

$$QP_{CTU\_prev} - \Delta \le QP_{CTU\_cur} \le QP_{CTU\_prev} + \Delta$$

$$QP_{pic} - 4 \le QP_{CTU\_cur} \le QP_{pic} + 4 \qquad (5.17)$$

where $\lambda_{CTU\_prev}$ and $\lambda_{CTU\_cur}$ are the $\lambda$ of the previous and current CTU respectively, $QP_{CTU\_prev}$ and $QP_{CTU\_cur}$ are the QP values of the previous and current CTU respectively, $\lambda_{pic}$ and $QP_{pic}$ are the $\lambda$ and QP of the frame respectively, $T_{remain}$ and $T_{thresh}$ are the number of bits remaining to code the current frame and low bit threshold respectively, whilst the parameters *A, B,* and $\Delta$ depend on the previous and current CTU type and are given in Table 5.6.

| Previous CTU Type | Current CTU Type | *A* | *B* | $\Delta$ |
|---|---|---|---|---|
| Background | Background | 1 | 1 | 1 |
| Foreground | Foreground | 2 | 2 | 2 |
| Edge | Edge | 2 | 2 | 2 |
| Edge | Foreground | 2 | 2 | 2 |
| Foreground | Edge | 2 | 2 | 2 |
| Background | Edge | 4 | 4 | 3 |
| Foreground | Background | 2 | 4 | 3 |
| Edge | Background | 2 | 4 | 3 |
| Background | Foreground | 4 | 4 | 3 |

Table 5.6: Adaptive clipping parameters

The parameters shown in Table 5.6 were empirically derived such that the salient characteristics of the depth map are retained whilst minimising frame bit exhaustion. The latter was also tackled through the use of a low bit threshold check at the start of processing of each CTU in the frame.

### 5.2.5 Results

In order to establish the performance of the proposed scheme, the HEVC reference software HM12.0 [118] was modified to implement the new rate control algorithm and configured for multi-view compression using the low-delay prediction structure, shown in Figure 5.3. The target bit rates used for testing were set as per the CTC described in [125] whilst the encoding parameters are shown in Table 5.8. As the focus of this work is on the depth map views, the texture views are encoded using the reference rate control algorithm [108] whilst the performance of the proposed depth map rate control algorithm was assessed, in terms of synthesised view PSNR improvement, BD-PSNR [128], and bit rate error as compared to the reference algorithm implemented in MV-HEVC. With reference to Table 5.7, the average

improvement of the synthesised view PSNR and BD-PSNR are 1.15% and 0.45dB respectively whilst bit rate error was reduced by an average of 0.2% when compared to the reference algorithm in MV-HEVC. Meanwhile, as shown in Figure 5.21, the proposed depth map rate control scheme yields a better R-D performance when compared to the reference algorithm used in MV-HEVC due to the refined bit budget allocation to the key frames and salient CTUs coupled with an improved R-λ model for the depth map.

| **Sequence / Resolution** | **Target bitrate (kbps)** | **% Δ PSNR** | **% Δ bitrate error** | **BD-PSNR (dB)** |
|---|---|---|---|---|
| *Balloons* *(1-**3**-5) 1024x768* | 466.7 | 1.01 | -0.61 | 0.55 |
| | 219.3 | 1.23 | -0.79 | |
| | 140.1 | 1.28 | -0.00 | |
| | 91.4 | 1.88 | -0.03 | |
| *Book* *(8-**9**-10) 1024x768* | 295.7 | 1.20 | -0.27 | 0.41 |
| | 152.1 | 0.91 | -0.18 | |
| | 101.5 | 1.29 | -0.13 | |
| | 69 | 1.13 | -0.09 | |
| *Kendo* *(1-**3**-5) 1024x768* | 597.5 | 0.76 | -0.22 | 0.3 |
| | 314.7 | 0.72 | -0.15 | |
| | 213.9 | 0.66 | -0.10 | |
| | 147.5 | 0.60 | -0.08 | |
| *Newspaper* *(2-**4**-6) 1024x768* | 370.1 | 1.47 | -0.65 | 0.66 |
| | 181.5 | 1.79 | -0.07 | |
| | 115.9 | 2.16 | -0.05 | |
| | 77.2 | 2.70 | -0.04 | |
| *PoznanHall* *(7-**6**-5) 1920x1088* | 205.8 | 1.01 | -0.28 | 0.39 |
| | 108.3 | 1.06 | -0.16 | |
| | 77.8 | 0.71 | -0.07 | |
| | 57 | 0.32 | -0.06 | |
| *PoznanStreet* *(5-**4**-3) 1920x1088* | 332.7 | 0.69 | -0.21 | 0.39 |
| | 150.2 | 1.12 | -0.34 | |
| | 96.7 | 0.82 | -0.11 | |
| | 66.4 | 1.04 | -0.02 | |
| **Average** | | **1.15** | **-0.20** | **0.45** |

Table 5.7: Rate Control Performance

(Synthesised view is indicated by the camera number shown in **bold**)

| Parameter | Value |
|---|---|
| GOP Size | 4 |
| Intra Period | 24 |
| LCU Level Rate Control | 1 |
| LCU size | 64 x 64 |
| Partition Depth | 4 |
| LCU separate RC Model | True |
| $T_{thresh}$ | 5% of frame allocation |

Table 5.8: Encoding parameters

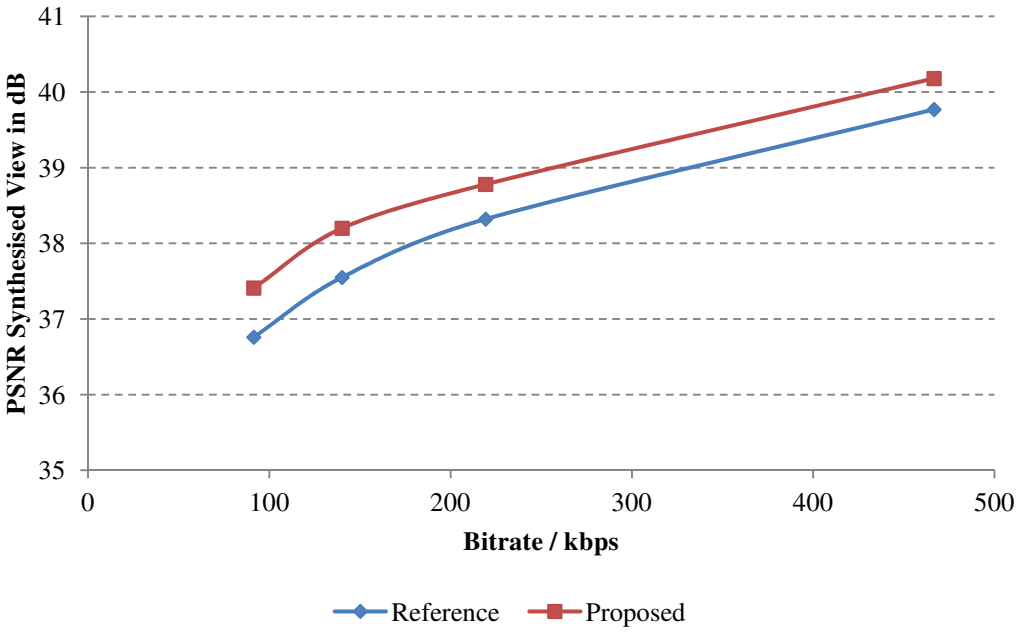



Figure 5.21: R-D plot - *Balloons* Synthesised View 3

## 5.3 Summary

The first part of this chapter focused on the design, implementation and testing of a cross-layer depth-texture rate allocation scheme. As opposed to other schemes found in literature, the proposed scheme is able to estimate, on-line, the depth-texture rate allocation ratio without the use of pre-encoding of frames. This is based on a statistical model formulated using both image characteristics metrics and the total available bit rate for the MVD stream. The statistical model formulation starts with a series of experiments showing the importance of finding the right balance between the

depth map view and texture view bit rate together with the fact that the optimal depth-texture ratio varies with both sequence characteristics and total bit rate. This is then followed by a statistical analysis yielding a multiple regression model composed of seven significant predictors and having a coefficient of determination, $R^2$, of 0.86. Moreover, a series of rigorous statistical tests were conducted in order to ensure model robustness prior to being tested on several test sequences. The proposed scheme is not only able to estimate accurately the optimal depth-texture rate allocation for test sequences which were not used in model formulation with an estimation error of 3.3% but is able to adapt to varying video characteristics and channel bandwidth.

In the second part of this chapter, the characteristics of depth map views were exploited to enhance the rate control performance for such views. Although the proposed scheme is based on an already established rate control algorithm, a number of enhanced features, including a refined key frame bit allocation and R-λ model coupled with an adaptive Lagrange multiplier and QP clipping function, were introduced. As evidenced by the simulation results shown in Table 5.7, the enhanced rate control scheme achieved an average improvement of the synthesised view PSNR and BD-PSNR of 1.15% and 0.45dB respectively when compared to the standard rate control scheme.

# Chapter 6 Mobile Video Transmission over LTE Networks

This chapter presents the foundations for the transmission of mobile video over an LTE network and starts with a description of the video bit stream, packetisation and the LTE Quality of Service (QoS) framework. The next section focuses on the eNB packet scheduler, which was discussed in chapter 2, and gives an overview of the relevant packet scheduling algorithms found in literature. Finally, in the last section the cross layer design concept, which was introduced in chapter 5, is elaborated with a description of the various types of cross-layer schemes together with a review of relevant works.

## 6.1 Video bit stream, packetisation and LTE QoS framework

### 6.1.1 Video bit stream and packetisation

The coded bit stream structure at the output of the HEVC video encoder inherits a number of structural elements from H.264/AVC and is composed of a series of network abstraction layer (NAL) units. As shown in Table 6.1, there are 64 different NAL unit types divided into two classes namely Video Coding Layer (VCL) NAL units, carrying coded slice video information, and non-VCL NAL units carrying control information used by the decoder to reconstruct the video sequence from the bit stream. The latter includes [70]:

- Video Parameter Set (VPS): This was introduced in HEVC and applies to all layers. It includes the number of layers, decoded picture buffer size and picture ordering parameters together with information related to temporal scalability amongst other parameters.
- Sequence Parameter Set (SPS): This contains parameters which apply to an entire coded video sequence and do not change from one picture to another. These include the coding tools used and associated parameters, the conformance parameters such as the profile, tier and level indicators, picture height and width in luma samples together with luma and chroma bit depth.
- Picture Parameter Set (PPS): This contains parameters which change for different pictures types within a video sequence such as reference indices,

initial QP values, chroma QP offsets, de-blocking filter controls and tile configurations.

| NAL unit type | Picture Type Category | Picture Type | Identifier | Short name description |
|---|---|---|---|---|
| VCL | Trailing non-IRAP | Non-TSA, non-STSA trailing | 0 | TRAIL_N |
| | | | 1 | TRAIL_R |
| | | Temporal sub-layer access (TSA) | 2 | TSA_N |
| | | | 3 | TSA_R |
| | | Step-wide temporal sub-layer (STSA) | 4 | STASA_N |
| | | | 5 | STSA_R |
| | Leading | Random access decodable | 6 | RADL_N |
| | | | 7 | RADL_R |
| | | Random access skipped leading | 8 | RASL_N |
| | | | 9 | RASL_R |
| | Intra random access point (IRAP) | Broken link access | 16 | BLA_W_LP |
| | | | 17 | BLA_W_RADL |
| | | | 18 | BLA_N_LP |
| | | Instantaneous decoding refresh | 19 | IDR_W_RADL |
| | | | 20 | IDR_N_LP |
| | | Clean random access | 21 | CRA |
| | Reserved | Reserved non-IRAP | 10-15 | RSV |
| | | Reserved IRAP | 22-23 | RSV |
| | | Reserved non-IRAP | 24-31 | RSV |
| Non-VCL | Parameter sets | Video parameter set | 32 | VPS_NUT |
| | | Sequence parameter set | 33 | SPS_NUT |
| | | Picture parameter set | 34 | PPS_NUT |
| | Delimiters | Access unit delimiter | 35 | AUD_NUT |
| | | End of sequence | 36 | EOS_NUT |
| | | End of bitstream | 37 | EOB_NUT |
| | Filler data | Filler data | 38 | FD_NUT |
| | Supplemental enhancement information (SEI) | Prefix SEI | 39 | PREFIX_SEI_NUT |
| | | Suffix SEI | 40 | SUFFIX_SEI_NUT |
| | Reserved | | 41-47 | RSV |
| | Unspecified | | 48-63 | UNSPEC |

Table 6.1: NAL Unit Classification

A VCL NAL unit contains a slice segment of a coded picture whilst a coded picture, which can be composed of multiple slice segments, together with any associated non-VCL NAL units is referred to as an HEVC access unit (AU) [129]. Similarly, in the case of layered extensions such as MV-HEVC, an AU contains all coded pictures from

the *n* views (layers) belonging to the same picture order count (POC) as shown in Figure 6.1. Extending this concept to the MVD case yields two separate bit streams which are then multiplexed for transmission by the underlying network.


Figure 6.1: MV-HEVC AU structure

The NAL unit described earlier is composed of a fixed two byte NAL header followed by a NAL payload referred to as raw byte sequence payload (RBSP). As shown in Figure 6.2, the header describes the NAL unit type, layer identifier, which is applicable to the scalable and MV/3D video coding extensions [130], and temporal sub-layer identifier. Thus, any media-aware network element (MANE) along the end-to-end transmission chain can determine the content priority with relative ease by inspecting the NAL unit header

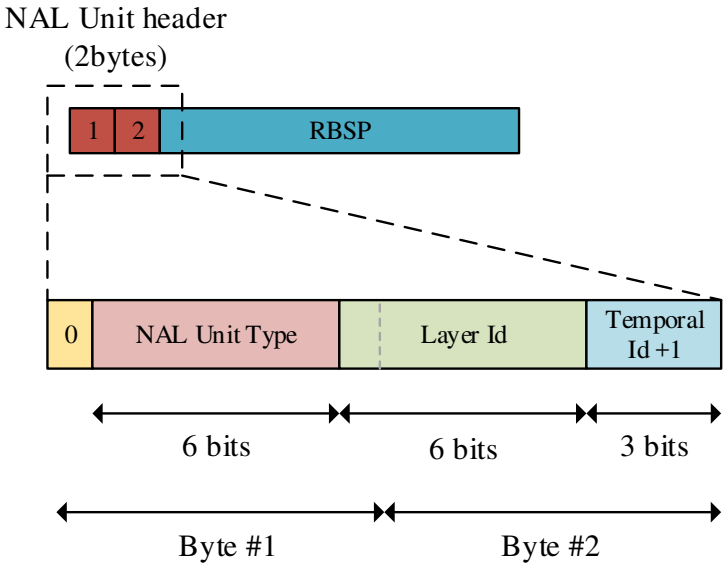



Figure 6.2: NAL Unit header

Whilst there exist several system layer technologies which integrate with HEVC, as illustrated in Figure 6.3, the most prominent delivery mechanisms over packet networks are based on MPEG-Dynamic Adaptive Streaming over HTTP (DASH) [131,132] and the Real Time Protocol (RTP) [133].

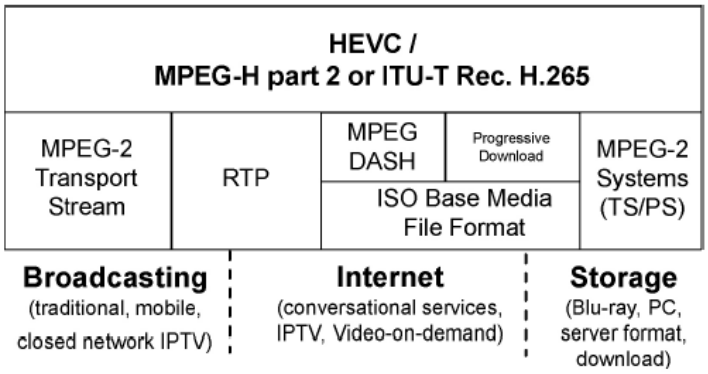



Figure 6.3: HEVC system layers [130]

The focus of this work is on RTP which has been widely adopted and used in various services, ranging from IPTV to conversational applications requiring relatively low latency such as video conferencing [130]. Whilst RTP can be used over a number of different transport protocols, it is usually operated in conjunction with User Datagram Protocol (UDP) which, in turn, is encapsulated in IP. Thus, considering a single RTP stream on a single media transport (SRST) and assuming that each slice is encoded in a

single NAL unit, the structure of the RTP header and associated RTP/UDP/IP encapsulated packet is shown in Figure 6.4.

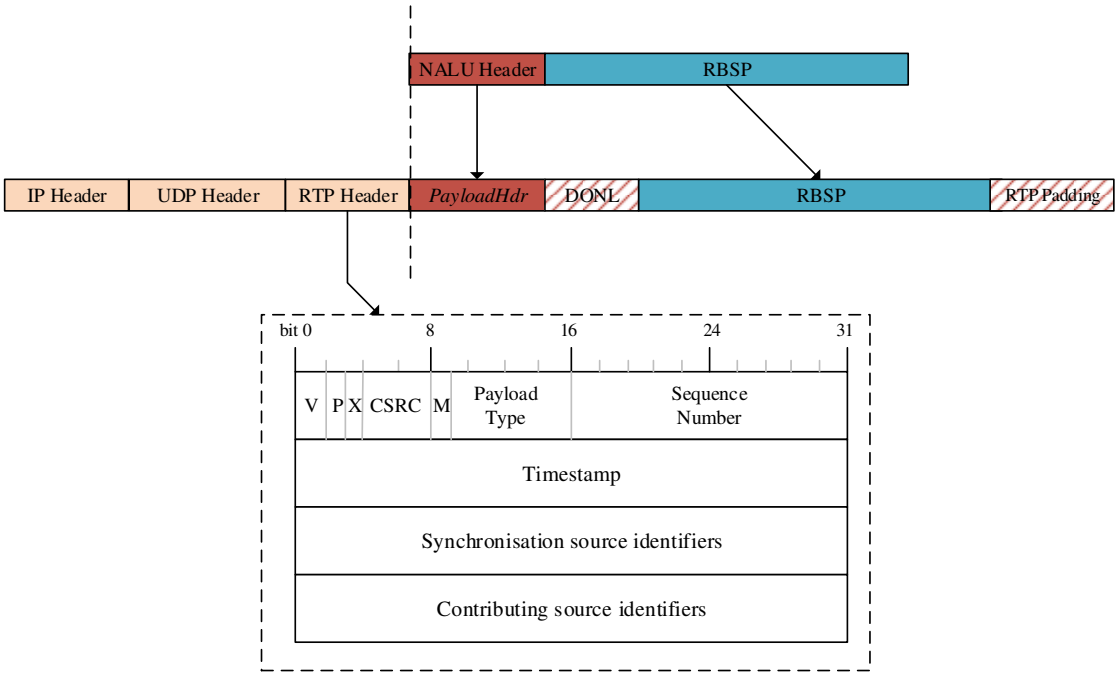



Figure 6.4: RTP header structure and RTP/UDP/IP encapsulation

Focusing on the NAL unit packet, the payload header, denoted by *PayloadHdr*, is a bit-exact copy of the NAL header, shown in Figure 6.2, whilst a detailed description of the remaining RTP/UDP/IP header fields is found in [133, 134]. A key parameter in the IP header, which is particularly relevant in this work, is the Differentiated Service Code Point (DSCP) value. The DSCP is a 6-bit field set by the transmitting entity to indicate the quality of service (QoS) level a packet receives in the network.

### 6.1.2 LTE QoS Framework

Starting with the first release of LTE, a comprehensive end-to-end LTE QoS framework, based on the Evolved Packet System (EPS) bearer model, has been defined in order to deliver demanding multimedia applications and services through the use of effective end-to-end management of network resources. With reference to Figure 6.5, LTE defines several bearers traversing various interface types with the most relevant to this discussion being the EPS bearer, which provides a logical data path between the UE and the packet data network gateway (P-GW). LTE defines two main classes of EPS bearers, namely default and dedicated data bearers. A default bearer is setup whenever a UE registers with the network thereby allowing continuous IP

connectivity. Although such bearers allow data transfer between the UE and EPS, they are assigned basic QoS capabilities without any service guarantees and are strictly used for best effort traffic. In case packet loss/delay sensitive traffic, such as real-time video, needs to be transported across an LTE network, a dedicated bearer having a more stringent QoS is setup, in addition to the default bearer, such that IP packets associated with this traffic type are treated accordingly.

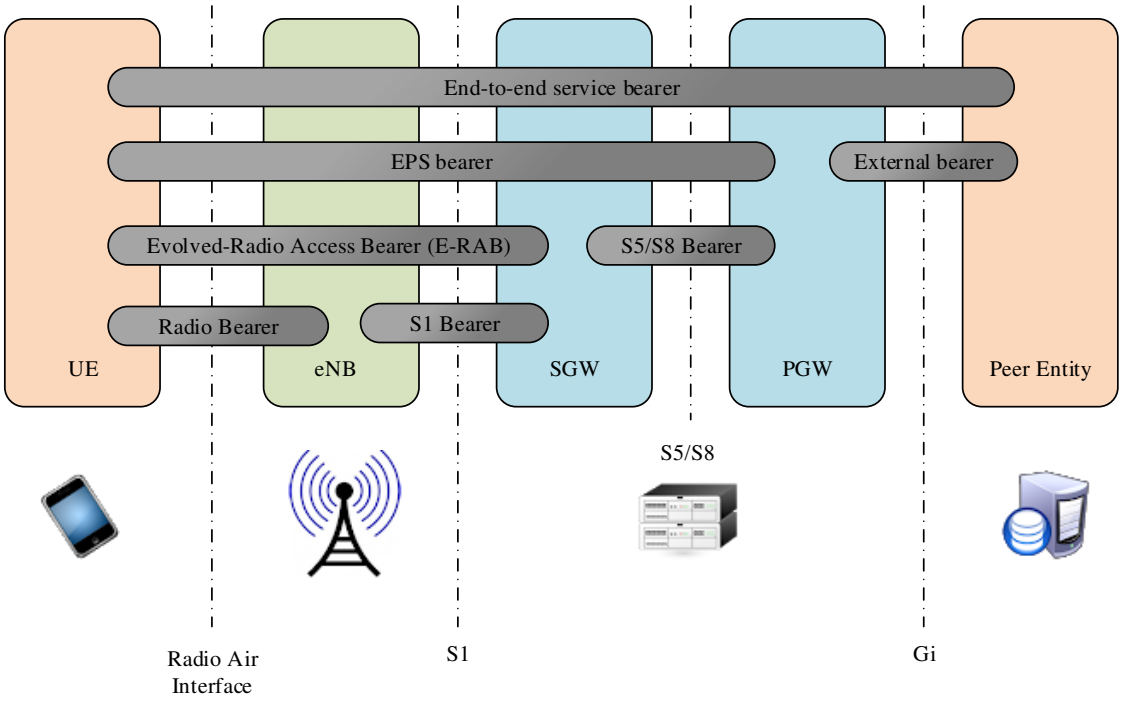



Figure 6.5: LTE bearer architecture

In order to ensure end-to-end QoS, each bearer in LTE is assigned a set of standardised QoS attributes which are then used by the LTE protocol layers between the UE and eNB to manage the traffic. The EPS bearer QoS depends on the data bearer class whereby default data bearers are assigned non-Guaranteed Bit Rate (non-GBR) QoS attributes whilst dedicated data bearers can be assigned either non-GBR or GBR QoS attributes [135]. The three types of QoS attributes which are provided to the eNB by the Mobility Management Entity (MME) within the EPC to enforce the QoS on the air interface include:

*QoS Class identifier (QCI)*: The QCI is a key attribute as it has a direct impact on the packet scheduling, queue management and link layer protocol configuration. A QCI refers to an index which identifies a set of established QoS parameter values for the

priority, packet delay and packet loss rate. As of LTE release 14, 3GPP has standardised 15 classes, shown in Table 6.2, nevertheless QCI 1-9 are widely adopted as these have been present as of LTE release 8. Although the values shown in Table 6.2 represent the 3GPP guideline for the above mentioned parameters, they also ensure a homogenous QoS level for traffic mapped to a given QCI in multi-vendor network deployments and in case of roaming subscribers.

| QCI | Resource Type | Priority | Packet Delay Budget | Packet Error Loss Rate | Example Services |
|---|---|---|---|---|---|
| 1 | GBR | 2 | 100ms | $10^{-2}$ | Conversational Voice |
| 2 | GBR | 4 | 150ms | $10^{-3}$ | Conversational Video (Live Streaming) |
| 3 | GBR | 3 | 50ms | $10^{-3}$ | Real Time Gaming, V2X messages |
| 4 | GBR | 5 | 300ms | $10^{-6}$ | Non-Conversational Video (Buffered Streaming) |
| 65 | GBR | 0.7 | 75ms | $10^{-2}$ | Mission Critical user plane Push to Talk voice |
| 66 | GBR | 2 | 100ms | $10^{-2}$ | Non-Mission-Critical user plane Push to Talk voice |
| 75 | GBR | 2.5 | 50ms | $10^{-2}$ | V2X messages |
| 5 | non-GBR | 1 | 100ms | $10^{-6}$ | IMS Signalling |
| 6 | non-GBR | 6 | 300ms | $10^{-6}$ | Video (Buffered Streaming) TCP-Based (for example, www, email, chat) |
| 7 | non-GBR | 7 | 100ms | $10^{-3}$ | Voice, Video (Live Streaming), Interactive Gaming |
| 8 | non-GBR | 8 | 300ms | $10^{-6}$ | Video (Buffered Streaming) TCP-Based (for example, www, email, chat) |
| 9 | non-GBR | 9 | 300ms | $10^{-6}$ | Video (Buffered Streaming) TCP-Based (for example, www, email, chat). Typically used as default bearer |
| 69 | non-GBR | 0.5 | 60ms | $10^{-6}$ | Mission Critical delay sensitive signalling |
| 70 | non-GBR | 5.5 | 200ms | $10^{-6}$ | Mission Critical Data |
| 79 | non-GBR | 6.5 | 50ms | $10^{-2}$ | V2X messages |

Table 6.2: Standardised QCI characteristics [135]

*Allocation and Retention Priority (ARP)*: In the case of congestion, this parameter may be used to select which bearers are allowed in the network and which need to be modified or dropped.

*Maximum Bit Rate (MBR) / Guaranteed Bit Rate (GBR)*: For GBR type bearers, the GBR indicates the bit rate that can be expected to be provided by the bearer whilst the MBR limits the bit rate provided by the bearer [17]. In the case of non-GBR type bearers an aggregated maximum bit rate (AMBR) parameter is used to control the bit

rate across all non-GBR bearers of a subscriber or a particular access point name (APN) [17]. Any excess traffic is discarded through the use of a rate-shaping function. As already highlighted in the previous section, the DSCP parameter in the IP header plays an important role in delivering an end-to-end QoS for multimedia applications. With reference to Figure 6.6, different traffic types are tagged with differing DSCP parameters by either the transmitting entity itself or a traffic classifier node. This allows the S-GW/P-GW to map traffic to the bearer associated with the traffic type QCI using a DSCP-to-QCI mapping table. Meanwhile, in the eNB, traffic from the different bearer types is placed into respective queues prior to scheduling on the radio interface based on their QoS service requirements.

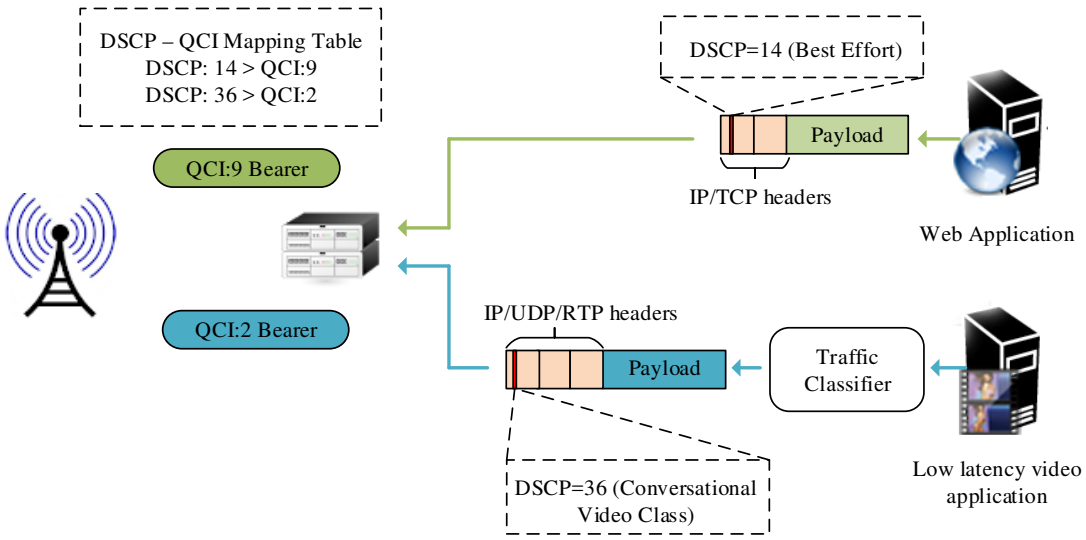



Figure 6.6: End-to-End QoS – DSCP to QCI mapping

## 6.2 Packet Scheduling

The packet scheduler residing in the MAC layer of the LTE protocol stack plays a pivotal role in the radio resource management and efficient delivery of video traffic as it needs to distribute fairly the available resources among the active UEs whilst adhering to their QoS requirements [136]. With reference to the generic model of the downlink packet scheduling function, shown in Figure 6.7, and considering one traffic flow per UE, the packet scheduling process at every TTI is composed of five main steps namely:

1) Determine the traffic flows which can be scheduled in the current TTI. In order for a flow to be selected as a candidate for scheduling it must have buffered data in the MAC layer and UE is in active state;
2) Collect channel state information from the UEs in the form of CQI reports;
3) For each flow, $j$, and resource block, $k$, compute a scheduling metric, $m_{j,k}$, indicating the transmission priority of a specific flow on a given resource block. Once the scheduling metric for all possible combinations are computed, the $k^{th}$ RB is allocated to the $j^{th}$ flow such that:

$$m_{j,k} = \max_{i \in j} \{m_{i,k}\} \tag{6.1}$$

   The scheduling metric, $m_{j,k,}$ depends on the scheduling strategy and performance requirement. However, it typically involves the CQI information reported by the UEs together with other flow specific indicators such as queue length, head of queue packet delay, service priority and past achieved throughput;
4) For each scheduled flow, compute the transport block size and the MCS to be used for the current TTI and UE channel conditions based on 3GPP defined tables [20];
5) Inform the scheduled UEs of their assigned resource blocks and selected MCS by sending downlink control information on the PDCCH.

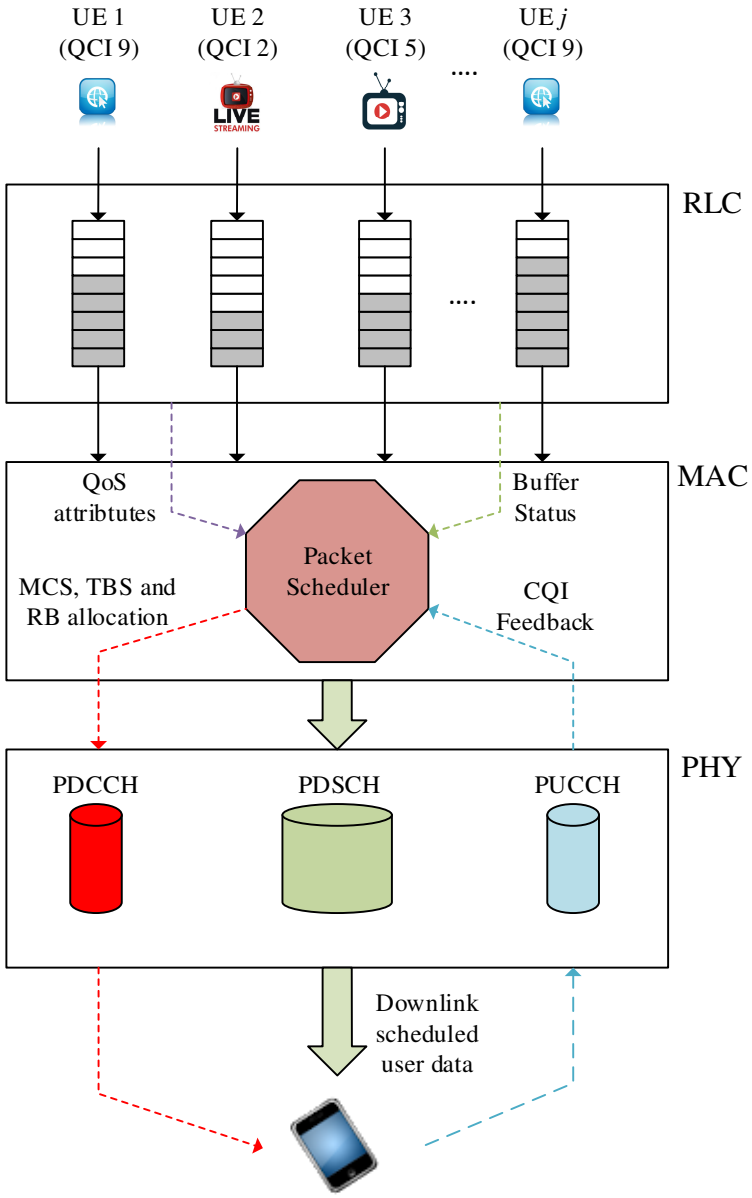



Figure 6.7: Generic packet scheduler model

Several packet scheduling algorithms have been proposed in literature achieving varying levels of performance, in terms of QoS adherence, spectral efficiency, resource assignment fairness, and computational complexity. Nevertheless, the large majority of the scheduling algorithms can be classified into three main types depending on their channel and QoS awareness.

### 6.2.1 Channel un-aware packet scheduling

Channel un-aware schedulers have been designed for wired networks and are based on the assumption that the transmission channel is time-invariant and error free [136], thereby intrinsically exhibit poor performance in wireless networks. One of the

simplest algorithms in this category is the round robin (RR) scheduling algorithm which cyclically schedules users in order to allocate the available radio resources fairly among all users, irrespective of their channel conditions. This leads to instances whereby users with poor channel conditions are still allocated resources. Thus, even though this algorithm ranks high in terms of resource assignment fairness, the system throughput is low as it cannot exploit multi-user (MU) diversity.

In [137], a blind equal throughput (BET) scheduling algorithm was proposed with the aim of achieving throughput fairness among the users. In this case, the scheduling metric for the $j^{th}$ flow is given by:

$$\begin{aligned} m_{j,k} &= \frac{1}{\bar{R}^j(t-1)} \\ \bar{R}^j(t) &= \beta \bar{R}^j(t-1) + (1-\beta) r^j(t) \end{aligned} \tag{6.2}$$

where $\bar{R}^j(t)$ is the past average throughput achieved by the $j^{th}$ flow until time $t$, $r^j(t)$ is the throughput achieved by $j^{th}$ flow at time $t$ and $\beta$ is a weighting factor between 0 and 1.

The use of the past average throughput measurement, $\bar{R}^j(t)$, allows the eNB to schedule users which experience low throughput caused by bad channel conditions thereby improving throughput fairness.

### 6.2.2 Channel aware - QoS un-aware packet scheduling

Channel aware scheduling schemes use CQI feedback reports sent by the UEs in order to exploit multi-user diversity and improve system performance. In [137], a maximum throughput (MT) algorithm was proposed whereby the eNB assigns RBs to users which experience the best channel conditions, thereby maximising the system throughput. Thus, in a MT scheduler, the scheduling metric for the $j^{th}$ flow is given by:

$$m_{j,k} = \arg\max_{i \in j} \{d_{i,k}(t)\} \tag{6.3}$$

where $d_{i,k}$ is the maximum instantaneous achievable data rate of the $i^{th}$ flow on the $k^{th}$ RB at time $t$.

Evidently, although the system throughput is maximised, the resource fairness ranks very low as users in cell-edge conditions will seldom be scheduled. In order to strike a balance between high spectral efficiency and throughput fairness, the scheduling metrics of the MT and BET scheduling schemes are merged resulting in the proportional fair (PF) scheduling metric given by:

$$m_{j,k} = \arg\max_{i \in j} \left\{ \frac{d_{i,k}(t)}{\overline{R^i(t-1)}} \right\}$$
$$\overline{R^i(t)} = \left(1 - \frac{1}{T}\right) \overline{R^i(t-1)} + \frac{1}{T} r^i(t) \tag{6.4}$$

where $\overline{R}^i(t)$ is the past average throughput achieved by the $i^{th}$ flow until time $t$, $r^i(t)$ is the throughput achieved by $i^{th}$ flow at time $t$ and $T$ defines the time window length to calculate the average throughput.

The PF scheduling scheme does not take into account any QoS requirements and thus it is specifically used for the non-real time (NRT) traffic class. Nevertheless, this scheme has attracted significant research attention and several enhancements have been proposed [138] with varying degrees of performance improvement at the expense of additional computational complexity. Meanwhile, a generalised proportional fair scheduler scheme is proposed in [139] whereby weighting factors $\alpha$ and $\beta$, controlling the trade-off between spectral efficiency and fairness, are introduced resulting in the scheduling metric given by:

$$m_{j,k} = \arg\max_{i \in j} \left\{ \frac{\left[d_{i,k}(t)\right]^{\alpha}}{\left[\overline{R^i(t-1)}\right]^{\beta}} \right\} \tag{6.5}$$

### 6.2.3 Channel aware - QoS aware packet scheduling

The classic packet scheduling algorithms discussed in the previous sections focus on the fairness, system spectral efficiency or the balance between these two factors. Nevertheless, the real-time (RT) flow QoS attributes, such as the stringent delay requirement in multi-view video transmission over LTE networks, is not addressed by these solutions. For this reason, several QoS aware packet scheduling schemes were

proposed in literature, with the most representative focusing on guaranteeing a bounded delay, this being one of the main requirements for the RT traffic class.

In [140], the authors merge the characteristics of the PF scheme with an exponential function of the end-to-end delay resulting in the EXP/PF scheme. This scheme is designed to handle both RT and NRT flows whereby NRT flows are handled using PF scheduling metrics whilst the RT flows scheduling metric is given by:

$$m_{j,k} = \arg\max_{i \in j} \left\{ \exp\left( \frac{\eta_i D_i - \overline{D_i}}{1+\sqrt{\overline{D_i}}} \right) \frac{d_{i,k}(t)}{\overline{R^i(t-1)}} \right\}$$
$$\overline{D_i} = \frac{1}{N_{RT}} \sum_{i=1}^{N_{RT}} \eta_i D_i \tag{6.6}$$

where $N_{RT}$ is the number of real-time flows in the radio sector, $\eta_i$ is a weighting parameter and $D_i$ is the head of line (HOL) delay.

The concept of shaping the scheduling metric with a function based on the HOL packet delay is studied further in [141]. Here the authors present a logarithmic based scheme (LOG-rule) and an improved version of the EXP/PF scheme (EXP-rule) whose scheduling metrics are given in [136]. As discussed in [142], the EXP-rule scheme is more robust as the scheduling metric of the EXP-rule increases at a faster rate as the HOL delay increases and it takes into account the delay of all the users in the radio sector.

In [143], a modified largest weighted delay first (M-LWDF) scheduling scheme considering the channel conditions and the HOL packet delay is presented. In this scheme, NRT traffic flows are handled using a PF scheduling scheme whilst the scheduling metric used for RT traffic flows is given by:

$$m_{j,k} = \arg\max_{i \in j} \left\{ \lambda_i D_i \frac{d_{i,k}(t)}{\overline{R^i(t-1)}} \right\} \tag{6.7}$$

where $\lambda_i$ is a weighting parameter.

This scheduling scheme has been widely used in various studies related to the transport of RT traffic, such as video, in view of its simplicity and good performance.

Moreover, a number of modifications have been proposed. In [144], a virtual token resource allocation algorithm which guarantees a minimum throughput for NRT traffic flows in the presence of RT traffic flows is presented whilst in [145] the M-LWDF is adapted to include the queue size. In the latter, although simulation results indicate an improvement in performance, the study considered a simple scenario composed of a single LTE cell whilst no information on the type of channel state feedback used is provided. In [146], the M-LWDF is enhanced such that the scheduling metric includes a distance factor in conjunction with the channel conditions reported by the UEs thereby allowing the eNB to provide a better service to cell-edge users. Although effective, the study does not clearly indicate how the distance between the UE and eNB is calculated and how the threshold to segregate UE between near-cell and cell-edge is determined.

A series of works focus on the use of multi-level QoS-aware scheduling schemes. A two level scheduling frame level scheduler (FLS) scheme based on control theory is presented in [142]. At the top level, a discrete time linear control model is used to estimate the amount of data that should be transmitted by each RT flow within a single radio frame to satisfy its delay requirement. This is then followed by a standard PF scheme to assign the RBs to the users thereby ensuring fairness among the RT flows. The accurate control of the delay provided by this algorithm outperforms several other schemes at the expense of an increased computational effort introduced by the control theory based estimation model at the top level.

In [147], the authors propose a time domain (TD) and frequency domain (FD) scheduler working in tandem. The TD scheduler classifies flows into GBR and non-GBR flows and selects a subset based on their QoS requirements and channel status. In the next level, the FD scheduler uses a PF and M-LWDF scheduler to assign resources to the NRT and RT traffic flows respectively. Meanwhile, an advanced frame level scheduling scheme is proposed by Skondras *et al* in [148]. In this work, the authors extend the work presented in [142] by allocating radio resources using a three level scheduling algorithm whereby the data quota estimation method used in [142] is employed in the first level whilst the second and third levels are based on the M-LWDF scheduling scheme.

## 6.3 Cross-layer design

The strict layered approach adopted by the mainstream mobile communication standards, such as 3GPP LTE, poses a limit on the efficient use of radio resources and effective delivery of video traffic. For this reason, a cross-layer design (CLD) network architecture, whereby the application (APP), MAC and PHY layer interact to optimise layer-specific parameters, has gained popularity in recent years. Several studies can be found in literature however the cross-layer adaptations typically employed in the downlink transmission of video traffic can be classified into three main types, namely [149]:

- Adaptation of the MAC/PHY layer parameters based on the video source information at the APP layer;
- APP layer bit rate adaptation based on the MAC/PHY layer information;
- Joint adaptation of the APP/MAC/PHY parameters.

With reference to the generic cross-layer video transmission system shown in Figure 6.8, the controller entity is responsible for sharing and adapting layer specific parameters using the knowledge of the radio channel conditions, in the form of CQI feedback reports, and source content information. A typical approach adopted in cross-layer design involves the formulation of a multi-dimensional optimisation problem with the objective of maximising the received video quality under a given set of constraints [149]. This is done by adapting:

- Source coding parameters in the APP layer including the quantisation parameter (QP) and coding mode, $S_p$;
- Resource allocation and scheduling parameters in the MAC layer, $RA_p$;
- Link adaptation parameters including the modulation and coding scheme (MCS) to be used in the PHY layer, $AMC_p$.

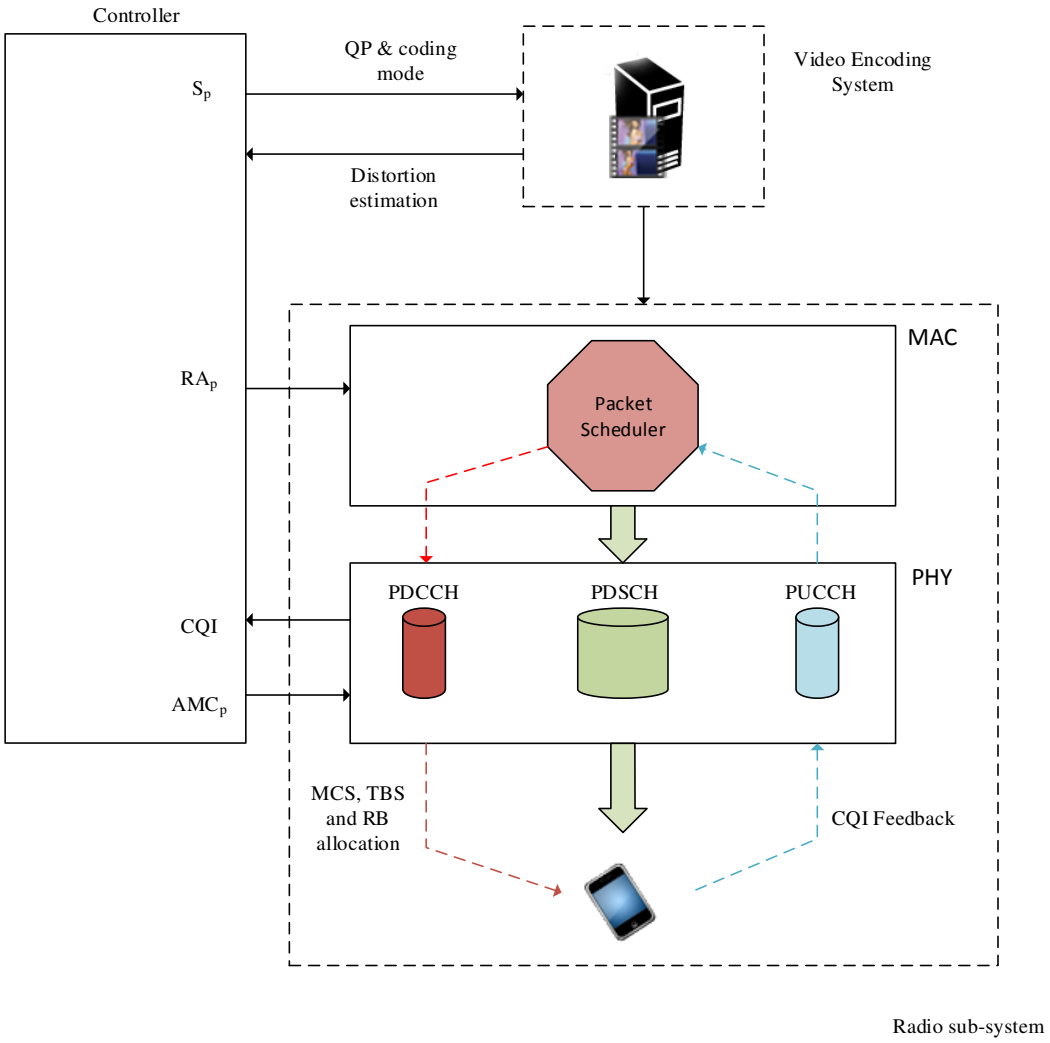



Figure 6.8: General cross-layer adaptation

Although such a multi-dimensional approach is very effective in improving radio resource management and the video quality experienced by users in challenging radio conditions, the computational effort required to solve such optimisation problems can be very high. Moreover, such cross-layer techniques typically employ recursive optimal per-pixel estimation (ROPE) algorithm [150] to accurately estimate the packet loss induced distortion, thereby further increasing the computational complexity. In order to manage the computational complexity of such cross-layer schemes, several works have been proposed employing various techniques ranging from the application of heuristics [151] to directed graphs [152].

Meanwhile, lean cross-layer design schemes characterised by simple and low complexity techniques have been proposed in literature. Such schemes depart from the

multi-objective optimisation framework and end-to-end distortion estimation process, seeking optimal parameter settings across multiple layers, and instead focus on prioritising packets based on the packet payload type. In [153], a content aware adaptive HARQ retransmission scheme for 3GPP LTE networks is proposed whereby the packet induced distortion, and hence its priority, is inferred based on the video frame type carried in the payload. In this work, the authors exploit the fact that not all frames in a GOP carry the same level of importance as losing an I frame causes a significantly different level of distortion when compared to losing a P frame. Similarly, losing a P frame in the anterior section of the GOP is more detrimental when compared to losing a P frame in the posterior section of the GOP. Thus, the media server hosting the video encoder exploits the various frame dependencies and relative frame position in the GOP to tag packets, by setting the differentiated service code point (DSCP) field in the IP header, based on their relative priority. This allows the RLC and MAC layers in the eNB to be aware of each packet priority and adapts the number of HARQ re-transmissions accordingly. Although the simulation results indicate an improvement in the video quality experienced by the users, the authors considered only two static users whilst no information is provided on the type of scheduler used.

A similar packet prioritisation technique is adopted in [154] for the transmission of 3D-HEVC content in wireless networks. In this work, the authors exploit the fact that CQI reports sent by the UEs are merely a recommendation to the eNB indicating the MCS and TBS to be used and hence the MAC layer can adapt these parameters based on the packet payload type. Thus, high priority packets, such as I frames and highly referenced frames, are typically assigned a robust MCS in order to improve protection and reduce the retransmission probability. Meanwhile, the lower priority packets are assigned an MCS based on the CQI feedback report sent by the UE. Although simulation results indicate an improvement in the high priority packet loss rate, the simulations have been carried out using a PF scheduling algorithm with no maximum end-to-end packet delay constraint.

Techniques based on PHY layer parameter adaptation such as those presented in [153-154], decrease the spectral efficiency of the radio network through either higher L1 retransmission rates or the use of more robust MCS. This might lead to a further

deterioration in system performance in an otherwise already congested network. Thus, in [155], the authors propose a cross-layer adaptive 3D-HEVC video transmission for buffered streaming services combining content awareness with packet scheduling and an active packet dropping technique. In the proposed scheme, the packets are first prioritised based on their dependencies whilst the eNB schedules high priority packets in lieu of lower priority packets being dropped. In order to mitigate the impact of the active dropping technique adopted by the proposed scheme, a sophisticated error concealment technique is used to provide graceful degradation in the presence of packet loss.

# Chapter 7 Cross-Layer CQI Feedback Adaptation for MVD Video Transmission in a Crowd Event Scenario

This chapter builds on the CQI feedback reduction techniques proposed in chapter 3 and describes a cross-layer CQI feedback adaptation scheme for the transmission of MVD video in a crowd event scenario. It starts with a performance evaluation of several QoS-aware packet scheduling algorithms in this scenario followed by the enhancement and testing of the M-LWDF algorithm to include MVD video content awareness. The third section focuses on the impact of the CQI feedback reduction techniques, presented in chapter 3, on the transmission of MVD video in a crowd event scenario. Finally, in order to alleviate the impact of the feedback reduction schemes, a cross layer CQI feedback adaptation scheme is proposed and tested using both objective and subjective testing methodologies.

## 7.1 Performance of MVD video transmission over LTE networks in a crowd event scenario

Although several studies have been conducted on the transmission of MVD video over LTE networks, these do not cater for a crowd event scenario [156-158]. Moreover, such studies generally assume perfect channel state information (CSI), involve a limited number of users or use legacy video compression schemes for the generation of video traffic. Thus, prior to the development of the proposed work, a comparative performance analysis of three packet scheduling algorithms, namely M-LWDF, EXP/PF and FLS, was carried out. This selection is based on the fact that both M-LWDF and EXP/PF packet scheduling algorithms have been widely adopted in various multimedia video transmission studies due to their efficacy and low complexity nature. Meanwhile, even though the FLS algorithm exhibits a higher computational complexity, studies have noted a marginal performance improvement when compared to the previously mentioned schemes.

### 7.1.1 Simulation environment, parameters and results

In order to establish the performance of the selected packet scheduling algorithms in a crowd event scenario, the simulation environment, described in section 3.2.2, together

with the selected packet scheduling schemes, were implemented and tested in a system level LTE simulator [55] over 50 random seeds using the simulation parameters shown in Table 7.1. With reference to the parameters shown in Table 7.1, UEs are stationary and each receives an MV-HEVC real-time encoded MVD stream composed of two texture and depth map views. Meanwhile, unlike other studies found in literature, a periodic wideband CQI feedback scheme, widely used in commercial networks particularly in crowd event scenarios, is adopted.

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | M-LWDF, EXP/PF, FLS |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| Traffic Model | 1 x Trace based MVD stream |
| Maximum end-to-end delay for video traffic, $\zeta$ | 100ms |
| Protocol Overhead [55] | RTP/UDP/IP with RoHC: 3 bytes<br>PDCP: 2 bytes<br>MAC and RLC: 5 bytes<br>CRC: 3 bytes |
| RLC Mode | UM / L1 ReTx: 3 |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz[55] |
| Fast Fading Realization | Jakes Model[55] |
| Number of UE's, $T_u$ | {25, 50, 70, 90, 110} |

Table 7.1: Simulation parameters for the comparison of packet scheduling algorithms

With reference to the system model shown in Figure 7.1, the MVD video traffic is generated by encoding the texture and depth map views of the *balloons* test sequence using the parameters shown in Table 7.2. The resulting texture and depth-map view bit streams are then packetized at the application layer and encapsulated within RTP, UDP and IP protocols prior to being sent to the eNB. Packets received by the eNB are then placed into the respective UE first-in-first-out (FIFO) queue at the RLC layer whilst waiting for scheduling by the MAC layer. Moreover, packets which remain enqueued at the eNB for a time period greater than the maximum allowed transmission delay are discarded.

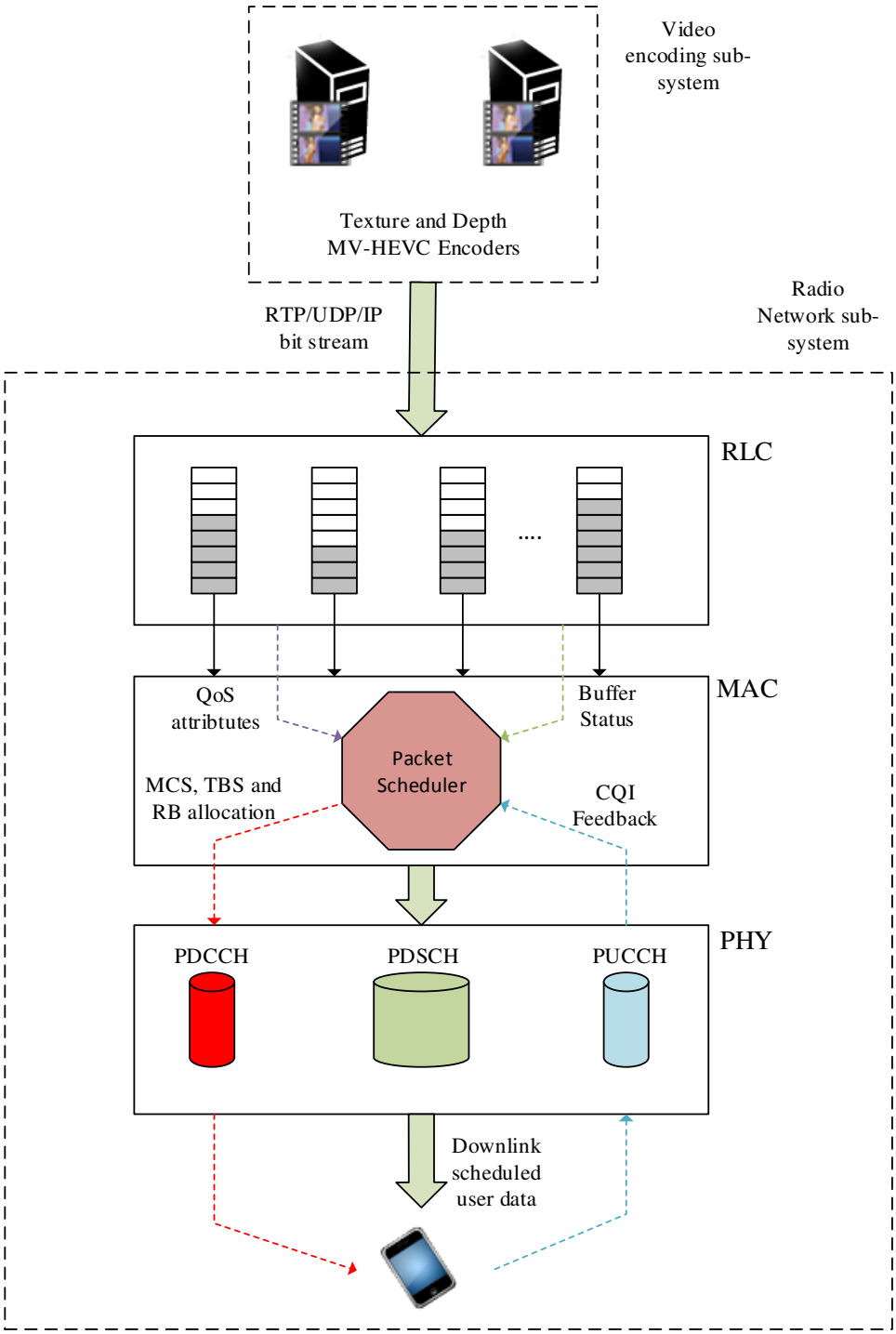
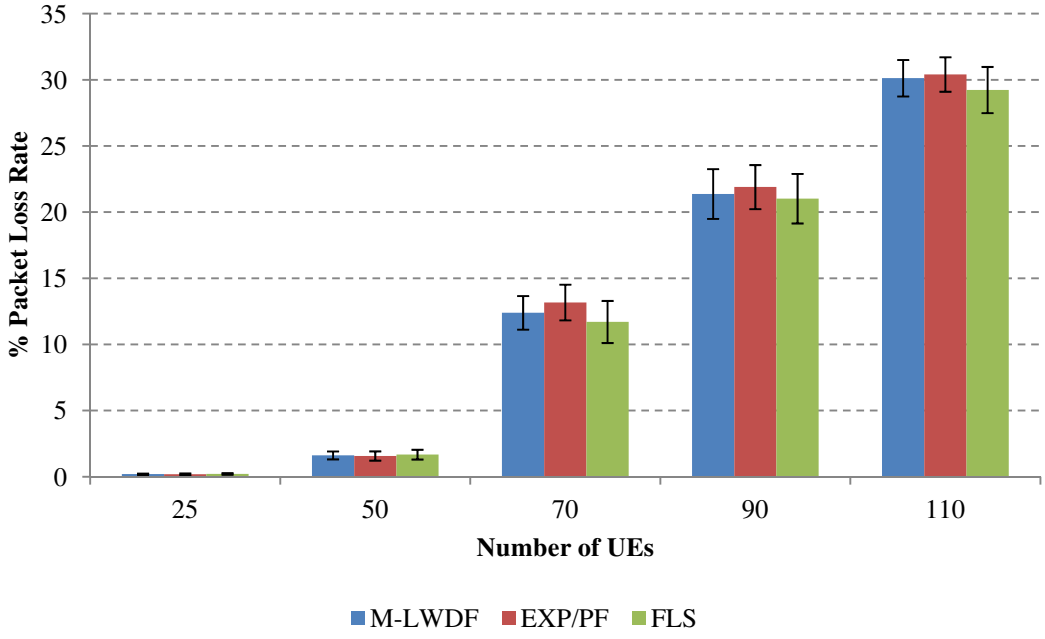



Figure 7.1: MVD Transmission over LTE system model

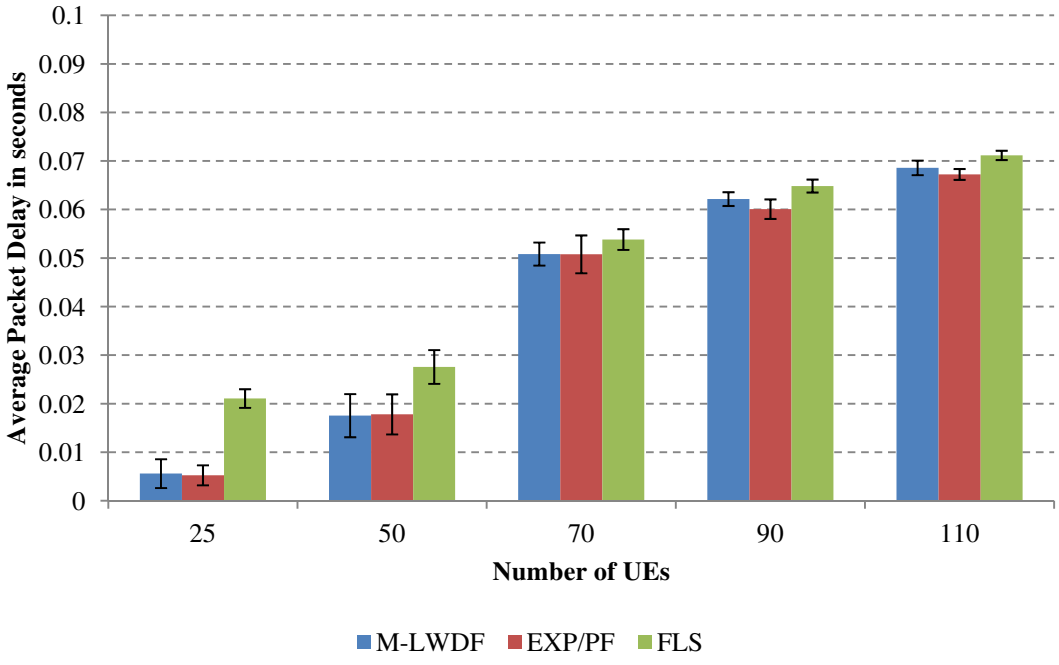
| Parameter | Value |
|---|---|
| GOP Size | 4 |
| Intra Period | 12 |
| Prediction Structure | Low latency (see Figure 5.3) |
| QP | Texture: 40 / Depth: 45 |
| Total bit rate | 675kbps |

Table 7.2: MVD encoding parameters

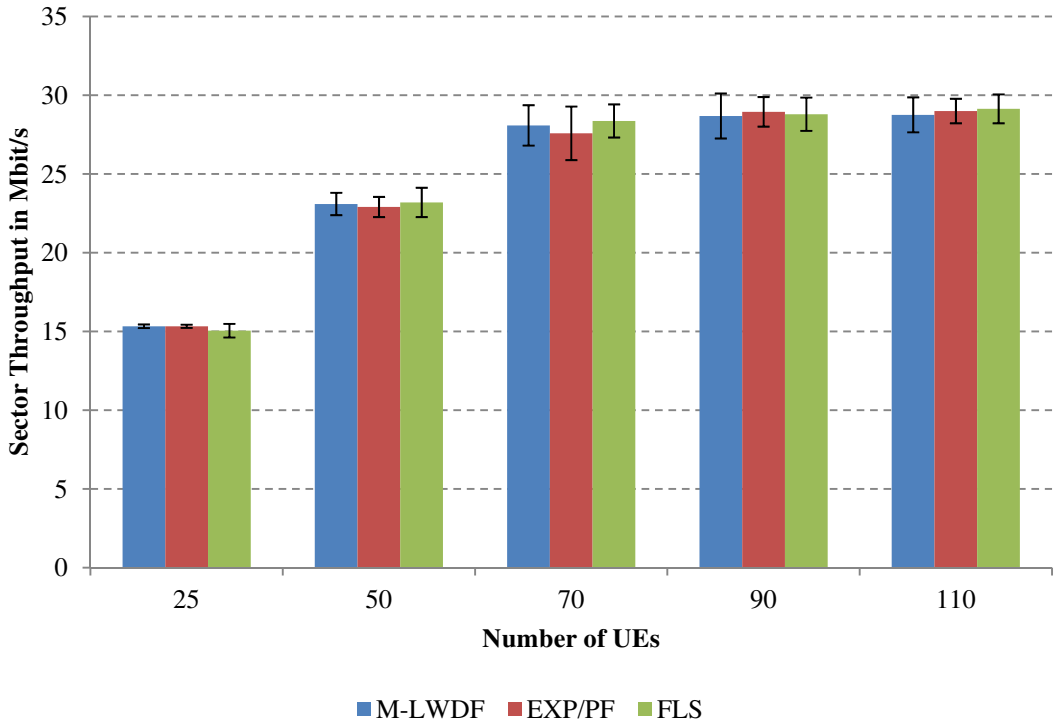
The performance of the selected packet scheduling algorithms is assessed along three main performance metrics namely, the packet loss rate (PLR), average packet delay and sector throughput. With reference to Figure 7.2(a), the PLR increases rapidly as the number of users in the radio sector exceeds 50. Moreover, it is evident that the three scheduling schemes are in general comparable, in terms of PLR, with the FLS scheme exhibiting a marginally better performance in high load conditions.




(a)




(b)




(c)

Figure 7.2: Performance metrics for the tested scheduling algorithms (a) % Packet Loss Rate, (b) Average packet delay, (c) Sector Throughput

It should be noted that in crowd event scenarios, the majority of the packet losses are due to excessive packet delays caused by bloated queues in the eNB leading to a high proportion of packet discards. As shown in Figure 7.2(b), the average delay exhibited by the evaluated packet scheduling schemes increases with traffic load, however the FLS scheme tends to have a marginally higher packet delay due to the proportional fair algorithm implemented at the lower layer of this scheme. Meanwhile, as expected, the sector throughput, shown in Figure 7.2(c), increases with traffic load until it tapers towards an asymptotic value.

Considering the above discussion on the relative performance of the selected packet scheduling schemes and the relatively higher computational complexity of the FLS scheme, the M-LWDF scheduling algorithm is selected as the baseline scheduler for the rest of the work. Moreover, based on the test conditions and MVD encoding parameters, shown in Table 7.1 and 7.2 respectively, the PLR increases rapidly beyond 50 users in the radio sector and thus, unless otherwise stated, this has been selected as the operating point for the rest of the work.

## 7.2 Enhanced content-aware M-LWDF packet scheduling algorithm

As shown in the previous section, congestion causes high delay build-up leading to significant packet loss thereby impacting the quality of the received MVD stream. However, as discussed in the previous chapter, not all packets have the same level of importance and thus it is beneficial to prioritise the scheduling of packets according to their relative importance with the aim of reducing the impact on the received video quality. A number of techniques have been discussed in chapter 6 nevertheless in this work the M-LWDF packet scheduling algorithm is enhanced through the use of RLC queue based weighted factors thereby allowing content awareness in the evaluation of the scheduling metric.

### 7.2.1 Proposed content-aware packet scheduling algorithm

With reference to Figure 7.3, the content-aware packet scheduling scheme is split into two phases; namely a time domain and a frequency domain scheduling phase. The time-domain scheduling phase is responsible for selecting and prioritising the flows (queues) to be scheduled in the current TTI, $n$, and is based on the calculation of a priority metric per flow $j$, denoted by $TD\_metric(n)_j$, given by (7.1). This is based on three main factors which are strongly correlated with the flow performance and are given by (7.2)-(7.4). Flows with a priority metric greater than $T_{TD\_Th}$ are considered as suffering from queue stalling and have a higher incidence of packet discards. For this reason, these flows are tagged as high priority flows in order to be given preference during PRB assignment in the frequency domain scheduling phase.

$$TD\_metric(n)_j = \mu w\_HOL(n)_j + (1-\mu)\left[\sigma \overline{w\_Q(n)_j} + (1-\sigma)\overline{PLR\_HPP(n)_j}\right] \quad (7.1)$$

where $\mu$ and $\sigma$ are tuning parameters controlling the relative importance of the three factors discussed and empirically set to 0.75 and 0.25 respectively, $w\_HOL(n)_j$ is the current weighted head of line (HOL) packet delay for queue $j$ determined through (7.2), whilst $\overline{w\_Q(n)_j}$ and $\overline{PLR\_HPP(n)_j}$ are the linearly smoothed weighted queue

size found using (7.3) and high priority packet (HPP) PLR for queue *j* calculated using (7.4) respectively.

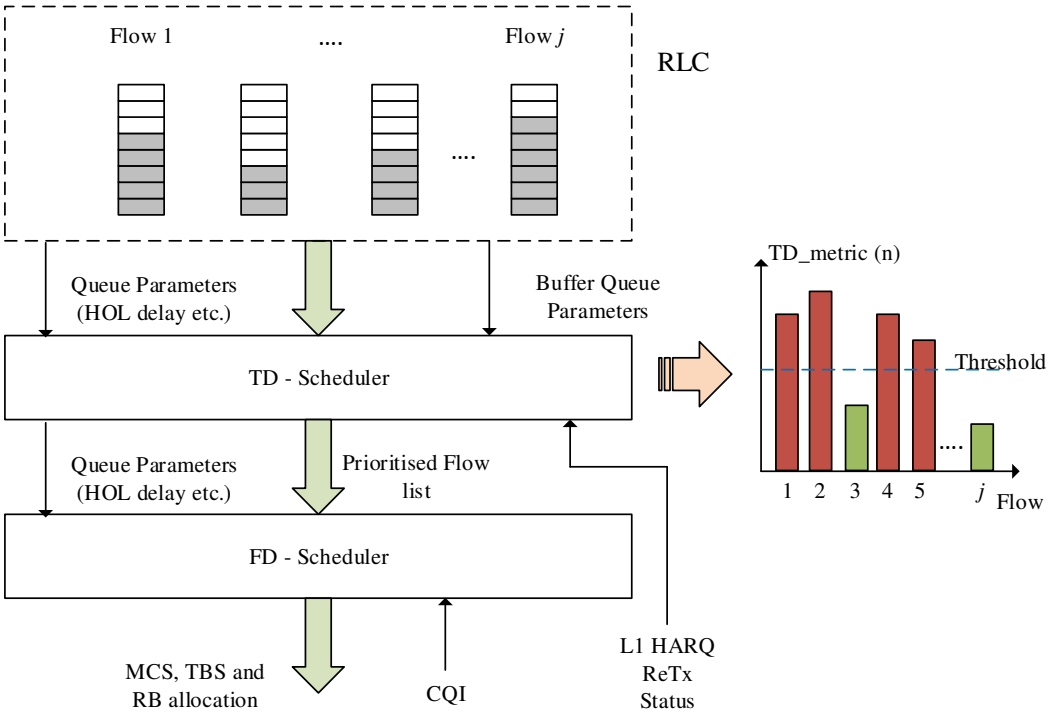



Figure 7.3: TD-FD Content-aware M-LWDF packet scheduling scheme

$$w\_HOL(n)_j = \frac{\left(1+\omega_{j,1}\right) HOL_j}{\zeta} \tag{7.2}$$

where $HOL_j$ is the head of line packet delay of flow *j*, $\zeta$ is the maximum end-to-end transmission delay for video packets set to 0.1 seconds whilst $\omega_{j,1}$ is a priority level weighting based on the content of the video packet for the head of line packet in flow *j*. The latter together with the definition of high priority packets will be discussed in further detail in section 7.2.2.

$$\begin{aligned} \overline{w\_Q(n)_j} &= 0.5 w\_Q(n)_j + 0.5 w\_Q(n-1)_j \\ w\_Q(n)_j &= \frac{\left(1+N_{Q\_hpp}(n)/N_Q(n)\right) Q(n)_i}{\psi} \end{aligned} \tag{7.3}$$

where $Q(n)_j$ is the $j^{th}$ queue size in bytes at the current TTI, $N_{Q\_hpp,j}(n)$ and $N_{Q,j}(n)$ are the number of HPP and total number of packets in queue *j* at the current TTI respectively whilst $\psi$ a normalising factor set to 20,000 bytes.

$$\overline{PLR\_HPP(n)_j} = 0.5PLR\_HPP(n)_j + 0.5PLR\_HPP(n-1)_j$$

$$PLR\_HPP(n)_j = \frac{\left(\frac{N_{tx\_hpp,j}(n) - N_{rx\_hpp,j}(n)}{N_{tx\_hpp,j}(n)}\right)}{\kappa} \tag{7.4}$$

where $N_{tx\_hpp,j}$ and $N_{rx\_hpp,j}$ are the number of successfully transmitted and received HPP calculated using layer 1 (L1) HARQ re-transmission status whilst $\kappa$ is a normalising factor set to 0.03.

The frequency domain (FD) scheduling phase is responsible for the assignment of PRBs to the users in the radio sector. The FD scheduling starts with the high priority flows, identified during the TD scheduling phase, and assigns a scheduling metric $w_{j,i}$, computed using (7.5), to the $j^{th}$ flow, for the $i^{th}$ PRB. With reference to (7.5), the scheduling weight is based on the M-LWDF scheme and uses the CQI reports sent by the UE to calculate the flow instantaneous throughput on $i^{th}$ PRB, $r_{j,i}(n)$, relative to the current achieved average throughput, $\overline{R_j}(n)$. In order to introduce content awareness at the PRB assignment stage, the head of line packet delay, $HOL_j(n)$, is modulated with a weighting factor based on both the priority level weighted head of line packet delay and the cumulative priority level weighted delay of subsequent packets. The latter is important as it allows the scheduler to consider not only the head of line packet but also packets which will be imminently scheduled.

$$w_{j,i}(n) = -\frac{\log \delta_j}{\zeta}\left[\left(1+\omega_{j,1}\right)HOL_j(n)\left(1+\sum_{p=2}^{5}\omega_{j,p}D_{j,p}(n)\right)\right]\frac{r_{j,i}(n)}{\overline{R_j}(n)} \tag{7.5}$$

$$\overline{R_j}(n) = 0.8\overline{R_j}(n-1) + 0.2\overline{R_j}(n) \tag{7.6}$$

where $\delta_j$ is the drop probability commonly set to 0.005, $\zeta$ is the maximum end-to-end transmission delay for video packets set to 0.1 seconds, $HOL_j$ is the head of line packet delay of flow $j$, $\omega_{j,p}$ is the priority level weighting of packet $p$ in the $j^{th}$ flow, and $D_{j,p}(n)$ is the delay of packet $p$ in the $j^{th}$ flow.

Once the scheduling metric is computed for all high priority traffic flows, across all the PRBs, the assignment process is started whereby each PRB is assigned to the flow having the highest scheduling metric. This process is repeated until no PRBs are left.

In case all high priority flows are scheduled and there are still available PRBs to be assigned, the process is restarted for the remaining flows.

### 7.2.2 MVD video packet prioritisation

In order to enable content awareness and allow the scheduler in the eNB to prioritise between the MVD video packets, the encoder needs to classify the MVD packet stream based on the induced PSNR degradation in case of packet loss. Several works have been proposed to estimate the packet loss induced distortion ranging from the application of recursive optimal per-pixel estimation (ROPE) to detailed analytical distortion models able to accurately estimate the distortion in the synthesised view [159]. Nevertheless, these methods either entail high computational effort or require pre-estimation of model parameters. In this work, a simple packet prioritisation scheme is adopted whereby the packet priority is generally based on the view type (texture or depth), layer (base or dependent view), frame type (key frame or other) and the position of the frame relative to the key frame. For the rest of the work, a tile-based coding configuration with three columns and three rows, as shown in Figure 7.4, is adopted in order to have a finer priority control for the depth map views. Moreover, considering a maximum slice size of 1500 bytes, texture tiles are generally composed of multiple slices (packets) whilst depth map tiles are composed of a single slice (packet).

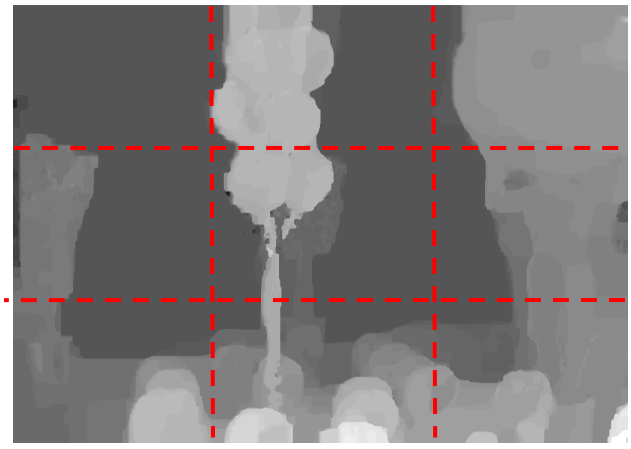
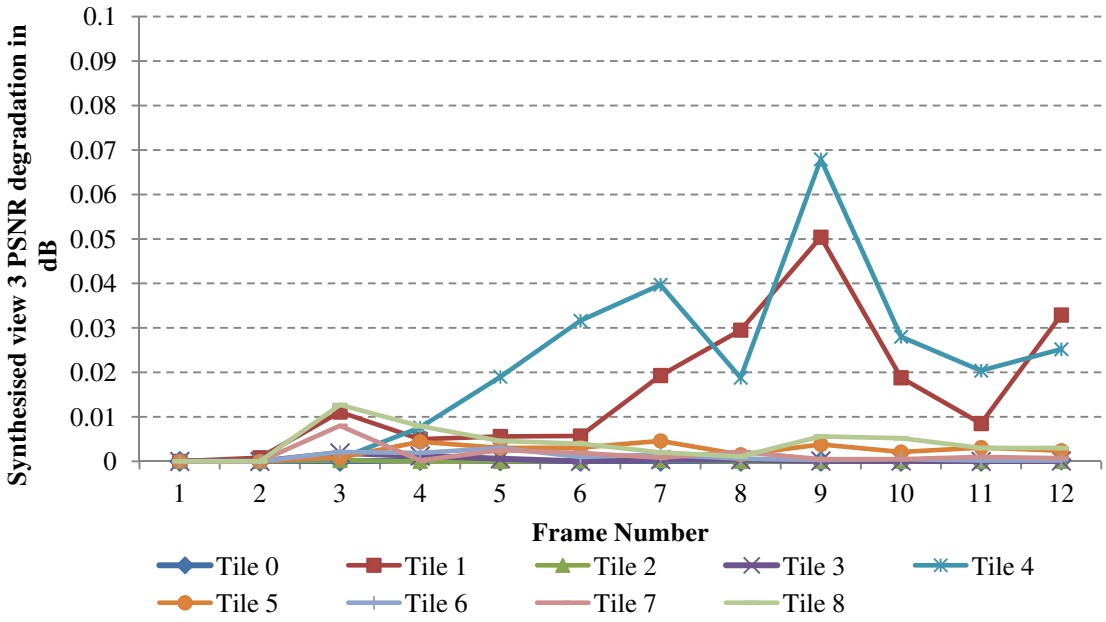


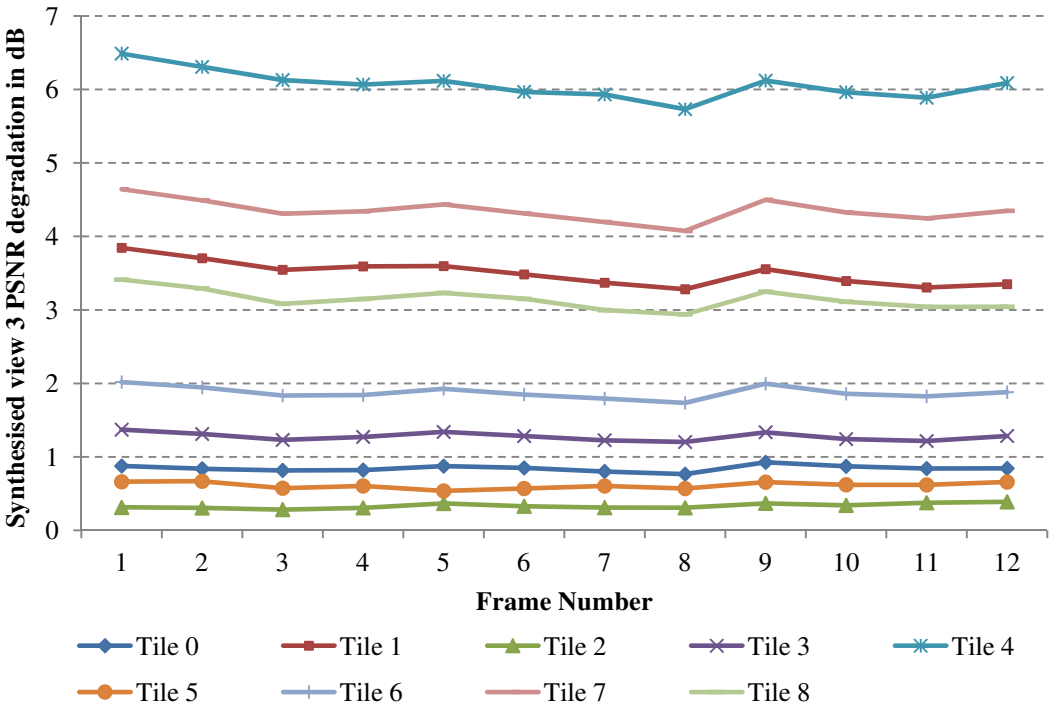
Figure 7.4: Tile configuration - *balloons* depth map view 1

Whilst the texture views are generally treated with higher priority when compared to depth map views, the priority of individual tiles in the depth map views depends on the degradation of the synthesised view PSNR resulting from packet loss effecting a tile.

In order to study the degradation of the synthesised view PSNR in the presence of depth map view tile loss, a series of experiments were carried out on various test sequences encoded using the low latency prediction structure shown in Figure 5.3. In these experiments, we iteratively simulate the loss of individual tiles in the depth map base view, thereby causing both temporal and inter-view error propagation, and compute the relative degradation in the synthesised view PSNR for each tile loss. Meanwhile, a simple tile copy error concealment strategy is used in order to minimise the impact on the synthesised view quality. Two tile loss scenarios were considered namely, tile losses in the I-key frame and tile losses in non-key frames when the I-key frame is received correctly. With reference to the synthesised view PSNR degradation for the *balloons* test sequence in these two scenarios, shown in Figure 7.5, a tile loss in a non-key frame generally results in a negligible degradation in PSNR showing that even a simple error concealment strategy can be effective for depth map views, given their characteristics. However, as expected, a tile loss in an I-key frame results in a significant degradation in the synthesised view PSNR due to the temporal and inter-view error propagation. Moreover, in the latter case, the degradation in the synthesised view PSNR varies widely in between tiles and thus, in this work, tiles in the depth map view key frames are segregated into two distinct priority levels, namely *depth_high* and *depth_low*. The top 5 ranking tiles in terms of synthesised view PSNR degradation are assigned to the *depth_high* category whilst the remaining are set to *depth_low*.




(a)




(b)

Figure 7.5: Synthesised view PSNR degradation in case of (a) Tile loss in non-key frame, (b) Tile loss in key frame

In order to determine the priority of the tiles in the depth map key frames, the degradation in the synthesised view PSNR needs to be first assessed for each tile. Thus the re-constructed texture and depth map key frames, which are already available in the respective MV-HEVC encoder, are exported to the traffic classifier whereby the packet loss-free synthesised view is rendered and stored. Following this process, the traffic classifier simulates individual tile losses in the depth map key frame iteratively, each time rendering a synthesized view and calculates the degradation in the synthesized view PSNR relative to the stored packet loss-free synthesised view. This process is carried out on-line for every key frame and allows the traffic classifier to effectively rank tiles according to their relative contribution to the synthesised view PSNR degradation in the case of packet loss.

The drawback of this method is that the tile loss assessment process needs to be repeated for every key frame thereby increasing the traffic classifier computational load. However, an analysis of various test sequences reveals that the depth map tile priority ranking does not vary significantly from one key frame to the next, as evidenced in Figure 7.6 for the *balloons* test sequence, and thus tile loss re-

assessments can be reduced leading to a corresponding reduction in computational load. Nevertheless, in order to improve robustness, this work evaluated the use of image statistics, specifically the sum of the depth map and texture view gradient statistics calculated on co-located tiles, to detect changes in the content thereby triggering a tile loss re-assessment. As shown in Figure 7.7, the variation in the computed image statistic for the *balloons* test sequence at frame 133, shown by the red dashed line, correlates with a change in the tile ranking shown in Figure 7.6. Thus, based on these observations and with reference to Algorithm 3 shown in Figure 7.8, the traffic classifier triggers a tile loss re-assessment and ranking procedure only when the variation in the image statistic is greater than a set deviation threshold, rather than at every key frame. Following a number of experiments on various test sequences, the deviation threshold in this work was set to 30% as this was found to strike a balance between the number of tile re-assessments and tile priority ranking error.

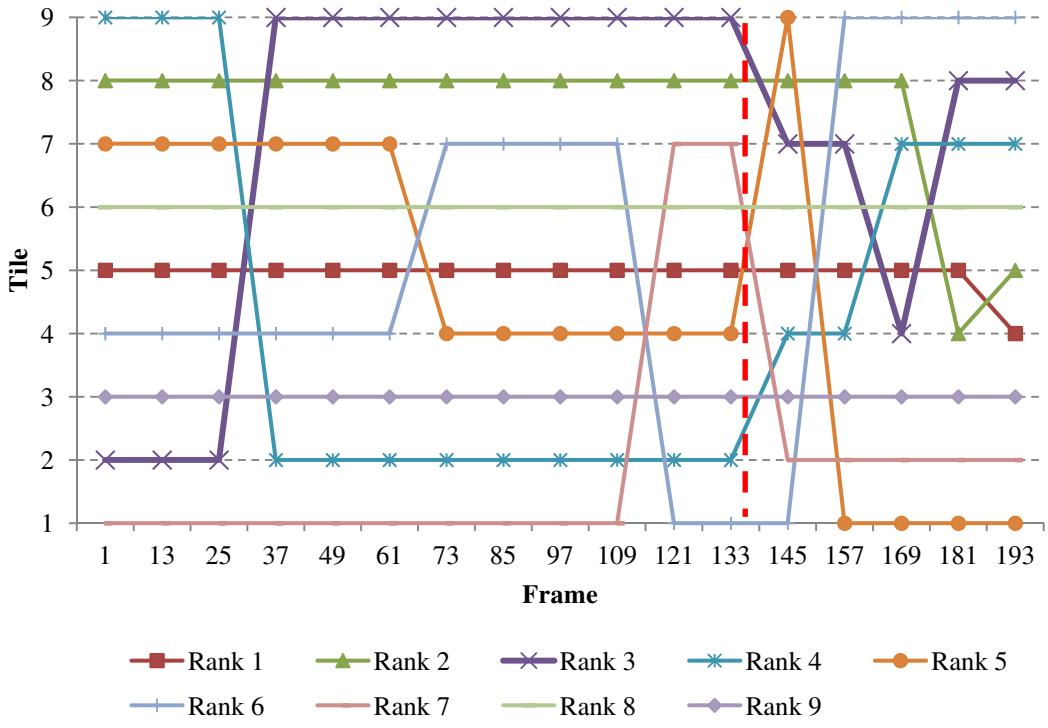



Figure 7.6: Tile loss ranking for *balloons* test sequence

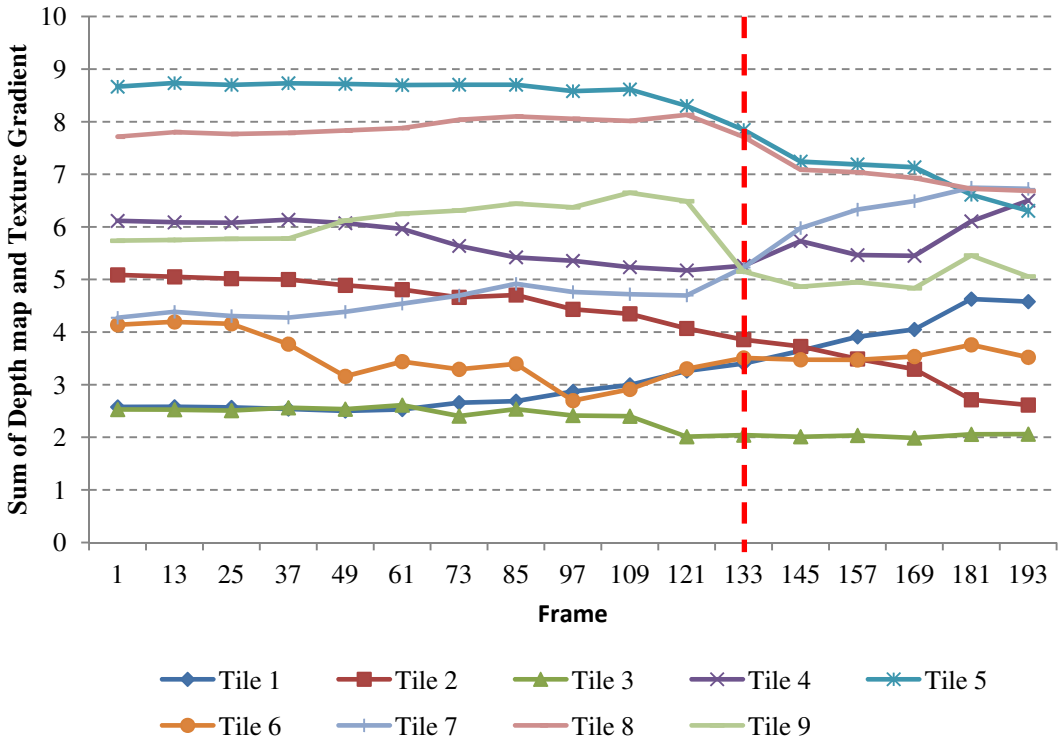



Figure 7.7: Image statistic variation per tile for *balloons* test sequence

## Algorithm 3: Tile Loss Re-assessment

1: for all depth map key frames do
2: if depth map key frame is the first in sequence then
3: Render and store the packet loss free synthesised view
4: for all tiles in the depth map key frame do
5: Simulate a tile loss in the depth map key frame
6: Render the synthesised view, calculate and store PSNR degradation
7: Compute and store the image statistic for the processed tile
8: end for
9: Evaluate tile ranking based on the PSNR degradation
10: else
11: for all tiles in the depth map key frame do
12: Compute and store the image statistic for the processed tile
13: if tile image statistic > *deviation_threshold* do
14: break loop and go to line 3
15: end if
16: end for
17: end if
18: end for

Figure 7.8: Algorithm 3 – Tile loss re-assessment

Considering the discussion on the MVD packet prioritisation scheme, the priority level weighting of packet $p$ in the $j^{th}$ flow, $\omega_{j,p}$, is set according to the values shown in Table 7.3, where P1 and P2 refers to the first 7 frames after a key frame and the last 4 frames in the GOP respectively. Once the traffic classifier determines the packet priority, this is indicated to the eNB scheduler by setting the appropriate DSCP parameter in the IP packet as discussed in chapter 6.

| Priority Level | Weighting $\omega_{j,p}$ | View Type | Layer (view) | Frame Type | Depth map Key Frame Tile Ranking | HPP |
|---|---|---|---|---|---|---|
| 1 | 2 | Texture | Base | Key Frame | | ✓ |
| 2 | 2 | Texture | Dependent | Key Frame | | ✓ |
| 3 | 1 | Texture | Base | Other - P1 | | ✓ |
| 4 | 1 | Depth | Base | Key Frame | *depth_high* | ✓ |
| | | | Dependent | Key Frame | *depth_high* | ✓ |
| | | Texture | Dependent | Other - P1 | | |
| 5 | 0.5 | Texture | Base | Other - P2 | | |
| | | | Dependent | Other - P2 | | |
| 6 | 0.5 | Depth | Base | Key Frame | *depth_low* | |
| | | | Dependent | Key Frame | *depth_low* | |
| | | | Base | Other - P1 | | |
| | | | Dependent | Other - P1 | | |
| | | | Base | Other - P2 | | |
| | | | Dependent | Other - P2 | | |

Table 7.3: MVD packet priority weighting

### 7.2.3 Simulation results

The modified content aware packet scheduling algorithm was tested in a system level LTE simulation over 50 random seeds and compared with the M-LWDF scheme using the crowd event scenario, described in section 3.2.2, and the simulation parameters shown in Table 7.4. The MVD test sequences used in these simulations, shown in Table 7.5, are encoded using the parameters given in Table 7.6 and prioritised using the traffic classifier based scheme discussed in the previous section.

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Content-Aware M-LWDF |
| TD-Scheduler Threshold, $T_{TD\_Th}$ | 0.6 |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| Traffic Model | 1 x Trace based MVD stream |
| Maximum end-to-end delay for video traffic, $\zeta$ | 100ms |
| Protocol Overhead [55] | RTP/UDP/IP with RoHC: 3 bytes<br>PDCP: 2 bytes<br>MAC and RLC: 5 bytes<br>CRC: 3 bytes |
| RLC Mode | UM / L1 ReTx: 3 |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | 50 |

Table 7.4: Simulation parameters – Content-aware M-LWDF

| Sequence | Total bit rate |
|---|---|
| *Balloons* | 674.7 kbps |
| *Book Arrival (Book)* | 656.6 kbps |
| *Kendo* | 662.8 kbps |
| *Newspaper (News)* | 660.8 kbps |
| *Poznan Hall* | 391.4 kbps |
| *Poznan Street* | 893.8 kbps |

Table 7.5: MVD test sequences bit rates

| Parameter | Value |
|---|---|
| GOP Size | 4 |
| Intra Period | 12 |
| Prediction Structure | Low latency (see Figure 5.3) |
| QP | Texture: 40 / Depth: 45 |
| Tile Configuration | 3 columns and 3 rows |

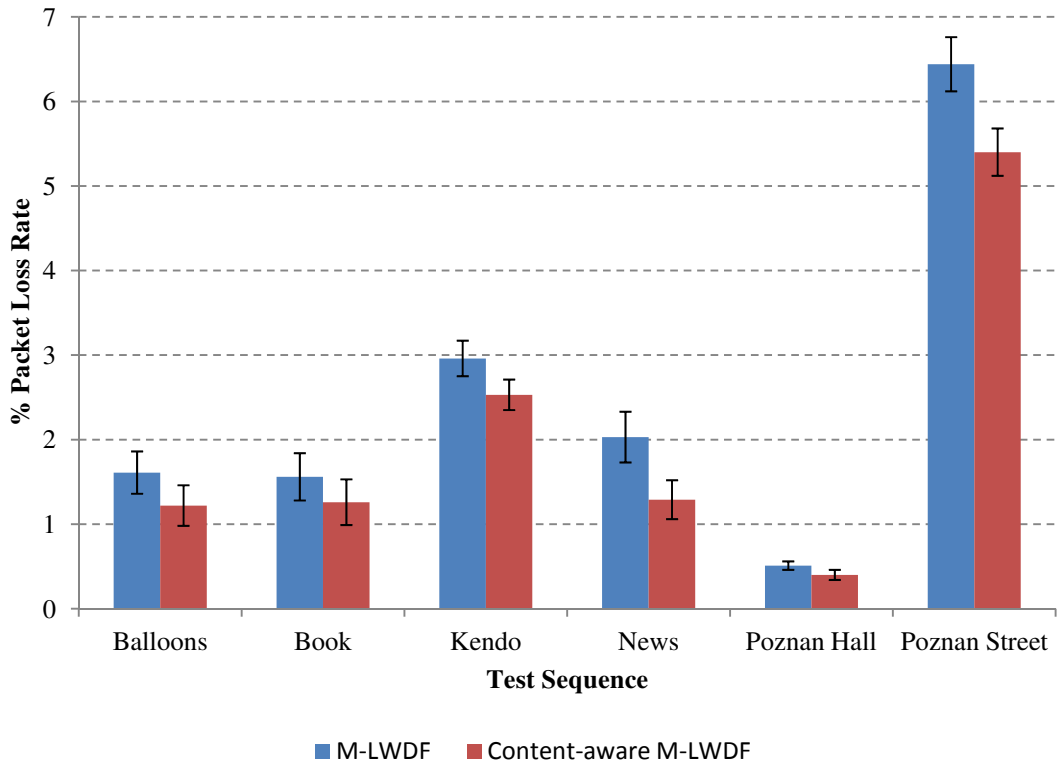
Table 7.6: MVD encoding parameters

Apart from the standard performance metrics namely, the packet loss rate (PLR), average packet delay and sector throughput, the proposed content aware M-LDWF scheme is also assessed on the high priority packet (HPP) PLR metric defined by:

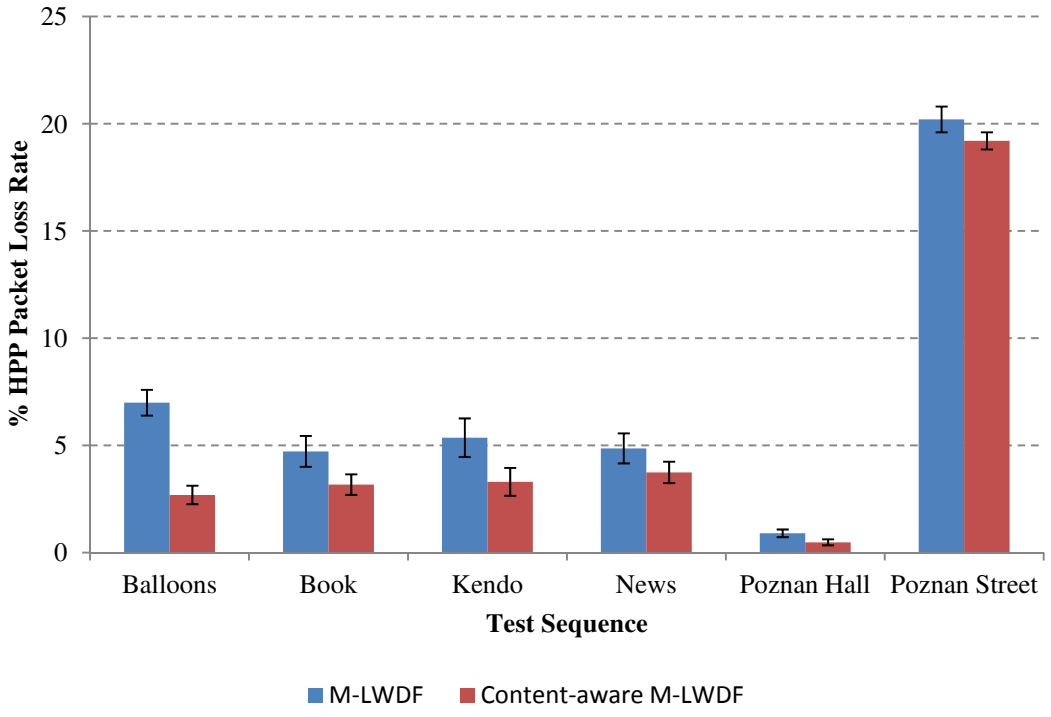
$$PLR_{HPP} = \frac{N_{tx\_hpp} - N_{rx\_hpp}}{N_{tx\_hpp}} \tag{7.7}$$

where $N_{tx\_hpp}$ and $N_{rx\_hpp}$ are the number of HPP packets transmitted by the eNB and successfully received by the UE respectively.

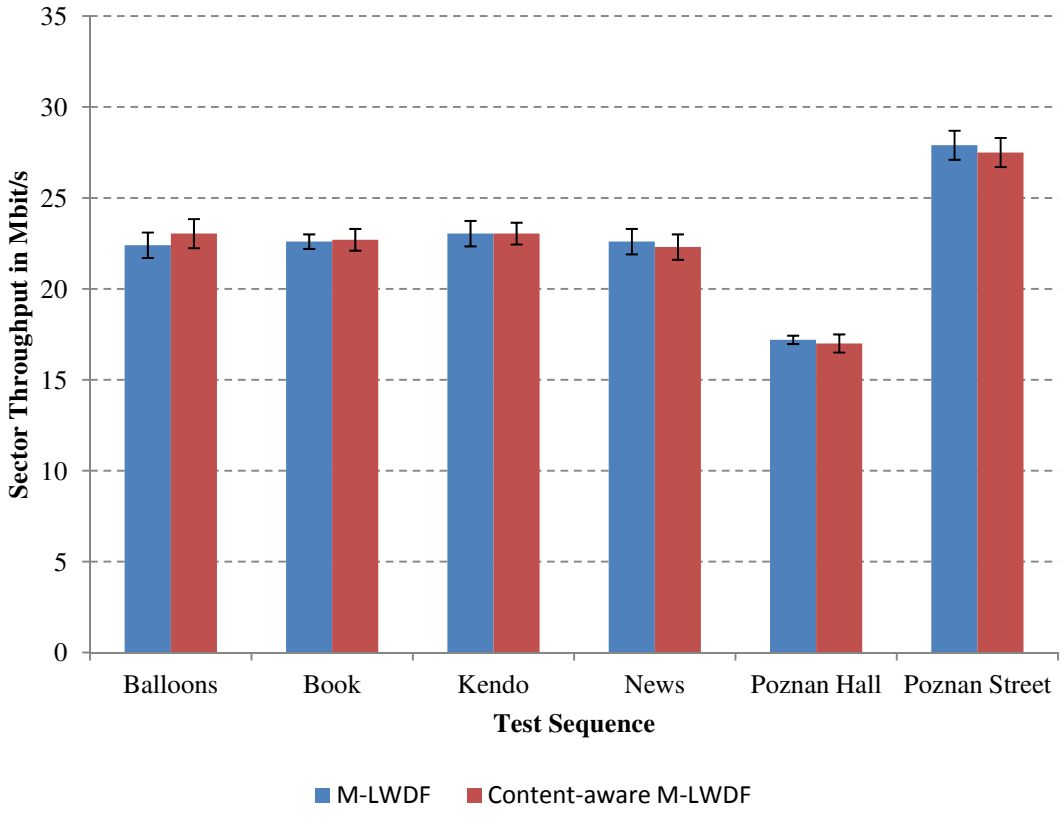
As shown in Figure 7.9, although the improvement in the average PLR is marginal, the high priority packet PLR of the proposed scheme is on average 1.74% better than that achieved by the standard M-LWDF and up to 4.3% for the *balloons* test sequence. This improvement is achieved by sacrificing low priority packets, predominantly in the depth map views, in lieu of improving the delivery of packets carrying more relevant video content.



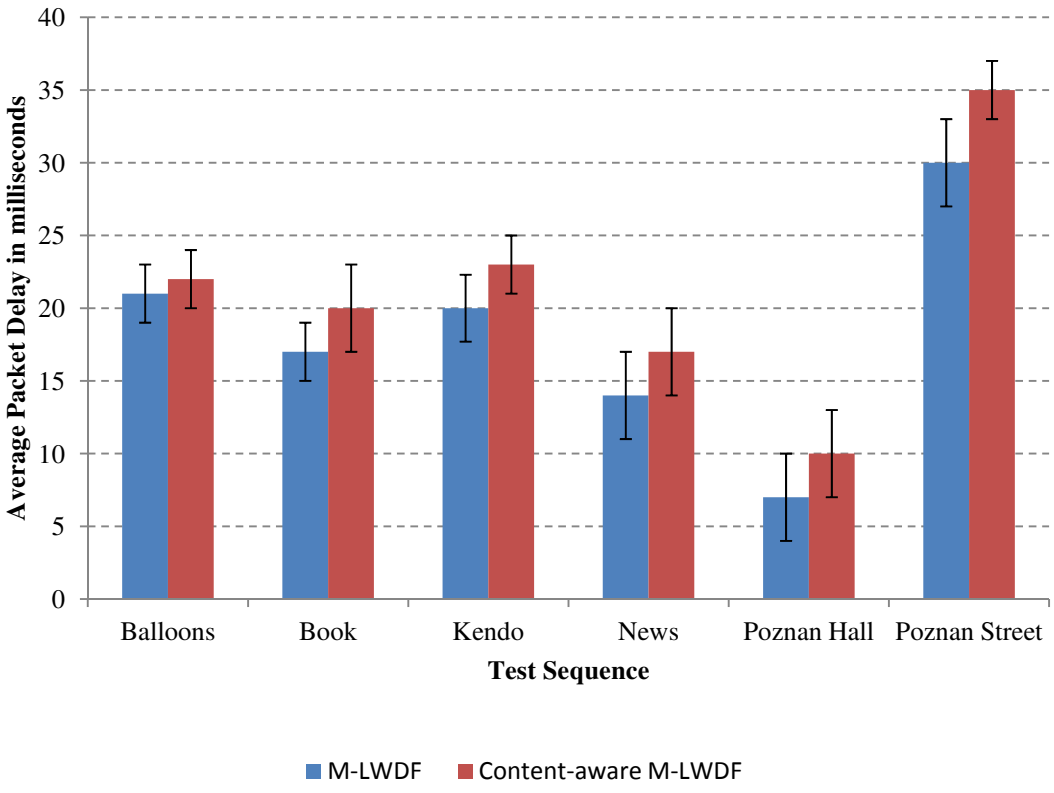

(a)


25
20
15
10
5
0
% HPP Packet Loss Rate
Balloons
Book
Kendo
News
Poznan Hall
Poznan Street
Test Sequence
M-LWDF
Content-aware M-LWDF

(b)


35
30
25
20
15
10
5
0
Sector Throughput in Mbit/s
Balloons
Book
Kendo
News
Poznan Hall
Poznan Street
Test Sequence
M-LWDF
Content-aware M-LWDF

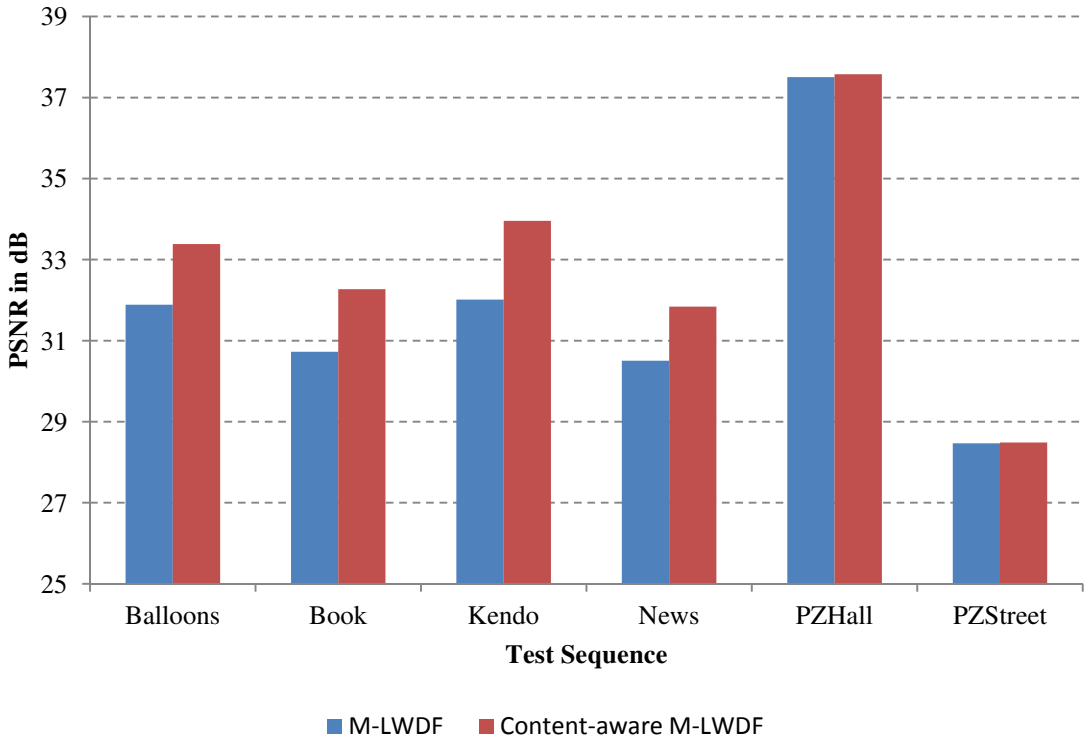
(c)



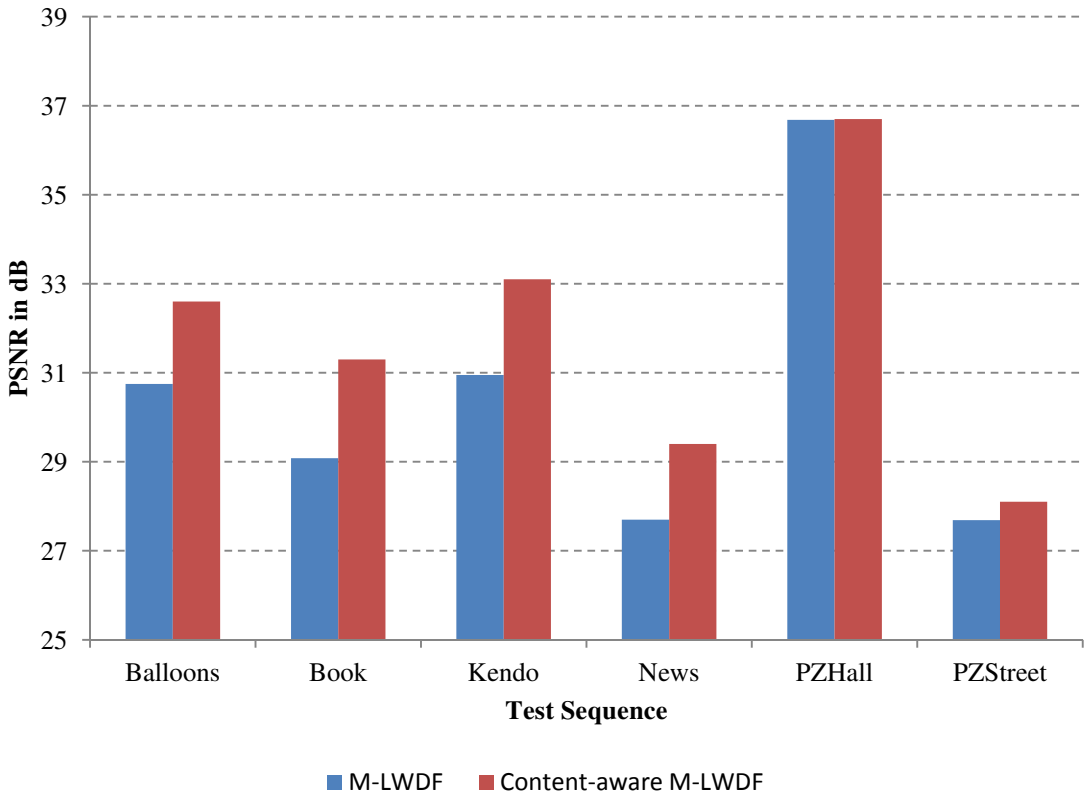

(d)

Figure 7.9: Performance of M-LWDF and Content-aware M-LWDF (a) PLR, (b) HPP PLR, (c) Sector Throughput, (d) Packet delay

Although the HPP PLR metric is indicative of the impact on the video quality, in order to assess the visual quality of the delivered video sequences, the corrupted bit streams are recovered and decoded using a tile copy error concealment technique. Using the decoded sequences of all the users in the radio sector, the PSNR and structural similarity metric (SSIM), shown in Figure 7.10 and 7.11 respectively, are calculated across all texture and synthesised views. As evidenced in Figure 7.10, the proposed scheme improves the texture and synthesised view quality, in loaded conditions, by an average of 1.13dB and 1.47dB respectively. This improvement is also corroborated by the SSIM results shown in Figure 7.11.

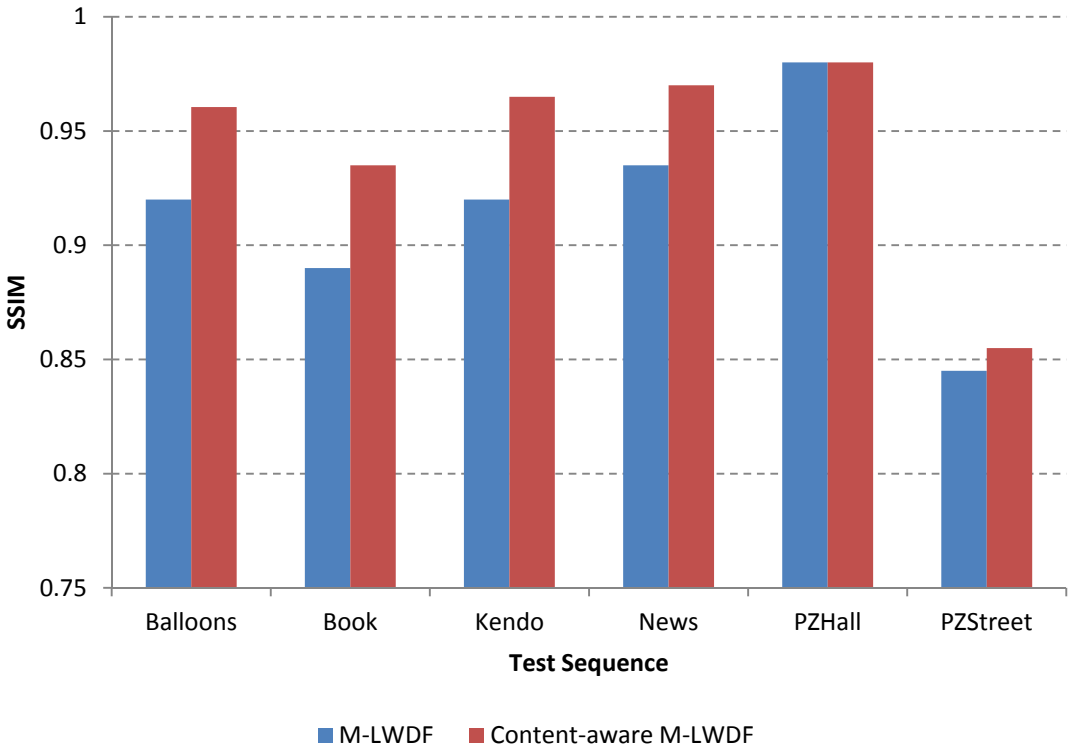



(a)

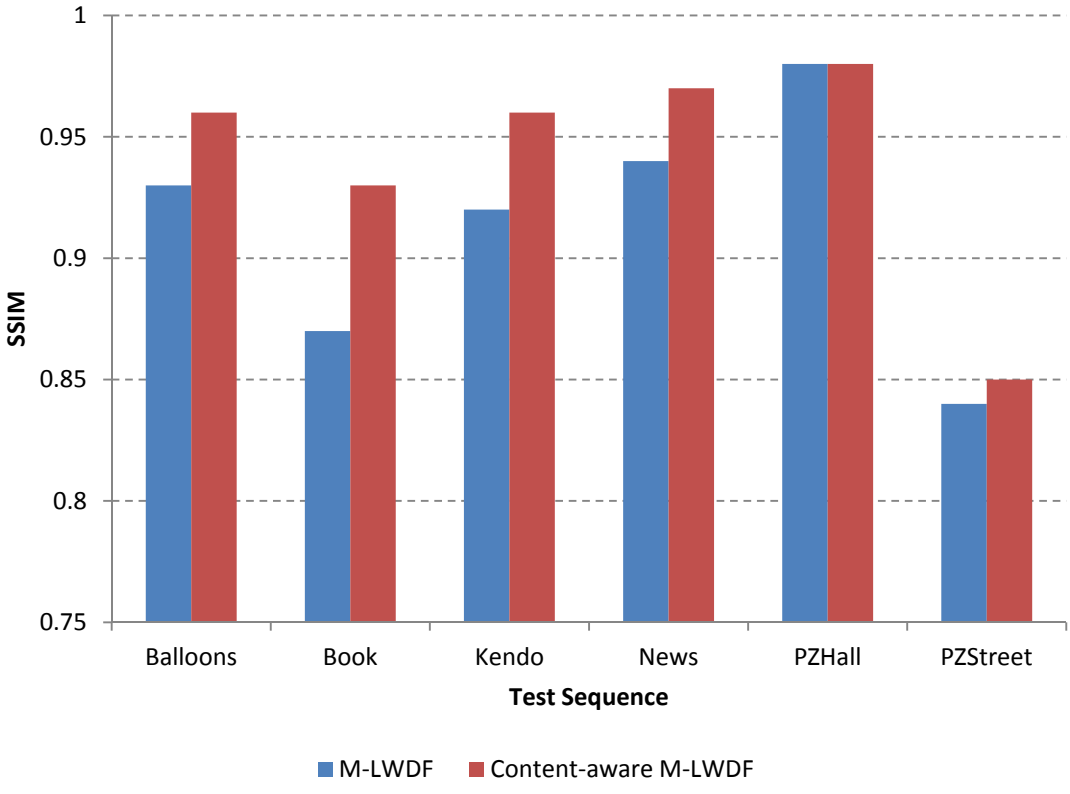



(b)

Figure 7.10: PSNR (a) Texture, (b) Synthesised views




(a)




(b)

Figure 7.11: SSIM (a) Texture, (b) Synthesised views

## 7.3 Impact of feedback reduction schemes on MVD video transmission in a crowd event scenario

The experiments carried out in the previous section focused on the use of a standard wideband CQI feedback reporting scheme however, as discussed in chapter 3, crowd event scenarios are characterised by high uplink feedback signalling overhead leading to increased uplink resource consumption. For this reason, two new complimentary schemes, namely a UE-assisted predictive filtering scheme and a UE clustering scheme, were introduced in chapter 3. Although both schemes are able to achieve significant feedback reduction, this comes at the expense of an increase in packet loss rate. Moreover, the simulation results shown in chapter 3, are based on a best effort infinite buffer traffic model and do not take into account the stringent requirements of MVD video transmission. Thus, in this section, the two complimentary CQI feedback signalling reduction schemes, proposed in chapter 3, are combined in order to study the impact on the performance of MVD video transmission when using the content aware M-LWDF packet scheduling algorithm.

### 7.3.1 System model

With reference to the system model shown in Figure 7.12, the UEs in the radio sector are first clustered into $k$ clusters and a cluster leader is elected for each cluster as described in section 3.3. In the combined scheme, the cluster leaders employ a UE assisted LMS based prediction filtering scheme, detailed in section 3.1, in order to achieve additional feedback signalling overhead savings through the reduction of cluster leader feedback reports. The inter-working of the proposed schemes is controlled by parameters $\tau_{UE}$, $\gamma$ and $|E_{max\text{-}cqi}|$ and is summarised in the state flow diagram shown in Figure 7.13.

The combined feedback reduction scheme is implemented and tested in a system level LTE simulator [55] over 50 random seeds using the simulation parameters shown in Table 7.7. In these tests, the MVD video traffic is generated by encoding the *balloons* test sequence using the parameters given in Table 7.6 and prioritised using the traffic classifier based scheme discussed earlier.

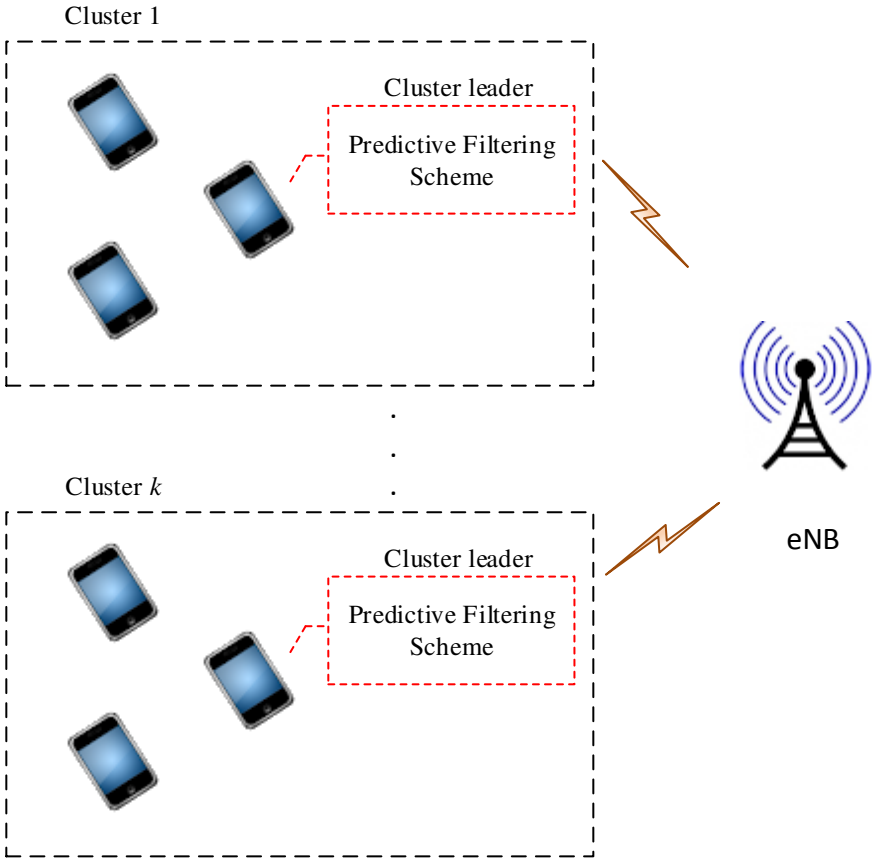



Figure 7.12: Combined feedback reduction scheme system model

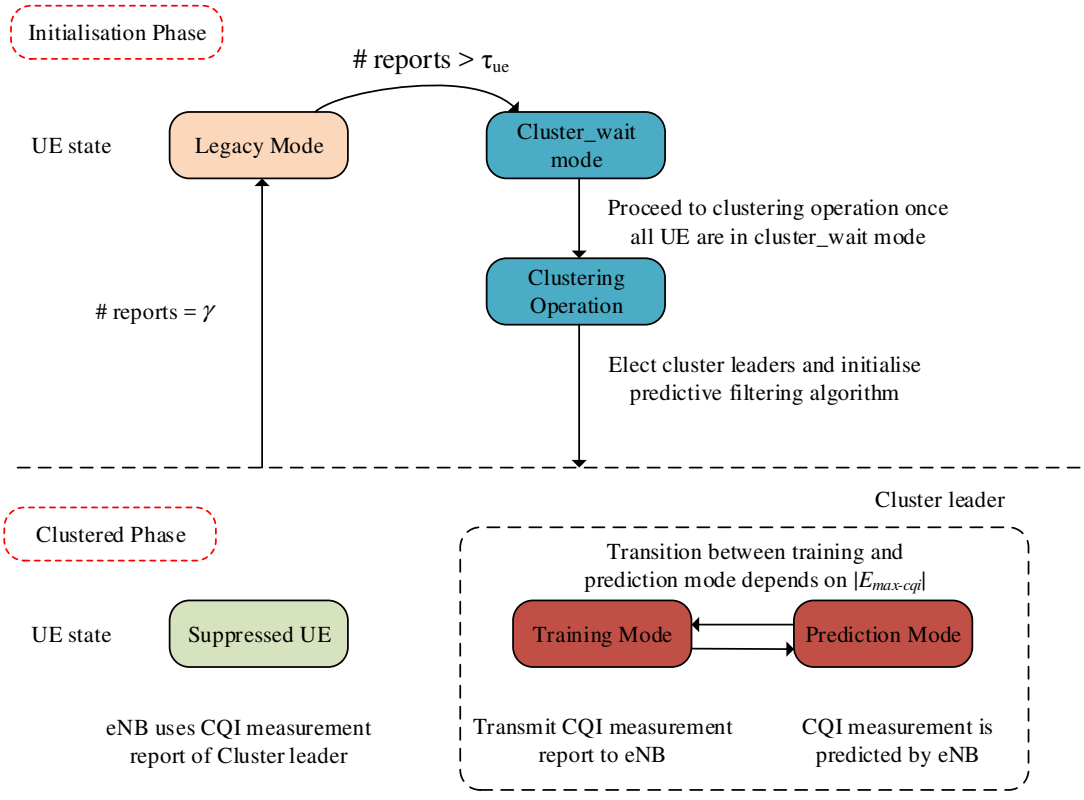



Figure 7.13: Combined feedback reduction scheme - State flow diagram

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Content-Aware M-LWDF |
| TD-Scheduler Threshold, $T_{TD\_Th}$ | 0.6 |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| CQI feedback reduction scheme | Combined clustering and LMS-based predictive filtering |
| Clustering parameters | $\tau_{UE}$ =100, γ=1200 |
| Predictive filtering parameter, $\|E_{max\text{-}cqi}\|$. | {0.5, 1.0, 1.5} |
| Traffic Model | 1 x Trace based MVD stream |
| Maximum end-to-end delay for video traffic, $\zeta$ | 100ms |
| Protocol Overhead [55] | RTP/UDP/IP with RoHC: 3 bytes<br>PDCP: 2 bytes<br>MAC and RLC: 5 bytes<br>CRC: 3 bytes |
| RLC Mode | UM / L1 ReTx: 3 |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | 50 |

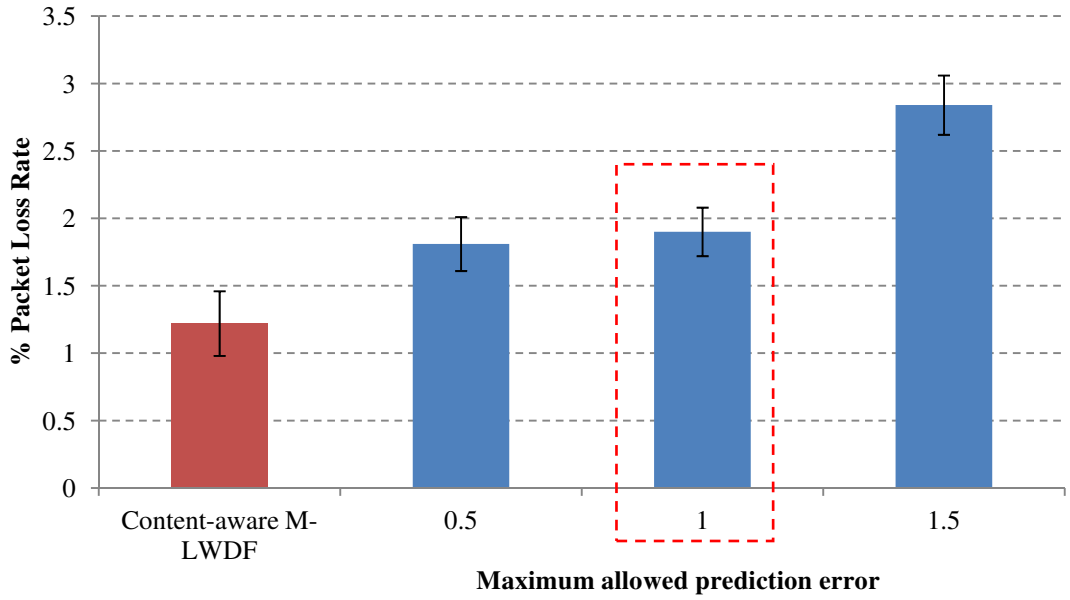
Table 7.7: Simulation parameters – Combined feedback reduction scheme

### 7.3.2 Simulation results

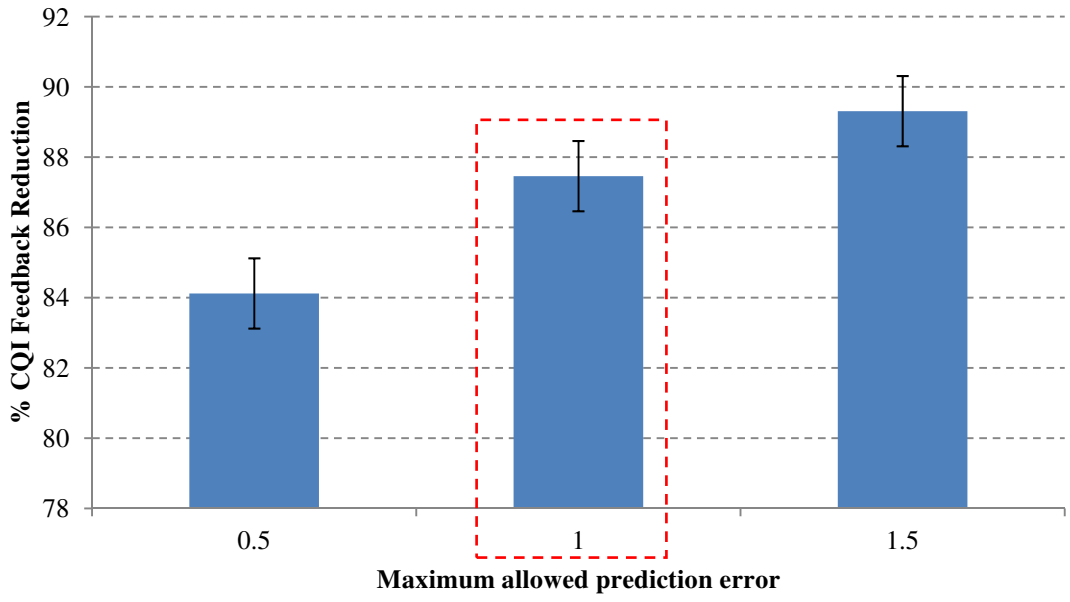
Whilst the clustering parameters $\tau$ and $\gamma$ are empirically set to 100 and 1200 respectively, the impact of the maximum allowed prediction error, $|E_{max\text{-}cqi}|$, on the average PLR and CQI feedback reduction is evaluated by varying $|E_{max\text{-}cqi}|$ between 0.5 and 1.5 in steps of 0.5. As expected, the reduction in CQI feedback, shown in Figure 7.14, increases with the maximum allowed prediction error at the expense of an increase in PLR when compared to the performance achieved by the content-aware M-LWDF scheme without the use of feedback reduction techniques. In the analysed scenario, this increase is primarily due to an increase in packet discards caused by packet delay timer expiry rather than physical errors. This, in turn is caused by the predictive filtering entity in the eNB feeding the scheduler a CQI value which is lower than the actual CQI value thereby leading to a relatively lower scheduling opportunity and the use of lower MCS when the UE is scheduled.

As shown in Figure 7.14(c), the use of the predictive filtering scheme reduces the cluster leader reports by an average of 9%. Moreover, considering that the PLR

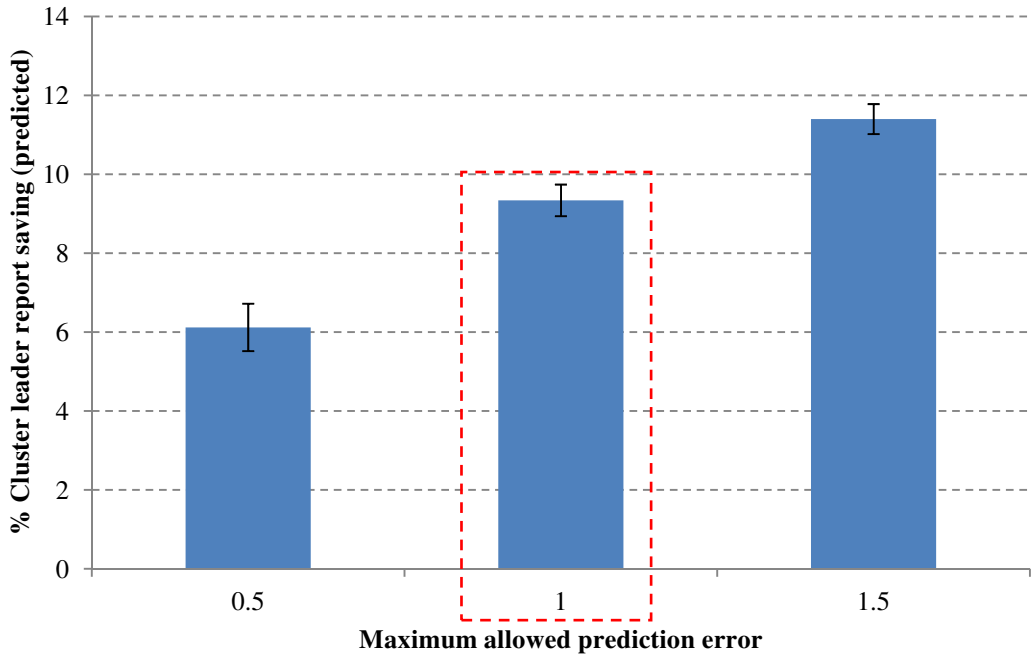
increases significantly when the maximum allowed prediction error is greater than 1, $|E_{max\text{-}cqi}|$ is set to 1 for the rest of the work as it yields the best trade-off between the selected metrics. As expected, even though the combined use of the CQI feedback reduction technique yields a saving in uplink feedback signalling of 87.5% at the selected operating point, the increase in the packet loss rate degrades the visual quality of the texture and synthesised views by approximately 2dB in terms of PSNR as evidenced in Figure 7.15.




(a)



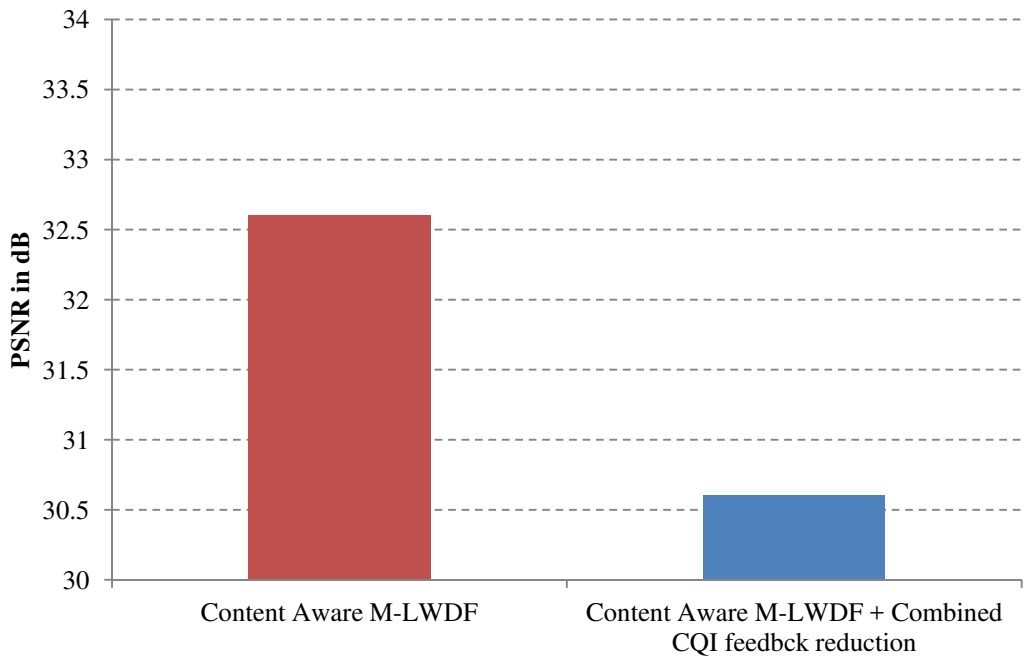

(b)




(c)

Figure 7.14: Content aware M-LWDF performance combined with CQI feedback reduction scheme – Variation of the maximum allowed prediction error (a) PLR, (b) CQI feedback reduction, (c) Cluster leader feedback report savings


(a)




(b)

Figure 7.15: Content aware M-LWDF performance combined with CQI feedback reduction scheme - PSNR (a) Texture views, (b) Synthesised views at selected operating point

## 7.4 Cross-layer CQI feedback adaptation for MVD video transmission in a crowd event scenario

In order to counter the impact of the proposed combined CQI feedback reduction scheme, a cross-layer technique is designed to adapt the CQI feedback by soliciting CQI reports (referred to as CQI boosts) from individual UEs. The eNB measures and tracks the high priority packet PLR and weighted HOL delay for all suppressed UEs at each cluster leader CQI reporting cycle using (7.4) and (7.2) respectively and compares these metrics to pre-set thresholds. This process allows the eNB to identify a subset of suppressed UEs which are either experiencing high priority packet loss or accumulating packet delays thereby becoming more susceptible to packet discards.

To prevent further degradation in these metrics, which have a direct bearing on the visual quality of the MVD video transmission, the eNB triggers CQI boosts from the identified subset of suppressed UEs through the use of a 1-bit CQI boost request (CBR) indicator within the physical downlink control channel (PDCCH). When one of the metrics exceeds its corresponding pre-set threshold, the eNB sets the CBR indicator to 1, thereby requesting the UE to report the actual CQI value in the uplink at

every CQI reporting cycle until the CBR indicator is reset to 0. The eNB resets the CBR indicator flag for a suppressed UE when the measured metrics fall below the pre-set threshold whilst considering some hysteresis to protect against oscillations between the normal and CQI boosting mode. Whenever the eNB receives an actual CQI value from a suppressed UE, as a result of a CQI boost request, the eNB uses this value instead of the cluster leader derived value for scheduling purposes.

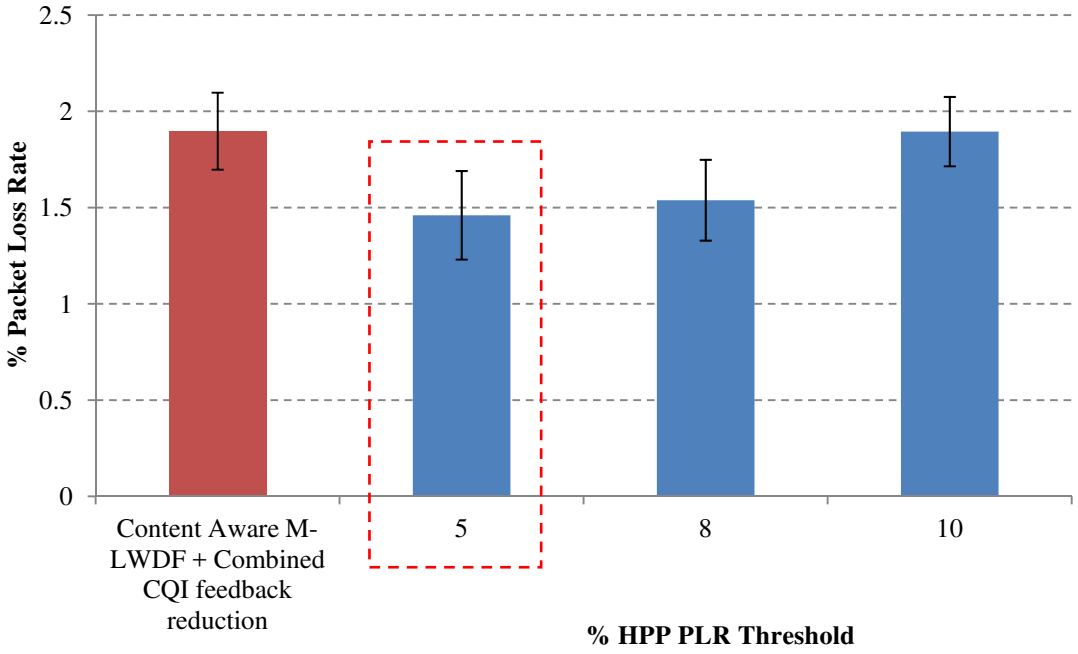
### 7.4.1 Simulation results

Whilst the weighted HOL delay threshold is empirically set to 80ms, the effect of the HPP PLR threshold is studied by integrating the cross-layer technique with the combined feedback reduction scheme, discussed in the previous section, and tested in a system level LTE simulator [55] over 50 random seeds using the simulation parameters shown in Table 7.8. Similar to previous parameter tuning experiments, the MVD video traffic is generated by encoding the *balloons* test sequence using the parameters given in Table 7.6 and prioritised using the traffic classifier based scheme discussed in section 7.2.2.

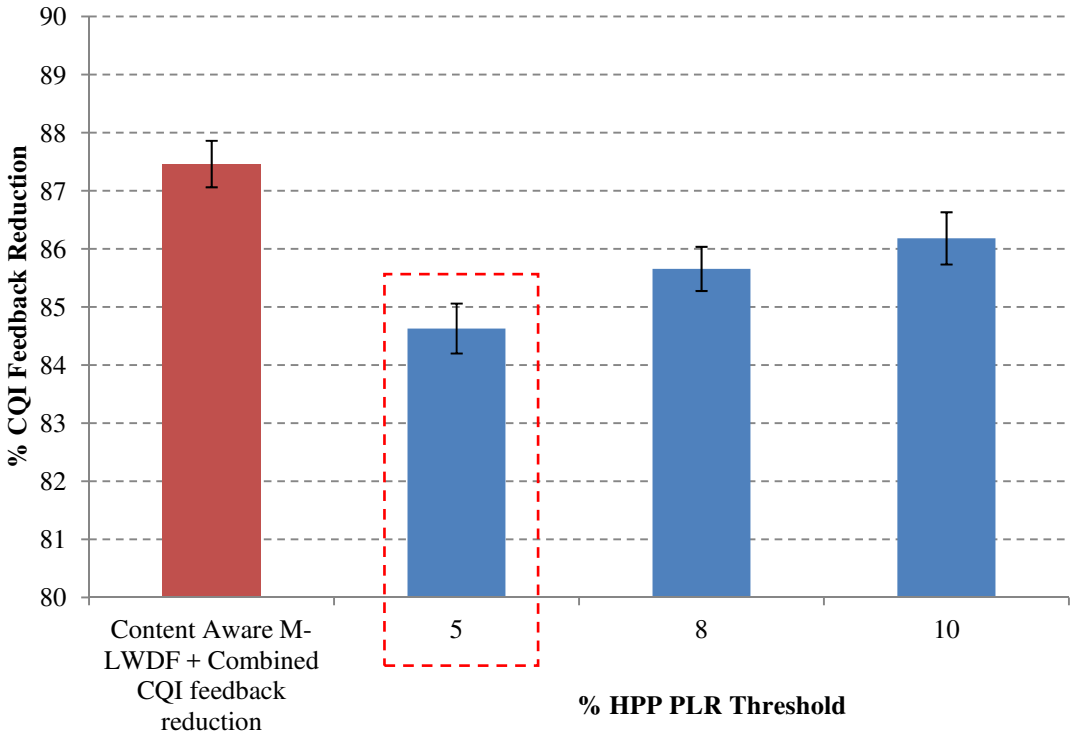
With reference to Figure 7.16(a-b), the use of the cross-layer CQI feedback adaptation scheme improves the packet loss rate, when compared to the same scheme without cross-layer adaptation, at the expense of a marginal degradation in CQI feedback reduction caused by the solicited CQI reports. As expected, the CQI boosts become less frequent for increasing values of HPP PLR threshold leading to a gradual degradation in packet loss rate whilst improving the CQI feedback reduction. Meanwhile, the sector throughput and average packet delay, shown in Figure 7.16(c-d) suffer a marginal degradation as the HPP PLR threshold increases, thus this was set to 5% for the rest of the work as it yields an acceptable trade-off between CQI feedback reduction and packet loss rate.

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Content-Aware M-LWDF |
| TD-Scheduler Threshold, $T_{TD\_Th}$ | 0.6 |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (20ms), wideband CQI (4bits) |
| CQI feedback reduction scheme | Combined clustering and LMS-based predictive filtering |
| Clustering parameters | $\tau$ =100, $\gamma$=1200 |
| Predictive filtering parameter, $\|E_{max\text{-}cqi}\|$. | 1.0 |
| Cross-layer HPP PLR threshold | {5%, 8%, 10%} |
| Cross-layer weighted HOL threshold | 80ms |
| Traffic Model | 1 x Trace based MVD stream |
| Maximum end-to-end delay for video traffic, $\zeta$ | 100ms |
| Protocol Overhead [55] | RTP/UDP/IP with RoHC: 3 bytes<br>PDCP: 2 bytes<br>MAC and RLC: 5 bytes<br>CRC: 3 bytes |
| RLC Mode | UM / L1 ReTx: 3 |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F [52] |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz [55] |
| Fast Fading Realization | Jakes Model [55] |
| Number of UE's, $T_u$ | 50 |

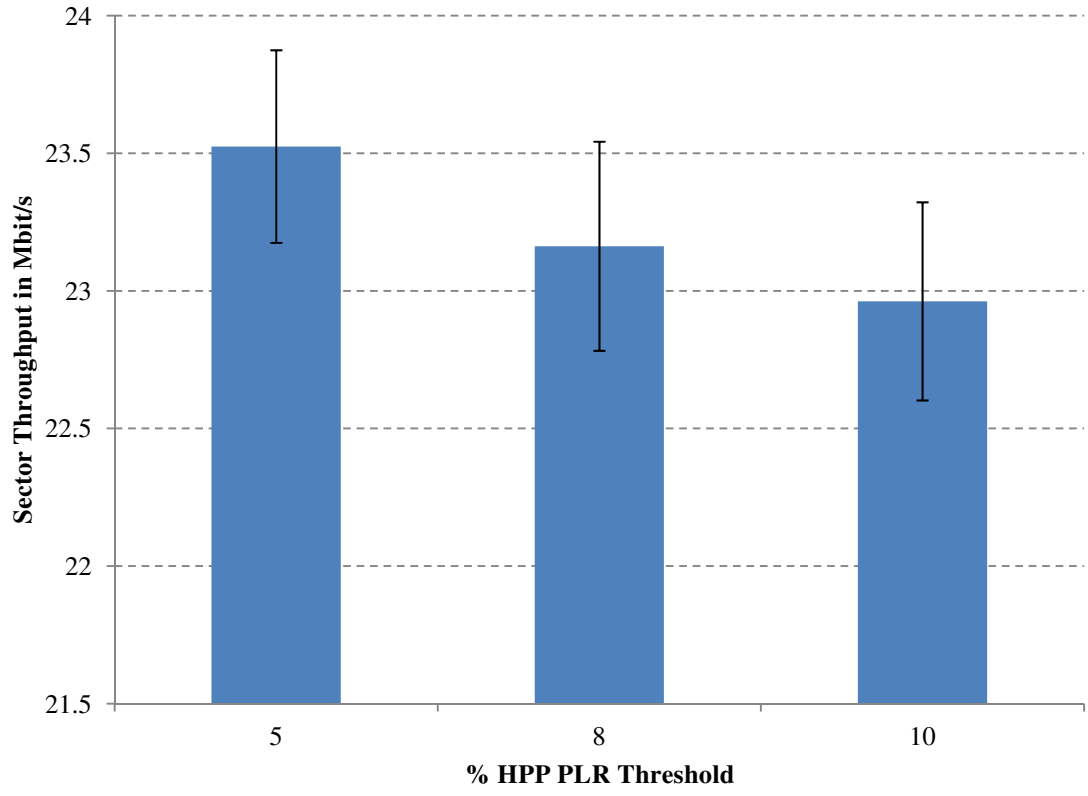
Table 7.8: Simulation parameters – Cross layer adaptation




(a)


90
89
88
87
86
85
84
83
82
81
80
% CQI Feedback Reduction
Content Aware M-LWDF + Combined CQI feedback reduction
5
8
10
% HPP PLR Threshold

(b)


24
23.5
23
22.5
22
21.5
Sector Throughput in Mbit/s
5
8
10
% HPP PLR Threshold

(c)

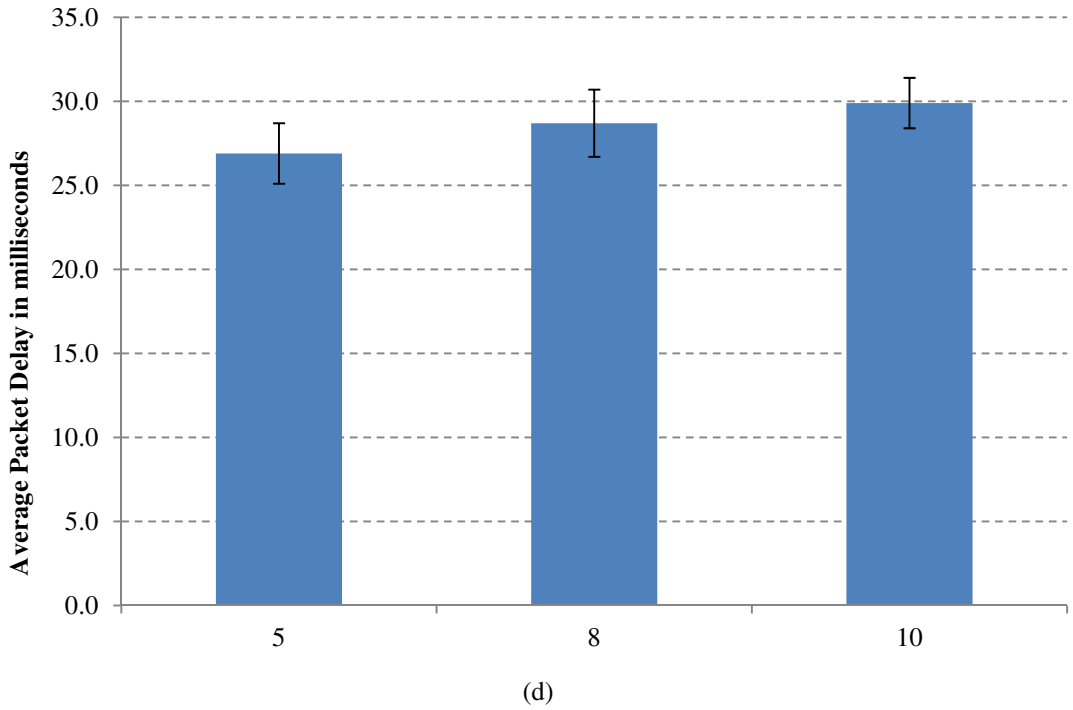
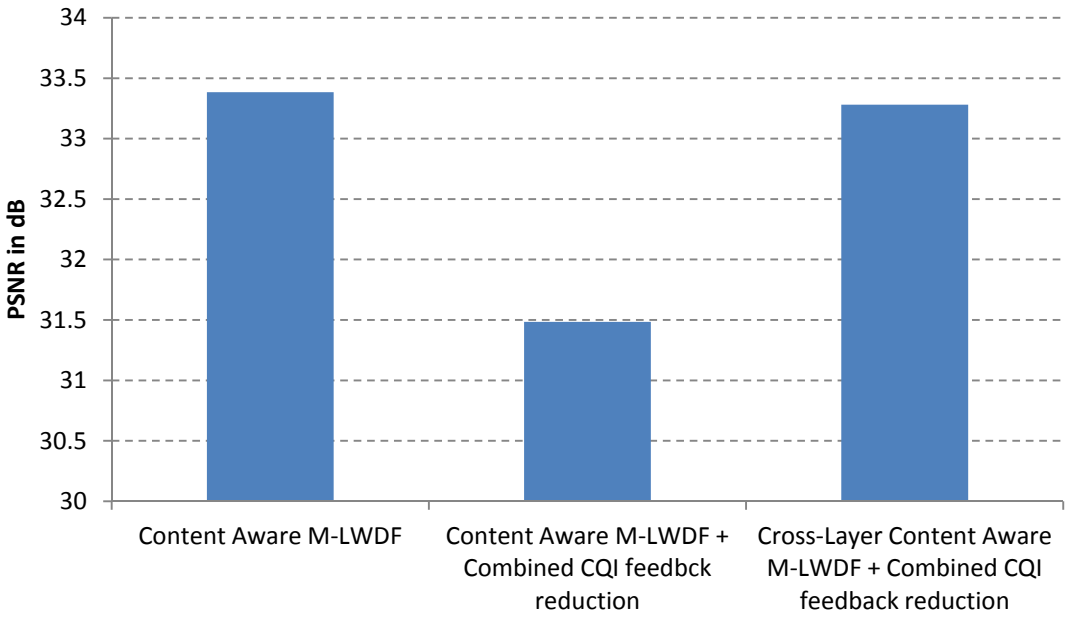



Figure 7.16: Cross layer CQI feedback adaptation system level performance metrics – *balloons* test sequence (a) PLR, (b) CQI feedback reduction, (c) Sector throughput, (d) Average packet delay

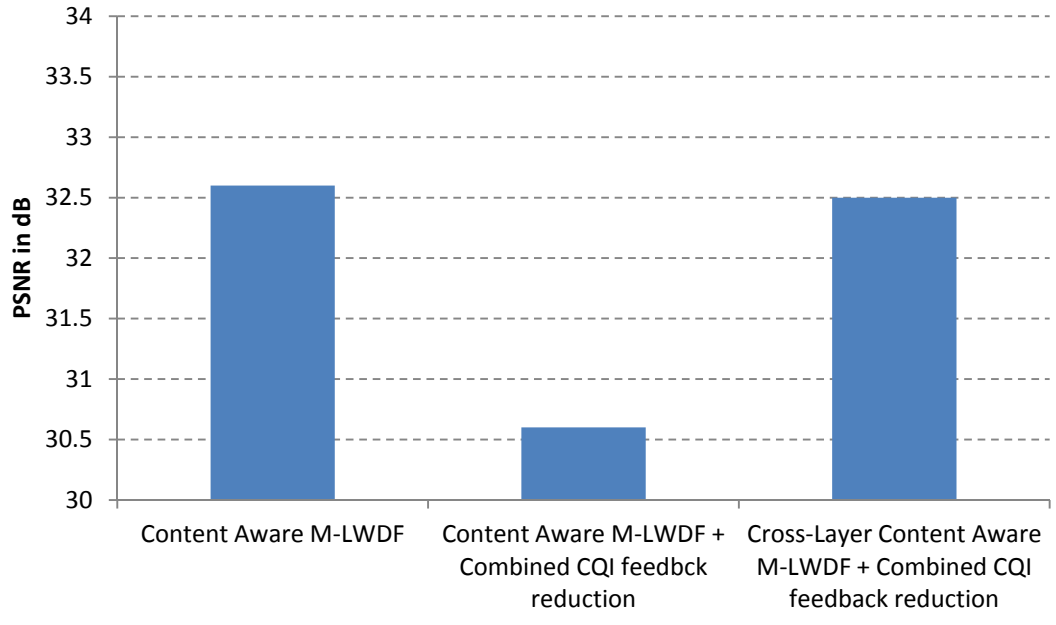
The improvements observed in the system level performance metrics at the selected operating point are reflected in the visual quality of the MVD stream. With reference to the comparative assessment shown in Figure 7.17 for the *balloons* test sequence, the use of the proposed cross layer technique results in a significant improvement in the texture and synthesised view PSNR, closely matching that of the content-aware M-LWDF without any feedback reduction techniques. Conducting the same experiment using the proposed cross layer technique at the selected operating point for the remaining test sequences reveals similar performance improvements for both system level and visual quality metrics as shown in Table 7.9 - 7.10 and Figure 7.18 respectively.

The average CQI feedback reduction achieved by the proposed cross-layer scheme is 84.1% whilst the PSNR is generally within 0.5dB from the content-aware M-LWDF scheme for all test sequences. Further observation reveals that the *PZStreet* test sequence exhibits a CQI feedback reduction of 78.5%, this being the lowest of all test sequences. This is caused by the fact that this test sequence exhibits a relatively high total bit rate causing severe congestion and leading to a high PLR. For this reason, the

proposed cross-layer scheme attempts to reduce the PLR, by soliciting CQI reports from a large number of UEs thereby lowering the CQI feedback reduction. Nevertheless, considering that the PSNR quality metric for this test sequence is below 30dB for both the texture and synthesised views in the two tested schemes, the most beneficial action in this case would be to reduce the traffic intensity and switch from MVD video to single view video transmission.




(a)




(b)

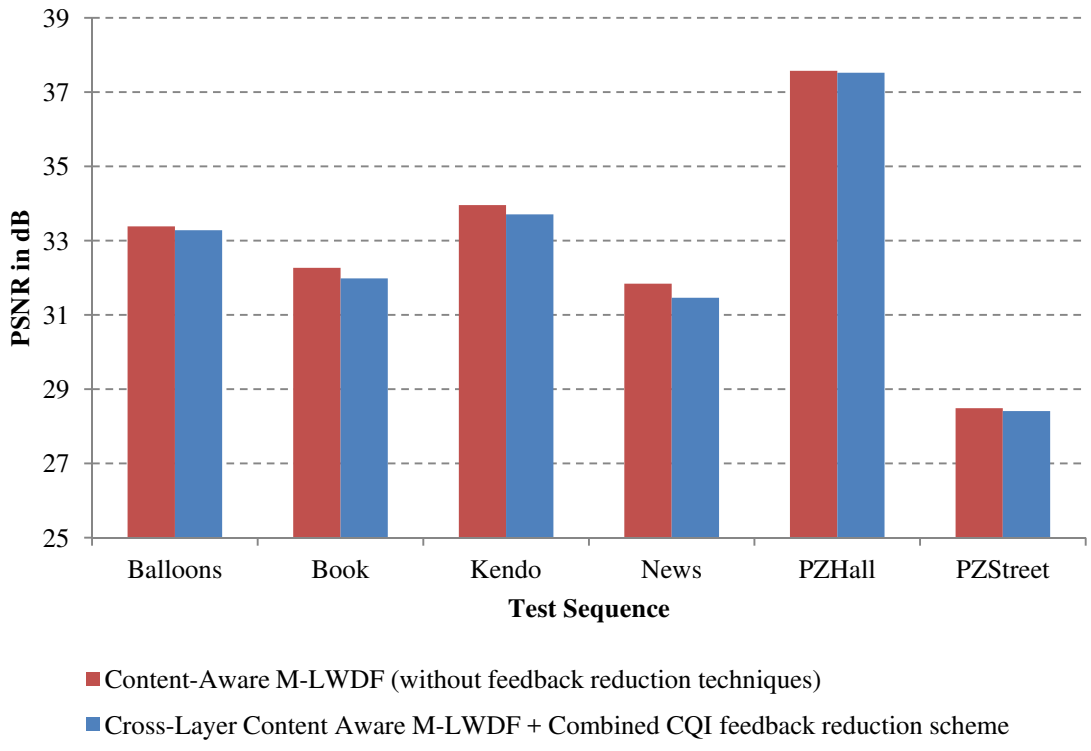
Figure 7.17: PSNR comparative assessment – *balloons* test sequence (a) Base view (b) Synthesised view

| Sequence | % PLR | % HPP PLR | Average Packet Delay / ms | Sector Throughput / Mbit/s | % CQI Feedback Reduction |
|---|---|---|---|---|---|
| *Balloons* | 1.46 (±0.23) | 3.25 (±0.47) | 27 (± 2) | 23.52 (±0.35) | 84.62 (±0.43) |
| *Book* | 1.35 (±0.4) | 3.36 (±0.42) | 28 (± 8) | 22.69 (±0.51) | 84.57 (±0.61) |
| *Kendo* | 2.66 (±0.61) | 3.41 (±0.52) | 28 (± 5) | 23.01 (±0.62) | 85.27 (±0.21) |
| *News* | 1.43 (±0.25) | 3.82 (±0.45) | 27 (± 7) | 22.77 (±0.81) | 84.28 (±0.18) |
| *PZHall* | 0.43 (±0.16) | 0.5 (±0.12) | 9 (± 3) | 17.23 (±0.23) | 87.20 (±0.37) |
| *PZStreet* | 5.72 (±0.96) | 19.42 (±1.1) | 44 (± 4) | 28.8 (±0.66) | 78.53 (±0.91) |

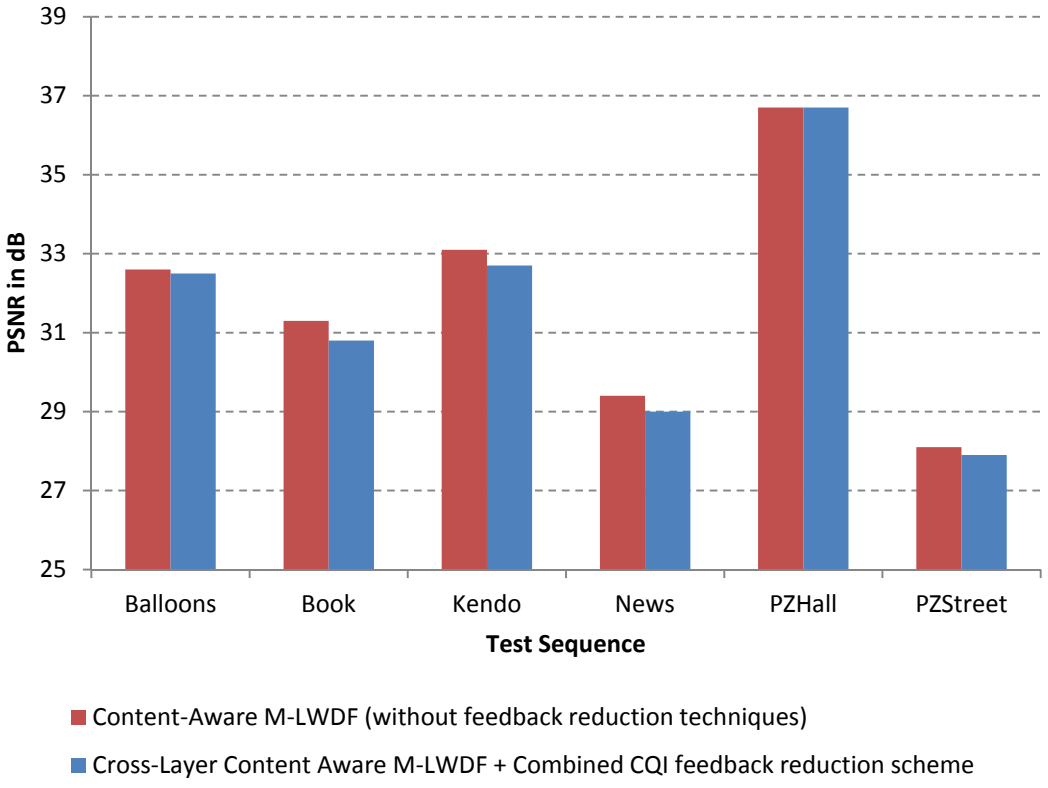
Table 7.9: Cross-layer Content-aware M-LWDF with Combined CQI feedback reduction scheme – System level performance metrics

| Sequence | % PLR | % HPP PLR | Average Packet Delay / ms | Sector Throughput / Mbit/s |
|---|---|---|---|---|
| *Balloons* | 1.22 (±0.24) | 2.69 (±0.43) | 27 (± 2) | 23.04 (±0.8) |
| *Book* | 1.26 (±0.27) | 3.17 (±0.48) | 27 (± 5) | 22.69 (±0.6) |
| *Kendo* | 2.53 (±0.18) | 3.30 (±0.65) | 28 (± 7) | 23.04 (±0.6) |
| *News* | 1.29 (±0.23) | 3.73 (±0.50) | 27 (± 3) | 22.34 (±0.7) |
| *PZHall* | 0.40 (±0.06) | 0.48 (±0.14) | 8 (± 8) | 17.01 (±0.5) |
| *PZStreet* | 5.42 (±0.28) | 19.23 (±0.40) | 46 (± 2) | 27.50 (±0.8) |

Table 7.10: Content-aware M-LWDF packet scheduling – System level performance metrics



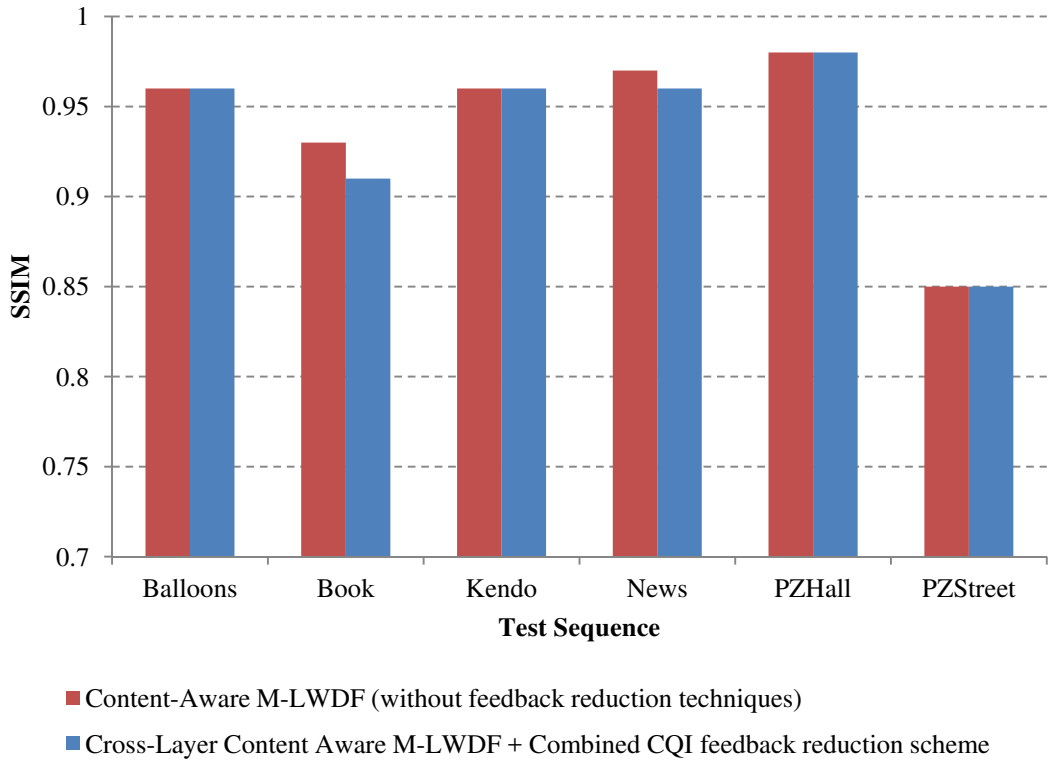

(a)


39
37
35
33
31
29
27
25
PSNR in dB
Balloons
Book
Kendo
News
PZHall
PZStreet
Test Sequence
Content-Aware M-LWDF (without feedback reduction techniques)
Cross-Layer Content Aware M-LWDF + Combined CQI feedback reduction scheme

(b)


1
0.95
0.9
0.85
0.8
0.75
0.7
SSIM
Balloons
Book
Kendo
News
PZHall
PZStreet
Test Sequence
Content-Aware M-LWDF (without feedback reduction techniques)
Cross-Layer Content Aware M-LWDF + Combined CQI feedback reduction scheme

(c)




(d)

Figure 7.18: Visual quality metrics comparative assessment of (a) Base view – PSNR (b) Synthesised view PSNR (c) Base view – SSIM (d) Synthesised view – SSIM

### 7.4.2 Subjective evaluation

Apart from the objective visual quality metrics, used in the previous sections to assess the efficacy of the proposed schemes, the end user experience whilst viewing video content is a pertinent factor. For this reason, a series of subjective evaluations were conducted according to the P.910 ITU-T standard [160] to evaluate the video quality of the MVD video transmission in a crowd event scenario. As the cross-layer content aware scheduling scheme achieves significant uplink feedback signalling reduction whilst exhibiting similar performance, in terms of PSNR and SSIM, to the content aware M-LWDF scheme, the cross-layer scheme together with the standard M-LWDF packet scheduling scheme were selected for subjective evaluation. In order to limit the number of subjective tests, the base and synthesised views of each test sequence were selected for this study. Moreover, as the system level simulations considered a large number of users receiving an MVD video stream, the subjective evaluation was conducted using the impaired test sequences received by two users representing the $25^{th}$ and $50^{th}$ percentile of the PSNR cumulative distribution function (CDF). Thus, each test sequence view is shown five times, in a random order, and includes:

- Original (compressed but without any transmission impairments);
- Standard M-LWDF packet scheduling ($25^{th}$ and $50^{th}$ percentile);
- Cross-layer content-aware M-LWDF packet scheduler with combined feedback reduction technique ($25^{th}$ and $50^{th}$ percentile).

With reference to the P.910 ITU-T standard, the subjective evaluations were carried out by asking participants to rate the quality of the video content based on the Absolute Category Rating with Hidden Reference (ACR-HR) method. In order to have a higher discriminative power the nine-level scale, shown in Table 7.11 has been used for quality rating.

| Quality Rating | Description |
|---|---|
| 9 | Excellent |
| 8 | Good |
| 7 | |
| 6 | Fair |
| 5 | |
| 4 | Poor |
| 3 | |
| 2 | Bad |
| 1 | |

Table 7.11: Nine-level quality rating scale for subjective evaluation

Whilst a minimum number of 15 participants are required for subjective evaluations, a total of 25 participants were enrolled to assess the visual quality of the various video sequences. The collected data was imported into IBM SPSS statistical package, however prior to any further statistical analysis, the participant's opinion scores were analysed in order to identify any outliers which might skew the results. No significant outliers were detected and thus the data gathered from the entire cohort, having the distribution shown in Table 7.12, was considered for the statistical analysis.

| **Participant information** | **Yes** | **No** | | | |
|---|---|---|---|---|---|
| Normal vision acuity | 100% | - | | | |
| Normal colour vision | 100% | - | | | |
| Experience in Video Coding, Transmission and Quality Assessment | 32% | 68% | | | |
| Gender | **Male** | **Female** | | | |
| | 72% | 18% | | | |
| Age Group | **18-24** | **25-34** | **35-44** | **45-54** | **55+** |
| | 64% | 20% | 16% | 0% | 0% |

Table 7.12: Participant distribution

In order to assess the reliability of the collected data, the Cronbach's alpha [69] coefficient is computed for each test sequence view using:

$$\alpha_{CB} = \frac{N\overline{c}}{\overline{v} + (N-1)\overline{c}} \tag{7.8}$$

where $\overline{c}$ is the average inter-item covariance among the items, $N$ is the number of items and $\overline{v}$ is the average variance.

The Cronbach's alpha coefficient measures the consistency of the participants used in the subjective evaluation and although it can take values between negative infinity and 1, any value greater than 0.8 indicates good internal consistency [69]. With reference to Table 7.13(a-b), the cross-layer content-aware M-LWDF with combined CQI feedback reduction improves the mean opinion score (MOS) of both the base view and synthesised view by an average of 0.77 and 0.85 respectively when compared to the standard M-LWDF scheduling scheme. This improvement is in accord with the objective metrics evaluated in the previous section.

| Sequence | Original | Standard M-LWDF | | Cross-Layer content-aware M-LWDF with combined feedback reduction | | Improvement | | Cronbach's Alpha |
|---|---|---|---|---|---|---|---|---|
| | | 25th per. | 50th per. | 25th per. | 50th per. | 25th per. | 50th per. | |
| *Balloons* | 6.83 (±1.08) | 1.89 (±0.90) | 3.06 (±0.80) | 3.16 (±1.01) | 4.61 (±1.08) | 1.27 | 1.55 | 0.972 |
| *Book* | 7.56 (±1.01) | 2.83 (±1.20) | 4.33 (±1.30) | 3.66 (±1.20) | 4.77 (±1.50) | 0.83 | 0.44 | 0.98 |
| *Newspaper* | 7.11 (±1.55) | 1.94 (±0.67) | 3.16 (±0.92) | 3.05 (±0.92) | 4.0 (±1.15) | 1.11 | 0.84 | 0.948 |
| *Poznan Hall* | 7.16 (±1.33) | 3.55 (±1.16) | 3.72 (±0.90) | 3.77 (±1.02) | 4.44 (±1.23) | 0.22 | 0.72 | 0.957 |
| *Poznan Street* | 7.27 (±1.28) | 1.77 (±0.84) | 2.47 (±0.91) | 1.94 (±0.84) | 2.55 (±1.08) | 0.17 | 0.08 | 0.957 |
| *Kendo* | 6.72 (±1.37) | 2.5 (±1.01) | 3.1 (±0.70) | 3.5 (±1.08) | 4.05 (±1.17) | 1 | 0.95 | 0.951 |

(a)

| Sequence | Original | Standard M-LWDF | | Cross-Layer content-aware M-LWDF with combined feedback reduction | | Improvement | | Cronbach's Alpha |
|---|---|---|---|---|---|---|---|---|
| | | 25th per. | 50th per. | 25th per. | 50th per. | 25th per. | 50th per. | |
| *Balloons* | 6.44 (±1.61) | 1.44 (±0.51) | 3.27 (±1.28) | 4.0 (±1.44) | 4.33 (±1.38) | 2.56 | 1.06 | 0.958 |
| *Book* | 6.22 (±1.64) | 1.72 (±1.25) | 3.72 (±1.28) | 3.5 (±1.12) | 3.83 (±1.12) | 1.78 | 0.11 | 0.934 |
| *Newspaper* | 4.22 (±1.31) | 1.55 (±0.82) | 2.22 (±0.89) | 2.77 (±0.74) | 3.16 (±0.66) | 1.22 | 0.94 | 0.874 |
| *Poznan Hall* | 6.66 (±1.22) | 3.55 (±1.26) | 3.61 (±1.05) | 3.5 (±1.08) | 3.72 (±0.75) | -0.05 | 0.11 | 0.970 |
| *Poznan Street* | 6.55 (±1.29) | 1.05 (±0.19) | 2.44 (±0.77) | 1.61 (±0.51) | 2.55 (±1.16) | 0.56 | 0.11 | 0.878 |
| *Kendo* | 5.58 (±1.07) | 2.27 (±0.81) | 3.72(±0.94) | 3.38 (±0.91) | 4.38 (±0.87) | 1.11 | 0.66 | 0.977 |

(b)

Table 7.13: Subjective quality assessment – MOS statistical results for (a) Base view, (b) Synthesised view

## 7.5 Summary

This chapter focused on the design, implementation and testing of a cross-layer CQI feedback adaptation scheme for the MVD video transmission in a crowd event scenario. It started with a comparative analysis of various packet scheduling schemes in a crowd event scenario whilst adopting a periodic wideband CQI feedback scheme. All tested packet scheduling schemes exhibited the same level of performance however the M-LWDF was adopted as the baseline packet scheduling algorithm for the rest of the work given that it was widely adopted in other studies and exhibits relatively lower computational cost.

In order to improve the visual quality experienced by the users receiving an MVD stream, an enhanced M-LWDF scheme based on RLC queue based weighted factors was explored. In this scheme, the MVD video packets are classified according to their relative importance such that the eNB steers the scheduling metric in order to reduce the high priority packet loss. This scheme was tested on various test sequences and whilst the gains in the average PLR is marginal, the HPP PLR was reduced by an average of 1.74% when compared to the M-LWDF and up to 4.3% for the *balloons* test sequence. The improvement in the system level performance metrics was also reflected in the visual quality metrics whereby the texture and synthesised view PSNR quality were improved by 1.13dB and 1.47dB respectively.

The third part of this chapter combined the two feedback reduction schemes, proposed in chapter 3 and evaluated the performance of the content aware M-LWDF scheduling scheme in conjunction with the combined feedback reduction scheme. As expected, tests on the *balloons* test sequence showed that although the combined feedback reduction scheme is able to reduce uplink feedback signalling by 87.5% this comes at the expense of 2dB degradation in the texture and synthesised view PSNR quality. In order to counter the effect of the combined feedback reduction scheme, the CQI feedback was adapted through the use of a cross-layer technique which allows the eNB to solicit CQI reports from individual UEs based on the HPP PLR and weighted HOL delay. The use of the proposed scheme allowed the eNB to tune the CQI feedback such that the texture and synthesised view PSNR quality approaches that of the content

aware M-LWDF whilst achieving an average of 84.1% in CQI feedback signalling reduction. A summary of the key performance indicators for the *balloons* test sequence is shown in Table 7.14(a-b). Moreover, subjective evaluations showed that the proposed cross-layer content-aware M-LWDF scheme coupled with the combined feedback reduction technique yielded a MOS improvement of up to 1.55 and 2.56 in the base view and synthesised view quality respectively when compared to the standard M-LWDF scheme.

| **Packet Scheduler** | **Enhancements** | | **% Feedback Reduction** | **% PLR** | **% HPP PLR** |
|---|---|---|---|---|---|
| | **Combined feedback reduction** | **Cross Layer Adaptation** | | | |
| M-LWDF | × | × | - | 1.61 | 6.99 |
| Content aware M-LWDF | × | × | - | 1.22 | 2.69 |
| | ✓ | × | 87.5 | 1.9 | 4.24 |
| | ✓ | ✓ | 84.6 | 1.46 | 3.25 |

(a)

| **Packet Scheduler** | **Enhancements** | | **Average Texture views PSNR (dB)** | **Average Synthesised view PSNR (dB)** | **Average Texture views SSIM** | **Average Texture views SSIM** |
|---|---|---|---|---|---|---|
| | **Combined feedback reduction** | **Cross Layer Adaptation** | | | | |
| M-LWDF | × | × | 31.88 | 30.75 | 0.92 | 0.93 |
| Content aware M-LWDF | × | × | 33.38 | 32.6 | 0.96 | 0.96 |
| | ✓ | × | 31.48 | 30.6 | 0.91 | 0.92 |
| | ✓ | ✓ | 33.28 | 32.5 | 0.96 | 0.96 |

(b)

Table 7.14: Key performance indicators for *balloons* test sequence

# Chapter 8 Conclusion

The growth in mobile video traffic coupled with the increasing demand for an immersive multimedia experience, is expected to have an impact on the ability of MNO to deliver a superior user experience, given the capacity constraints of present networks. This issue is further accentuated in uplink limited crowd event scenarios due to the higher user propensity to upload multimedia content on social media platforms coupled with excessive uplink signalling overhead. To this end, in order to address this high uplink signalling overhead problem whilst minimising view quality distortion, this thesis investigated the use of lean cross-layer design techniques for the transmission of low latency MVD video content over an LTE network.

## 8.1 Summary of work

The first section of this thesis addressed the high uplink signalling overhead problem through the introduction of two novel schemes. The first scheme initially considered a standard mobility scenario using a UE-assisted predictive filtering technique based on adaptive filters. Both the LMS and NLMS based schemes presented in this work yield a significant feedback reduction in excess of 90% when compared to the 3GPP standardised method however, the LMS based scheme ranks better, with a PLR within 5% of the 3GPP standardised scheme in both low and high mobility scenarios. This represents an improvement in CQI feedback signalling reduction of more than 10% over the work presented in [49]. The proposed scheme exhibits a low computational load and memory footprint. Nevertheless, a reduced complexity scheme, yielding an average computation load reduction of up to 35% when compared to the LMS based scheme, is proposed. However, this comes at the expense of a marginal degradation in the system performance metrics, when compared to the LMS based scheme, and thus it might not be suitable for low loss, low latency applications.

In order to address the crowd event scenario, this work considered a typical sport event scenario whereby users are seated in a three-tier football stadium partitioned into several areas with a number of radio sectors. In such a scenario, a wideband CQI feedback mechanism coupled with extended reporting period is generally adopted by

MNO with the intention to mitigate the prohibitively high signalling overhead. Nevertheless, measurements collected from a live commercial network show that this is usually not enough and thus the LMS based predictive filtering scheme is adapted for the wideband feedback case. System level simulation results indicate the efficacy of the proposed technique with an average CQI feedback reduction of 87% and a marginal degradation in PLR and sector throughput when compared to the 3GPP standard wideband scheme.

The second proposed scheme applies directly to the crowd event scenario whereby the slow varying radio channel condition of static users in the same radio sector is exploited by using a CQI clustering technique. This scheme yields a feedback reduction of 88.2% with only a minor degradation in PLR and sector throughput when compared to the 3GPP standard wideband scheme. Although this scheme ranks better than the wideband LMS based predictive filtering scheme, both schemes could be considered as potential candidates for use in crowd event scenarios. As a result of this uplink signalling overhead reduction, the dimensioning of the PUCCH region could be relaxed thereby allowing more capacity for users in the radio sector to upload data over the PUSCH.

In the second section, the depth-texture view bit allocation problem was addressed through the introduction of a cross-layer MV-HEVC based depth-texture bit rate allocation estimation scheme for an FVV application. Opposed to previous work in this field, the proposed scheme does not require pre-encoding and utilises a formulated statistical model based on both the image characteristics and the total available bit rate for the MVD stream. The proposed scheme is able to estimate the optimal depth-texture ratio for video test sequences, which were not used in the model formulation, with an estimation error of 3.3%. This results in an average synthesised view PSNR improvement of 1.2dB, for the *Newspaper* test sequence, when compared to a commonly used fixed depth-texture rate allocation of 20% [92].

Apart from the depth-texture view bit allocation problem, this thesis investigated whether additional performance gains could be achieved through a custom tuned rate control algorithm for the depth map views. In this work, we exploit the depth-map views characteristics to enhance the standard R-$\lambda$ model based rate control scheme.

These enhancements, which include a refined key frame bit allocation and R-λ model coupled with an adaptive Lagrange multiplier and QP clipping function, yield an average improvement of the synthesised view PSNR and BD-PSNR of 1.15dB and 0.45dB respectively when compared to the standard rate control scheme.

Finally, in the last section, an enhanced content-aware scheduling scheme based on the M-LWDF scheduling algorithm was presented. The enhancement is achieved through the use of RLC queue based weighted factors in the calculation of the scheduling metric together with a relatively low complexity MVD packet prioritisation scheme. System level simulations conducted on various test sequences indicate an improvement in the system level metrics resulting in an up lift in both texture and synthesised view PSNR by an average of 1.13dB and 1.47dB respectively when compared to the standard M-LWDF packet scheduling scheme. As expected, although the combined use of the CQI feedback reduction schemes, presented in the first section of this thesis, is able to reduce the uplink feedback signalling by 87.5%, this comes at the expense of a significant degradation in view quality thereby impacting the user experience. Thus, a lean cross-layer technique was explored in order to adapt the CQI feedback by allowing the eNB to solicit CQI reports from individual UEs based on the HPP PLR and weighted HOL delay metrics. This technique has not only improved the texture and synthesised view PSNR quality, approaching that of the content aware M-LWDF scheme without any CQI feedback reduction applied, but has also achieved an average of 84.1% in uplink feedback signalling overhead reduction.

## 8.2 Future research avenues

Regardless of the fact that the contributions in this thesis address the objectives set out in this research journey, there exist a number of future research avenues that could be explored within the presented cross-layer MVD video transmission framework. These include:

Enriched channel state information feedback and Sounding Reference Signals (SRS)

This work focused on the reduction of uplink signalling caused by CQI feedback. However, in addition, UEs typically report other information including the PMI, RI

and PTI. Moreover, the increased penetration of multi-band carrier aggregation UEs and higher order MIMO increases further the amount of feedback signalling. In view of the channel state correlation that exists in between sub-bands, the application of clustering techniques coupled with lossy or lossless compression schemes presents a possible research avenue. Meanwhile, uplink channel-aware schedulers rely on SRS for efficient uplink resource allocation. Nevertheless, the use of SRS consumes additional uplink resources and thus a study of the proposed techniques in the light of these additional overheads can be carried out.

Energy-aware cluster leader rotation and cluster size adaptation

In the proposed CQI clustering scheme, the cluster leader is elected solely based on its centrality. This implies that a UE might repeatedly get re-elected resulting in a battery energy usage disparity between the UEs. Thus, a cluster leader rotation mechanism coupled with an energy-aware cluster leader election can be investigated. Moreover, in this work, the number of users remained static during the clustering operation, however one would need to consider the case whereby users are added or removed from cluster depending on their activity.

Moving crowd scenario

All the results presented in this thesis assume a typical crowd event scenario whereby users are static. However, this is not always the case as in public demonstrations a crowd can move slowly along an open area. Thus, a CQI dynamic clustering scheme can be investigated in order to re-trigger a clustering operation as the cluster leader become increasingly sub-optimal rather than after a set number of feedback cycles. Nevertheless, such re-clustering needs to seek a balance between system performance, user experience and signalling overhead reduction. Moreover, slow moving crowd scenarios introduce the possibility of handovers between neighbouring radio sectors. Considering that UEs might dwell in the handover zone for an appreciable amount of time, reporting both the CQI and RSRP of the serving and neighbouring sectors, the use of RSRP measurements coupled with sector load measurements to estimate CQI feedback reports of individual or group of UEs can be investigated.

5G Deployment

The emerging standalone (SA) 5G New Radio (NR) specifications, standardised in June 2018, introduces several new network functions including Software Defined Networking (SDN), network slicing and Mobile Edge Computing (MEC). The latter is particularly relevant as it brings both content and intelligence in the radio access network and closer to the UE. Apart from the evident reduction in latency, resulting in improved interactivity, the use of traffic breakout and video caching techniques in the edge nodes lead to lower mobile backhaul traffic. Meanwhile, with sufficient channel state feedback from the UEs, a radio analytics application operating on the mobile edge host might allow a fast channel bandwidth prediction and reporting to the back-end video server. For these reasons, the use of the CQI feedback reduction techniques proposed in this work can also be studied and extended in context of a MEC platform.

Evidently, the potential research avenues indicated above coupled with the deployment of 5G networks, expected in 2019, and the eventual development of a "beyond H.265" standard is expected to deliver a truly immersive multimedia experience in the years ahead.

# References


[1] K. Leswing. (2017). 3D TV is dead [Online]. Available: http://www.businessinsider.com/3d-tv-is-dead-2017-1.

[2] C. Lee, A. Tabatabai, and K. Tashiro, "Free viewpoint video (FVV) survey and future research direction," *APSIPA Transactions on Signal and Information Processing*, vol. 4, p. e15, 2015.

[3] J. Habig. (2016). Is 360 video worth it? [Online]. Available: https://www.thinkwithgoogle.com/advertising-channels/video/360-video-advertising/.

[4] M. Tanimoto, M. P. Tehrani, T. Fujii and T. Yendo, "Free-Viewpoint TV," in *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 67-76, 2011.

[5] R. Suenaga, K. Suzuki, T. Tezuka, M. P. Tehrani, K. Takahashi, and T. Fujii, "A practical implementation of free viewpoint video system for soccer games," in *Three-Dimensional Image Processing, Measurement, and Applications,* 2015.

[6] A. Smolic, "3D video and free viewpoint video - From capture to display," in *Pattern Recognition*, vol. 44, no. 9, pp. 1958-1968, 2011.

[7] B.Bal, G. Dureja, "Hawk eye: a logical innovative technology use in sports for effective decision making," in *Sport Science Review*, vol. 21, no. 1-1, pp. 107– 119, 2012.

[8] Canon. (2017). Canon announces development of the Free Viewpoint Video System virtual camera system that creates an immersive viewing experience [Online]. Available: https://www.canon-europe.com/press-centre/press-releases/2017/09/free-viewpoint-video-system/.

[9] G. Lafruit, M. Domanski, K. Wegner, et al, "New visual coding exploration in MPEG: Super-Multiview and Free Navigation in Free viewpoint TV," in *Proc. of the Electronics Imaging Conf.: Stereoscopic Displays and Application*, 2016, pp. 1-9.

[10] Cisco. (2017). Cisco visual networking index forecast report 2016-2021 [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html.

[11] Bain Insights. (2017). Mobile Operators: How to keep up with customer demand [Online]. Available: https://www.forbes.com/sites/baininsights/2017/08/30/mobile-operators-how-to-keep-up-with-customer-demand/#1640a87a5ad0.

[12] A. Vetro, S. Uea, and A. Smolic, “Towards a 3D video format for auto-stereoscopic displays”, in *Proc. Of SPIE Conference on Applications of Digital Image Processing XXXI*, vol. 7073, Sep. 2008.

[13] 3GPP TS36.300 V8.12.0 (2010), “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) - Overall Description”.

[14] 3GPP TS36.323 V8.6.0 (2009), “Evolved Universal Terrestrial Radio Access (E-UTRA) – PDCP”.

[15] K. Sandlund, G. Pelletier, L. Jonsson, "RObust Header Compression (ROHC): Requirements on TCP/IP Header Compression", RFC 5795, March 2010.

[16] 3GPP TS36.332 V8.7.0 (2009), “Evolved Universal Terrestrial Radio Access (E-UTRA) – Radio Link Control (RLC) Protocol specification”.

[17] A. Elnashar, M. A. El-saidny, M.R. Sherif, “LTE Air Interface and Procedures,” in *Design, Deployment and Performance of 4G LTE Networks – A Practical Approach*, 1st ed., West Sussex: John Wiley & Sons Ltd., 2014, pp. 47-101.

[18] 3GPP TS 36.321 V8.8.0 (2009), “Evolved Universal Terrestrial Radio Access (E-UTRA) – Medium Access Control (MAC) Protocol Specification”.

[19] S. Sesia, I. Toufik, M. Baker, *The UMTS Long Term Evolution – From Theory to Practice*. West Sussex: John Wiley & Sons Ltd., 2009.

[20] 3GPP TS36.213 V13.0.0 (2016), “Evolved Universal Terrestrial Radio Access (E-UTRA) – Physical Layer Procedures”.

[21] F. Khan. *LTE for 4G Mobile Broadband – Air Interface Technologies and Performance*. Cambridge: Cambridge University Press, 2009.

[22] G. Berardinelli, L. Ruiz de Temino, S. Frattasi, M. Rahman, and P. Mogensen, “OFDMA vs. SC-FDMCA: Performance comparison in local area IMT-A scenarios," in *Wireless communications*, IEEE, vol. 15, pp. 64-72, Oct. 2008

[23] H. Holma, A. Toskala, “Introduction to OFDMA and SC-FDMA and to MIMO in LTE,” in *LTE for UMTS-OFDMA and SC-FDMA based radio access*, 1st ed., West Sussex: John Wiley & Sons Ltd., 2009.

[24] 3GPP, R1-061777, “DCT based CQI reporting scheme”, LG Electronics, RAN1 LTE Ad Hoc, Cannes, France, June, 2006.

[25] 3GPP, R1-070187, “DCT partitioning for CQI reporting”, Panasonic, RAN1 Meeting 47, Sorrento, Italy, January, 2007.

[26] A. Haghighat, G. Zhang and Z. Lin, “Full-band CQI feedback by Haar compression in OFDMA systems,” in *Proc. of the IEEE 70th Vehicular Technology Conference Fall* (*VTC 2009-Fall*), September 2009, pp. 1-5.

[27] M. Kang and K. S. Kim, “Performance Analysis and Optimization of Best-M Feedback for OFDMA Systems,” in *IEEE Communications Letters*, vol. 16, no. 10, pp. 1648-1651, October 2012.

[28] H. H. Alyazidi and I. Kostanic, “OFDMA feedback optimization in 4G-LTE systems,” *IEEE 17th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, Barcelona, 2012, pp. 70-74.

[29] S. N. Donthi and N. B. Mehta, “Performance analysis of subband-level channel quality indicator feedback scheme of LTE,” in *National Conference on Communications (NCC)*, Chennai, 2010, pp. 1-5.

[30] M. Pauli, U. Wachsmann, and S. Tsai, “Quality determination for a wireless communications link”, U.S. Patent Office publication, Patent (US 2004/0219883, November 2004.

[31] E. Westman, “Calibration and evaluation of the exponential effective SINR mapping (EESM) in 802.16,” Master’s thesis, The Royal Institute of Technology (KTH), Stockholm, Sweden, September 2006.

[32] C. Fa-tang and T. Gen-lin, “A novel MCS selection criterion for supporting AMC in LTE system,” in *Proc. of the International Conference on Computer Application and System Modelling* (*ICCASM*), 2010, vol. 6, pp. 598-603.

[33] H. Dai, Y. Wang, C. Shi and W. Zhang, “The evaluation of CQI delay compensation schemes based on Jake’s model and ITU scenarios,” in *Proc. of the IEEE 76th Vehicular Technology Conf.* (*VTC 2012-Fall)*, September 2012, pp. 1-5.

[34] T. Werthmann, M. Kaschub and M. Proebster, “Simple channel predictors for lookahead scheduling,” in *Proc. of the IEEE 75$^{th}$ Vehicular Technology Conf.* (*VTC 2012-Spring)*, May 2012, pp.1-6.

[35] X. Xu, M. Ni, R. Mathar, “Improving QoS by predictive channel quality feedback for LTE,” in *Proc. Int. Conf. Software, Telecom. and Computer Networks*, September 2013, pp. 1-5.

[36] M. Ni, X. Xu and R. Mathar, "A channel feedback model with robust SINR prediction for LTE systems," in *Proc. Euro. Conf. on Ant. and Prop. (EuCAP)*, Gothenburg, 2013, pp. 1866-1870.

[37] M. A. Awal and L. Boukhatem, "Cross-layer CQI feedback window adaptation for OFDMA systems under feedback budget constraint," *in Proc of the IEEE Symposium on Computers and Communications (ISCC)*, Kerkyra, 2011, pp. 242-247.

[38] M.A. Awal and L. Boukhatem, "Effect of feedback prediction on OFDMA system throughput," in *Proc. 5th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2009.

[39] T. Ekman, "Prediction of mobile radio channels," Ph.D. dissertation, Uppsala University, 2002.

[40] H. Oh and H. M. Kim, "An adaptive determination of channel information feedback period in OFDMA systems," in *10th International Symposium on Consumer Electronics (ISCE)*, 2006.

[41] S. Iijima, R. Zhou and I. Sasase, "Reduction of feedback overload by exploiting adaptive channel in OFDMA systems," in *Pacific Rim Conference on Communications, Computers and Signal Processing*, August 2007.

[42] D. Gesbert, M.S. Alouini, "How much feedback is multi-user diversity really worth?," in *Proc. of Int. Conf. Comm.*, June 2004, pp. 234–238.

[43] L. Sivridis, X. Wang, and J. Choi, "Radio resource management for fast fading environments," *Journal of Comm.*, vol. 6, issue 7, pp. 540-548, October 2011.

[44] K. Hoon, and H. Youngnam, "An opportunistic channel quality feedback scheme for proportional fair scheduling," *IEEE Communications Letter*, vol. 11, issue 3, June 2007, pp. 501–503.

[45] S. Kittipiyakul, A. Boonkajay, and T. Aphichartsuphapkhajorn, "Performance of low-feedback rate, gradient-based OFDMA subcarrier allocation with partial channel information," in *Proc. Symp. Wireless Perv. Comm.*, May 2010, pp. 511–521.

[46] J.Jeon, K. Son, H.Lee, and S. Chong, "Feedback reduction for multiuser OFDM systems," in *Trans. Vehicular Tech.*, vol. 59, issue 1, Jan. 2010, pp. 160–169.

[47] L. Sivridis, J. He, "A strategy to reduce the signaling requirements of CQI feedback schemes," in *Journal Wireless Pers. Comm.*, vol. 70, issue 1, May 2013, pp. 85-98.

[48] M. Abdulhasan, M. Salman, C. Ng, N. Noordin, S. Hashim, and F. Hashim, "An Adaptive Threshold Feedback Compression Scheme Based on Channel Quality Indicator (CQI) in Long Term Evolution (LTE) System," in *Journal Wireless Pers. Comm.*,vol. 82, issue 4, pp. 2323-2349, June 2015.

[49] A. Chiumento, M. Bennis, C. Desset, L. Van der Perre, and S. Pollin, "Adaptive CSI and feedback estimation in LTE and beyond: a Gaussian process regression approach," in *EURASIP Journal Wireless Comm. Networking*, Dec. 2015.

[50] A. Chiumento, C. Desset, S. Pollin, L. Van der Perre, and R. Lauwereins, "Impact of CSI Feedback Strategies on LTE Downlink and Reinforcement Learning Solutions for Optimal Allocation," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, Jan. 2017, pp. 550-562.

[51] P. Legg and G. Peters, "Management of channel quality reporting in highly loaded LTE networks," in *IEEE International Conference on Communications (ICC)*, London, 2015, pp. 1892-1896.

[52] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska and M. Rupp, "Simulating the Long Term Evolution physical layer," in *Proc. European Signal Processing Conference*, Glasgow, 2009, pp. 1471-1478.

[53] J. C. Ikuno, M. Wrulich and M. Rupp, "System Level Simulation of LTE Networks," in *IEEE Vehicular Technology Conference*, Taipei, 2010, pp. 1-5.

[54] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice Hall, 2004.

[55] G. Piro, L. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open source framework", in *IEEE Transactions Vehicular Technology*, vol. 60, no.2, February 2011.

[56] "IEEE Standard for Floating-Point Arithmetic," in *IEEE Std. 754-2008*, vol., no., pp.1-70, Aug. 29 2008.

[57] M. Lotfizad, and H.S. Yazdi, "Modified Clipped LMS Algorithm," in *EURASIP Journal on Adv. Sig. Proc.*, 2005.

[58] Rohde & Schwarz. (2018). Swissqual Field Test Equipment [Online]. Available:http://www.swissqual.com/en/products/benchmarking/diversity-benchmarker-2/

[59] ETSI TR 102.581, "Speed Processing, Transmission and Quality Aspects (STQ); A study on the Minimum Additional Required Attenuation on the Antenna Path of the Field Test Equipment", 2007.

[60] M. Cordina, C.J. Debono, "Robust Predictive Filtering Schemes for Sub-band CQI Feedback Compression in 3GPP LTE Systems," IET Communications, vol. 11, no. 11, pp. 1797 - 1807, September 2017.

[61] E-Architect. (2017). Manchester United Stadium: Old Trafford Ground [Online]. Available: https://www.e-architect.co.uk/manchester/manchester-united-stadium.

[62] FIFA. (2017). FIFA-football stadiums Technical recommendation and requirements [Online]. Available: https://www.scribd.com/doc/100501692/FIFA-Football-Stadiums-Technical-recommendation-and-requirements-5th-edition.

[63] Pordata. (2017). Market share of leading operator in mobile telecommunications [Online]. Available: http://www.pordata.pt/en/Europe/Market+share+of+leading+operator+in+mobile+telecommunications-1502.

[64] LiveAnalytics. (2017). Mobile behaviour and viewing technology – May 2014 [Online]. Available: https://slideshare.net/LiveAnalytics/mobile-us-final

[65] Commscope. (2017). Commscope Stadium Antenna Model: CNLPX3055F [Online]. Available: http://www.commscope.com

[66] P. Kapustka. (2017). Verizon: U.S. Bank Stadium DAS already seeing more traffic than Super Bowl 51 [Online]. Available: https://www.mobilesportsreport.com/2017/12/verizon-u-s-bank-stadium-das-already-seeing-more-traffic-than-super-bowl-51/

[67] M. Cordina, and C. Debono, "A CQI feedback clustering technique for signalling traffic reduction during crowd events in 3GPP LTE systems," in *Proc. Int. Symp. On Pers., Indoor, and Mobile Radio Comm. (PIMRC'17)*, Canada, October, 2017.

[68] G. Hamerly, C. Elkan, "Learning the *k* in *k*-Means," in *Proc. Int. Conf. Neural Info. Proc. Sys.*, Canada, 2003, pp. 281-288.

[69] M. Norusis. *IBM SPSS Statistics 19 Advanced Statistical Procedures Companion*. New Jersey: Prentice Hall, 2011.

[70] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[71] M. T. Pourazad, C. Doutre, M. Azimi and P. Nasiopoulos, "HEVC: The New Gold Standard for Video Compression: How Does HEVC Compare with H.264/AVC?," in *IEEE Consumer Electronics Magazine*, vol. 1, no. 3, pp. 36-46, July 2012.

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

[73] S. Wenger and M. Horowitz, "FMO: Flexible macroblock ordering," in JVT-C089, May 2002.

[74] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth and M. Zhou, "An Overview of Tiles in HEVC," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 969-977, Dec. 2013.

[75] A. Fuldseth, M. Horowitz, S. Xu, A. Segall, and M. Zhou, "Tiles," in JCTVC-F335, July 2011.

[76] Brian, Marshall. (2017). How 3-D Glasses Work [Online]. Available: http://science.howstuffworks.com/3-d-glasses2.htm

[77] A. Vetro, "Frame Compatible Formats for 3D Video Distribution," in *IEEE International Conference on Image Processing (ICIP)*, September 2010, pp. 2405-2408.

[78] C. Theobalt, G. Ziegler, M. Magnor, H. Seidel, "Model-based free-viewpoint video: Acquisition rendering and encoding" in *Proc. Picture Coding Symp.*, vol. 5, 2004, pp. 1-6.

[79] S. Wrmlin, E. Lamboray, M. Gross, "3D video fragments: Dynamic point samples for real-time free-viewpoint video," in *Computers and Graphics Special Issue on Coding Compression and Streaming Techniques for 3D and Multimedia Data*, vol. 28, no. 1, 2004, pp. 3-14.

[80] A. Vetro, A. M. Tourapis, K. Muller and T. Chen, "3D-TV Content Storage and Transmission," in *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 384-394, June 2011.

[81] A. Vetro and D. Tian, "Analysis of 3D and multiview extensions of the emerging HEVC standard," in *SPIE Optical Engineering Applications,* 2012.

[82] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro and Y. K. Wang, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35-49.

[83] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multiview video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, pp. 201-204.

[84] G. Tech, K.Wegner, Y. Chen, M. M. Hannuksela, and J. Boyce, “MV-HEVC draft text 5,” in Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-E1004, 5th Meeting: Vienna, Aug., 2013.

[85] L. Zhang, Y. Chen, and M. Karczewicz, “Disparity vector based advanced inter-view prediction in 3D-HEVC,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 1632–1635.

[86] H. Liu, J. Jung, J. Sung, J. Jia, and S. Yea, “3D-CE2.h: Results of illumination compensation for inter-view prediction,” in Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3VB0045, 2nd Meeting: Shanghai, China, Oct., 2012.

[87] L. Zhang, Y. Chen, and L. Liu, “3D-CE5.h: Merge candidates derivation from disparity vector,” in Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-B0048, 2nd Meeting: Shanghai, China, Oct., 2012.

[88] L. Zhang, Y. Chen, X. Li, and M. Karczewicz, “CE4: Advanced residual prediction for multiview coding,” in Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-D0117, 4th Meeting: Incheon, Korea, Apr., 2013.

[89] K.Müller, P.Merkle, G. Tech, and T. Wiegand, “3D video coding with depth modelling modes and view synthesis optimization,” in *Proc. Asia-Pacific Signal and Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Oct. 2012.

[90] A. Vetro, Y. Chen and K. Mueller, “HEVC-Compatible extensions for advanced coding of 3D and multiview video," in *2015 Data Compression Conference*, Snowbird, UT, 2015, pp. 13-22.

[91] E. Bosc, F. Racape, V. Jantet, P. Riou, M. Pressigout and L. Morin, “A study of depth/texture bit-rate allocation in multi-view video plus depth compression,” in *Ann. Telecommun.*, pp. 615-625, Dec. 2013.

[92] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV,” in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems*, 2004, vol. 5291.

[93] K. Klimaszewski, K. Wegner and M. Domanski, “Distortions of synthesized views caused by compression of views and depth maps,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Potsdam, 2009, pp. 1-4.

[94] Y. Morvan, D. Farin, and P. De With, “Joint depth/texture bit-allocation for multi-view video compression”, in *Proc. of Picture Coding Symposium*, Nov. 2007.

[95] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model," in *Signal Proc.: Image Comm.*, vol. 24, no. 8, pp. 666-681, Sep. 2009.

[96] G. Cheung, V. Velisavljevic and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," in *IEEE Trans. on Image Proc.*, vol.20, no.11, Nov. 2011.

[97] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding", in *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 21, no. 4, Apr. 2011.

[98] S. Ma, W. Gao, and Y, Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," in *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.

[99] Y. Liu, Q. Huang, S. Ma, D. Zhao,W. Gao, S. Ci, H. Tang, "A novel rate control technique for multiview video plus depth based 3D video coding," in *IEEE Trans. Broadcast.*, vol. 57, no 2, pp. 562–571,2011.

[100] F. Shao, G. Jiang, W. Lin, M. Yu, Q. Dai, "Joint bit allocation and rate control for coding multi-view video plus depth based 3D video", in *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1843–1854, 2013.

[101] Y. Harshalatha, and B. Prabir, "SSIM-based joint-bit allocation for 3D video coding," in *Proc. of Multimedia Tools and Applications*, Nov. 2017, pp. 1-19.

[102] G. Zhu, G. Jiang, M. Yu, F. Li, F. Shao and Z. Peng, "Joint video/depth bit allocation for 3D video coding based on distortion of synthesized view," in *IEEE international Symposium on Broadband Multimedia Systems and Broadcasting*, Seoul, 2012, pp. 1-6.

[103] C. Yang, P. An, L. Shen, "Adaptive bit allocation for 3D video coding," in *Circuits, systems, and signal processing*, vol. 36, no. 5, pp 2102-2124, 2016.

[104] M. Cordina, and C. Debono, "An adaptive texture-depth rate allocation estimation technique for low latency multi-view video plus depth transmission," in *Proc. of the IEEE Visual Comm. and Image Proc.*, Nov. 2013.

[105] K. Klimaszewski, K. Wegner, and M. Domanski, "Video and depth bitrate allocation in multiview compression," in *Int. Conf. on Systems, Signals and Image Proc.*, May 2014.

[106] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin and S. Rahardja, "*Adaptive Basic Unit Layer Rate Control for JVT"*, JVT-G012r1, Pattaya, Thailand, Mar. 2003.

[107] H. Choi, J. Nam, J. Yoo and D. Sim, “Rate control based on unified RQ model for HEVC,” in JCT-VC H0213, 8th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC29/WG 11, San Jose, CA, USA, 2012.

[108] B. Li, H. Li, L. Li and J. Zhang, “Rate control by R-lambda model for HEVC,” in JCT-VC K0103, 11th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Shanghai, CN, 2012.

[109] J. Si, S. Ma and W. Gao, “Adaptive rate control for HEVC,” in JCT-VC-I0433, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, CH, 2012.

[110] L. Sun, O. Au and W. Dai, “An adaptive frame complexity based rate-quantization model for intra-frame rate control of high efficiency video coding (HEVC),” APSIPA ASC, 2012, 1–6.

[111] B. Li, D. Zhang, H. Li, and J. Xu, “QP determination by lambda value”, in JCTVC-I0426, Geneva, CH, 27 April - 7 May 2012.

[112] X. Wang and M. Karczewicz, “Intra frame rate control based on SATD,” in JCT-VC M0257, Incheon, 2013.

[113] M. Wang, K. Ngi Ngan and H. Li, “An efficient frame-content based Intra frame rate control for high efficiency video coding,” in *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 896-900, July 2015.

[114] J. Si, S. Ma and W. Gao, “Efficient bit allocation and CTU level rate control for high efficiency video coding,” in *Proc. of Picture Coding Symposium*, pp.89-92, Dec. 2013.

[115] J. Si, S. Ma and W. Gao, “Laplace distribution based CTU level rate control for HEVC,” in Proc. of Visual Communications and Image Processing, pp. 1-6, Nov. 2013.

[116] B. Lee, M. Kim, and T. Q. Nguyen, “A frame-level rate control scheme based on texture and non-texture rate models for high efficiency video coding,” in *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 24, pp. 465–479, Mar. 2014.

[117] M. Xu, X. Deng, S. Li, and Z. Wang, “Region-of-Interest based conversational HEVC coding with hierarchical perception model of face,” *IEEE J. Sel. Topics Signal Process.*, vol.8, no.3, pp. 475-489, Jun. 2014.

[118] “HEVC reference software,” [Online]. Available: http:// hevc.kw.bbc.co.uk/trac/browser/jctvc-htm/tags/HM-12.0

[119] W. Lim, and S. Sooyoun, “Multi-view Rate Control based on HEVC for 3D Video Services,” in *Journal of the Inst. Of Elect. And Info. Engineers*, vol. 8, no. 8, 2013, pp. 245-249.

[120] A. Fiengo, G. Chierchia, M. Cagnazzo and B. Pesquet-Popescu, "Convex optimization for frame-level rate allocation in MV-HEVC," in *IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, 2016, pp. 2157-2161.

[121] W. Lim, D. Sim and I. V. Bajić, “JCT3V – Improvement of the rate control for 3D multi-view video coding,” ISO/IEC JTC1/SC29/WG11, JCT3V-C0090, Geneva, Switzerland, Jan. 2013.

[122] S. Tan, J. Si, S. Ma, S. Wang, W. Gao, “Adaptive Frame Level Rate Control in 3D-HEVC,” in *Visual Communication and Image Processing (VCIP)*, Malta, Dec. 2014.

[123] X. Wang, S. Kwong, H. Yuan, Y. Zhang, Z. Pan, “View synthesis distortion model based frame level rate control optimisation for multiview depth video coding,” in *Journal of Sig. Proc.*, vol. 112, iss. C, July 2015, pp. 189-198.

[124] J. Lei, Z. Li, T. Zhu, X. He, L. You, C. Hou, “Region-based bit allocation and rate control for depth video in HEVC,” in *Journal of Mult. Tools and App.*, vol. 76, no. 3, February 2017, pp. 4179-4195.

[125] Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG, “Common test conditions of 3DV core experiments,” JCT3V-G1100, San Jose, USA, Tech. Rep., Jan. 2014.

[126] J. F. Canny, “A computational approach to edge detection,” in *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679- 697, 1986.

[127] M. Cordina, C.J. Debono, “A Depth Map Rate Control Algorithm for HEVC Multi-view Video plus Depth,” *Proc. of the IEEE International Conference on Multimedia and Expo (ICME 2016) - 22nd International Packet Video Workshop (PV)*, July 2016.

[128] G. Bjøntegaard, “Calculation of average PSNR difference between RD-curves,” Doc. VCEG-M33, Apr. 2001

[129] M. Budagavi, G. Sullivan, V. Sze, High Efficiency Video Coding (HEVC): Algorithms and Architectures. Springer, 2014.

[130] T. Schierl, M. M. Hannuksela, Y. K. Wang and S. Wenger, “System Layer Integration of High Efficiency Video Coding,” in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1871-1884, Dec. 2012.

[131] E. Thomas, M. O. van Deventer, T. Stockhammer, A. C. Begen, M. L. Champel and O. Oyman, “Application of SAND Technology in DASH-Enabled Content Delivery Networks and Server Environments,” in *SMPTE Motion Imaging Journal*, vol. 127, no. 1, pp. 48-54, Feb. 2018.

[132] D. Yun and K. Chung, “DASH-based Multi-view Video Streaming System,” in *IEEE Trans. Circuits Syst. for Video Technol.*, vol. PP, no. 99, pp. 1-1.

[133] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications, IETF STD 0064, RFC 3550 [Online]. Available: http://tools.ietf.org/html/rfc3550

[134] D. Comer, Internetworking with TCP/IP: Principles, Protocols, and Architectures – Volume 1. New Jersey: Prentice Hall, 2000.

[135] 3GPP TS23.203 V12.6.0 (2014), “Technical Specification Group services and system aspects: Policy and charging control architecture”.

[136] S. Sulthana, R. Nakkeeran, “Study of downlink scheduling algorithm in LTE networks," in *Journal of Networks*, vol. 9 no. 12 pp. 3381-3391 Dec. 2014.

[137] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen and M. Moisio, “Dynamic packet scheduling performance in UTRA Long Term Evolution downlink,” in *3rd International Symposium on Wireless Pervasive Computing (ISPWC)*, Santorini, 2008, pp. 308-313.

[138] R. Kwan, C. Leung, and J. Zhang, “Proportional Fair Multiuser Scheduling in LTE,” in *IEEE Signal Process. Letters*, vol. 16, no. 6, pp. 461-464, June 2009.

[139] C. Wengerter, J. Ohlhorst, and A. von Elbwart, “Fairness and Throughput Analysis for Generalised Proportional Fair Frequency Scheduling in OFDMA,” in *Proc. of IEEE Vehicular Technology Conference Spring (VTC 2005-Spring)*, Stockholm, Sweden, vol. 3, pp. 1903 – 1907, May 2005.

[140] J. Rhee, J. Holtzman, and D. K. Kim, “Scheduling of Real/non-real time services: Adaptive EXP/PF Algorithm,” in *Proc. of IEEE Vehicular Technology Conference (VTC 2003 –Spring)*, Jeju, Korea, vol.1, pp. 462 – 466, Apr. 2003.

[141] B. Sadiz, R. Madan, and A. Sampath, “Downlink Scheduling for Multiclass Traffic in LTE,” in *Eurasip Journal. Of Wireless Comm. Net*, vol. 2, pp. 9-13, Oct. 2009.

[142] G. Piro, L. A. Grieco, G. Boggia, R. Fortuna and P. Camarda, “Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks,” in *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 1052-1065, Oct. 2011.

[143] P. Ameigeiras, J, Wigard, and P. Mogensen, “Performance of the M-LWDF scheduling algorithm for streaming services in HSDPA,” in *Proc. of IEEE Vehicular Technology Conference (VTC 2004 –Fall)*, Los Angeles, USA, vol. 2, 2004, pp. 462-466.

[144] M. Iturralde, T. A. Yahiya, A. Wei, and A. Beylot, “Performance Study of Multimedia Services using virtual token mechanism for resource allocation in LTE networks,” in *Proc. of IEEE Vehicular Technology Conference (VTC 2011 – Fall)*, San Francisco, USA, 2011, pp. 1-5.

[145] D. Nguyen, H. Nguyen, E. Renault, “Performance evaluation of E-MQS scheduler with mobility in LTE heterogeneous network,” in *IEEE Int. Conf. on Comm.*, France, May 2017, pp. 1-6.

[146] H. Rashic Chayon, K. Bin Dimyati, H. Ramiah, and A. Wasif Reza, “Enhanced Quality of Service of Cell-edge user by Extending Modified Largest Weighted Delay First Algorithm in LTE Networks,” in *Symmetry*, vol. 9, 2017, pp. 81-95.

[147] Y. Farzaneh and A. Mardi, “A QoS-Aware Downlink Packet Scheduler using Token Bucket Algorithm for LTE Systems,” in *22nd Iranian Conf. on Elect. Eng. (ICCE)*, 2014, pp. 1775-1780.

[148] E. Skondras, A. Michalas, A. Sgora and D. D. Vergados, “A downlink scheduler supporting real time services in LTE cellular networks," in *6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Corfu, 2015, pp. 1-6.

[149] T. Efimushkina, M. Gabbouj, “A cross-layer adaptation based video downlink transmission over LTE: Survey,” in *Computer and Info. Science*, vol. 279, 2014.

[150] R. Zhang, S. L. Regunathan and K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” in *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966-976, June 2000.

[151] X. Cheng and P. Mohapatra, “Quality-optimized downlink scheduling for video streaming applications in LTE networks,” in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, 2012, pp. 1914-1919.

[152] H. Luo, S. Ci, D. Wu, J. Wu and H. Tang, “Quality-driven cross-layer optimized video delivery over LTE,” in *IEEE Communications Magazine*, vol. 48, no. 2, pp. 102-109, February 2010.

[153] B. Tirouvengadam, R. Radhakrishnan and A. Nayak, “CAAHR: Content aware adaptive HARQ retransmission scheme for 4G/LTE network,” in *Fourth International Conference on Ubiquitous and Future Networks (ICUFN)*, Phuket, 2012, pp. 456-461.

[154] A. Jassal, B. Oztas, M. T. Pourazad and P. Nasiopoulos, “A packet prioritization scheme for 3D-HEVC content transmission over LTE networks,” in *IEEE International Conference on Communication Workshop (ICCW)*, London, 2015, pp. 1788-1793.

[155] B. Oztas, M. Pourazad, P. Nasiopoulos, V. Leung, “A media-aware transmission framework for 3D-HEVC over LTE Networks,” in *Proc. of the 6th Symp. On Devel. And Anal. Of Intell. Vehic. Net. and Appl. (DIVANet)*, Malta, 2016.

[156] G. Piro, C. Ceglie, D. Striccoli and P. Camarda, “3D Video transmissions over LTE: A performance evaluation,” *Eurocon 2013*, Zagreb, 2013, pp. 177-185.

[157] M. M. Nasralla, C. T. E. R. Hewage and M. G. Martini, “Subjective and objective evaluation and packet loss modelling for 3D video transmission over LTE networks," in *International Conference on Telecommunications and Multimedia (TEMU)*, Heraklion, 2014, pp. 254-259.

[158] C. Ceglie, G. Piro, D. Striccoli, “Performance evaluation of 3D video streaming services in LTE-Advanced networks,” in Journal of Wireless Networks, vol. 20, no. 8, pp. 2255-2273, Nov. 2014.

[159] P. Gao and W. Xiang, “Modeling of packet-loss-induced distortion in 3-D synthesized views,” in *Visual Communications and Image Processing (VCIP)*, Singapore, 2015, pp. 1-4.

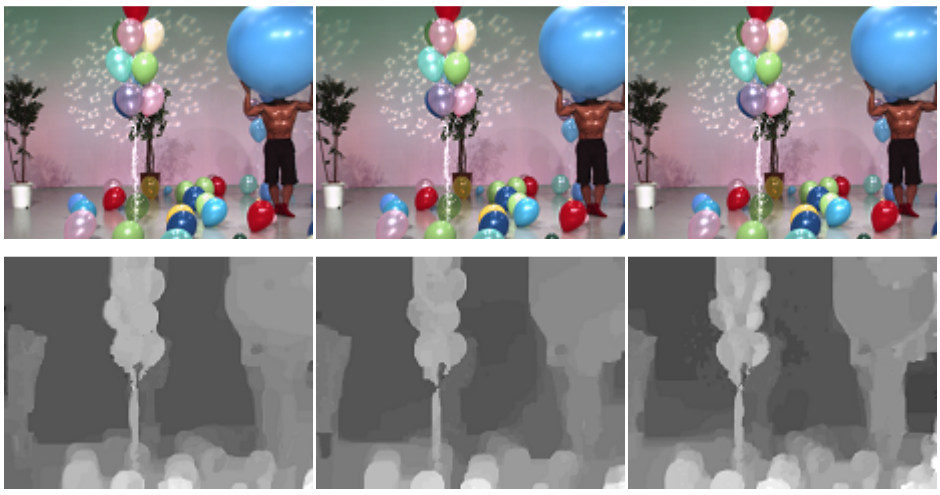
[160] International Telecommunications Union. (2016). P.910: Subjective video quality assessment methods for multimedia applications [Online]. Available: https://www.itu.int/rec/T-REC-P.910/en

# Appendix A: Test Video Sequences

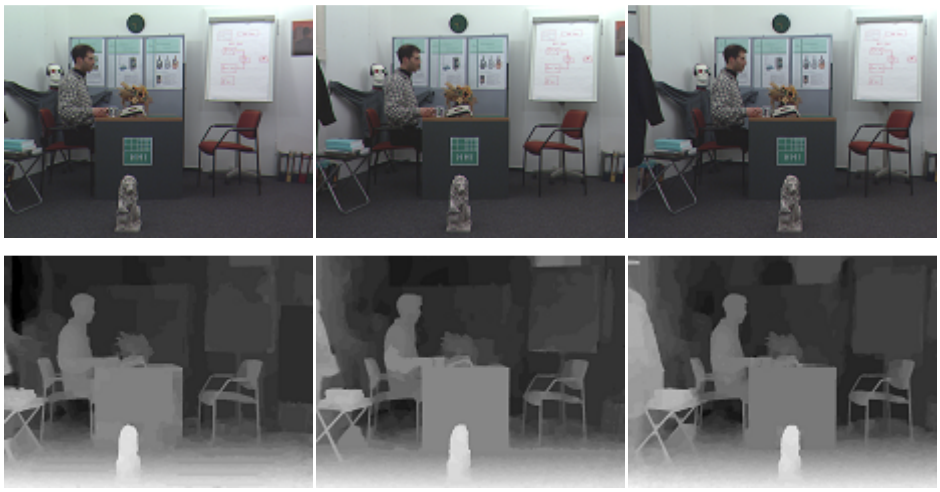
In this thesis, a series of MVD video test sequences have been used to evaluate the performance of the proposed schemes. These short MVD video test sequences are widely used by the standardisation bodies and academia in the study of multimedia systems. The key characteristics and screenshots of the MVD test sequences used in this thesis are shown in Table A.1 and Figure A.1 (a-f) respectively.

| Test Sequence | Provider | Resolution | Frame Rate | Baseline/Camera Array | Cameras (L-C-R) |
|---|---|---|---|---|---|
| Balloons | Nagoya University | 1024 x 768 | 30 | 5cm / 1D | 1-3-5 |
| Book Arrival | Fraunhofer-HHI | 1024 x 768 | 15 | 6.5cm / 1D | 6-8-10 |
| Kendo | Nagoya University | 1024 x 768 | 30 | 5cm / 1D | 1-3-5 |
| Newspaper | Gwangju Institute of Science and Technology (GIST) | 1024 x 768 | 30 | 5cm / 1D | 2-4-6 |
| Poznan Hall | Poznan University | 1920 x 1088 | 25 | 13.75cm / 1D | 5-6-7 |
| Poznan Street | Poznan University | 1920 x 1088 | 25 | 13.75cm / 1D | 3-4-5 |

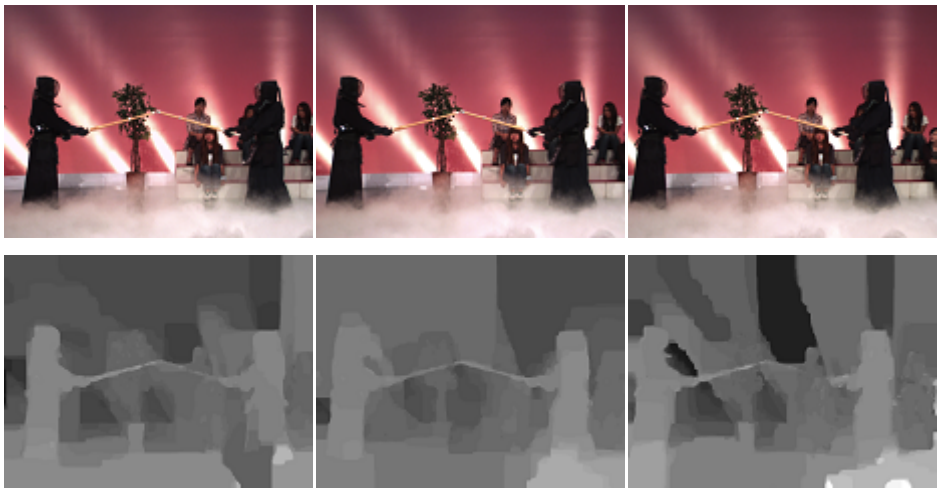
Table A.1: Test Video Sequences Characteristics (L - Left, C - Centre, R – Right)



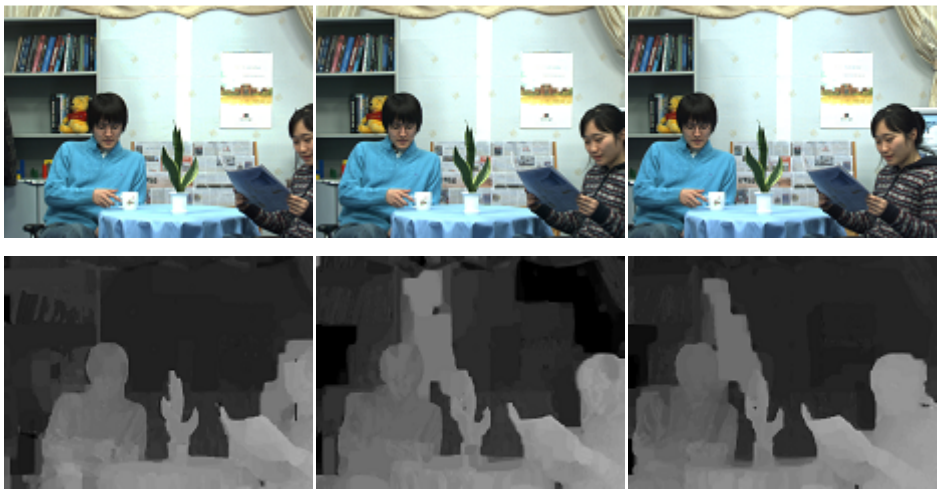
(a)



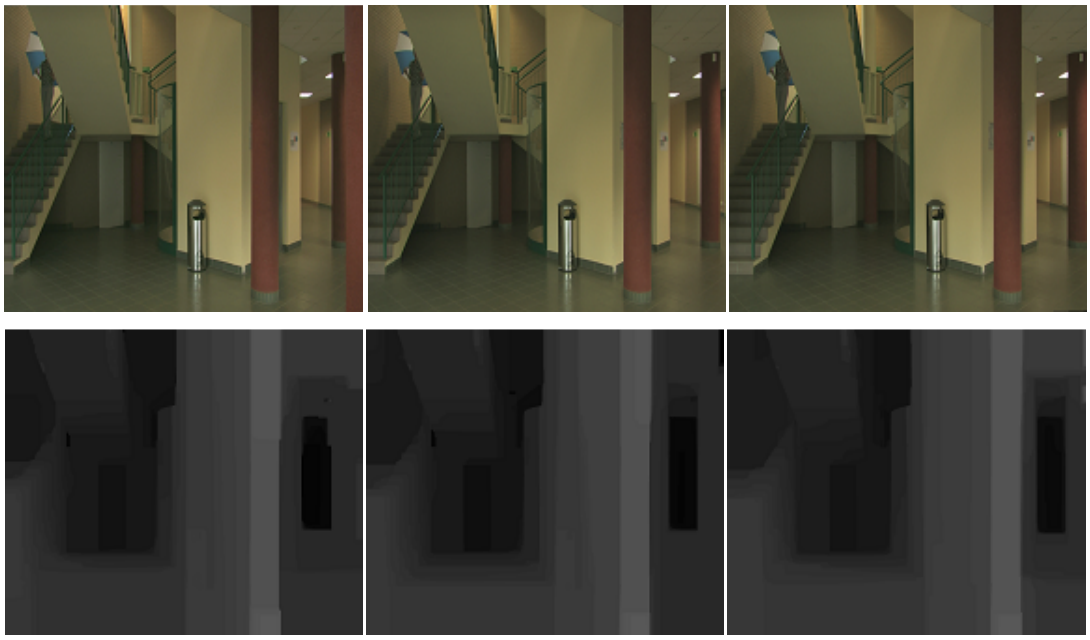
(b)



(c)



(d)



(e)

(f)

Figure A.1: MVD Test sequence screenshots (Left – Centre – Right) (a) Balloons, (b) Book Arrival, (c) Kendo, (d) Newspaper, (e) Poznan Hall, (f) Poznan Street

# Appendix B: Subjective Quality Evaluation Questionnaire

# Cross-Layer Design for Multi-view Video plus Depth Transmission over LTE Networks in Crowd Event Scenarios – Subjective Testing

UNIVERSITY OF MALTA
L-Università ta' Malta

**Date: _____**

Dear Sir/Madam,

Thank you for participating in this experiment relating to my research, "Cross-Layer Design for Multi-view Video plus Depth Transmission over LTE Networks in crowd event scenarios". Kindly fill in the below by encircling one of the available options before starting the assessment.

**Age Group:** 18-24 25-34 35-44 45-54 55+

**Gender:** M / F

**Normal visual acuity (e.g, with glasses if worn):** Yes / No

**Normal colour vision:** Yes / No

**Video Coding/Processing and/or Multimedia Transmission Experience:** Yes / No

**Assessment Location:** UoM / Other

In this experiment, you will be shown six (6) sets of video sequences. Each test sequence is split into 2 views (View 1 and View 2) and each view will be shown 5 times with varying degrees of quality. The first sequence being shown does not necessarily mean it is the best quality sequence, since they will be shown to you at random. Each time a sequence is shown, you should judge its quality by using one of the nine levels of the following scale. Do not base your rating on the content of the scene or the quality of the acting. Observe carefully the entire video sequence before making your judgement.

| | |
|---|---|
| 9 | Excellent |
| 8 | Good |
| 7 | |
| 6 | Fair |
| 5 | |
| 4 | Poor |
| 3 | |
| 2 | Bad |
| 1 | |

**Video Quality Rating**
Please mark the appropriate box with a

**Test Sequence: Balloons View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Balloons View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Book View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Book View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Newspaper View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Newspaper View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Poznan Street View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Poznan Street View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Poznan Hall View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Poznan Hall View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Kendo View 1**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

**Test Sequence: Kendo View 2**

| | Rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Reference** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **A** | | | | | | | | | |
| **B** | | | | | | | | | |
| **C** | | | | | | | | | |
| **D** | | | | | | | | | |
| **E** | | | | | | | | | |

# Appendix C: Publications

# A CQI Feedback Clustering Technique for Signaling Traffic Reduction During Crowd Events in 3GPP LTE Systems

Mario Cordina
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
m.r.cordina@ieee.org

Carl J. Debono
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
c.debono@ieee.org

***Abstract*—Crowd events such as public demonstrations, sports events and mass entertainment gatherings pose additional challenges to mobile network operators due to the asymmetric traffic profile. In such crowd events, networks are typically uplink limited whilst also requiring a significant amount of signaling information to be sent by the user equipment (UE) to the evolved Node B (eNB). This signaling includes channel quality indicator (CQI) feedback reports which provide necessary information for the selection of the downlink transmission parameters such as the modulation and coding scheme. In this work, the authors address this crowd event signaling problem by considering a common sports event scenario, whereby users are seated in a football stadium, and propose a novel CQI clustering scheme to reduce the feedback report signaling. The proposed scheme was implemented and tested in an LTE system level simulator, which was purposely modified to model the scenario, and has shown efficacy with an overall CQI feedback signaling reduction of up to 91% whilst maintaining stable sector throughput, when compared to the standard third generation partnership project (3GPP) CQI feedback mechanism.**

***Index Terms*—3GPP LTE, frequency selective scheduling, clustering techniques, crowd event handling**

## I. Introduction

According to Cisco, the average mobile traffic is expected to increase five-fold by 2020 when compared to 2015 [1]. This increase in traffic refers to normal business day operating conditions. However, mobile network operators have to also contend with crowd events. During such events a large number of users, typically in the tens of thousands, are usually gathered in a confined geographical area posing significant challenges to the mobile network operators tasked with providing an acceptable quality of service. As opposed to normal operating conditions, the traffic profile during such crowd events is uplink dominated as users generate and share multimedia content, in the form of photos and video, on social media platforms. Meanwhile to fully exploit the benefits of frequency domain scheduling (FDS) [2], each user equipment (UE) in the radio sector has to report a channel quality indicator (CQI) feedback report to the eNB. This generates significant uplink signaling overhead whilst increasing the risk of uplink control channel congestion.

In this paper, we propose a novel CQI clustering scheme to reduce signaling traffic in a typical crowd event scenario whereby users are seated in a football stadium. To the best knowledge of the authors, this work represents the first study on the application of clustering techniques to CQI feedback signaling reduction in crowd events. Even though the focus of this work is on the 3GPP LTE standard, being the cellular network technology commonly deployed worldwide, it can be applied to any current and future OFDMA-based cellular network technology.

The rest of the paper is organized as follows: Section 2 provides a brief description of CQI feedback reporting and signaling reduction techniques; Section 3 gives an overview of the system model together with a description of the proposed scheme; Section 4 first elaborates on the formulation of a statistical model used to estimate the number of clusters in the radio sector and then proceeds to a description of the simulation environment, parameter selection and results, while Section 5 provides some comments and conclusions.

## II. Related Work and Our Contribution

In LTE, the UE uses the downlink reference signals transmitted by the eNB [2] to measure the signal-to-interference noise ratio (SINR) for each PRB $k$, given by

$$\gamma_k = \frac{P_{i,k} G_{i,k}}{\sum_{j \neq i} P_{j,k} G_{j,k} + n_k} \quad (1)$$

where $P_i$ and $G_i$ are the transmit power and antenna gain of the serving base station $i$ whilst $P_j$ and $G_j$ are the transmit power and antenna gain of the interfering base

978-1-5386-3531-5/17/$31.00 ©2017 IEEE

station $j$, and $n_k$ is the additive Gaussian noise. The CQI value is then derived from the measured SINR by using a chipset vendor specific mapping table which gives the highest modulation coding scheme (MCS) that the UE can decode with a block error rate not exceeding 10% [2]. Several feedback signaling reduction schemes have been proposed in literature, which can be divided into three categories, namely time domain, frequency domain and threshold based techniques. Time domain techniques form part of the 3GPP LTE standard and are split into two sub-categories, namely periodic and aperiodic reporting [3]. These techniques are commonly used in commercial networks as they allow network operators to control the amount of CQI feedback reports to be sent by the UE to the eNB by explicitly defining when CQI feedback reports need to be transmitted and their periodicity.

Frequency domain based feedback signaling reduction techniques are divided into three sub-categories namely full-band, wideband, and subband compression techniques. In full-band compression techniques, a UE uses a mathematical transform such as the Haar wavelet transform [3] and the discrete cosine transform (DCT) [4] to calculate a small set of dominant transform coefficients which are then transmitted to the eNB. Even though such techniques allow the eNB to reconstruct a good approximation of the entire system bandwidth quality, their efficacy depends on the delay spread of the channel, thereby reducing their applicability.

Wideband compression, which forms part of the 3GPP LTE standard, requires a UE to transmit a single CQI value derived from the effective SINR $\gamma_{eff}$ given by:

$$\gamma_{eff} = -\lambda \log\left(\frac{1}{N}\sum_{k=1}^{N} e^{-\frac{\gamma_k}{\lambda}}\right) \quad (2)$$

where $N$ represents the number of PRBs to group, $\lambda$ is a calibration factor that is MCS dependent [7] and $\gamma_k$ is the instantaneous SINR of the $k^{th}$ PRB. The effective SINR is computed using the effective exponential signal-to-noise ratio mapping (EESM) [5, 6] as different PRBs undergo different fading characteristics. In the case of wideband compression, $N$ is set to the total number of PRBs defined by the system bandwidth. Even though wideband compression is widely used in live commercial networks due to its simplicity and high compression factor, it precludes the eNB to have rich channel state information across the whole frequency band thereby limiting system performance.

Subband compression techniques have attracted significant research attention and several proposals can be found in literature including Best-M individual [8] and Best-M average [8]. In addition, the 3GPP LTE standard defined two sub-band compression schemes namely eNB-configured sub-band feedback and UE-selected sub-band feedback. These techniques allow the eNB to have richer channel state information across a wider frequency band, leading to a better system performance when compared to wideband CQI reporting. Nevertheless, this comes at the expense of a higher signaling overhead on the physical uplink shared channel (PUSCH) or physical uplink control channel (PUCCH), depending on whether or not the UE has uplink data scheduled to be transmitted [2], making them unsuitable for crowd event scenarios.

Inspired by the work found in [9-11], the authors in [12] apply a threshold-based signaling reduction technique whereby users which exceed a set threshold, based on the likelihood that the user will be scheduled, are candidates to use full-band feedback whilst the rest use wideband frequency domain compression. The use of CQI prediction techniques was investigated in [13-15] whilst in [16] the authors explore the use of spatial correlation of a number of users in a moving vehicle to reduce the CQI feedback signaling. In this work, the authors assume that a group of users which are in close proximity of each other, experience the same SINR conditions and hence a CQI feedback report received from a representative UE in the group can directly be applied to the rest of the users in the group. Apart from the additional complexity required to collect accurate user location information, the assumption that a group of users in close proximity of each other experience the same radio conditions might hold true only for the wanted signal, $S$, but not the interferers, $I$. The wanted signal, $S$, received by the group of users is transmitted from a single location and thus undergoes the same radio impairments in terms of path loss, shadow and multi-path fading. However, the same cannot be said for the interferer, $I$, which is usually composed of several interfering signals transmitted from different locations. Hence the SINR of users in close proximity of each other cannot be assumed to be similar and thus, grouping the users solely based on their spatial proximity is not enough.

Our contribution in this work is two-fold. First, we exploit the slow varying radio channel conditions of static users located in the same radio sector and propose a novel CQI clustering scheme to reduce feedback signaling. As opposed to [16], we use CQI sample based statistics to group UEs into several clusters. Secondly, a statistical model to estimate the number of clusters based on the active users in the radio sector is formulated. In this work, we focus on wideband reporting, this being widely used in live commercial networks, however the proposed scheme can be applied to the subband reporting scheme with minor modifications.

## III. System Model and Proposed Scheme

### A. System Model

A common strategy used by mobile network operators in tackling crowd events is to partition the crowd into smaller areas by introducing additional radio

sectors. For the case of this study, consider a three-tier football stadium, shown in Fig. 1, which has been partitioned into several areas with a number of radio sectors. The rest of this work will focus on the three radio sectors in the south stand, shown in Fig. 1, whereby sector B is the sector under study whilst the adjacent sectors A and C contribute to the interference.


Figure 1: Three-tier football stadium and radio sectors under study [17]

The seating configuration parameters for each area, which are aligned with the FIFA football stadium technical guidelines [18], are defined in Table I whereby passage ways refers to zones used for spectator movement. Using this seating configuration, each area has a seating capacity of 1540 spectators. Considering a 40% operator market share [19], 54% of spectators who actively share content in such crowd events [20] and a 40% simultaneous user ratio, the maximum number of simultaneous active UEs, at any one time, in the sector is approximately 130.

TABLE I. SEATING CONFIGURATION PARAMETERS

| Parameter | Value |
|---|---|
| Area dimension | 30m x 30m |
| Front/Back passage ways | 1m |
| Lateral passage ways | 2m |
| Inter-seat spacing in x-direction | 0.5m |
| Inter-seat spacing in y-direction | 1m |

In order to minimize the cell overlap, a sculpted radiation pattern antenna having a sharp -3dB cut-off point is typically used in such scenarios [21]. These antennas are typically fixed to the underside of the stadium roof facing the seating area to be covered. The use of this antenna allows the signal to be focused in a specific area of the crowd whilst minimizing inter-cell interference. The signal level, in dBm, provided by sector B at 0.5m above ground level (AGL), is shown in Fig. 2 whilst the relevant simulation parameters used in this path loss study are provided in Table II.

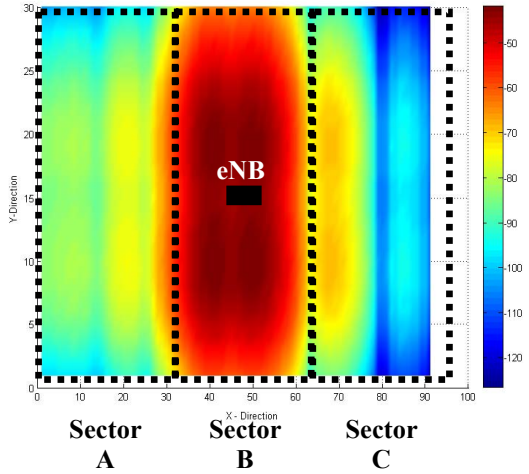



Figure 2: Sector B signal level in dBm @ 0.5m AGL (Elevation)

*B. Proposed Scheme*

With reference to Fig. 3, the operation of the proposed scheme is divided in three phases namely: Initialization, cluster formation and clustered phase. During the initialization phase, the UEs operate in *legacy_mode* and periodically report the wideband CQI measurement to the eNB to allow the eNB to build a CQI report history for each active UE. The eNB keeps track of the number of CQI reports received per UE and when a UE reaches $\tau_{UE}$ reports, the eNB updates the status of the UE to *cluster_wait* mode. In this mode, the UE continues to report the CQI measurements however the eNB will overwrite old CQI measurement values thereby ensuring that the eNB retains the latest $\tau_{UE}$ reports for each UE. When all active UEs are in *cluster_wait* mode, the eNB moves on to the cluster formation phase. During this phase, the eNB will:

1. For each UE, compute the mean $\mu_{ue}$ and variance $\sigma_{ue}^2$ using the last $\tau_{UE}$ CQI measurement reports and normalize the data.
2. Apply a data clustering algorithm on the two dimensional data vector, composed of the computed normalized statistics in step 1, to cluster users into *k* clusters and assign a cluster ID to each UE. In this work, we selected the *k*-means algorithm, being one of the most commonly used and robust data clustering techniques.

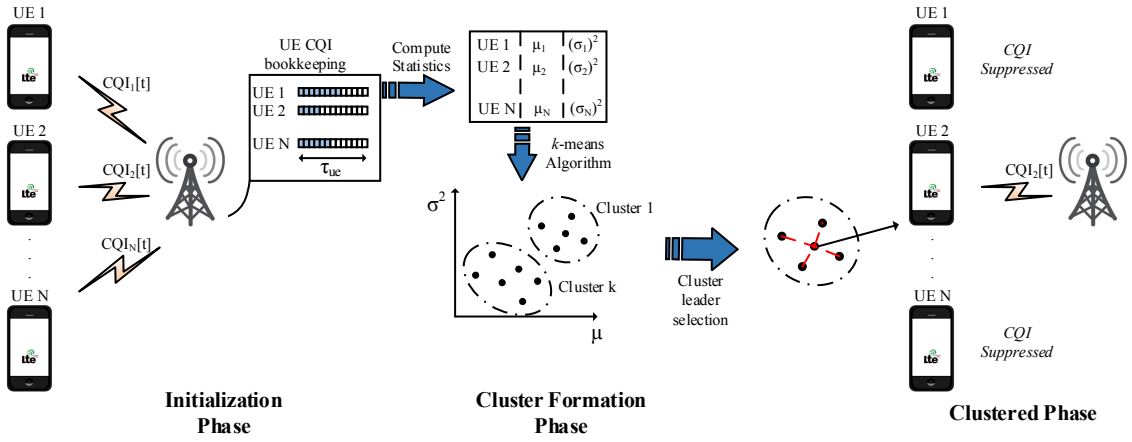



Figure 3: Proposed CQI feedback clustering scheme

3. For each cluster; calculate the centrality of each UE within the cluster using (3). The centrality of a UE, $C_i$, defines how close a UE, $i$, is to the other UEs, $j$, in the same cluster, based on the computed normalized statistics. The UE with the lowest centrality value (i.e. highest centrality) is elected as cluster leader, $UE_{CL}$. This process ensures that the cluster leader CQI reports are representative of the whole cluster.

$$C_i = \sum_{j \neq i} \sqrt{\left(\mu_i - \mu_j\right)^2 + \left(\sigma_i^2 - \sigma_j^2\right)^2} \qquad (3)$$

Following the cluster formation phase, the eNB moves to the clustered phase and remains in this mode until it receives $\gamma_{CL}$ CQI measurement reports from all the cluster leaders. After which, the eNB moves back to the initialization phase and restarts the process. In the clustered phase, the cluster leader will continue sending CQI reports to the eNB and the eNB will use these reports as representative CQI reports for the other UEs in the cluster. Meanwhile, the other UEs in the cluster will be set to *CQI_Suppressed* mode indicating that these UEs will not transmit CQI measurement reports.

## IV. Statistical Model Formulation and Results

A critical input parameter for the majority of data clustering techniques is the required number of clusters, $k$. Choosing $k$ automatically is notoriously difficult and is usually either fixed or else an iterative technique to merge/split clusters is adopted [22] at the expense of increased computational complexity. In this work, we exploit domain-specific knowledge to formulate a statistical model able to estimate the number of clusters based on the active users in the radio sector. In this respect, the cluster formation algorithm, described in Section 3, was first implemented in MATLAB® and tested using CQI measurements extracted from a large number of UEs, simulated in an LTE system level simulator [23], for various values of $k$. The parameters used in these simulations are provided in Table I and II.

TABLE II. Simulation Parameters

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic (2ms), wideband CQI (4bits) |
| Traffic Model | Full Buffer |
| Mobility | All UE's are stationary |
| eNB output power | 43dBm |
| Frequency Re-use | 1 |
| Antenna Height | 20m |
| Antenna Type | Commscope CNLPX3055F |
| Propagation Model | Micro Cell<br>L=24 + 45log(d+20) @ 2GHz |
| Fast Fading Realization | Jakes Model |
| Number of Active UE's, M | {50, 70, 90, 110, 130} |
| $\tau_{UE}$ and $\gamma_{CL}$ | 100 and {3000, 5000, 8000} |

With reference to Fig. 4, it is evident that increasing the number of clusters, $k$, would result in more compact clusters as the sum of within-cluster point-to-cluster centroid distance, $D$, decreases. Nevertheless, the increase in $k$ implies an increased number of cluster leaders transmitting CQI reports to the eNB. In order to strike a balance between these two opposing factors, a cost function relating the normalized CQI report transmission cost, $\eta$, and cluster compactness metric, $D$, is used to understand the variation of the total cost, $\phi$, with the number of clusters $k$. Considering the reporting period and CQI message length shown in Table II, the cost function is given by:

$$\begin{aligned} \phi(k) &= \alpha \cdot D + \beta \cdot \eta \\ \eta &= 2000k \quad \text{(bits/sec)} \end{aligned} \qquad (4)$$

where $\alpha$ and $\beta$ are weighting constants empirically set to 15 and 1.2 respectively.

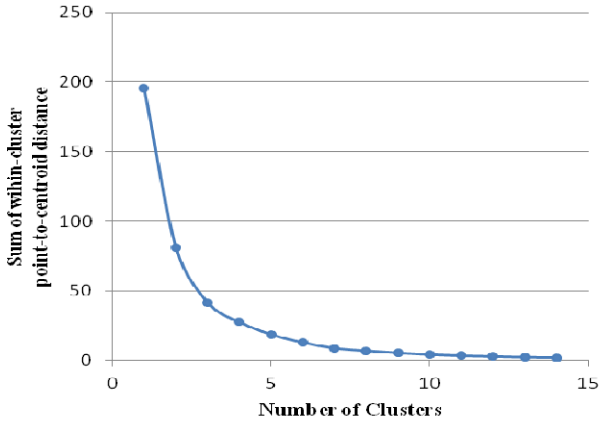



Figure 4: Variation in cluster compactness with number of clusters (number of active UEs = 130)

Plotting the cost function on a logarithmic scale for different number of active UEs, $M$, yields a set of upward concave parabolas shown in Fig. 5. Further inspection shows that an optimal number of clusters, $k_{opt}$, exist for different values of $M$.

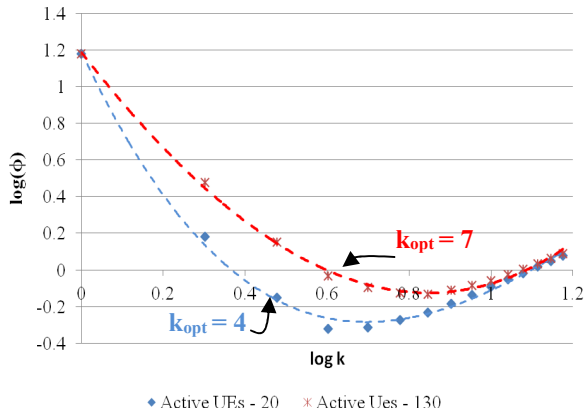



Figure 5: Cost function dynamics

Repeating this analysis for various values of $M$ and finding the optimal number of clusters for each case reveals the inverse relationship shown in Fig. 6. As the number of active UEs increases, the optimal number of clusters rises until it reaches an asymptotic value. This is expected as the CQI transmission cost increases linearly with $k$ thereby enforcing a hard limit on the optimal number of clusters. Meanwhile, at low values of $k$, the inverse relationship will return a progressively lower number of clusters. Nevertheless, experimental results show that below 15 active users (which does not constitute a crowd event in any case), the proposed clustering scheme benefits diminish and hence it would be more appropriate to switch back to the legacy scheme whereby UEs report the CQI directly to the eNB. The data was further analysed in IBM SPSS© statistical package and an inverse model, given by (5) and having a coefficient of determination, $R^2$, of 0.966 was derived using the least square estimation technique [24].

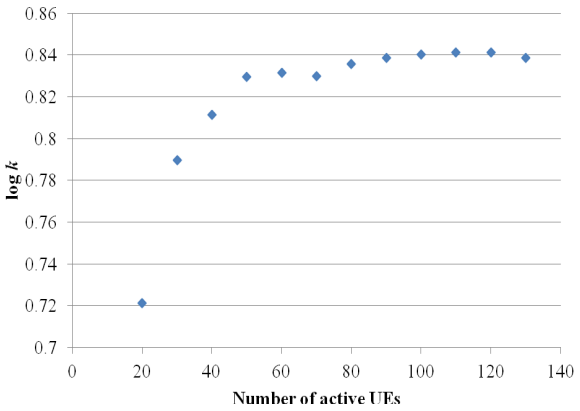



Figure 6: Inverse relationship between log($k$) and $M$

$$\log(k) = -2.61 \cdot \left(\frac{1}{M}\right) + 0.867 \qquad (5)$$

The proposed scheme described in Section 3 together with the derived statistical model given by (5) were implemented in LTE-Sim [23] to verify the efficacy in terms of average cell throughput and feedback reduction. Monte Carlo simulations, using the parameters given in Table I and II were carried out over 50 random seeds in order to achieve statistical relevant results. The average sector throughput and feedback reduction, when compared to the 3GPP standardized wideband feedback scheme, for various values of $M$ are shown in Fig. 7 and Table III respectively.

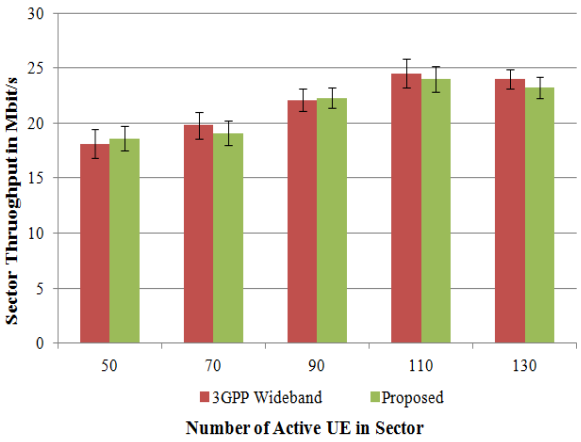



Figure 7: Sector throughput comparison for various number of Active UEs in sector (standard deviation is indicated by vertical bars)

The proposed CQI feedback clustering scheme manages to achieve an average CQI feedback reduction of 88.2% when compared to the 3GPP wideband scheme. This level of feedback reduction was achieved whilst maintaining approximately the same level of performance, in terms of sector throughput, as evidenced in Fig. 7. Moreover, with reference to Table III, the feedback reduction gradually increases as the clustering becomes more effective with increasing number of active UEs.

TABLE III. FEEDBACK REDUCTION WITH RESPECT TO 3GPP WIDEBAND SCHEME

| Active UEs | % Feedback Reduction |
|---|---|
| 50 | 83.1 |
| 70 | 86.9 |
| 90 | 89.1 |
| 110 | 90.4 |
| 130 | 91.4 |

The effect of the parameter $\gamma_{CL}$ on the performance of the proposed scheme was also investigated for various values of $\gamma_{CL}$, shown in Table IV. Intuitively, as $\gamma_{CL}$ increases, we prolong the clustered phase thereby achieving a marginally improved feedback reduction. However, this comes at the expense of degradation in sector performance due to the user clustering and cluster leader selection becoming increasingly sub-optimal.

TABLE IV. EFFECT OF PARAMETER $\gamma_{CL}$ ON THE PERFORMANCE OF THE PROPOSED SCHEME (ACTIVE UE'S = 90)

| $\gamma_{CL}$ | 3000 | 5000 | 8000 |
|---|---|---|---|
| Sector Throughput in Mbit/s | 22.3 | 22.2 | 21.1 |
| % CQI Feedback Saving | 89.1 | 90.3 | 90.9 |

## V. CONCLUSION AND FUTURE WORK

In this paper we presented a novel CQI feedback scheme for signaling reduction during crowd events. The key contribution lies in the use of a commonly used data clustering technique within the CQI feedback management domain coupled with a statistical model to estimate, online, the number of clusters based on the active users in the radio sector. Experimental results based on Monte Carlo simulations showed that the proposed technique provides up to 91% CQI feedback signaling reduction whilst maintaining the same level of sector throughput when compared to the standardized 3GPP wideband scheme. Future work will focus on the use of a cross-layer technique to adapt the CQI feedback scheme in relation to the traffic type being downloaded by the user.

## REFERENCES


[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper", http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, accessed 15 April 2017.

[2] H. Holma, and A. Toskala, "LTE for UMTS – OFDMA and SC-FDMA Based Radio Access", Wiley, 2009.

[3] A. Haghighat, G. Zhang and Z. Lin, "Full-band CQI feedback by Haar compression in OFDMA systems," in *Proc. of the IEEE 70th Vehicular Technology Conference Fall* (*VTC 2009-Fall*), September 2009, pp. 1-5.

[4] R1-061777, "DCT based CQI reporting scheme", LG Electronics, RAN1 LTE Ad Hoc, Cannes, France, June, 2006.

[5] M. Pauli, U. Wachsmann, and S. Tsai, "Quality determination for a wireless communications link", U.S. Patent Office publication, Patent (US 2004/0219883, November 2004.

[6] E. Westman, "Calibration and evaluation of the exponential effective SINR mapping (EESM) in 802.16," Master's thesis, The Royal Institute of Technology (KTH), Stockholm, Sweden, September 2006.

[7] C. Fa-tang and T. Gen-lin, "A novel MCS selection criterion for supporting AMC in LTE system," in *Proc. of the International Conference on Computer Application and System Modeling* (*ICCASM*), 2010, vol. 6, pp. 598-603.

[8] M. Kang, and K. Soon, "Performance analysis and optimization of best-M feedback for OFDMA systems", *IEEE Communications Letter*, vol. 16, issue 10, October 2012, pp.1648-1651.

[9] D. Gesbert, M.S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proc. of Int. Conf. Comm.*, June 2004, pp. 234–238.

[10] L. Sivridis, X. Wang, and J. Choi, "Radio resource management for fast fading environments," *Journal of Comm.*, vol. 6, issue 7, October 2011, pp. 540-548.

[11] K. Hoon, and H. Youngnam, "An opportunistic channel quality feedback scheme for proportional fair scheduling," *IEEE Communications Letter*, vol. 11, issue 3, June 2007, pp. 501–503.

[12] L. Sivridis, J. He, "A Strategy to reduce the signaling requirements of CQI feedback schemes," *Journal Wireless Pers. Comm.*, vol. 70, issue 1, May 2013, pp. 85-98.

[13] T. Werthmann, M. Kaschub and M. Proebster, "Simple channel predictors for lookahead scheduling," in *Proc. of the IEEE 75th Vehicular Technology Conf.* (*VTC 2012-Spring)*, May 2012, pp. 1-6.

[14] H. Dai, Y. Wang, C. Shi and W. Zhang, "The evaluation of CQI delay compensation schemes based on Jake's model and ITU scenarios," in *Proc. of the IEEE 76th Vehicular Technology Conf.* (*VTC 2012-Fall)*, September 2012, pp. 1-5.

[15] X. Xu, M. Ni, R. Mathar, "Improving QoS by predictive channel quality feedback for LTE," in *Proc. Int. Conf. Software, Telecom. and Computer Networks*, September 2013, pp. 1-5.

[16] W. Lee, D. Cho, "CQI feedback reduction based on spatial correlation in OFDMA System," in P*roc. of the 68th IEEE Vehicular Technology Conf.* (*VTC 2008- Fall*), September 2008.

[17] "Manchester United Stadium: Old Trafford Ground", https://www.e-architect.co.uk/manchester/manchester-united-stadium. Accessed: 10 June 2017

[18] "FIFA-football stadiums Technical recommendation and requirements," https://www.scribd.com/doc/100501692/FIFA-Football-Stadiums-Technical-recommendation-and-requirements-5th-edition. Accessed: 10 June 2017.

[19] "Market share of leading operator in mobile telecommunications," http://www.pordata.pt/en/Europe/Market+share+of+leading+operator+in+mobile+telecommunications-1502. Accessed: 10 June 2017.

[20] "Mobile behaviour and viewing technology – May 2014," https://slideshare.net/LiveAnalytics/mobile-us-final. Accessed: 10 June 2017.

[21] "Commscope Stadium Antenna Model: CNLPX3055F", Antenna patterns and datasheet available: http://www.commscope.com. Accessed:10 June 2017.

[22] G. Hamerly, C. Elkan, "Learning the K in K-Means - Neural Information Processing Systems,", MIT Press, 2003.

[23] G. Piro, L. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open source framework", *IEEE Transactions Vehicular Technology*, vol. 60, no.2, February 2011.

[24] M. Norusis, IBM SPSS Statistics 19 Advanced Statistical Procedures Companion, Pearson, February 2011.

IET Communications

Research Article

IET Journals
The Institution of Engineering and Technology

# Robust predictive filtering schemes for sub-band CQI feedback compression in 3GPP LTE systems

ISSN 1751-8628
Received on 14th February 2017
Revised 19th April 2017
Accepted on 2nd May 2017
doi: 10.1049/iet-com.2017.0153
www.ietdl.org

*Mario Cordina*[1] ✉, *Carl J. Debono*[1]

[1]*Department of Communications and Computer Engineering, University of Malta, Msida, Malta*

✉ *E-mail: m.r.cordina@ieee.org*

**Abstract:** The third generation partnership project (3GPP) long-term evolution (LTE) cellular system offers high data rate capabilities by leveraging several techniques including link adaptation and frequency selective scheduling. These techniques rely on accurate channel quality indicator (CQI) feedback reports that are sent by the user equipment (UE) to the evolved node B, a process which results in high signalling overhead. In this study, the authors propose a UE-assisted sub-band feedback compression scheme based on a predictive filtering technique to reduce this signalling overhead. Four schemes based on adaptive filters have been designed, implemented and tested in an LTE system level simulator. Simulation results indicate that the proposed compression scheme has shown efficacy with an overall CQI feedback signalling reduction of up to 92.5% whilst maintaining stable sector throughput, when compared with the standard 3GPP CQI feedback mechanism. Although the proposed scheme exhibits low computational and memory complexity, a reduced-complexity scheme achieving an average computational load reduction of up to 35% is also presented.

## 1 Introduction

### 1.1 Background and motivation

Contemporary wireless communication standards such as long-term evolution (LTE) standardised by the third generation partnership project (3GPP) in 2008 and further enhanced through subsequent releases [1], offer significant improvement in the coverage, capacity and flexibility of mobile cellular systems compared with legacy technologies. This is realised through several fundamental changes across all layers of the communication stack. Primarily, at the physical layer, the system adopts an orthogonal frequency division multiple access (OFDMA) technique which splits the available bandwidth into orthogonal physical resource blocks (PRB), thereby offering increased capacity and robust performance in a multipath frequency selective channel. The high peak data rate capability is achieved through the use of several techniques including multiple-input multiple-output, adaptive modulation and coding and frequency-domain scheduling (FDS) [2]. These rely on accurate channel state information, predominantly in the form of channel quality indicator (CQI) feedback reports, sent by the user equipment (UE) to the evolved node B (eNB) on the physical uplink shared channel (PUSCH) or physical uplink control channel (PUCCH) depending on whether or not the UE has uplink data scheduled to be transmitted [2]. This CQI feedback is derived from the signal-to-interference noise ratio (SINR) measurements of the downlink reference signals transmitted by the eNB. For each PRB $k$, the received SINR is given by

$$\gamma_k = \frac{P_{i,k}G_{i,k}}{\sum_{j\neq i}P_{j,k}G_{j,k} + n_k} \tag{1}$$

where $P_i$ and $G_i$are the transmit power and antenna gain of the serving base station $i$ whilst $P_j$ and $G_j$ are the transmit power and antenna gain of the interfering base station $j$, and $n_k$ is the additive Gaussian noise. The resulting SINR is then mapped to a discrete CQI value, using a UE chipset vendor specific mapping table, which indicates the highest modulation coding scheme (MCS) that the UE can decode with a block error rate not exceeding 10%. The reported CQI value does not only indicate the downlink channel quality but also considers the capabilities of the UE receiver thereby allowing a UE with an advanced receiver architecture to report a higher CQI value for the same downlink channel quality.

To take full advantage of FDS, each UE must feedback a CQI value for each PRB to the eNB. This is known as full-band feedback and although it allows the eNB to gain knowledge of the channel conditions on each PRB, thereby maximising the downlink performance, it is not practical due to the significant uplink signalling overhead that needs to be transported over the bandwidth limited physical channels. This is particularly relevant in high traffic areas, where uplink control channel congestion can occur, leading to a rapid degradation of the overall sector performance. This problem is further exacerbated by the expected increase in smartphone data traffic on mobile networks. According to Cisco, the average smartphone traffic in 2020 will be almost five times greater than 2015 [3] and thus this feedback will constitute a real and significant problem for mobile network operators.

### 1.2 Related work and our contribution

The challenge to design low overhead CQI feedback schemes has garnered significant research attention in recent years. The majority of the techniques found in literature can be categorised into three main groups namely time domain, frequency domain and threshold/prediction based techniques. The overarching goal is to strike a balance between achieving accurate channel estimation with limited uplink feedback whilst maintaining an adequate system performance.

To control the uplink signalling overhead, the CQI feedback is generally limited in both the time and the frequency domain. In the time domain, CQI reporting can either be periodic or aperiodic [4] whilst in the frequency domain, several feedback compression techniques were proposed which can be broadly divided into three categories: wide-band, full-band and sub-band compression techniques. In wide-band compression, a UE transmits a single CQI value for all the PRBs in the bandwidth. In view of frequency selective fading, different PRBs undergo different fading characteristics and therefore when grouping the instantaneous SINR of the PRBs, the effective SINR, $\gamma_{\text{eff}}$, needs to be computed using the effective exponential signal-to-noise ratio mapping [5, 6] given by

$$\gamma_{\text{eff}} = -\lambda \log\left(\frac{1}{N}\sum_{k=1}^{N} e^{-(\gamma_k/\lambda)}\right) \quad (2)$$

where $N$ represents the number of PRBs to group, $\lambda$ is a calibration factor that is MCS dependent [7] and $\gamma_k$ is the instantaneous SINR of the $k$th PRB found using (1). For wide-band compression, $N$ is the total number of PRBs defined by the system bandwidth [2]. The effective SINR is then mapped to a discrete CQI value using the chipset vendor specific SINR-CQI mapping table stored in the UE. Although this technique is attractive due to its simplicity and high compression factor, it comes at the expense of lower system performance due to the lack of channel quality detail. Nevertheless, wide-band compression was included in the 3GPP LTE standard as of Release 8 and is widely used in live commercial networks.

Full-band compression techniques generally involve the use of mathematical transforms such as the discrete cosine transform [8, 9] and the Haar wavelet transform [10] to reduce the signalling overhead. Their advantage is that the eNB can reconstruct a good approximation of the entire system bandwidth quality using a relatively small number of transform coefficients reported by the UE. Nevertheless, the reconstruction quality depends on the delay spread of the channel.

In sub-band compression, a UE transmits a selected set of CQI values to the eNB. Such techniques include Best-M individual [11] and Best-M average [11]. The 3GPP LTE standard defines two sub-band compression techniques:

- *eNB configured sub-band feedback [2]*: Here, the system bandwidth is divided into $q$ sub-bands of $N$ consecutive PRBs and the UE reports a wide-band CQI value together with a differentially encoded CQI value for each sub-band.
- *UE-selected sub-band feedback [2]*: Similar to eNB-configured sub-band feedback, the system bandwidth is divided into $q$ sub-bands of $N$ consecutive PRBs. The UE then selects $M$ preferred sub-bands and reports a wide-band CQI value together with a differentially encoded CQI value reflecting the average quality of the $M$ preferred sub-bands. The position of the selected $M$ preferred sub-bands also needs to be reported thereby lowering some of the gains.

The third category encompasses techniques which make use of either channel prediction or event-based triggering to reduce the feedback. A number of predictive channel quality methods can be found in literature however these are all aimed at mitigating the impact of CQI delay on the system performance without any consideration to the CQI compression problem. In [12], the authors propose an autoregressive prediction filter, whilst a normalised least mean square (NLMS) filter was applied to compensate for CQI delay in [13]. Similarly, in [14, 15] the authors propose several CQI prediction schemes based on Wiener filter, cubic spline extrapolation and short-term average. In [16], Chiumento *et al.* propose an adaptive channel quality estimation method based on Gaussian process (GP) regression at the base station. This GP-based CQI prediction is exploited in a dual-control technique which makes use of active learning to determine an optimal prediction time for each user, achieving a 77% signalling overhead reduction when compared with the 3GPP standard eNB-configured sub-band feedback compression technique with a packet loss rate (PLR) of 10%. To the best of the authors' knowledge, this represents the only work which makes use of channel quality prediction towards CQI feedback signalling reduction however it does not report the computational complexity of the dual-control technique. In [17], the authors extend the work found in [18–22] and propose a threshold-based signalling reduction technique. In this scheme users which exceed a set threshold, based on the likelihood that the user will be scheduled, are candidates to use full-band feedback whilst the rest use wide-band frequency-domain compression. Out of all the candidates allowed to use full-band feedback, only those which have the lowest average SINR will be requested by the network to use periodic full-band feedback. Although this scheme achieves marginal gains, it is mostly suitable for low-mobility scenarios where channel conditions are largely stable. Meanwhile, in [23], the authors propose an adaptive threshold feedback compression scheme to address the shortcomings of threshold-based approaches.

In this paper, we propose two sub-band feedback compression schemes based on a UE-assisted predictive filtering technique. As opposed to previous work, we exploit short-term channel prediction for feedback reduction. This is accomplished through the use of simple, low computational complexity adaptive filters in a predictive filtering algorithm involving the UE. Two types of adaptive filters were designed and evaluated using either the reported CQI or SINR measurements and the efficacy of each scheme was assessed in terms of the downlink system performance and feedback reduction. The best scheme was further evaluated in terms of computational complexity and memory footprint leading to the design and evaluation of a reduced-complexity scheme. Tests were also carried out in a live commercial LTE network to establish the effectiveness of the proposed schemes in real-world channel conditions. To the best of the authors' knowledge, this work represents the first study on the application of a UE-involved prediction filtering technique to feedback compression which has also been tested in a live commercial network. The proposed schemes can be applied to any OFDMA-based cellular network technology, however, in this work, we focus on the 3GPP LTE standard; this being the most advanced cellular network technology available.

The rest of the paper is organised as follows. Section 2 describes the proposed system model and algorithm, Section 3 gives a description of the simulation environment, parameter selection, complexity analysis and results, and a detailed account of the field measurement setup and results, while Section 4 provides some comments and conclusions.

## 2 System model and algorithm design

With reference to Fig. 1, assume that the system bandwidth composed of $L$ PRBs, where $L$ is defined by the 3GPP Release 8 standard [1], is divided into $q$ sub-bands of $N$ consecutive PRBs. For each PRB, a UE conducts SINR measurements denoted by $\gamma_L$ and calculates the effective SINR of each sub-band, denoted by $\gamma_q^{\text{sub}}$, using (2). Two sub-band feedback compression schemes were designed using either the reported effective SINR (fed directly to the adaptive filter as shown by the dotted path marked with an asterisk) or CQI measurements (where the effective SINR is mapped to a discrete CQI value in the UE as described in [24] prior to transmission). The advantage of transmitting the effective SINR is that the eNB is able to reconstruct an un-quantised view of the channel conditions experienced by the UE. However, the CQI value does not only indicate the downlink channel quality but it also takes into account the capabilities of the UE receiver and therefore a UE offering a better receiver performance can report a higher CQI value for the same effective SINR measurement.

The SINR to CQI mapping is UE chipset vendor dependent and thus reporting the effective SINR directly would introduce additional complexities as the eNB needs to either keep a database of SINR-CQI mapping tables, one per UE type, resulting in additional memory complexity in the eNB or interrogate the UE to exchange the SINR-CQI mapping table leading to a marginal increase in signalling. Now consider a stream of measurements for the $q$th sub-band, $x[k]$, that has to be reported by the UE. Also, consider that the required accuracy of the measurements at the eNB is $\pm E_{\text{max-sinr}}$ and $\pm E_{\text{max-cqi}}$ depending on whether the effective SINR or CQI measurements are used, respectively. Then, using the proposed predictive filtering algorithm, instead of reporting the stream $x[k]$, we selectively report some elements of the measurement stream such that the eNB is able to reproduce the complete measurement stream for the $q$th sub-band within the given accuracy. This is achieved by maintaining a set of predictors, one for each sub-band, in the UE and eNB which are an exact replica of each other. These predictors are based on adaptive filter theory and compute an estimate of the next element in the measurement stream $\hat{x}[k]$, given some previous elements. As shown in Fig. 2, for the effective SINR measurement scheme, if the predicted value $\hat{x}[k]$ differs from the actual value $x[k]$ by more than

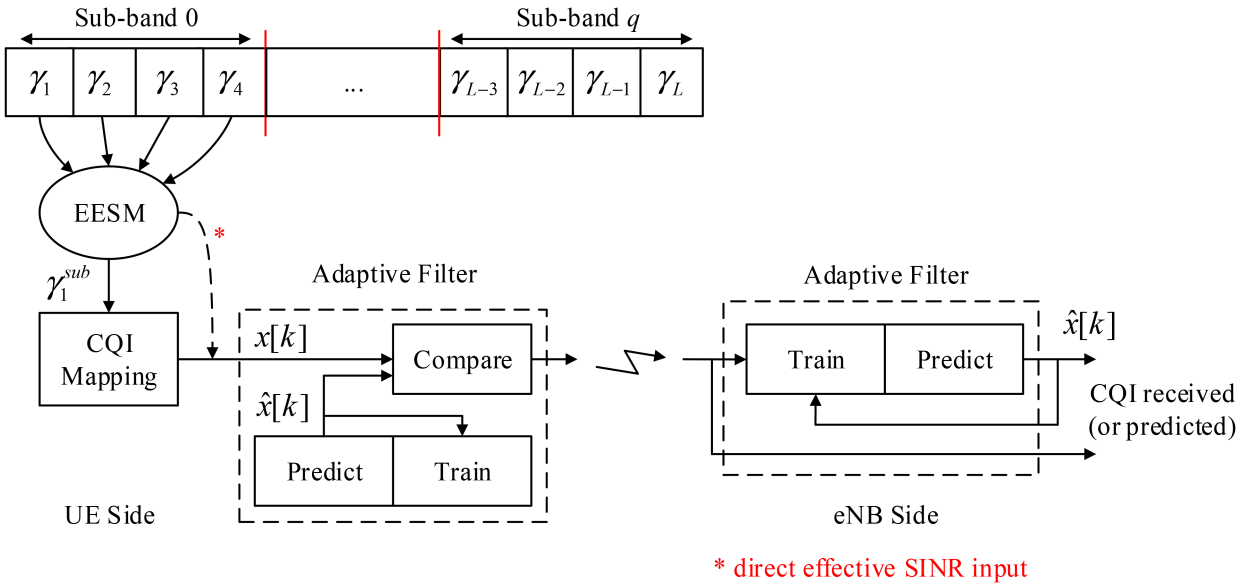



**Fig. 1** *Prediction filtering technique (N = 4) showing the report for sub-band 0*

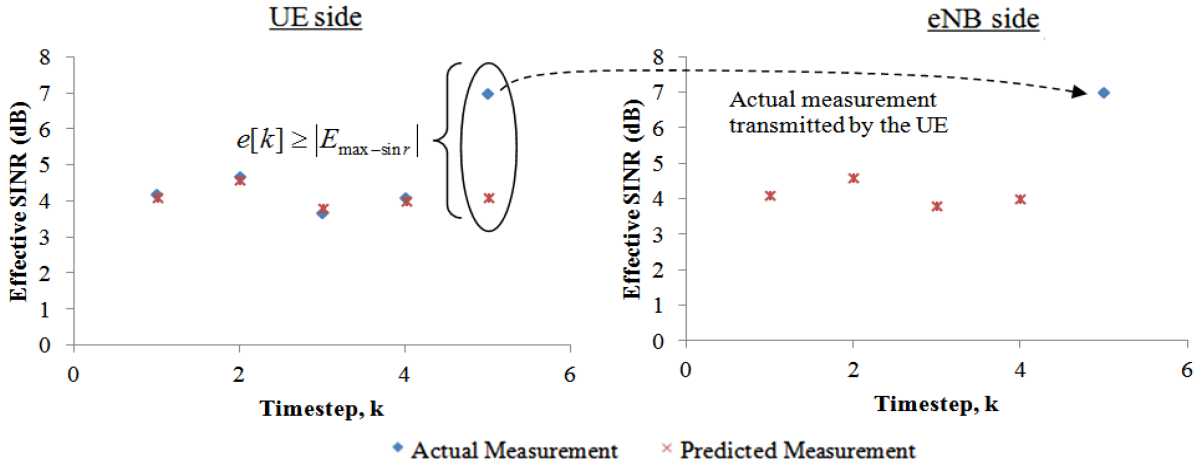



**Fig. 2** *Effective SINR channel feedback reduction*

the error budget, $\pm E_{\text{max-sinr}}$, then $x[k]$ is transmitted; otherwise the actual measurement is not reported and the eNB uses its prediction, $\hat{x}[k]$. Thus, feedback reduction could be achieved by exploiting short-term channel predictions.

### 2.1 Adaptive filtering algorithms

The choice of the adaptive filtering algorithm depends on the trade-off of several factors such as rate of convergence, robustness, computation complexity and numerical properties. In this work, we selected two well-known adaptive filter algorithms: namely, the least mean square (LMS) and the NLMS algorithms. Despite their simplicity, these algorithms provide good performance in a wide variety of applications. The fundamental equations for the computation of the output, error and LMS weight updating are given by (3)–(5), respectively

$$y[k] = \underline{\boldsymbol{w}}^{\mathrm{T}}[\boldsymbol{k}]\underline{\boldsymbol{x}}[k] \tag{3}$$

$$e[k] = d[k] - y[k] \tag{4}$$

$$\underline{\boldsymbol{w}}[\boldsymbol{k}+1] = \underline{\boldsymbol{w}}[k] + \mu\underline{\boldsymbol{x}}[k]e[k] \tag{5}$$

where $y[k]$ is the output of the adaptive filter, $e[k]$ is the prediction error, $d[k]$ is the desired signal, $\mu$ is the step size parameter, $M$ is the filter order whilst $\underline{\boldsymbol{w}}[k]$ and $\underline{\boldsymbol{x}}[k]$ denote the $M \times 1$ column vectors

$$\underline{\boldsymbol{w}}[k] = [w_1[k], w_2[k], w_3[k], \ldots, w_M[k]]^{\mathrm{T}} \tag{6}$$

$$\underline{\boldsymbol{x}}[k] = [x[k-1], x[k-2], x[k-3], \ldots, x[k-M]]^{\mathrm{T}} \tag{7}$$

The main difference between the LMS and NLMS algorithms is in the weight updating equation. The LMS algorithm is sensitive to the scaling of its input $x[k]$ thereby making it difficult to choose an appropriate step size that guarantees convergence [25]. Alternatively, the NLMS algorithm normalises the step size parameter with the power of the input signal. The NLMS adaptive filter updates the weight according to

$$\underline{\boldsymbol{w}}[\boldsymbol{k}+1] = \underline{\boldsymbol{w}}[k] + \frac{\mu_T}{\alpha + \| x[k] \|^2}\underline{\boldsymbol{x}}[k]e[k] \tag{8}$$

where $\mu_T$is the step size parameter and $\alpha$ is a constant to ensure that $\underline{\boldsymbol{w}}[\boldsymbol{k}+1]$ is bounded even when the input is very small. An in-depth explanation of the LMS/NLMS algorithm can be found in [25–27].

### 2.2 Algorithm description

The operation of the predictive filtering algorithm is divided into two phases as shown in Algorithm 1 (see Fig. 3). This is a generic algorithm and applies for both the LMS and NLMS adaptive filter cases. In training mode, measurements $x[k]$ are reported by the UE to the eNB. Meanwhile, both the UE and the eNB use the last $M$ measurements to compute $\hat{x}[k]$ for the upcoming measurement and

```
1: Initialisation:
2: k ← 0
3: mode ← training
4: txCounter ← 0
5: for k do
6:    if mode = training then
7:       UE transmits measurement to eNB
8:       txCounter ← txCounter + 1
9:       if txCounter ≥ M then
10:         UE and eNB uses last M measurements to compute:
11:         Prediction, x̂[k]
12:         Prediction error, e[k]
13:         Update filter weights, w
14:         if txCounter ≥ T_max & e ≤ error budget for β times then
15:               errCounter ← 0
16:               mode ← prediction
17:         end if
18:      end if
19:   else
20:         UE and eNB use last M measurements to compute prediction, x̂[k]
21:         UE and eNB shift x̂[k], into prediction filter
22:         UE uses x̂[k] and x[k] to compute prediction error, e[k]
23:      if e ≥ error budget then
24:               errCounter ← errCounter + 1
25:               if errCounter = β then
26:                  mode ← training
27:               end if
28:      else
29:               errCounter ← errCounter -1
30:               if errCounter < 0 then
31:                  errCounter ← 0
32:               end if
33:         end if
34:      end if
35:   k ← k + 1
36: end for
```

**Fig. 3** *Algorithm 1: Prediction filtering algorithm*

update the filter coefficients $w$ based on the actual prediction error using (5) or (8) depending on the type of adaptive filter used. This ensures that the prediction filters in the UE and eNB are synchronised. As long as the average prediction error over the past $M$ measurements is above the error budget or the number of measurements reported to the eNB is lower than $T_{\max}$, the UE operates in training mode and transmits $x[k]$. Otherwise, the UE will switch to prediction mode. In this mode, the UE computes $\hat{x}[k]$ at each time step $k$ and compares it to the actual measurement. Meanwhile, the eNB is also in prediction mode computing $\hat{x}[k]$ at each time step $k$ and feeding this value to its own prediction filter. If the prediction error is within the error budget, the UE discards the measurement and feeds the prediction filter with $\hat{x}[k]$ instead of $x[k]$ to ensure that the state of the prediction filters at both sides of the link remain consistent. This causes $e[k]$ to be zero and thus the filter weights remain unchanged, saving computational overhead. If at a time instant $k$ the UE observes that the average prediction error over the past $M$ measurements exceeded the error budget for $\beta$ consecutive times, it will report the reading $x[k]$ to the eNB and switch back to training mode.

## 3 Results

### *3.1 Prediction filtering parameter tuning*

Prior to evaluating the efficacy of the proposed solution, the parameters of the adaptive filters need to be tuned. These include:

- *Step size* $\mu$ *or* $\mu_T$: This is a critical parameter as it tunes the convergence speed.
- *Filter length* $M$: This affects the computational load and memory footprint of the filters. The LMS algorithm requires $2M + 1$ multiplications and $2M$ additions per prediction, whereas the NLMS algorithm requires $3M + 1$ multiplications and $2M$ additions per prediction.

In addition to the above, the hysteresis parameter $\beta$, which provides protection against oscillations between the prediction and training state, and the maximum allowed prediction error, $|E_{\text{max-sinr}}|$ and $|E_{\text{max-cqi}}|$, need to be studied in relation to (i) the average error between the actual and predicted SINR/CQI, (ii) the proportion of SINR/CQI reports sent by the UE from the total measurements and (iii) the total number of computations required. To tune these parameters, the predictive filtering algorithm described in Section 2, was first implemented in MATLAB® and tested using measurements extracted from a large number of UEs simulated in an LTE system level simulator [24]. The parameters used for the LTE simulations are provided in Table 1 whilst the optimal parameter values for both the LMS and NLMS solutions when using the effective SINR or CQI measurements are given in Table 2. These values were empirically determined by considering trade-offs between (i), (ii) and (iii) above. The effect of varying $\mu$ on the proportion of CQI measurements reported to the eNB and the average error of the predicted CQI for the LMS-based prediction filtering scheme are shown in Figs. 4*a* and b, respectively. Selecting a $\mu$ of 0.00005 yields the best trade-off between the selected metrics. A similar analysis was carried out to determine the other parameters for the remaining schemes.

### *3.2 Simulation results*

The four predictive filtering schemes were implemented in LTE-Sim [24] to verify their efficacy in terms of average cell throughput, feedback reduction and PLR. Monte Carlo simulations, using the parameters given in Tables 1 and 2 were carried out over 50 random seeds in order to achieve statistical relevant results. The mean and standard deviation (shown in parenthesis) of the feedback reduction and sector throughput for different UE speeds and cell loading are shown in Table 3 and Figs. 5*a* and *b*, respectively.

All of the proposed schemes deliver significant feedback signalling reduction at the expense of a marginal degradation in sector throughput and increase in packet loss when compared with the baseline configuration using the 3GPP standardised eNB configured sub-band feedback method. Moreover, results in Table 3 show that CQI-based prediction filtering schemes are on average 0.9 and 6.7% better than SINR-based schemes for low- and high-mobility scenarios, respectively.

The increase in feedback reports transmitted to the eNB for SINR-based techniques in high-mobility scenarios is due to the degraded temporal correlation within each sub-band leading to the UEs to prolong the training phase. This problem is somewhat mitigated in CQI-based prediction filtering schemes because of the discretisation function performed by the SINR-CQI mapping. This allows the CQI prediction filtering scheme to statistically achieve the same level of performance, in terms of the proportion of feedback transmitted, for both low and high mobility. Moreover, NLMS-based predictive filtering schemes are generally marginally better than their LMS counterparts particularly in high-mobility scenarios. As shown in Fig. 5*a*, CQI-based prediction filtering schemes are consistently better in terms of sector throughput when compared with SINR-based schemes owing to better short-term channel predictions. Meanwhile, in low-mobility scenarios, the sector throughput achieved by both the LMS and NLMS CQI-based prediction filtering schemes is within 5% of the baseline configuration with the CQI-LMS-based scheme having an edge over the NLMS-based scheme, whilst in high-mobility scenarios, all the proposed schemes achieve a comparable sector throughput.

The mean and standard deviation (shown in parenthesis) of the PLR for the various schemes are shown in Table 4. As expected, the proposed schemes suffer from an increase in PLR; however the CQI-based schemes rank better than the SINR-based schemes in all scenarios. In particular, in low-mobility scenarios, the CQI-LMS-based prediction filtering scheme offers the best performance and is within 5% of the baseline configuration whilst achieving comparable performance in high-mobility scenarios. Considering the above discussion together with the computational complexity of the adaptive filters, the CQI-LMS-based prediction filtering scheme achieves the best compromise between the selected performance metrics.

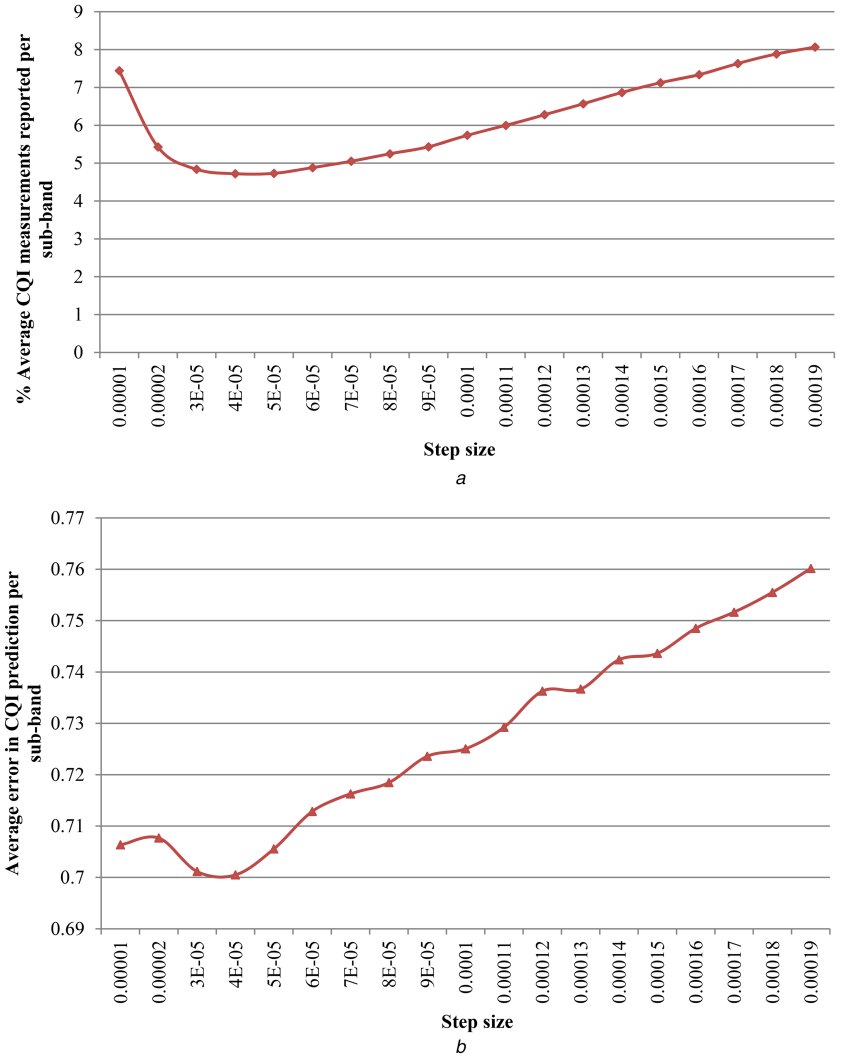



**Fig. 4** *LMS adaptive filter parameter tuning*

***(a)*** Average proportion of CQI measurements reported to eNB from total per sub-band, ***(b)*** Average error of the predicted CQI value per sub-band

**Table 1** Simulation parameters

| Parameter | Value |
|---|---|
| system bandwidth | 10 MHz (50 PRB) |
| scheduler | proportional fair |
| number of antennas | 1 (Tx and Rx) |
| CQI type | periodic, eNB configured sub-band size = 5 |
| traffic model | full buffer |
| frequency | 2 GHz |
| frequency re-use | 1 |
| cellular layout | hexagonal grid, 19 cells |
| inter-site distance | 500 m |
| UE speed | 3 km/h, 120 km/h |
| mobility model | random direction |
| UE per cell | 10–40 UEs |
| $T_{max}$ | 100 |
| propagation model | typical urban $L = 128.1 + 37.6\log(d)$ at 2 GHz |
| fast fading realisation | Jakes model |

**Table 2** Optimal parameter set for the evaluated predictive filtering schemes

| Parameter | CQI-LMS | CQI-NLMS | SINR-LMS | SINR-NLMS |
|---|---|---|---|---|
| filter order $M$ | 4 | 4 | 6 | 6 |
| step size ($\mu$ or $\mu_T$) | 0.00005 | 0.01 | 0.000008 | 0.0025 |
| hysteresis $\beta$ | 3 | 3 | 4 | 4 |
| $E_{max}$ | 1.5 | 1.5 | 4 | 4 |

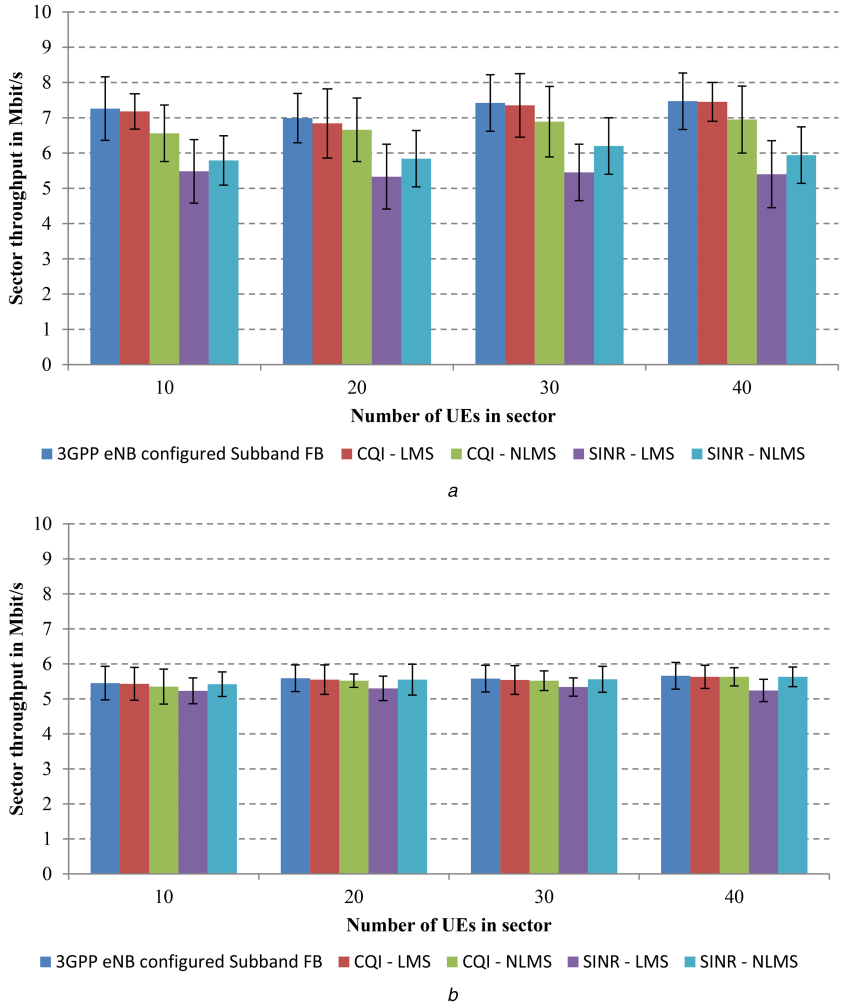



**Fig. 5** *Sector throughput comparison (standard deviation is indicated by vertical bars)*
***(a)*** UE speed: 3 km/h, ***(b)*** UE speed: 120 km/h

**Table 3** Feedback reduction

| UE speed | Prediction filtering scheme | Number of UEs in sector | | | | Average |
|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | |
| 3 km/h (low mobility) | CQI-LMS | 91.8% (±0.4) | 92.4% (±0.5) | 92.6% (±0.2) | 92.6% (±0.2) | 92.7% (±0.4) |
| | CQI-NLMS | 92.6% (±1.2) | 93.1% (±0.4) | 93.1% (±0.4) | 93.1% (±0.6) | |
| | SINR-LMS | 92.1% (±0.7) | 91.4% (±1.1) | 91.5% (±0.6) | 91.8% (±1.0) | 91.8% (±0.4) |
| | SINR-NLMS | 91.7% (±1.4) | 92.1% (±0.9) | 91.2% (±1.1) | 92.5% (±0.8) | |
| 120 km/h (high mobility) | CQI-LMS | 92.6% (±0.2) | 92.4% (±0.2) | 92.5% (±0.2) | 92.4% (±0.2) | 93.1% (±0.7) |
| | CQI-NLMS | 94% (±0.2) | 93.6% (±0.1) | 93.6% (±0.1) | 93.9% (±0.1) | |
| | SINR-LMS | 86.2% (±1.2) | 85.1% (±0.8) | 85.1% (±0.4) | 85.3% (±0.4) | 86.4% (±1.1) |
| | SINR-NLMS | 86.8% (±1.3) | 87.3% (±1.0) | 87.9% (±0.8) | 87.5% (±0.7) | |

**Table 4** Packet loss rate

| Prediction filtering scheme | UE speed | |
|---|---|---|
| | 3 km/h (low mobility) | 120 km/h (high mobility) |
| 3GPP eNB configured sub-band | 2.5% (±0.3) | 17.4% (±1.7) |
| CQI-LMS | 6.7% (±0.7) | 17% (±2.3) |
| CQI-NLMS | 11.6% (±1.4) | 19.9% (±1.7) |
| SINR-LMS | 20.9% (±1.2) | 21.2% (±2.0) |
| SINR-NLMS | 15.1% (±1.1) | 27.4% (±3.2) |

### *3.3 Complexity analysis of the CQI-LMS-based prediction filtering scheme*

The LMS adaptive filter algorithm requires $2M+1$ multiplications and $2M$ additions per measurement when operating in the training phase. Considering that an LMS adaptive filter is required for each user in the cell per sub-band, then the additional computational load in the eNB and UE are

$$C_{\text{enb}} = \frac{1000}{\tau_p}(4M+1)N_{\text{sub}}T_u \tag{9}$$

$$C_{ue} = \frac{1000}{\tau_p}(4M+1)N_{\text{sub}} \tag{10}$$

where $M$ is the filter order, $N_{\text{sub}}$ is the number of sub-bands, $T_u$ is the number of simultaneous active users in the cell and $\tau_p$ is the feedback reporting period which is commonly set to 20 ms in commercial LTE networks. Using the parameters listed in Tables 1 and 2 for the LMS-CQI-based scheme and assuming $T_u$ to be 200, the proposed scheme results in a modest additional 1.7 Mflops and

**Table 5** eNB and UE memory footprint

| Variable | eNB memory footprint (bytes) | UE memory footprint (bytes) |
|---|---|---|
| $k$ | 4 | 4 |
| mode | 1 | 1 |
| txCounter | 4 | 4 |
| errorCounter | 4 | 4 |
| $x[k]$ | 4 | 4 |
| $e[k]$ | 4 | 4 |
| filter weights size $M$ | $4M$ | $4M$ |
| sub-bands | 10 | 10 |
| users | 200 | 1 |
| total memory footprint | 74 kbytes | 0.37 kbytes |

8.5 Kflops in the eNB and UE, respectively. This calculation assumes that the algorithm is always operating in training phase. However, in practice, the majority of the time it is in the prediction phase as evidenced by the significant feedback reduction achieved. In this phase, the computational complexity in the eNB and UE, given by (11) and (12), is reduced significantly as no weight update operations are carried out

$$C_{\text{enb}} = \frac{1000}{\tau_p}(2M-1)N_{\text{sub}}T_u \tag{11}$$

$$C_{ue} = \frac{1000}{\tau_p}(2M)N_{\text{sub}} \tag{12}$$

Thus, considering a worst-case scenario where the algorithm operates in the training phase for 20% of the time, the overall computational complexity is reduced to 0.9 Mflops and 4 Kflops in the eNB and UE, respectively.

An analysis of the memory requirements was also carried out. With reference to Algorithm 1 and using the IEEE 754-2008 single precision floating point format [28] for all stored variable requirements, the memory footprint required for both the UE and eNB is very low as shown in Table 5. It should be noted that the variable named *mode* is essentially a binary flag to indicate whether the algorithm is operating in prediction or training mode. However, as the use of bit flags induces the use of a mask to read/write from/to memory, the authors opted for a more computationally efficient method to represent this variable by using a byte representation thereby avoiding additional masking operations.

### *3.4 Reduced-complexity CQI-LMS-based prediction filtering scheme*

Even though the complexity of the proposed scheme is very low compared with other operations carried out by the eNB baseband processor (such as the fast Fourier transform and its inverse commonly used to synthesise the OFDM symbols in a radio frame), the authors studied the possibility of further lowering this and its effect on the overall performance. As most of the computations are carried out in the weight update process, the use of a sign function, defined by (13), to clip the input vector in the weight update function given by (14) was investigated. When the tap input vector for time sample $k$ is less than the specified threshold $\delta$, $\underline{\boldsymbol{x}}[k]$ will be equal to zero and no coefficient adaptation for the corresponding weights needs to be performed leading to a reduction in the computational load

$$\text{sign}(\underline{\boldsymbol{x}}[\boldsymbol{k}], \delta) = \begin{cases} \underline{\boldsymbol{x}}[k], & \text{if } x[k] \geq \delta \\ 0, & \text{if } x[k] < \delta \end{cases} \tag{13}$$

$$\underline{\boldsymbol{w}}[k+1] = \underline{\boldsymbol{w}}[k] + \mu e[k]\,\text{sign}(\underline{x}[k], \delta) \tag{14}$$

Using the parameter set for the CQI-LMS-based prediction filtering scheme given in Table 2, simulations were carried out in MATLAB® to study the effect of $\delta$ in relation to (i) CQI feedback, (ii) the computational load saving when compared with the CQI-LMS-based prediction filtering scheme and (iii) the average error between the actual and predicted CQI values. The effects of varying this parameter on (i), (ii) and (iii) are shown in Figs. 6*a*–*c*, respectively, where setting $\delta$ to 4 yields the best trade-off between the selected metrics. With reference to Figs. 6*a*–*c*, as the threshold $\delta$ increases, a larger proportion of the input vector $\underline{\boldsymbol{x}}[k]$ will be equal to zero leading to a reduction in the computational load. However, this comes at the expense of an increase in the CQI prediction error as the adaptive filter is not updated regularly which would in turn result in an increase in CQI measurements being reported to the eNB.

Similar to the analysis carried out in Section 3.2, the sign clipped CQI-LMS-based predictive filtering scheme was implemented in LTE-Sim [24] and Monte Carlo simulations, using the parameters in Table 1, and were carried out over 50 random seeds. A comparison of the mean and standard deviation (shown in parenthesis) of the PLR, feedback reduction, sector throughput and computational load reduction for different UE speeds and cell loading with the CQI-LMS-based prediction filtering scheme are shown in Tables 6 and 7 and Fig. 7, respectively.

These results show an average computation load reduction of 35 and 26% at 3 and 120 km/h, respectively. However, this saving comes at the expense of a marginal degradation in packet loss, feedback reduction and sector throughput performance. A summary of the identified key performance indicators for the proposed schemes compared with both the 3GPP eNB configured sub-band scheme and the work in [16], which represent the latest work in this field, is shown in Table 8. The feedback reduction and throughput degradation is computed with reference to the 3GPP eNB configured sub-band feedback scheme, whereas the complexity reduction of the sign clipped LMS-CQI prediction filtering scheme is computed with reference to the LMS-CQI prediction filtering scheme. The proposed LMS-CQI and sign clipped LMS-CQI prediction filtering schemes are on average 14% better when compared with [16]. This improvement comes at the expense of degradation in PLR which is particularly notable at high mobility. Nevertheless, at high mobility, the proposed techniques result in statistically the same level of performance when compared with the 3GPP eNB configured sub-band feedback scheme.

The scheme in [16] makes use of two algorithms working in tandem. The first algorithm estimates the CQI using a GP regression technique which requires the use of a multivariate optimisation algorithm to determine a set of hyper-parameters. Although details of this algorithm are missing, the complexity of this process on its own requires several matrix operations per measurement for each user and sub-band. The second algorithm uses an adaptive dual-control method for the dynamic CQI feedback time window assignment. This requires several iterations per measurement for each user resulting in this scheme to rank high on computational costs and memory complexity.

### *3.5 Testing using field test measurements*

Field test measurements were carried out on a live commercial LTE network using a purposely setup test vehicle equipped with industry standard test equipment [29]. In accordance with [30], the UE is connected to a pair of omni-directional antennas placed in a roof box on top of the test vehicle through a 12 dB attenuator to compensate for the antenna gain and vehicle penetration loss. The field test route, shown in Fig. 8*a*, was purposely selected and is characterised by dense urban environment and high data traffic usage yielding a rich multipath channel with highly varying radio conditions. The CQI values reported to the network throughout the drive test were collected and used as input to the CQI-LMS-based prediction filtering schemes described in Section 2. Out of the 13,944 CQI measurements reported when using the 3GPP standardised method, only 2383 were transmitted. This represents a saving of ∼83% in feedback signalling whilst maintaining a CQI error of 1.2. A sample of the actual and predicted measurements, shown in Fig. 8*b*, indicates the operation of the two phases described in Section 2. During prediction, measurements are not

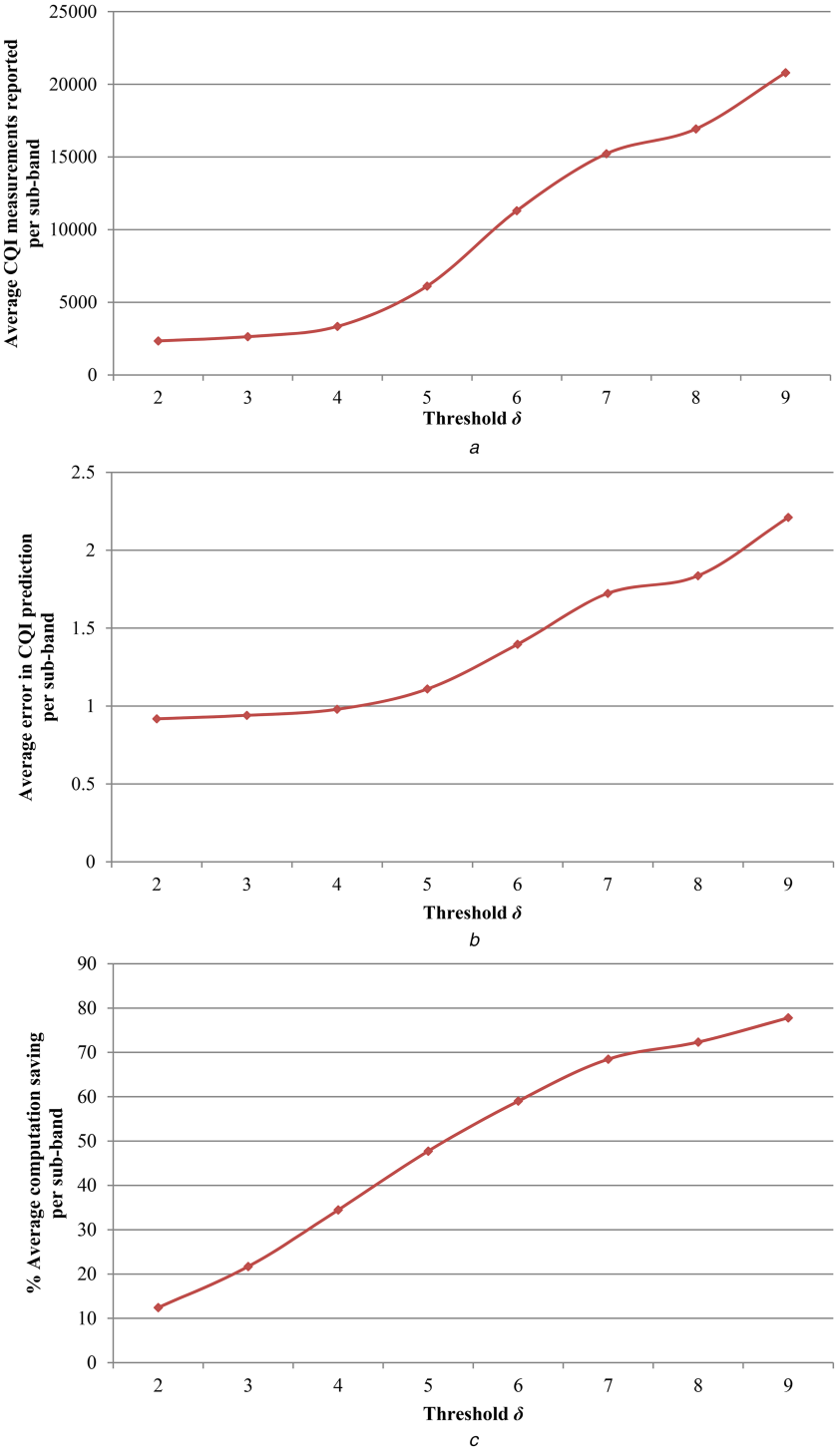



**Fig. 6** *Sign clipped LMS adaptive filter threshold parameter tuning*

***(a)*** Average CQI measurements reported per sub-band, ***(b)*** Average error of the predicted CQI value per sub-band, ***(c)*** Average % computation saving relative to the standard LMS adaptive filter per sub-band

transmitted and the predictors track the channel conditions while keeping record of the error between the actual and predicted measurements. The error gradually increases between samples 5 and 9 as a result of a fast fade in channel conditions. This triggers the algorithm to switch back to training mode during which measurements are transmitted to the eNB and the adaptive filters are re-trained using the latest channel measurements. Prediction mode is resumed as of sample 14 which marks the point at which the error falls below the error budget threshold and transmission of feedback ceases.

## 4 Conclusion

In this paper, we presented a UE-assisted sub-band feedback compression scheme based on predictive filtering techniques. It exploits short-term channel prediction through the use of low computational complexity adaptive filters at the UE and eNB side, to reduce the feedback signalling overhead. Four schemes were designed and tested with the CQI-LMS-based predictive filtering scheme providing the best compromise between the identified performance metrics with up to 92.5% CQI feedback reduction whilst maintaining the same level of sector throughput compared with the standardised 3GPP eNB configured sub-band feedback scheme and keeping the PLR within 10%. A reduced-complexity scheme achieving an average computational load reduction of 35% in low mobility is also presented. This was achieved at the expense of a marginal degradation in performance compared with the CQI-

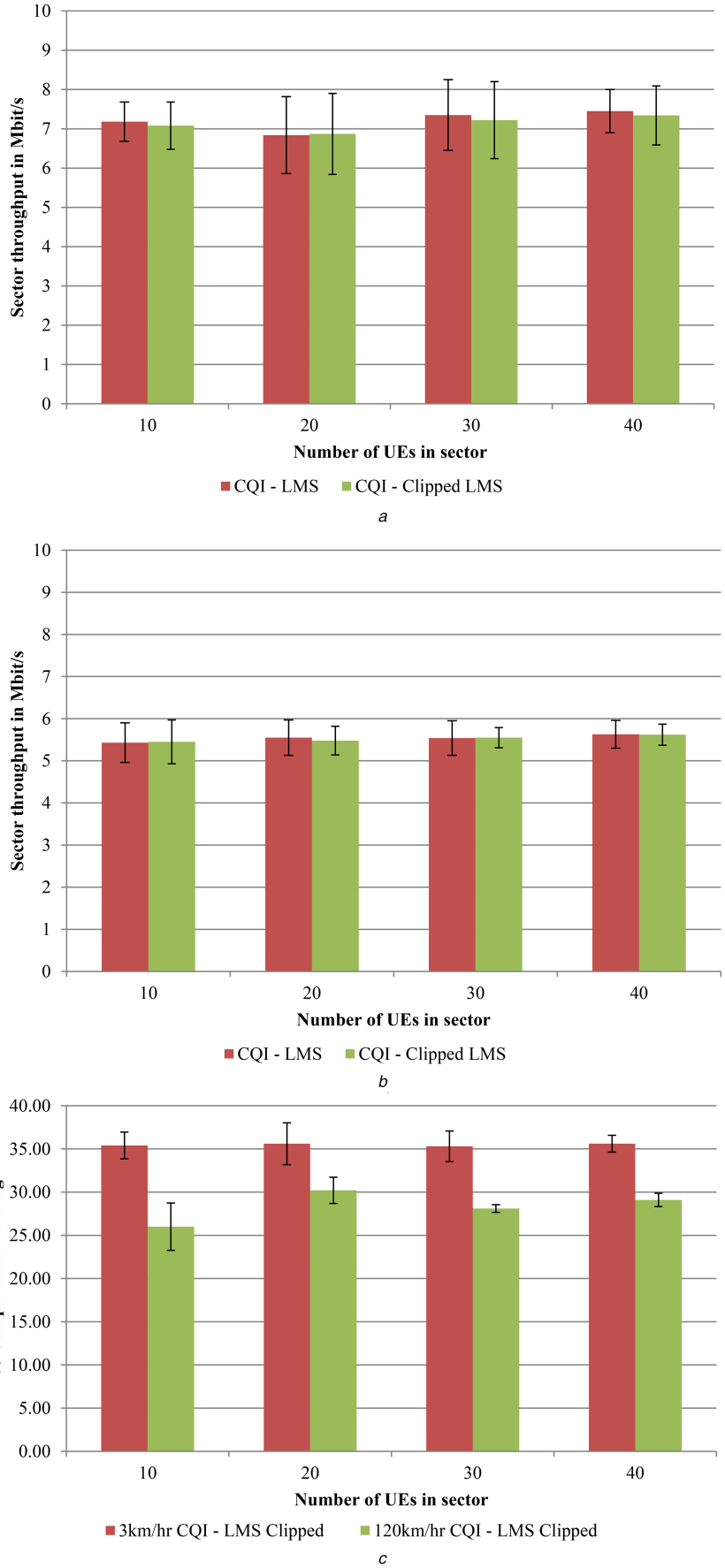



**Fig. 7** *Sector throughput and computational load comparison (standard deviation is indicated by vertical bars)*
***(a)*** Sector throughput at 3 km/h, ***(b)*** Sector throughput at 120 km/h, ***(c)*** % computational load reduction

LMS-based scheme. A comparative analysis showed that the proposed techniques represent a significant improvement compared with [16]. This reduction in the uplink CQI feedback allows the relaxation of the PUCCH dimensioning leaving more resources for the user data to be sent over the PUSCH coupled with substantial UE power saving. Future work will study the use of content

**Table 6** Packet loss rate

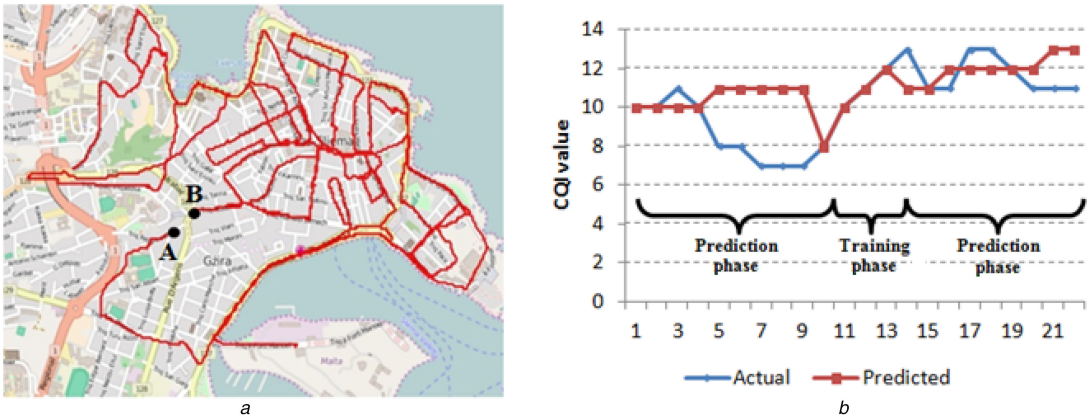
| Prediction filtering scheme | UE speed | |
|---|---|---|
| | 3 km/h (low mobility) | 120 km/h (high mobility) |
| CQI-LMS | 6.7% (±0.7) | 17% (±2.3) |
| sign clipped CQI-LMS | 8.6% (±1.3) | 17.2% (±1.3) |

**Table 7** Feedback reduction

| UE speed | Prediction filtering scheme | Number of UEs in sector | | | | Average |
|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | |
| 3 km/h (low mobility) | CQI-LMS | 91.8% (±0.4) | 92.4% (±0.5) | 92.6% (±0.2) | 92.6% (±0.2) | 92.40% (±0.38) |
| | sign clipped CQI-LMS | 89.4% (±0.3) | 90.1% (±0.5) | 90.2% (±0.3) | 89.9% (±0.3) | 89.90% (±0.36) |
| 120 km/h (high mobility) | CQI-LMS | 92.6% (±0.2) | 92.4% (±0.2) | 92.5% (±0.2) | 92.4% (±0.2) | 92.50% (±0.1) |
| | sign clipped CQI-LMS | 88.5% (±0.3) | 88.9% (±0.3) | 89.3% (±0.2) | 89.2% (±0.1) | 89.00% (±0.36) |

**Table 8** Comparison of the key performance indicators

| Scheme | Mobility characteristics | Feedback reduction | PLR | Throughput degradation | Complexity |
|---|---|---|---|---|---|
| 3GPP eNB configured sub-band feedback | low mobility | — | 2.5% (±0.3) | — | — |
| | high mobility | — | 17.4% (±1.7) | — | — |
| Chiumento *et al.* [16] | mixed – generally low mobility | 77% | 10% | focus of study is on PLR | high. See text for further details |
| LMS-CQI prediction filtering scheme | low mobility | 92.4% (±0.4) | 6.7% (±0.7) | 1.1% (±0.7) | 0.9 Mflops and 4 Kflops in the eNB and UE, respectively |
| | high mobility | 92.5% (±0.1) | 17% (±2.3) | 0.6% (±0.2) | |
| sign clipped LMS-CQI prediction filtering scheme | low mobility | 89.9% (±0.4) | 8.6% (±1.3) | 2.2% (±0.5) | −35.5% |
| | high mobility | 89% (±0.4) | 17.2% (±1.3) | 0.8% (±0.8) | −28.4% |




*a* *b*

**Fig. 8** *Field test measurements*
***(a)*** Test route – starting at point A and stopping at point B, ***(b)*** Actual versus predicted CQI measurements during field testing

awareness in the proposed technique to control the PLR during the transmission of critical data.

## 5 References


[1] 3rd Generation Partnership Project: 'Evolved universal terrestrial radio access (E-UTRA): physical layer procedures (Release 8)', 2008

[2] Holma, H., Toskala, A.: '*LTE for UMTS – OFDMA and SC-FDMA based radio access*' (Wiley, 2009)

[3] Cisco Visual Networking Index: 'Global mobile data traffic forecast update, 2015–2020 white paper'. Available at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, accessed 15 August 2016

[4] Sesia, S., Toufik, I., Baker, M.: '*LTE – the UMTS long term evolution: from theory to practice*' (Wiley, Chichester, 2009)

[5] Pauli, M., Wachsmann, U., Tsai, S.: 'Quality determination for a wireless communications link'. U.S. Patent 2004/0219883, November 2004

[6] Westman, E.: 'Calibration and evaluation of the exponential effective SINR mapping (EESM) in 802.16'. Master thesis, The Royal Institute of Technology (KTH), Stockholm, Sweden, 2006

[7] Fa-tang, C., Gen-lin, T.: 'A novel MCS selection criterion for supporting AMC in LTE system'. Proc. Int. Conf. Computer Application and System Modeling, October 2010, vol. **6**, pp. 598–603

[8] 3GPP: 'R1-061777 – DCT based CQI reporting scheme'. LG Electronics, RAN1 LTE Ad Hoc, Cannes, Franes, June 2006

[9] 3GPP: 'R1-070187 – DCT partitioning for CQI reporting'. Panasonic, RAN1 Meeting 47, Sorrento, Italy, January 2007

[10] Haghighat, A., Zhang, G., Lin, Z.: 'Full-band CQI feedback by Haar compression in OFDMA systems'. Proc. Conf. Vehicular Technology, September 2009, pp. 1–5

[11] Kang, M., Soon, K.: 'Performance analysis and optimization of best-M feedback for OFDMA systems', *IEEE Commun. Lett.*, 2012, **16**, (10), pp. 1648–1651

[12] Werthmann, T., Kaschub, M., Proebster, M., *et al.*: 'Simple channel predictors for lookahead scheduling'. Proc. Conf. Vehicular Technology, May 2012, pp. 1–6

[13] Dai, H., Wang, Y., Shi, C., *et al.*: 'The evaluation of CQI delay compensation schemes based on Jake's model and ITU scenarios'. Proc. Conf. Vehicular Technology, September 2012, pp. 1–5

[14] Xu, X., Ni, M., Mathar, R.: 'Improving QoS by predictive channel quality feedback for LTE'. Proc. Int. Conf. Software, Telecommunications and Computer Networks, September 2013, pp. 1–5

[15] Ni, M., Xu, X., Mathar, R.: 'A channel feedback model with robust SINR prediction for LTE systems'. Proc. Conf. Antennas & Propagation, April 2013, pp. 1866–1870

[16] Chiumento, A., Bennis, M., Desset, C., *et al.*: 'Adaptive CSI and feedback estimation in LTE and beyond: a Gaussian process regression approach', *EURASIP J. Wirel. Commun. Netw.*, 2015, **1**, pp. 1–14

[17] Sivridis, L., He, J.: 'A strategy to reduce the signaling requirements of CQI feedback schemes', *J. Wirel. Pers. Commun.*, 2013, **70**, (1), pp. 85–98

[18] Gesbert, D., Alouini, M.S.: 'How much feedback is multi-user diversity really worth?'. Int. Conf. Communications, June 2004, pp. 234–238

[19] Sivridis, L., Wang, X., Choi, J.: 'Radio resource management for fast fading environments', *J. Commun.*, 2011, **6**, (7), pp. 540–548
[20] Hoon, K., Youngnam, H.: 'An opportunistic channel quality feedback scheme for proportional fair scheduling', *IEEE Commun. Lett.*, 2007, **11**, (3), pp. 501–503
[21] Kittipiyakul, S., Boonkajay, A., Aphichartsuphapkhajorn, T.: 'Performance of low-feedback rate, gradient-based OFDMA subcarrier allocation with partial channel information'. Proc. Symp. Wireless Pervasive Computing, May 2010, pp. 511–521
[22] Jeon, J., Son, K., Lee, H.-W., *et al.*: 'Feedback reduction for multiuser OFDM systems', *Trans. Veh. Technol.*, 2010, **59**, (1), pp. 160–169
[23] Abdulhasan, M., Salman, M., Ng, C., *et al.*: 'An adaptive threshold feedback compression scheme based on channel quality indicator (CQI) in long term evolution (LTE) system', *J. Wirel. Pers. Commun.*, 2015, **82**, (4), pp. 2323–2349
[24] Piro, G., Grieco, L., Boggia, C.F., *et al.*: 'Simulating LTE cellular systems: an open source framework', *Trans. Veh. Technol.*, 2011, **60**, (2), pp. 498–513
[25] Haykin, S.: '*Adaptive filter theory*' (Prentice Hall, 2004)
[26] Sayed, A.H.: '*Fundamentals of adaptive filtering*' (Wiley, 2003)
[27] Haykin, S., Widrow, B.: '*Least mean square adaptive filter*' (Wiley-Interscience, 2003)
[28] IEEE 754-2008: 'IEEE standard for floating-point arithmetic', 2008
[29] 'Swissqual Field Test Equipment'. Available at http://www.swissqual.com/en/products/benchmarking/diversity-benchmarker-2/, accessed 15 August 2016
[30] ETSI TR 102.581: 'Speed processing, transmission and quality aspects (STQ); a study on the minimum additional required attenuation on the antenna path of the field test equipment', 2007

# A Support Vector Machine Based Sub-Band CQI Feedback Compression Scheme for 3GPP LTE Systems

Mario Cordina, and Carl J. Debono
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
m.r.cordina@ieee.org, c.debono@ieee.org

***Abstract*— Contemporary wireless communication standards, such as the long term evolution (LTE) standard, exploit several techniques, including link adaptation and frequency selective scheduling (FSS), to offer high data rate services. The efficacy of these techniques rely on the evolved Node B (eNB) having accurate channel state information through the use of a high signaling overhead process whereby channel quality indicator (CQI) feedback reports are sent by the user equipment (UE) to the eNB. In this work, we exploit a machine learning technique to address this problem and propose a novel sub-band CQI feedback compression scheme based on support vector machines to reduce this signaling overhead. The proposed compression scheme was implemented and tested in an LTE system level simulator and has shown efficacy with an overall CQI feedback signaling reduction of up to 88.7% whilst maintaining stable sector throughput, when compared to the standard third generation partnership project (3GPP) CQI feedback mechanism.**

***Keywords—CQI feedback; 3GPP LTE; machine learning; frequency selective scheduling; support vector machine***

## I. INTRODUCTION

The third generation partnership project (3GPP) long term evolution (LTE) system, initially standardized in 2008 and further improved in subsequent 3GPP releases [1], achieves high data rates through the use of several techniques which include frequency domain scheduling (FDS) [2] and adaptive modulation and coding (AMC) [2]. These techniques rely on the evolved Node B (eNB) to acquire accurate channel state information in the form of channel quality indicator (CQI) feedback reports sent by the user equipment (UE) to the evolved Node B (eNB). To be able to fully exploit the advantages of FDS, each UE in the sector has to report a CQI value for each physical resource block (PRB) to the eNB. This feedback process ensures that the eNB obtains an accurate state of the channel conditions across all the PRBs. However, this generates a significant uplink signaling overhead. This is particularly concerning in hotspot traffic areas, such as shopping malls, stadiums and event arenas, where the number of UEs in the sector is generally high thereby giving rise to congestion in the uplink control channel, leading to degradation in the overall sector performance. Moreover, considering that according to [3] the average smartphone traffic in 2020 will be close to 5 times greater than 2015, this uplink feedback overhead will be a significant problem in mobile networks.

The design of low overhead CQI feedback techniques has attracted significant research attention. Several schemes were proposed in literature with different levels of compression and system performance whilst a sub-set of these were selected by 3GPP and included in the LTE release standard [2]. In this paper, we propose a novel sub-band CQI feedback compression scheme based on a machine learning technique. To the best knowledge of the authors, this work represents the first study on the application of support vector machines (SVM) to CQI feedback compression. Although the focus of this work is on the 3GPP LTE standard, being the cellular network technology at the forefront of mobile network systems, this work can be applied to any OFDMA-based cellular network technology.

The rest of the paper is organized as follows: Section 2 provides a brief description of CQI feedback reporting and compression techniques; Section 3 describes the proposed scheme together with an overview of support vector machines; Section 4 gives a description of the simulation environment, parameter selection and results, while Section 5 provides some comments and conclusions.

## II. RELATED WORK AND OUR CONTRIBUTION

In LTE, the CQI feedback is calculated by the UE using the signal-to-interference noise ratio (SINR) measurements of the downlink reference signals transmitted by the eNB. The received SINR for each PRB $k$ is given by [2]:

$$\gamma_k = \frac{P_{i,k} G_{i,k}}{\sum_{j \neq i} P_{j,k} G_{j,k} + n_k} \tag{1}$$

where $P_i$ and $G_i$ are the transmit power and antenna gain of the serving base station $i$ whilst $P_j$ and $G_j$ are the transmit power and antenna gain of the interfering base station $j$, and $n_k$ is the additive Gaussian noise. The measured SINR is then mapped to a discrete CQI value using a chipset vendor specific mapping table. This gives the highest modulation coding

scheme (MCS) that the UE can decode with a block error rate not exceeding 10% [2].

In order to control the signaling overhead, several techniques have been proposed in literature, which can be grouped into three main types, namely threshold/prediction based, time domain (periodic or aperiodic reporting [4]) and frequency domain based techniques. The first category use either channel prediction or event based triggering to reduce the feedback. In [5], Chiumento *et al.* propose the use of a Gaussian process (GP) regression at the base station to adaptively estimate the channel quality. This GP-based CQI prediction method is integrated in a dual-control technique utilizing active learning to determine an optimal prediction time for each user and is reported to achieve a 77% signalling overhead reduction when compared to the 3GPP standard eNB-configured sub-band feedback compression technique. Inspired by the work found in [6-10], the authors in [11] apply a threshold-based signalling reduction technique whereby users which exceed a set threshold, based on the likelihood that the user will be scheduled, are candidates to use full-band feedback whilst the rest use wide-band frequency domain compression. An additional selection criterion is applied on the candidates earmarked to use full-band feedback such that only those which have the lowest average SINR will be requested by the network to use periodic full-band feedback. The reported simulation results show that this scheme achieves marginal gains and is mostly suitable for low mobility scenarios, where channel conditions are largely stable. Meanwhile, in [12], the authors address the shortcomings of threshold-based approaches by proposing an adaptive threshold feedback compression scheme.

Several frequency domain based feedback compression techniques were proposed in recent years and these can be broadly divided into three sub-categories namely wide-band, full-band and sub-band compression techniques.

### A. Wide-band Compression

In wide-band compression, a UE transmits a single CQI value for all the PRBs in the bandwidth. As different PRBs undergo different fading characteristics, the effective SINR $\gamma_{eff}$ is computed and transmitted to represent a number of PRBs. The effective exponential signal-to-noise ratio mapping (EESM) [13,14] is given by:

$$\gamma_{eff} = -\lambda \log\left(\frac{1}{N}\sum_{k=1}^{N} \mathrm{e}^{-\frac{\gamma_k}{\lambda}}\right) \quad (2)$$

where $N$ represents the number of PRBs to group, $\lambda$ is a calibration factor that is MCS dependent [15] and $\gamma_k$ is the instantaneous SINR of the $k^{th}$ PRB. In the case of wide-band compression, $N$ is set to the total number of PRBs defined by the system bandwidth. The effective SINR is then mapped to a discrete CQI value as discussed earlier. Even though wide-band compression is attractive due to its simplicity and high compression factor, it offers a lower system performance due to the lack of channel quality detail. Nevertheless, it forms part of the 3GPP LTE standard and is widely used in live commercial networks.

### B. Full-band Compression

Full-band compression techniques generally involve the use of mathematical transforms such as the discrete cosine transform (DCT) [16][17] and the Haar wavelet transform [18] to reduce the signaling overhead. Although these techniques allow the eNB to compute a good approximation of the entire system bandwidth quality using a relatively small number of transform coefficients reported by the UE, the reconstruction quality strongly depends on the delay spread of the channel. Full-band CQI reporting is not included in the 3GPP LTE standard thereby limiting its practical applicability.

### C. Sub-band Compression

In sub-band compression, a UE transmits a selected set of CQI values to the eNB. Such techniques include Best-M individual [19] and Best-M average [19] together with the 3GPP defined sub-band compression techniques, namely:

- eNB-configured sub-band feedback [2]: The system bandwidth is divided into $q$ sub-bands of $N$ consecutive PRBs and the UE reports a wideband CQI value together with a differentially encoded CQI value for each sub-band.
- UE-selected sub-band feedback [2]: Similar to eNB-configured sub-band feedback, the system bandwidth is divided into $q$ sub-bands of $N$ consecutive PRBs. The UE selects $M$ preferred sub-bands and reports a wide-band CQI value together with a differentially encoded CQI value reflecting the average quality of the $M$ preferred sub-bands. The position of the selected $M$ preferred sub-bands also needs to be reported thereby increasing the signaling overhead.

Similar to wide-band compression, the effective SINR needs to be computed using (2) and then mapped to a discrete CQI value as discussed earlier. In general, sub-band compression techniques allow the eNB to have a richer channel state information across a wider frequency band thereby leading to a better system performance when compared to wide-band CQI reporting. However, this comes at the expense of a higher signaling overhead on the physical uplink shared channel (PUSCH) or physical uplink control channel (PUCCH) depending on whether or not the UE has uplink data scheduled to be transmitted [2].

In this work, the authors explore the use of SVM in sub-band CQI feedback compression. SVMs have already been applied in various areas of wireless communication systems ranging from uplink and downlink LTE channel interpolation [20-22] to LTE uplink channel extrapolation for the prediction of channel state information (CSI) [23]. In contrast to previous work, we apply the functional approximation properties of SVM to model the sub-band channel response. The support vector weights together with their positions are then transmitted to the eNB, instead of the individual sub-band CQI values, to recover the entire sub-band channel response within a configurable margin of error. The same technique can also be applied to the full-band reporting scenario however this work will focus on the sub-band reporting; this being standardized by the 3GPP LTE standard.

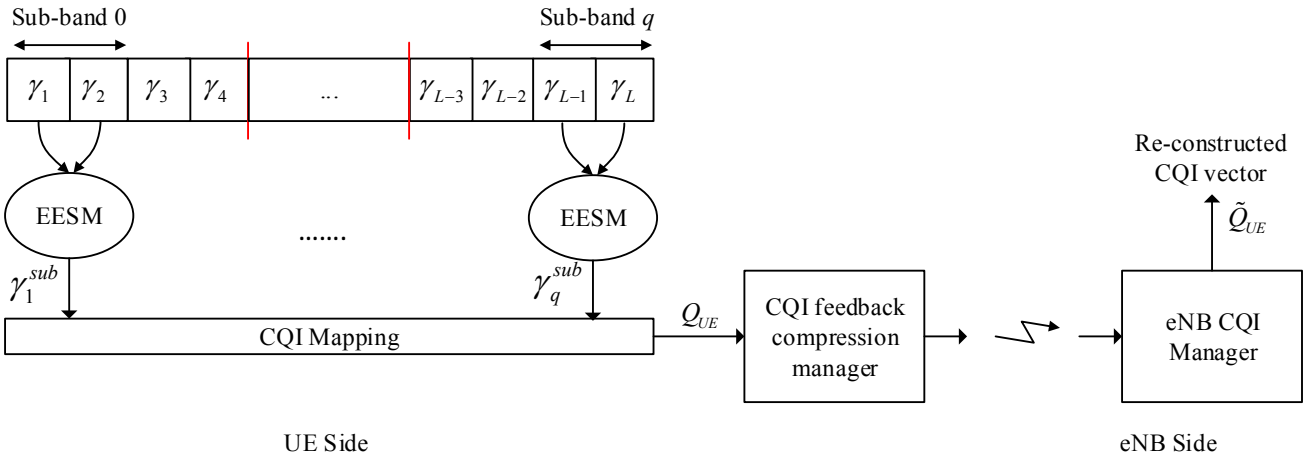



Figure 1: SVM Compression System Model (N=2)

## III. Proposed Scheme

With reference to the system model shown in Fig. 1, consider that the system bandwidth composed of $L$ PRBs, where $L$ is defined by the 3GPP Release 8 standard [1], is divided into $q$ sub-bands of $N$ consecutive PRBs. A UE measures the SINR of each PRB, denoted by $\gamma_L$, and calculates the effective SINR of each sub-band, denoted by $\gamma_q^{sub}$, using (2). This effective SINR, $\gamma_q^{sub}$, is then mapped to a discrete CQI value using an SINR-CQI mapping table defined in [24], $CQI_q^{sub}$, and aggregated to form the CQI vector, $\mathbf{Q}_{UE}$, for further processing by the CQI feedback compression manager. The role of the feedback compression manager is to:

- Determine the best compression method to use given the properties of $\mathbf{Q_{UE}}$. In case the CQI vector exhibits flat channel conditions, we bypass the SVM compression stage and transmit a single CQI value to the eNB. Otherwise, SVM compression is used.
- In case SVM compression is used, we first normalize $\mathbf{Q_{UE}}$ in the range {0,1} by dividing each element in the vector by 15 prior to feeding it to the SVM. The output of the SVM is a set of support vector weights, $\mathbf{w}$, together with their corresponding locations which are transmitted to the eNB. In case the SVM does not yield non-zero support vector weights (i.e. approximation function cannot be fitted), a UE-selected sub-band feedback method is adopted. The eNB CQI manager uses the received information to reconstruct the CQI vector, $\mathbf{Q_{UE}}$, and forwards it to the resource scheduler.

The SVM technique was developed by Vapnik *et al* [25] to primarily address classification problems however; this was later extended to the regression domain [26]. In this work, we use this functionality to approximate the measured CQI vector, $\mathbf{Q_{UE}}$, with:

$$f(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{v} w_i \phi_i(\mathbf{x}) \tag{3}$$

where the functions $\phi_i(\mathbf{x})$ are referred to as kernel functions, $v$ is the number of support vectors, $\mathbf{w}$ is the support vector weights which needs to be determined and $\mathbf{x}$ is the input vector. Using Vapnik's linear loss function [26] with $\varepsilon$-insensitivity zone, the approximation error is given by:

$$\left|\mathbf{Q_{UE}} - f(\mathbf{x},\mathbf{w})\right| = \begin{cases} 0 \text{ if } \left|\mathbf{Q_{UE}} - f(\mathbf{x},\mathbf{w})\right| \le \varepsilon \\ \left|\mathbf{Q_{UE}} - f(\mathbf{x},\mathbf{w})\right| - \varepsilon, \text{ otherwise} \end{cases} \tag{4}$$

where $\varepsilon$ is a configurable parameter defining the maximum allowed error. Thus, the aim of the SVM, is to find the smallest number of weights (support vectors) such that the approximation function has at most $\varepsilon$ deviation from the measured CQI vector $\mathbf{Q_{UE}}$. In this work, the widely used Gaussian kernel function (in 1 dimension) is selected, as it is able to map the input data into a higher dimensional space and was found to outperform other kernel functions (such as the linear kernel) in channel estimation scenarios [23]. Therefore, the function to be modeled is given by:

$$f(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N} w_i e^{-\frac{1}{2}\frac{(\mathbf{x}-c_i)^2}{\lambda_i}} \tag{5}$$

where $c_i$ are the Gaussians centers and $\lambda_i$ is the Gaussian width.

The SVM network is modeled as a Radial Basis Function (RBF) network where the input to the network, $\mathbf{x} = (x_1, x_2, ...., x_q)$, is the sub-band position whilst the output is the measured CQI vector, $\mathbf{Q_{UE}}(\mathbf{x}) = (CQI_1^{sub}, ...., CQI_q^{sub})$. With reference to the detailed mathematical treatment of support vector regression found in [26], the RBF design matrix is given by (6) whilst the non-linear optimization problem to be solved is given by (7).

$$G = \begin{pmatrix} G(x_1, c_1) & G(x_1, c_q) \\ G(x_q, c_1) & G(x_q, c_q) \end{pmatrix} \quad (6)$$

$$\min\left\{-0.5\boldsymbol{\alpha}^T \mathbf{H}\boldsymbol{\alpha} + \mathbf{f}^T\alpha\right\} \quad (7)$$

where the Hessian matrix $\mathbf{H}$ and vector $\mathbf{f}$ are defined by (8) and (9) respectively.

$$\mathbf{H} = \begin{pmatrix} G & -G \\ -G & G \end{pmatrix} \quad (8)$$

$$\mathbf{f} = \begin{pmatrix} \varepsilon - CQI_1^{sub} \\ \varepsilon - CQI_q^{sub} \\ \varepsilon + CQI_1^{sub} \\ \varepsilon + CQI_q^{sub} \end{pmatrix} \quad (9)$$

The solution to (7) is a column vector $\boldsymbol{\alpha}$ composed of *2q* Lagrange multipliers, which are then used to compute the weight vector given by (10)

$$\mathbf{w} = \boldsymbol{\alpha}^* - \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 - \alpha_{q+1} \\ \alpha_q - \alpha_{2q} \end{pmatrix} \quad (10)$$

The non-linear optimization problem in (7) can be solved using either quadratic programming techniques [26] or a non-negative least square method. As the support vector weights need to be computed on the UE side, the authors opted to use the latter to reduce the number of computations. The non-zero weight values calculated in (10) are the support vectors which need to be transmitted together with their corresponding position to the eNB to reconstruct the approximation of the CQI vector, $\mathbf{Q}_{\mathbf{UE}}$.

## IV. Results

Two important SVM parameters that need to be tuned in relation to (i) the normalized mean squared error between the actual and reconstructed CQI vector, and (ii) the average number of reports sent by the UE per CQI vector measurement are:

- Maximum allowed error $\varepsilon$, defining the insensitivity zone.
- Gaussian width, $\lambda$.

To tune these parameters, the proposed scheme, described in section III, was first implemented in MATLAB® and tested using SINR measurements extracted from a large number of UEs simulated in an LTE system level simulator [24]. The parameters used for the LTE simulations are provided in Table I whilst the optimal SVM parameter values are given in Table II. These optimal parameters were empirically determined by considering trade-offs between (i) and (ii) above.

TABLE I. Simulation Parameters

| Parameter | Value |
|---|---|
| System Bandwidth | 10MHz (50 PRB) |
| Scheduler | Proportional Fair |
| Number of Antennas | 1 (Tx and Rx) |
| CQI Type | Periodic, eNB configured sub-band size = 2 |
| Traffic Model | Full Buffer |
| Frequency | 2GHz |
| Frequency Re-use | 1 |
| Cellular Layout | Hexagonal grid, 19 cells |
| Inter-site distance (ISD) | 500m |
| UE speed | 3km/hr, 120km/hr |
| Mobility model | Random Direction |
| UE per Cell | 10-40 UEs |
| $T_{max}$ | 100 |
| Propagation Model | Typical Urban L= 128.1+37.6log(*d*) @ 2GHz |
| Fast Fading Realization | Jakes Model |

TABLE II. Optimal Parameter Set

| Maximum allowed error $\varepsilon$ | Gaussian width $\lambda$ |
|---|---|
| 0.2 | 200 |

The effect of varying these parameters on the average number of reports sent by the UE per CQI vector measurement and the normalized mean squared error between the actual and reconstructed CQI vector are shown in Fig. 2 and 3 respectively. As expected, as the insensitivity zone increases, the number of reports sent by the UE per CQI vector measurement decreases as we are allowing a larger error margin. Conversely, the normalized mean squared error between the actual and reconstructed CQI vector increases. Meanwhile, the Gaussian width parameter was found to be weakly linked to the selected metrics and was set to 200 for the rest of the studies whilst the maximum allowed error, $\varepsilon$, was set to 0.2 as this yields the best tradeoff between the selected metrics.

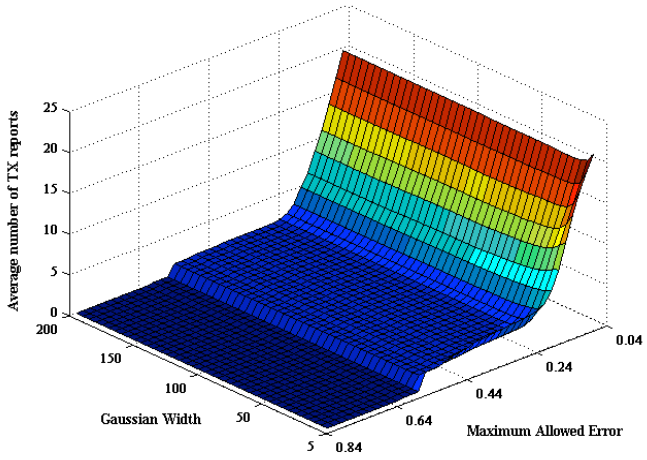



Figure 2: Average number of transmitted reports per CQI measurement

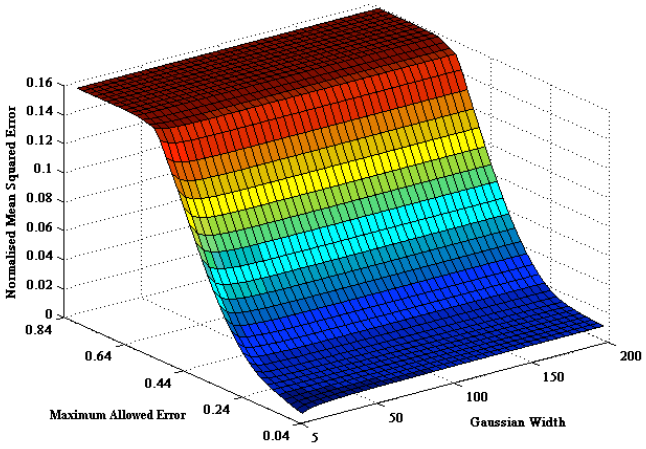



Figure 3: Normalized mean squared error between the actual and reconstructed CQI vector

The proposed SVM based sub-band CQI feedback compression scheme was implemented in LTE-Sim [24] to verify its efficacy in terms of average cell throughput and feedback reduction. Monte Carlo simulations, using the parameters given in Table I and the optimal parameter set for the proposed scheme in Table II, were carried out over 50 random seeds in order to achieve statistical relevant results. The average sector throughput and feedback reduction, when compared to the 3GPP standardized eNB-configured sub-band feedback, for the different simulation scenarios are shown in Fig. 4-5 and Table III respectively.

TABLE III. FEEDBACK REDUCTION

| | | Number of UEs in Sector | | | |
|---|---|---|---|---|---|
| **UE Speed** | **Sub-band Feedback scheme** | **10** | **20** | **30** | **40** |
| 3km/hr | 3GPP UE-selected | 80.1 | 79.9 | 80.0 | 80.0 |
| | Proposed | 86.9 | 86.9 | 87.0 | 87.1 |
| 120km/hr | 3GPP UE-selected | 80.1 | 80.0 | 80.2 | 79.9 |
| | Proposed | 88.6 | 88.8 | 88.7 | 88.8 |

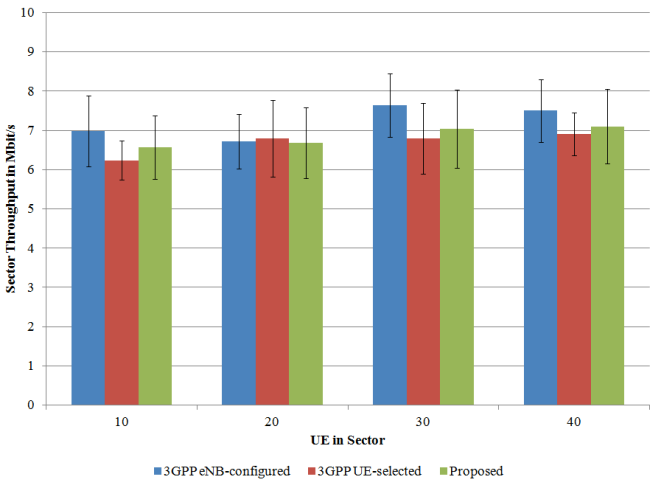



Figure 4: Sector throughput comparison for UEs moving at 3km/hr

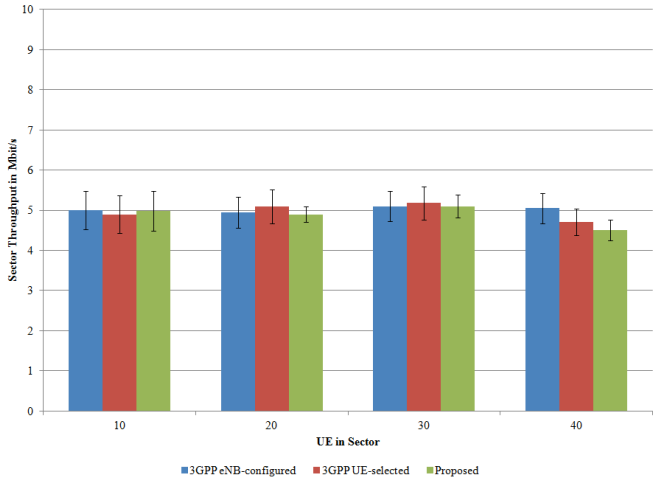



Figure 5: Sector throughput comparison for UEs moving at 120km/hr

The proposed SVM based sub-band CQI feedback compression scheme manages to achieve an average CQI feedback reduction of 87% and 88.7% for UEs moving at 3km/hr and 120km/hr respectively. This level of feedback reduction was achieved whilst maintaining approximately the same level of performance, in terms of sector throughput, as the 3GPP eNB configured and UE-selected sub-band feedback scheme. Moreover, with reference to Fig. 6-7, the CQI feedback compression manager is using SVM compression for more than 80% of the CQI measurements. At higher mobility, the channel variations increase resulting in the CQI feedback manager to use SVM compression for even a higher proportion of the CQI measurements leading to a marginal improvement in feedback reduction when compared to the lower mobility scenario.

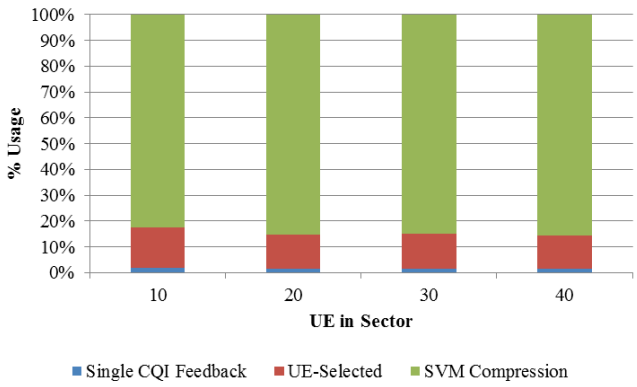



Figure 6: CQI feedback manager operating mode for UEs moving at 3km/hr

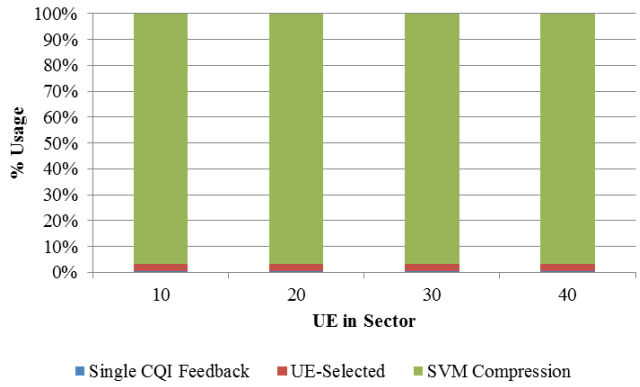



Figure 7: CQI feedback manager operating mode for UEs moving at 120km/hr

## V. Conclusion and Future Work

In this paper we presented a novel sub-band CQI feedback compression scheme based on a machine learning technique. The key contribution lies in the use of a support vector machine, at the UE, designed to reduce the CQI feedback signaling overhead. Experimental results based on Monte Carlo simulations showed that the SVM-based technique provides up to 88.7% CQI feedback reduction whilst maintaining the same level of sector throughput when compared to the standardized 3GPP eNB configured and UE selected sub-band feedback scheme. This represents an improvement of approximately 7% when compared to the 3GPP UE selected sub-band feedback scheme and close to 10% when compared to the most recent work on CQI feedback compression proposed by Chiumento *et al* [5] that achieved 77% reduction in CQI feedback. This reduction in the uplink CQI feedback will not only reduce the potential uplink control channel congestion in hotspot areas but also results in substantial UE power saving. Future work will study the use of the proposed technique in the transport of 3D video together with a thorough assessment of the computational complexity and application in a 5G scenario.

## References


[1] 3rd Generation Partnership Project, "Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Layer Procedures (Release 8)", TS 2008.

[2] H. Holma, and A. Toskala, "LTE for UMTS – OFDMA and SC-FDMA Based Radio Access", Wiley, 2009.

[3] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper", http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, accessed 15 April 2017.

[4] S. Sesia, I. Toufik, and M. Baker, "LTE- The UMTS long term evolution: from theory to practice", Wiley, Chichester, 2009.

[5] A. Chiumento, M. Bennis, C. Desset, L. Van der Perre, and S. Pollin, "Adaptive CSI and feedback estimation in LTE and beyond: a Gaussian process regression approach," *Journal on Wireless Communications and Networking*, Springer, 2015.

[6] D. Gesbert, M.S. Alouini, "How much feedback is multi-user diversity really worth?". Int. Conf. Comm, June 2004, pp. 234–238.

[7] L. Sivridis, X. Wang, and J. Choi, "Radio resource management for fast fading environments," Journal of Comm., vol. 6, issue 7, Oct. 2011, pp. 540-548.

[8] K. Hoon, and H. Youngnam, "An opportunistic channel quality feedback scheme for proportional fair scheduling," IEEE Commun. Lett , vol. 11, issue 3, June 2007, pp. 501–503.

[9] S. Kittipiyakul, A. Boonkajay, and T. Aphichartsuphapkhajorn, "Performance of low-feedback rate, gradient-based OFDMA subcarrier allocation with partial channel information," Proc. Symp. Wireless Perv. Comm., May 2010, pp. 511–521.

[10] J. Jeon, J, K. Son, H.W. Lee, and S. Chong, "Feedback reduction for multiuser OFDM systems," IEEE Trans. Vehicular Tech., vol. 59, issue 1, Jan. 2010, pp. 160–169.

[11] L. Sivridis, J. He, "A Strategy to reduce the signaling requirements of CQI feedback schemes," Journal Wireless Pers. Comm., vol. 70, issue 1, May 2013, pp. 85-98.

[12] M. Abdulhasan, M. Salman, C. Ng, N. Noordin, S. Hashim, and F. Hashim, "An Adaptive Threshold Feedback Compression Scheme Based on Channel Quality Indicator (CQI) in Long Term Evolution (LTE) System," Journal Wireless Pers. Comm.,vol. 82, issue 4, June 2015, pp. 2323-2349.

[13] M. Pauli, U. Wachsmann, and S. Tsai, "Quality determination for a wireless communications link", U.S. Patent Office publication, Patent (US 2004/0219883, Nov. 2004.

[14] E. Westman, "Calibration and evaluation of the exponential effective SINR mapping (EESM) in 802.16," Master's thesis, The Royal Institute(of Technology (KTH), Stockholm, Sweden, Sep. 2006.

[15] C. Fa-tang and T. Gen-lin, "A novel MCS selection criterion for supporting AMC in LTE system," in *Proc. of the International Conference on Computer Application and System Modeling* (*ICCASM*), 2010, vol. 6, pp. 598-603.

[16] R1-061777, "DCT based CQI reporting scheme", LG Electronics, RAN1 LTE Ad Hoc, Cannes, Franes, June, 2006.

[17] R1-070187, "DCT partitioning for CQI reporting", Panasonic, RAN1 Meeting 47, Sorrento, Italy, January, 2007.

[18] A. Haghighat, G. Zhang and Z. Lin, "Full-band CQI feedback by Haar compression in OFDMA systems," in *Proc. of the IEEE 70th Vehicular Technology Conference Fall* (*VTC 2009-Fall*), 2009, pp. 1-5.

[19] M. Kang, and K. Soon, "Performance analysis and optimization of best-M feedback for OFDMA systems", *IEEE Communications Letter*, vol. 16, issue 10, October 2012, pp.1648-1651.

[20] A. Djouama, L. Myoung-Seob, and F.Y. Ettoumi, "Channel Estimation in Long Term Evolution Uplink Using Minimum Mean Square Error-Support Vector Regression.", Wireless Personal Communications vol. 79.3, pp. 2291-2304, 2014.

[21] M.J. Fernández-Getino Garcia, J.L. Rojo-Álvarez, F. Alonso-Atienza, M. Martínez-Ramón, "Support vector machines for rubust channel estimation in OFDM," in IEEE Signal Processing Letters, vol. 13.7, pp. 397-400, 2006.

[22] A. Charrada, "Support Vector Machines Regresion for MIMO-OFDM Channel Estimation," in IAES International Journal of Artifical Intelligence, vol. 1.4, pp. 214-224, 2012.

[23] A. Djouama, E. Zochmann, S. Patschenr, M. Rupp and F.Y. Ettoumi, "Predicting CSI for Link Adaptation Employing Support Vector Regression for Channel Extrapolation", in Proc. of the International Workshop on Smart Antennas (WSA), 2016.

[24] G. Piro, L. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: An open source framework", *IEEE Transactions Vehicular Technology*, vol. 60, no.2, February 2011.

[25] V. Vapnik, "The nature of statistical learning theory," Springer, Science & Business Media, 2000.

[26] V. Kecman, "Learning and Soft Computing," MIT Press, Cambridge, MA, USA, 2001.

# A CROSS-LAYER MV-HEVC DEPTH-TEXTURE RATE ALLOCATION ESTIMATION TECHNIQUE IN 3GPP LTE SYSTEMS

Mario Cordina
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
m.r.cordina@ieee.org

Carl J. Debono
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
c.debono@ieee.org

## ABSTRACT

This paper presents a cross-layer depth-texture target bit rate allocation estimation technique for the transmission of Multiview High Efficiency Video Coding (MV-HEVC) texture plus depth content over 3GPP LTE systems. The proposed technique is based on a statistical model which exploits the texture and depth map image characteristics to estimate the optimal depth-texture rate allocation to be used by the codec's rate control algorithm. Experiments using standard test sequences show the effectiveness of the proposed technique as the model is able to estimate, on-line, the optimal depth-texture rate allocation with a mean absolute estimation error of 3.3% and a standard deviation of 2.2%. In addition, the proposed cross-layer architecture allows the depth-texture rate allocation to be adapted to both the video content characteristics and the available bandwidth offered by the wireless network making it suitable for the transmission of multiview 3D video to mobile devices.

***Index Terms***— Multiview video plus depth, 3GPP LTE systems, HEVC, depth-texture rate allocation, cross-layer architecture

## 1. INTRODUCTION

According to the latest Cisco visual networking index forecast report, video traffic accounted for 55% of the total mobile data traffic in 2015 and is expected to reach 75% by 2020 [1]. Meanwhile, Free-View Video (FVV) [2] based on the multiview video plus depth representation (MVD) [3], which allows the flexible rendering of novel views through the use of depth image-based rendering techniques (DIBR) [4], has garnered significant commercial and research interest and will inevitably enter the mobile ecosystem. This increase in multiview 3D video traffic coupled with the explosive growth in video data traffic is expected to exacerbate the pressure on the capacity constraints of current wireless cellular networks, thereby making it more difficult for mobile network operators to adhere to the Quality of Service (QoS) expected by the increasingly demanding subscriber base.

The MVD representation uses a limited set of texture views coupled with a matching set of per-pixel depth map views which are compressed by exploiting the redundancy in space, time and in between views [5,6]. Since the compressed MVD stream is usually transported over a bandwidth constrained network, such as a wireless cellular network, a rate control scheme functioning across the depth-texture and frame layer is necessary. The latter has attracted substantial research interest and several schemes have been proposed, however only a limited number of works exist on the optimal depth-texture rate allocation problem. Even though the depth map is treated as side-information and used only for novel view rendering, studies have shown that coding the depth map views at a low bit rate introduces artefacts which can have a significant impact on the quality of the rendered novel views [7]. Therefore,under the constraint of the total available bandwidth provided by the underlying wireless network, the correct balance between depth and texture rate allocation is critical.

In this work, we focus on the depth-texture rate allocation problem in cellular networks. As opposed to previous works, we formulate a statistical model which exploits the texture and depth map image characteristics to accurately estimate, on-line, the optimal depth-texture rate allocation. Our studies focus on the MV-HEVC compression standard due to its computational efficiency when compared to its extension, 3D-HEVC. Moreover, since in our application multiview 3D video is intended to be transported over a 3GPP LTE based cellular network, a cross-layer approach is adopted such that the depth-texture rate allocation is adapted to both the varying video content characteristics and available bandwidth of the underlying wireless network. The model was developed using three test sequences and later successfully tested using previously unseen test sequences exhibiting differing video characteristics. Moreover, real-world field measurements taken from a live commercial LTE network were used to verify the adaptation of the optimal depth-texture rate allocation when subjected to both varying video characteristics and available bandwidth.

The rest of the paper is organized as follows: Section 2 gives a summary of related work; Section 3 describes the proposed cross-layer MV-HEVC depth-texture rate allocation; Section 4 gives a description of the statistical

978-1-5090-6067-2/17/$31.00 

model formulation and performance results while Section 5 provides some comments and conclusions.

## 2. RELATED WORK

Although an optimal depth-texture bit rate allocation is critical for the correct rate control of MVD streams, this problem has attracted a relatively small number of research works in recent years. Moreover, the majority of the proposed solutions either require off-line processing or assume a fixed depth-texture ratio [4]. In the latter, although simplistic and suitable for on-line depth-texture rate allocation, it cannot guarantee an optimal novel view rendering quality especially when the video sequence characteristics vary. In [8], Morvan *et al.* combine the texture and depth rate-distortion (R-D) curves yielding a single R-D surface. The authors then propose a fast hierarchical optimization algorithm based on an orthogonal search pattern which exploits the smooth monotonic properties of the R-D surface to find the optimal depth-texture rate allocation ratio. Apart from the off-line analysis required to generate the R-D surface, the authors assume that a full reference assessment [9] of the synthesized view is always possible. In [10], the authors relax the full reference assessment constraint by proposing a complex distortion model incorporating the video coding, depth quantization and geometry induced distortion. This model is then used in conjunction with an off-line joint rate allocation method to find the correct balance between the depth and texture bit rate which maximizes the novel view synthesis quality.

In [11], Cheung *et al.* derive a cubic distortion model, whose parameters are established using off-line processing, and show that the optimal quantization levels for the texture and depth map is equivalent to the shortest path in a specially constructed 3-D trellis. Another depth-texture bit rate allocation solutionwas proposed by Yuan *et al.* [12]. Here, the authors formulate a simple distortion model for the synthesized view based on the average distortion of the left and the right view and the average distortion of the compressed depth maps. Using this model together with the fractional rate-quantisation model [13], a joint bit allocation problem is formulated as a constrained optimization problem and resolved using the Lagrangian multiplier method. In comparison with [11], the solution proposed by Yuan *et al.* also requires off-line processing to determine the model parameters, thereby limiting these solutions to non-real time applications.

The depth-texture rate allocation problem in H.264/MVC was studied by Cordina *et al.* [14]. Here, the authors use the prediction mode distribution of the macro blocks (MBs) at the discontinuity regions of the depth map video to estimate the optimal depth-texture ratio. Although the model exhibits a low estimation error and is suitable for low latency applications, its formulation is intrinsically tied with the MVC encoder and does not provide any input related to the available bandwidth. Moreover, to establish the prediction mode distribution of the MBs at the discontinuity region, one must start encoding the video sequence at some arbitrary quantization parameter (QP) value. In [15], the authors propose a simplistic model to directly select the QP of the depth map view given the quantization parameter of the texture video. Similar to [14], the authors focus on H264/MVC and the available bandwidth does not feature as one of the control inputs.

## 3. OPTIMAL DEPTH-TEXTURE RATE ALLOCATION

### 3.1. System model and problem formulation

Consider the radio network subsystem part of an MV-HEVC based 3D video delivery system shown in Figure 1. Further assume that the MV-HEVC video encoder, residing in the core network, is able to simultaneously encode multiple texture and depth map views. Without losing generality, consider the encoding of two texture views, view 0 and view 2, coupled with their corresponding depth map views whilst the synthesized view, view 1, is generated in the mobile terminal (UE) through DIBR techniques. As shown in Figure 1, the combined MVD stream is transmitted to the UE over a 3GPP LTE Release 8 compliant evolved Node B (eNB). In addition, as per 3GPP standards, the UE periodically reports a channel quality feedback to the eNB, which the scheduler uses to effectively determine the maximum bit rate (channel bandwidth) supported by the UE at the current radio and loading conditions denoted by $R_c$. This feedback is generally used by the eNB lower layers to determine the number of resource blocks to be assigned to the UE, however in this study we consider a cross-layer approach thereby making the available bandwidth available to the application layer to be exploited by the encoder.

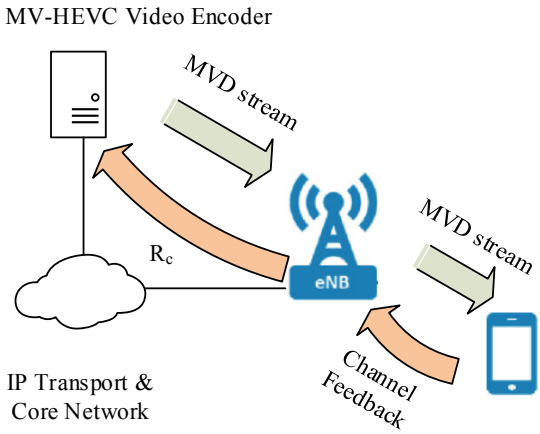



**Fig. 1.** System Model

Thus, the objective of this work is to formulate a statistical model which estimates, on-line, the optimal

depth-texture rate allocation ratio, $O_{DT}$, as a function of $\Phi_T$, $\Phi_D$ and $R_c$ given by:

$$O_{DT} = f\left(\Phi_T, \Phi_D, \alpha R_c\right) \quad (1)$$

where $\Phi_T$ and $\Phi_D$ represent the texture and depth map views image statistics respectively and $\alpha$ is the lower layers overhead factor which is typically 0.18 [16]. Using the model defined by (1), the target bit rates for the texture and depth map views that should be used by the encoder's rate control algorithm are given by:

$$\begin{aligned} R_d &= O_{DT}\alpha R_c \\ R_t &= (1 - O_{DT})\alpha R_c \end{aligned} \quad (2)$$

where $R_d$ and $R_t$ represent the bit rate allocated to the depth map views and texture views respectively.

### 3.2. Experimental setup and initial observations

Inspired by the works in [14] and [17], a number of experiments were carried out to determine the optimal depth-texture rate allocation ratio for various test sequences exhibiting different video characteristics and total bit rate. The latter is defined as the sum of the encoded bit rate of the texture and depth map views and is usually bounded by the channel bandwidth. In order to increase the number of data samples for the statistical model formulation, each test sequence was divided into chunks, whose size was set equal to the intra frame period (which was set to 12 in this work). For each test sequence, each chunk was encoded using MV-HEVC reference software HM11.0 [18] whilst the depth-texture ratio and total bit rate were adjusted by varying the QP between 20 and 44 in steps of 2 for both the texture and the depth map views. In view of our interest in real-time video applications, a low latency prediction structure described in [19] was used in this work. The test sequences used for these experiments, defined in the Common Test Conditions (CTC) [20], are shown in Table 1, where *L*, *C* and *R* represent the left, rendered and right views respectively, whilst the test setup is described in [17].

**Table 1**. Test sequence parameters and characteristics
($^V$ indicates the sequence used for model verification only)

| Sequence | Image Size | Cameras (L-C-R) | Characteristics (frame rate, baseline) |
|---|---|---|---|
| *Balloons* | 1024 x 768 | 1-3-5 | 30 fps; 5cm in 1D plane |
| *Book Arrival* | 1024 x 768 | 6-8-10 | 15 fps; 6.5cm in 1D plane |
| *Kendo* | 1024 x 768 | 1-3-5 | 30 fps; 5cm in 1D plane |
| *Newspaper*$^V$ | 1024 x 768 | 2-4-6 | 30 fps; 5cm in 1D plane |
| *Poznan Hall*$^V$ | 1920 x 1088 | 5-6-7 | 25 fps; 13.75cm in 1D plane |
| *Poznan Street*$^V$ | 1920 x 1088 | 3-4-5 | 25 fps; 13.75cm in 1D plane |

With reference to Figure 2, the plot of the synthesized view Peak Signal-to-Noise Ratio (PSNR) with depth-texture ratio for the *Balloons* video sequence at a given total bit rate follows a downward concave parabola yielding an optimal depth-texture ratio (shown in red) of 40.3%. Moreover, as shown in Figure 2, any large deviations from this optimal value (shown in green) will result in 1.1dB degradation in the synthesized view's PSNR, thereby emphasizing the importance to estimate the optimal depth-texture ratio with a low margin of error.

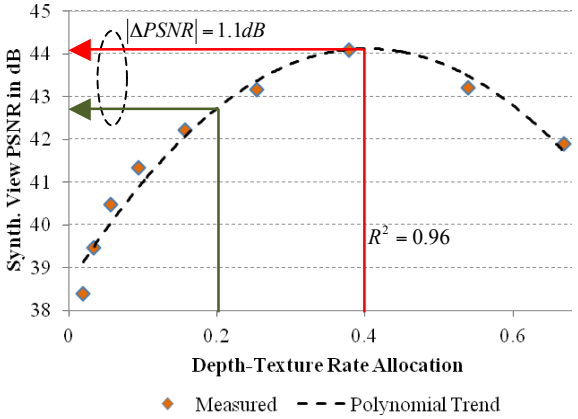



**Fig. 2.** *Balloons* - PSNR vs depth-texture ratio
($R^2$ represents the coefficient of determination)

As shown in Figure 3, the optimal depth-texture ratio is a function of both the total bit rate and sequence characteristics as different sequences impart contrasting trends. Further inspection reveals that the optimal depth-texture ratio trends towards an asymptotic value which never exceeds 60%. This is expected, as at high available bandwidths any further increase in the depth map bit rate would not translate in an improvement in the synthesized view's PSNR. Conversely, at very low bit rates, the allocated depth map bit rate is very low necessitating a return to standard 2D video, whereby the texture views only are transmitted. In order to reflect these constraints in (1) the following bounds are applied to the estimated optimal depth-texture rate calculated by the statistical model:

$$\begin{aligned} &if\ \Phi_{DT} > 0.6 \\ &\quad \Phi_{DT} = 0.6 \\ &elseif\ \Phi_{DT} < 0.1 \\ &\quad 3D \rightarrow 2D\ switch \\ &otherwise\ use\ \Phi_{DT} \end{aligned} \quad (3)$$

In order to model the sequence dependent characteristics, the authors considered a number of texture and depth map image statistics. These included the gradient

of the depth map view, $G_D$, and texture view, $G_T$, together with the depth map entropy ratio [17], $E_{DT}$, given by:

$$E_{DT} = \frac{E_D}{E_D + E_T} \tag{4}$$

where $E_D$ and $E_T$ are the average entropy of the depth map and texture views respectively.

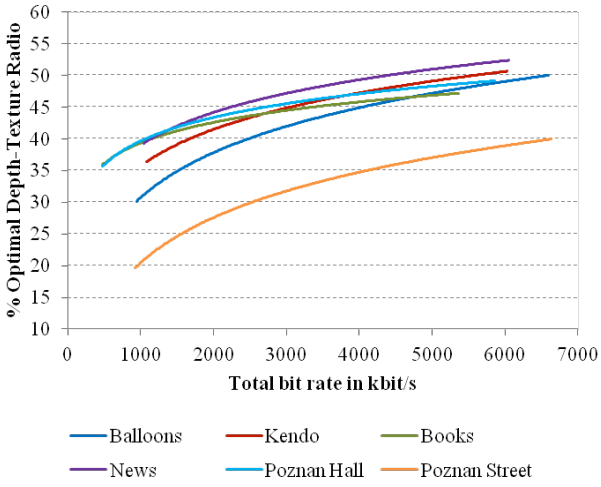



**Fig. 3.** Optimal depth-texture ratio with total bit rate

## 4. STATISTICAL MODEL FORMULATION AND RESULTS

### 4.1. Multi-regression model

The average chunk values of the image statistics for all test sequences used during model formulation, shown in Table 1, together with their optimal depth-texture ratio and total bitrate were parsed, collated and imported into IBM SPSS© statistical package for analysis. In this study, the authors opted to use a multi-regression model, as this family of models are the most commonly used due to their well-known statistical properties. Multi-regression models are composed of several predictors (image statistics and bit rate) which are used to model a single dependent variable (optimal depth-texture ratio) having a normal distribution. Thus, prior to fitting the multi-regression model, the dependent variable was checked for normality using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.568 validating the normality criteria. The multi-regression model was formulated using SPSS© which makes use of the least square estimation method to estimate the parameter values. Some predictors were not included in the final statistical model as they were either not correlated with the variation in the optimal depth-texture ratio or found to be correlated with each other and thus were removed to reduce multi-collinearity effects. The parsimonious model given by (5) is composed of 7 significant predictors whereas the coefficient of determination, $R^2$, of the formulated statistical model is 0.86.

$$\begin{aligned} O_{DT} &= 34.516 \\ &+ G_D\left[25.714 + 0.000999\alpha R_c G_D^{-2}\right] \\ &- G_T[3.899 - 0.004603\alpha R_c - 0.085502\alpha R_c G_T^{-2}] \\ &- bitrate E_{DT}[0.006149 E_{DT}^{-2} - 0.0623] \end{aligned} \tag{5}$$

where $O_{DT}$ is the estimated optimal depth-texture ratio, $G_D$ and $G_T$ are the average chunk gradient of the depth map views and texture views respectively, $E_{DT}$ is the depth map entropy ratio given by (4) whilst $\alpha R_c$ is the total bit rate available at the application layer.

### 4.2. Model Diagnostics Analysis

Following the formulation of the statistical model, a number of diagnostic checks were carried out to detect any model uncertainty:

*Standardized Residual Distribution* [21]: The standardized residuals are checked for normality using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.805, validating the normality criteria. Moreover, a plot of the Cook's distance [21] with the unstandardized predicted value, shown in Figure 4, indicates a relatively low number of influential points (points having relatively high Cook's distance) thereby reducing model uncertainty.

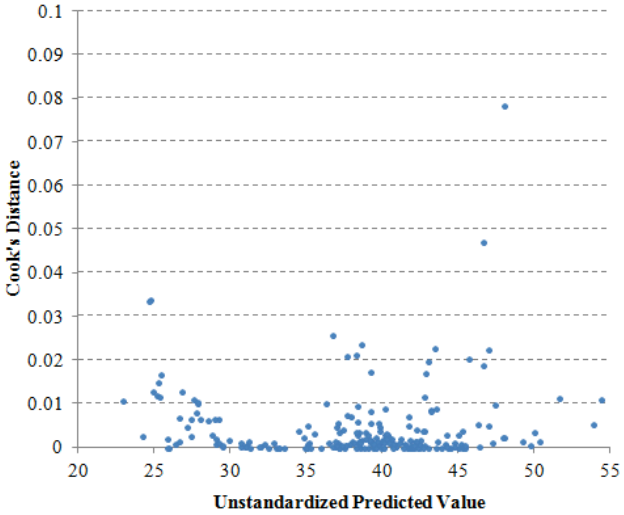



**Fig. 4.** Cook's distance plot

*Studentized Residual* [21]: Apart from the model's coefficient of determination, $R^2$, model fitness is checked using a plot of the studentized residual against the unstandardized predicted optimal depth-texture ratio as shown in Figure 5. Since, the majority of the points are distributed across a horizontal band, indicating homoscedasticity, and lie within the standard ±2 limit, the model is adequate for such data fit with a very small number of outliers (shown in red).

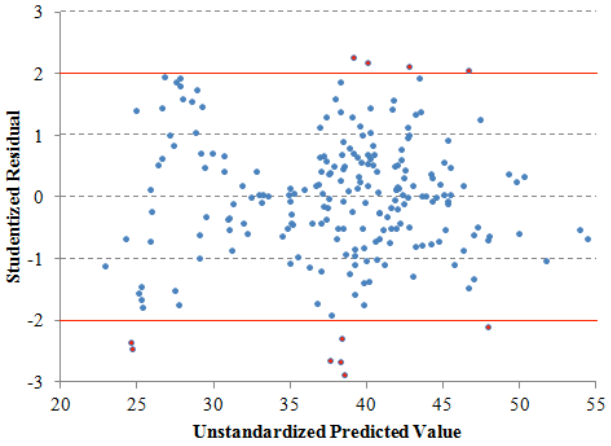



**Fig. 5.** Studentized residual plot

### 4.3. Model Verification Results

The proposed statistical model was integrated in the MV-HEVC HM11.0 reference encoder software to verify the efficacy of the proposed solution by calculating, on-line, the optimal depth-texture ratio for various average bitrates. The plots of the actual and estimated optimal depth-texture ratio for the test sequences used during model formulation and verification are shown in Figure 6 and 7 respectively.

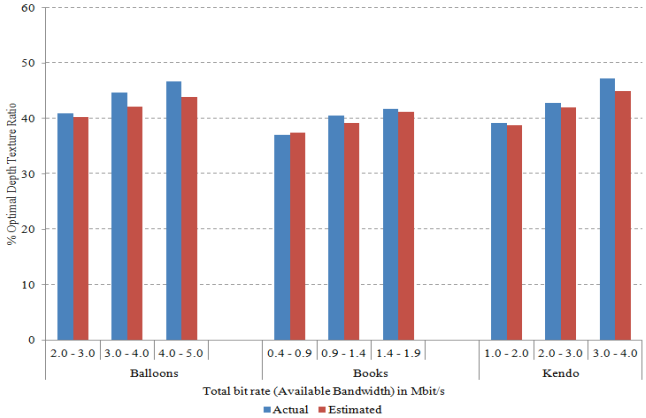



**Fig. 6.** Optimal depth-texture ratio – model formulation

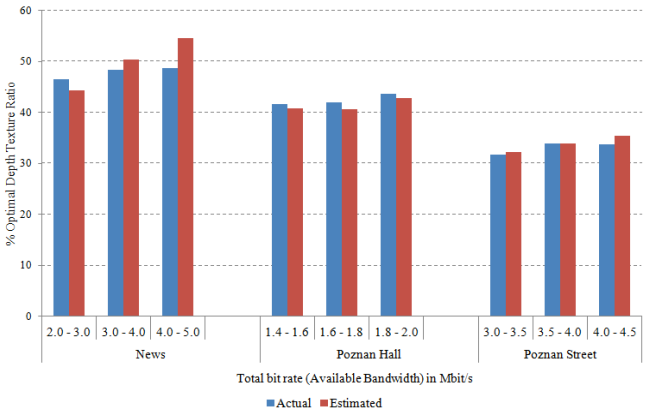



**Fig. 7.** Optimal depth-texture ratio – model verification

With reference to Table 2, the formulated statistical model exhibits an absolute mean estimation error of 2.2% and standard deviation of 2% for the test sequences used during the model formulation. Moreover, with reference to Figure 7, the proposed model is able to estimate the optimal depth-texture ratio for the *Newspaper*, *Poznan Hall* and *Poznan Street* verification test sequences with an absolute mean estimation error of 3.3% and standard deviation of 2.3%.

**Table 2.** Model estimation error

| Test Sequence | % Absolute Mean Estimation error | % Standard Deviation of the Estimation error |
|---|---|---|
| Model Formulation | 2.2 | 2.0 |
| Model Verification | 3.3 | 2.3 |

The impact of the estimation error and associated standard deviation on the synthesized view PSNR was evaluated for various test sequences at different total bit rates and this was found to be generally less than 0.05dB and at worse 0.1dB. As shown in Figure 8, the synthesized view PSNR for the *Newspaper* sequence using the proposed model was found to be on average 1.2 dB better, for various total bit rates, when compared to the commonly used 20% fixed depth-texture rate case [4].

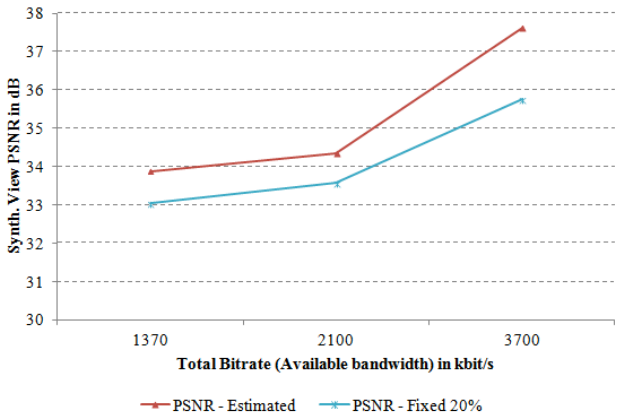



**Fig. 8.** Synthesized view PSNR – proposed vs fixed at 20%

With reference to the system model shown in Figure 1, in order to evaluate the adaptation of the optimal depth-texture ratio through the use of the cross-layer approach, a series of 120 frames (10 chunks) from 3 different test sequences were combined and encoded using the MV-HEVC framework integrated with the proposed statistical model. Field test measurements were carried out in a live commercial LTE network using a purposely setup test vehicle equipped with industry standard test equipment [22]. The channel quality feedback reported to the network by the UE throughout the field test were collected and elaborated to estimate the average physical layer channel bandwidth supported by the UE. Considering a lower layer overhead

factor $\alpha$ of 0.18, the application layer bandwidth available for the encoder, averaged over a chunk, is shown in Figure 9. As evidenced in Figure 9, the optimal depth-texture ratio was adapted at every intra frame period ensuring good tracking in the presence of varying video characteristics (abrupt test sequence change and intra sequence changes) and channel bandwidth.

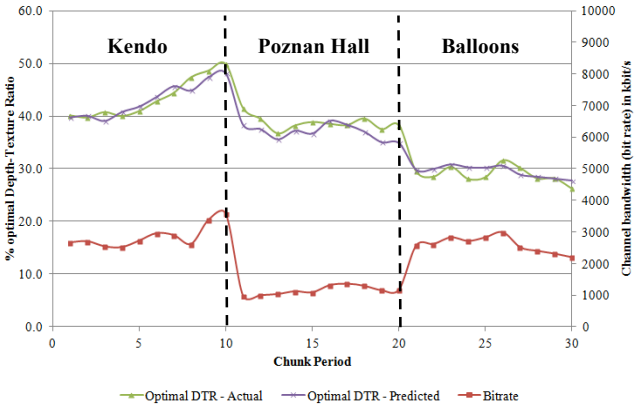



**Fig. 9.** Scene change handling of the proposed technique

## 5. CONCLUSION

In this paper, we presented a cross-layer depth-texture target bit rate allocation estimation technique for the transmission of multiview high efficiency video coding texture plus depth content over 3GPP LTE systemsbased on a statistical multi-regression model. Simulation results on test sequences exhibiting a variety of video characteristics have shown that the proposed technique is able to estimate, on-line, the optimal depth-texture ratio with an absolute mean estimation error of 3.3% for the test sequences not used during model formulation. Moreover, the PSNR of the synthesized view using the proposed technique is on average 1.1dB better when compared to the commonly used fixed depth-texture ratio of 20% whilst allowing good tracking performance in the presence of varying video characteristics and channel bandwidth.

## 6. REFERENCES


[1] "Cisco visual networking index forecast report", [Online]. Available:http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html

[2] A. Smolic, and P. Kauff, "Interactive 3-D Video representation and coding technologies," in *Proc. of the IEEE*, vol. 93, no.1, pp. 98-110, Jan 2005.

[3] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Process.*, pp. 201-204, Sep. 2007.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291.

[5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. Circuits & Syst. for Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[6] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. on Broadcast.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[7] A. Vetro, S. Uea, and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays", in *Proc. Of SPIE Conference on Applications of Digital Image Processing XXXI*, vol. 7073, Sep. 2008.

[8] Y. Morvan, D. Farin, and P. De With, "Joint depth/texture bit-allocation for multi-view video compression", in *Proc. of Picture Coding Symposium*, Nov. 2007.

[9] ITU-R BT.1683, Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference. Geneva, June 2004.

[10] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model," *Signal Process.: Image Commun.*, vol. 24, no. 8, pp. 666-681, Sep. 2009.

[11] G. Cheung, V. Velisavljevic and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. on Image Proc.*, vol.20, no.11, Nov. 2011.

[12] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding", *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 21, no. 4, Apr. 2011.

[13] S. Ma, W. Gao, and Y, Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Cicuits Syst. for Video Technol.*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.

[14] M. Cordina, and C. Debono, "An adaptive texture-depth rate allocation estimation technique for low latency multi-view video plus depth transmission," in *Proc. of the IEEE Visual Comm. and Image Proces.*, Nov. 2013.

[15] K. Klimaszewski, K. Wegner, and M. Domanski, "Video and depth bitrate allocation in multiview compression," in *Int. Conf. on Systems, Signals and Image Proc.*, May 2014.

[16] "LTE downlink bitrates", [Online]. Available: https://www.lte-bullets.com

[17] E. Bosc, F. Racape, V. Jantet, P. Riou, M. Pressigout and L. Morin, "A study of depth/texture bit-rate allocation in multi-view video plus depth compression," in *Ann. Telecommun.*, pp. 615-625, Dec. 2013.

[18] "HEVC reference software," [Online]. Available: http://hevc.kw.bbc.co.uk/trac/browser/jctvc-htm/tags/HM-11.0

[19] M. Cordina, and C. Debono, "A depth map rate control algorithm for HEVC multi-view video plus depth," in *Proc. of the IEEE Int. Conf. on Multi. and Expo Workshops,* July 2016.

[20] Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG, "Common test conditions of 3DV core experiments," JCT3V-G1100, San Jose, USA, Tech. Rep., Jan. 2014.

[21] M. Norusis, IBM SPSS Statistics 19 Advanced Statistical Procedures Companion, Pearson, Feb. 2011.

[22] "Swissqual Field Test Equipment", [Online]. Available: http://www.swissqual.com/en/products/benchmarking/diversity-benchmarker-2

# A DEPTH MAP RATE CONTROL ALGORITHM FOR HEVC MULTI-VIEW VIDEO PLUS DEPTH

Mario Cordina
Department of Communication and Computer Engineering
University of Malta
Msida, Malta
m.r.cordina@ieee.org

Carl J. Debono
Department of Communication and Computer Engineering
University of Malta
Msida, Malta
c.debono@ieee.org

## ABSTRACT

In this paper, we present a depth map rate control algorithm for the High Efficiency Video Coding (HEVC) Multi-View Video plus Depth (MVD) representation. The proposed algorithm is based on a bit allocation refinement technique for key frames together with a depth-map inspired R-λ model coupled with an adaptive clipping algorithm designed to exploit the depth map characteristics. This results in improved video quality of the synthesized views while maintaining depth map rate control accuracy. This scheme was tested using various standard test sequences and has shown efficacy with an average improvement of the synthesized view Peak Signal-to-Noise Ratio (PSNR) and Bjøntegaard Delta PSNR (BD-PSNR) of 1.15% and 0.45dB respectively, whilst achieving a bit rate error reduction of 0.2% when compared to the reference rate control algorithm implemented in MV-HEVC.

***Index Terms***— HEVC, Rate Control, Multi-view video, Adaptive clipping, View synthesis, Depth maps

## 1. INTRODUCTION

The rapid development of digital multimedia spurred an increased interest in three-dimensional (3D) technology which has reached commercialization in a variety of products and applications ranging from 3D television (3DTV), gaming, and medical imaging. The latest video coding standard developed by the Joint Collaborative Team on Video Coding (JCT-VC) is the High Efficiency Video Coding (HEVC) standard which was finalized in April 2013. The multi-view (MV) version, MV-HEVC, was released in October 2014 and it extends the HEVC standard to handle multi-view video (MVV), through the use of inter-layer prediction whilst adding depth map support. The latter allows the encoding of the multi-view video plus depth (MVD) representation [1]. This representation is attractive as it allows the adaptive rendering of a continuum of views at the decoder, through the use of depth-image-based rendering (DIBR) techniques [2], and realizes free-view video (FVV) which allows the viewer to interactively select an arbitrary viewing position in relation to the 3D scene being watched. In MVD, a set of texture views together with a corresponding set of per-pixel depth map views are independently compressed using MV-HEVC by exploiting the redundancy in space, time and in-between views [3,4]. To further improve the compression of the MVD representation, a second extension, 3D-HEVC, was released in February 2015. This version allows the joint coding of texture and depth map views through the introduction of several new coding tools such as view synthesis prediction/optimization and depth map modes, which improve the coding efficiency at the expense of increased computational complexity.

As the compressed MVD stream is usually transported over a bandwidth limited network, a rate control algorithm operating at the view, depth-texture, group-of-pictures (GOP), frame, and coding unit (CU) level is essential. At the view and depth-texture level, the bit rate ratios are usually fixed with the priority given to the base view and texture respectively. For the remaining levels, the same rate control algorithms are usually used for both the texture and depth map views. However, these algorithms are designed and optimized for rate controlling the texture views and do not take into consideration the depth map characteristics, at the detriment of the synthesized view quality. Although the depth map videos are only used for virtual view rendering, any artifacts introduced during coding and rate controlling these depth videos at a low bit rate, can have a significant impact on the quality of the synthesized views [5]. In this work, we propose a new depth map specific rate control algorithm. Thus, the encoder will have the standard rate control algorithm operating on the texture view together with the proposed algorithm, operating in parallel, on the depth map view. The latter is based on a bit allocation refinement technique for key frames together with a depth-map centric R-λ model and an adaptive quantization and λ parameter clipping algorithm designed to exploit the depth map characteristics.

The rest of the paper is organized as follows: Section 2 is a summary of related work; Section 3 describes the proposed rate control algorithm; Section 4 is a description of the test setup and the resulting performance of the proposed algorithm, while Section 5 provides some comments and conclusions.

978-1-4799-7082-7/15/$31.00 ©2015 IEEE

## 2. RELATED WORK

Rate control is a very active research topic and several algorithms have been proposed for the legacy video coding standards. However, rate control techniques in HEVC and its extensions are limited. Two main rate control proposals have been adopted for HEVC, namely: the unified rate quantization (URQ) algorithm and the R-λ model based rate control, defined in JCTVC-H0213 [6] and JCTVC-K0103 [7] respectively. The URQ rate control scheme is based on the quadratic model used in earlier video coding standards and has been further improved through the enhancement of the model accuracy. In [8], the authors use a per pixel gradient value in the R-Q model, while the sum of absolute transformed differences (SATD) is adopted in [9].

The R-λ model is based on the studies carried out by Li *et al* in [10]. In this work the bit rate $R$ was found to have a stronger relationship with the Lagrange multiplier $\lambda$ than the quantization parameter QP. Moreover, in [7], the authors show that the rate-distortion (R-D) model is best characterized using a hyperbolic function given by:

$$D(R) = CR^{-K} \tag{1}$$

where $C$ and $K$ are model parameters related to the source characteristics.

The Lagrange multiplier $\lambda$ is defined as the slope of the R-D curve and thus:

$$\lambda = -\frac{\partial D}{\partial R} = CKR^{-K-1} = \alpha R^{\beta} \Rightarrow R = \left(\frac{\lambda}{\alpha}\right)^{-\beta} = \alpha_1 \lambda^{\beta_1} \tag{2}$$

where $\alpha_1$ and $\beta_1$ are model parameters related to the source characteristics.

The rate control algorithm proposed in [10] became the reference algorithm for HEVC and has since been further improved in [11] and [12] to better handle the intra-frame rate control. In [13] the authors propose an efficient bit allocation method considering the HEVC hierarchical coding structure whilst in [14] a Laplace distribution based code tree unit (CTU) rate control for HEVC is proposed.

In MV-HEVC for the MVD representation, the same R-λ rate control algorithm used in HEVC is used across all the texture and depth map views regardless of the fact that the texture and depth map videos have strikingly different source characteristics. Similarly, for 3D-HEVC, the rate control algorithms developed for HEVC have been extended and applied for both texture and depth map views. In [15], the depth map is exploited to improve the accuracy of the mean absolute difference (MAD) by considering the inter-view disparity in the texture view MAD estimation. In [16], the authors propose an adaptive frame level rate control algorithm for 3D-HEVC based on a new initial quantization parameter decision and bit allocation scheme. Although the results look promising, the authors do not indicate any improvements in the synthesized view quality.

In all the rate control works reviewed for MV-HEVC and 3D-HEVC, no specific treatment is given to the depth map view. A rate control algorithm which is specifically designed for the depth map view will result in a beneficial improvement in both the rate control accuracy of the depth map views, and the quality of the synthesized views.

## 3. PROPOSED DEPTH MAP RATE CONTROL ALGORITHM

As shown in [17], the 3D-HEVC coding standard has a coding efficiency gain of 21% when compared to MV-HEVC. However, this comes at the expense of a 20% increase in computational complexity. Thus, although the proposed depth map rate control algorithm can be applied to both MV-HEVC and 3D-HEVC coding standards, we study its application to MV-HEVC due to our interest in low latency, low computational complexity video coding. Moreover, without loss of generality, the 3-view low latency prediction structure shown in Fig. 1 is used in this work.

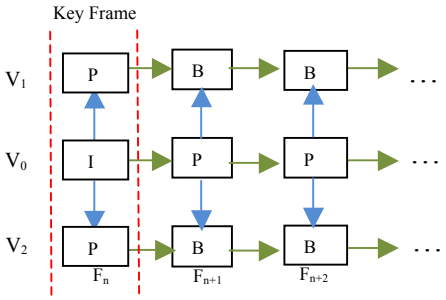



Fig. 1: Low Delay Prediction Structure

The proposed depth map rate control algorithm is based on the general R-λ model and is illustrated in Fig. 2. Here the rate controller is applied to each depth map view independently and operates at three main levels, namely: group of pictures, frame, and coding tree unit level.

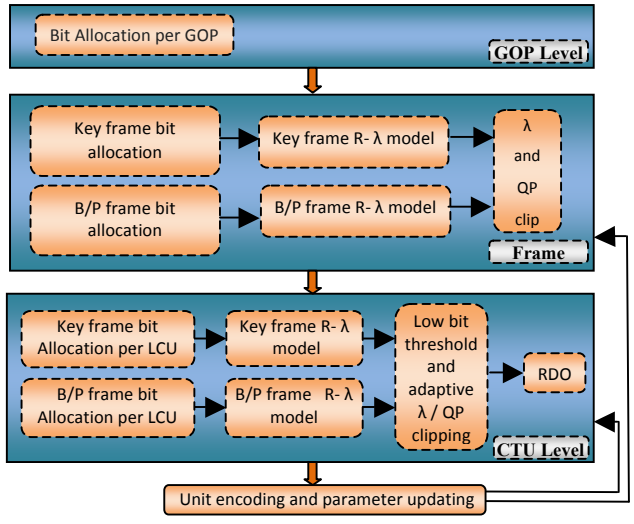



Fig. 2: Proposed depth map rate control algorithm

### A. GOP-level bit allocation

The bit allocation at the GOP level is similar to that in [7] and is given by:

$$T_{GOP} = N_{GOP}\left(\frac{\left(\frac{R_{target}}{f}\right)\left(N_{S,CODED}+S_W\right)-T_{S,CODED}}{S_W}\right) \quad (3)$$

where $R_{target}$ is the target bit rate, $f$ is the frame rate, $N_{GOP}$ is the number of frames in GOP, $N_{S,CODED}$ is the number of pictures in sequence already coded, $T_{S,CODED}$ is the bit cost of the coded pictures in the sequence, whilst $S_W$ is the smoothing window size. $S_W$ is set to 40 and used to allow the controller to change the bit rate smoothly and achieve $R_{target}$ after $S_W$ frames.

### B. Frame-level bit allocation

At the frame level, the inter-B/P frame bit allocation and R-λ model used in the proposed work is analogous to that in [7] and is given by (4) and (5) respectively.

$$T_{avgpic} = \frac{T_{GOP}-T_{G,CODED}}{\sum_{NotCodedPics}\omega_i}\cdot\omega_{CurPic} \quad (4)$$

$$\lambda_{pic} = \alpha_{B,pic}\left(\frac{T_{avgpic}}{P_x}\right)^{\beta_{B,pic}} \quad (5)$$

where $T_{avgpic}$ is the initial bit budget allocated per frame, $T_{GOP}$ is the target number of bits in a GOP and is computed using (3), $T_{G,CODED}$ is the bit cost of already coded frames in the current GOP, $\omega$ is the bit allocation weighting factors for the different hierarchical levels as defined in [7], $\alpha_{B,pic}$ and $\beta_{B,pic}$ are the R-λ model parameters initially set to 3.2003 and -1.367 respectively and adaptively tuned during the parameter update procedure described in [7], whilst $P_x$ is the number of pixels in a frame.

Different from the standard rate control algorithm, the bit allocation of the key frames, shown in Fig. 1, is refined such that these are given importance in relation to the other frames in the GOP. Thus, the refined bit budget allocation of the key frames is given by:

$$T_{key} = 0.25*\left(\frac{C}{T_{avgpic}}\right)^{\gamma} \quad (6)$$

where $C$ is the complexity measure based on the sum of absolute transformed difference [11] and $\gamma$ is a tuning parameter controlling the amount of bit budget allocated to the key frames from the total. For the intra frames, this parameter was set to 0.5582 [11], while for the P-key frames it was empirically set by encoding several depth map test sequences at various target bit rates and evaluating several quality metrics which include the standard deviation and average peak signal-to-noise ratio (PSNR) of the synthesized view together with the average PSNR of the posterior frames in the sequence. As shown in Fig. 3 for the *Balloons* test sequence encoded at a target bit rate of 110kbps, as $\gamma$ increases, the average PSNR of the synthesized view increases, however, as expected, the average PSNR of the posterior frames degrades rapidly after $\gamma = 0.5$, since the number of bits available for the posterior frames decreases. Considering these factors and after analysing the variation of $\gamma$ for various test sequences, the value of $\gamma$ was empirically set to 0.41 as this is the best trade-off between the identified quality metrics.

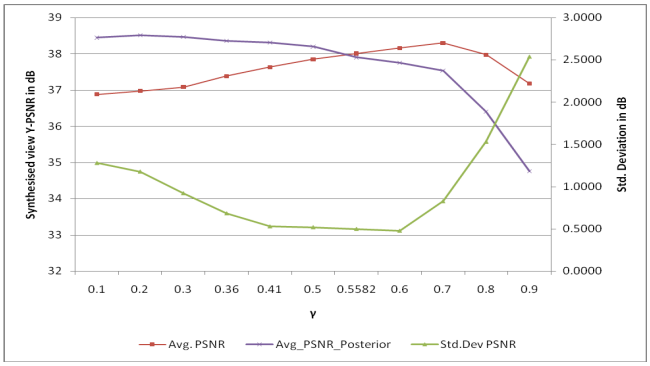


Fig. 3: Variation of tuning parameter for the *Balloons* test sequence

Moreover, experiments on texture and depth map views on the test sequences shown in Table 3 reveal that depth maps have on average 26% of CTUs with zero complexity cost whilst generally texture views do not have any CTUs with zero complexity cost. This is not surprising as depth maps have a large number of flat areas when compared to texture views. For this reason, the key-frame R-λ model, given by (7), needs to be adapted for the depth maps accordingly.

$$\lambda_{pic} = \alpha_{K,pic}\left(\frac{C}{\left(T_{key}/P_x\right)}\right)^{\beta_{K,pic}} \quad (7)$$

where $C$ is the complexity measure based on SATD [11], $T_{key}$ is the refined bit budget of the key frame, $P_x$ is the number of pixels in a frame, whilst $\alpha_{K,pic}$ and $\beta_{K,pic}$ are model parameters which are empirically set to fit the hyperbolic model.

To fit the hyperbolic model, an experiment was carried out whereby the MV-HEVC encoder was configured such that the depth map frames across the three views are encoded as key frames. Test sequences, shown in Table 3, were encoded and the average bits per pixel (bpp), frame complexity $C$, and $\lambda_{pic}$ at four QPs (34, 39, 42, 45 - as defined in the Common Test Conditions (CTC) [18]) were used to model the relationship between $\lambda_{pic}$ and ($C/bpp$) for the depth map key frames across the 3 views. Fig. 4 shows the relationship between these two variables and the model

parameter estimates for the *Balloons* test sequence across the 3 views. Taking the average value of the model parameter estimates for the various test sequences yields $\alpha_{K,pic} = 0.028$, and $\beta_{K,pic} = 1.9495$.

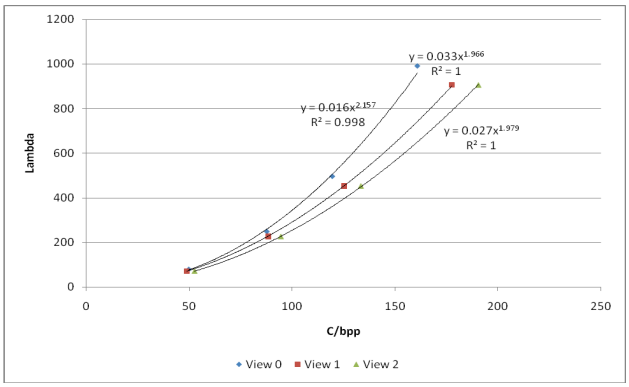



Fig. 4: R-λ modelling for the *Balloons* test sequence

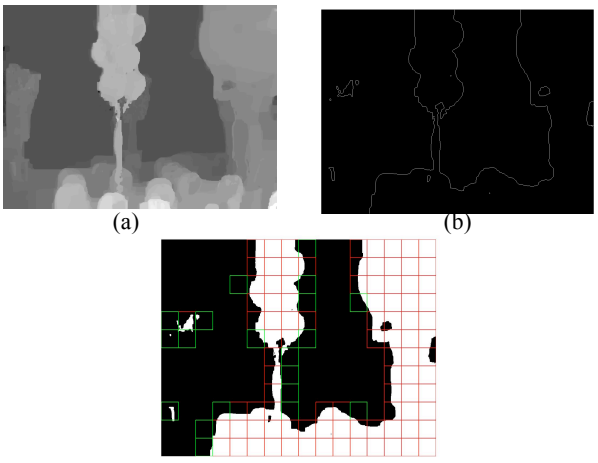
Using the estimated $\lambda_{pic}$, the QP for the frame is then calculated using [10]:

$$QP = 4.2005 \ln(\lambda) + 13.7122 \quad (8)$$

## C. CTU-level bit allocation

At the CTU level, the bit budget allocated to the frame is distributed across the CTUs in the frame depending on their cost complexity. The R-λ models used at the frame level, given by (5) and (7), are re-used at the CTU level. However, the R-λ model parameter estimates of the CTUs in a frame are initially set to the frame level model parameter estimates and then adaptively tuned, on a CTU level basis, during the CTU parameter update procedure described in [7]. Using the CTU level bit allocation and R-λ model, the $\lambda_{CTU}$ is estimated, clipped and then used to calculate $QP_{CTU}$ using (8), which is also clipped prior to being used for unit encoding. In the proposed rate control scheme, we introduce an algorithm which adaptively tunes the clipping range of $\lambda_{CTU}$ and $QP_{CTU}$ based on the type of CTU being processed. In doing so, we allow a wider clipping range for the CTUs which are in the foreground and edges in the depth map.

Inspired by the work in [19], CTUs in a frame are classified into 3 categories namely foreground, background, and edge type CTU depending on their characteristics using the algorithm shown in Fig. 6. This algorithm takes as input a binary edge map generated by an edge detection technique. Several edge detection techniques can be found in literature however, in this work, a Canny edge filter based on Laplacian edge detection is used due to its low complexity. Here the edges are detected by searching for zero crossings in the second derivative of the image. In order to improve the noise immunity, the image is first Gaussian smoothed prior to applying the Laplacian filter yielding the Laplacian of Gaussian (LoG) detection which forms the basis of Canny edge detection [20]. Thus, taking the original video depth map shown in Fig. 5(a), the binary edge map at the output of the Canny edge filter is shown in Fig. 5(b). This binary edge map is then processed together with the original video depth map using the CTU mask generation algorithm given in Fig. 6 to generate the mask shown in Fig. 5(c), where CTUs marked in red are treated as foreground while those marked in green are treated as edges. The rest are treated as background CTUs.



(a) (b)

(c)

Fig. 5: CTU mask generation for *Balloons* – Camera 1

```
Algorithm 1 CTU Mask Algorithm
for All binary edge map frames do
    for Each LCU in binary edge map do
        Split LCU into 16 16x16 Sub-blocks
        for Each Sub-block do
            Check whether sub-block contains an edge
            if Sub-block contains an edge then
                Mark sub-block as containing an edge
            end if
        end for
        for Each sub-block containing an edge in LCU do
            Check whether neighbouring sub-blocks contain an edge
            if Neighbouring sub-blocks containing an edge is 3 or more then
                Flag LCU as an Edge
            end if
        end for
    end for
end for

for All depth map frames do
    Extract foreground mask by depth map thresholding
    for Each LCU in foreground mask do
        if LCU in foreground mask is also marked as edge LCU then
            Mark LCU as Edge
        end if
        if LCU in foreground mask is not marked as edge LCU then
            Mark LCU as Foreground
        end if
    end for
    for All the remaining LCU not in foreground mask do
        Mark LCU as Background
    end for
end for
```

Fig. 6: CTU mask Algorithm

Using the CTU mask, the clipping range of $\lambda_{CTU}$ and $QP_{CTU}$ is then adapted based on the current and previous CTU type as shown in (9).

If $T_{remain} < T_{thresh}$ then:

$$\lambda_{CTU\_prev} \cdot 2^{-\frac{1}{3}} \le \lambda_{CTU\_cur} \le \lambda_{CTU\_prev} \cdot 2^{\frac{1}{3}}$$
$$\lambda_{pic} \cdot 2^{-\frac{2}{3}} \le \lambda_{CTU\_cur} \le \lambda_{pic} \cdot 2^{\frac{2}{3}}$$
$$QP_{CTU\_prev} - 1 \le QP_{CTU\_cur} \le QP_{CTU\_prev} + 1$$
$$QP_{pic} - 2 \le QP_{CTU\_cur} \le QP_{pic} + 2 \quad (9)$$

Else

$$\lambda_{CTU\_prev} \cdot 2^{-\left(\frac{A}{3}\right)} \le \lambda_{CTU\_cur} \le \lambda_{CTU\_prev} \cdot 2^{\left(\frac{B}{3}\right)}$$
$$\lambda_{pic} \cdot 2^{-\frac{2}{3}} \le \lambda_{CTU\_cur} \le \lambda_{pic} \cdot 2^{\frac{2}{3}}$$
$$QP_{CTU\_prev} - \Delta \le QP_{CTU\_cur} \le QP_{CTU\_prev} + \Delta$$
$$QP_{pic} - 4 \le QP_{CTU\_cur} \le QP_{pic} + 4$$

where $\lambda_{CTU\_prev}$ and $\lambda_{CTU\_cur}$ are the $\lambda$ of the previous and current CTU respectively, $QP_{CTU\_prev}$ and $QP_{CTU\_cur}$ are the QP of the previous and current CTU respectively, $\lambda_{pic}$ and $QP_{pic}$ are the $\lambda$ and QP of the frame respectively, $T_{remain}$ and $T_{thresh}$ are the number of bits remaining to code the current frame and low bit threshold respectively, whilst the parameters *A, B,* and $\Delta$ depend on the previous and current CTU type and are given in Table 1.

**Table 1.** Adaptive clipping parameters

| Previous CTU Type | Current CTU Type | *A* | *B* | $\Delta$ |
|---|---|---|---|---|
| Background | Background | 1 | 1 | 1 |
| Foreground | Foreground | 2 | 2 | 2 |
| Edge | Edge | 2 | 2 | 2 |
| Edge | Foreground | 2 | 2 | 2 |
| Foreground | Edge | 2 | 2 | 2 |
| Background | Edge | 4 | 4 | 3 |
| Foreground | Background | 2 | 4 | 3 |
| Edge | Background | 2 | 4 | 3 |
| Background | Foreground | 4 | 4 | 3 |

The parameters shown in Table 1 were empirically selected such that the salient characteristics of the depth map are retained whilst minimizing frame bit exhaustion. The latter was also tackled through the use of a low bit threshold check at the start of processing of each CTU in the frame.

## 4. SIMULATION ENVIRONMENT AND RESULTS

The proposed rate control algorithm was implemented in HEVC reference software HM11.0 [21], which was configured for multi-view compression using the low-delay prediction structure, shown in Fig. 1, and with the encoding parameters shown in Table 2. Moreover, the target bit rates for the rate control algorithm were set as per the CTC described in [18]. The performance of the proposed rate control algorithm, in terms of synthesized view PSNR improvement, Bjøntegaard Delta PSNR (BD-PSNR) [22], and bit rate error improvement, was assessed and compared with the latest reference algorithm implemented in MV-HEVC [11]. As shown in Table 3, the average improvement of the synthesized view PSNR and BD-PSNR are 1.15% and 0.45dB respectively. Moreover, this improvement in visual quality is achieved whilst reducing the bit rate error by an average of 0.2% when compared to the reference algorithm in MV-HEVC. These gains are achieved through better bit budget allocation to the key frames and salient CTUs together with an improved R-λ model for the depth map videos. Moreover, the complexity of the proposed algorithm is similar to that of the reference algorithm in MV-HEVC with a marginal increase in encoding time of 0.1% owing to the low complexity CTU mask algorithm used.

**Table 2.** Encoding parameters

| | |
|---|---|
| GOP Size | 4 |
| Intra Period | 24 |
| LCU Level Rate Control | 1 |
| LCU size | 64 x 64 |
| Partition Depth | 4 |
| LCU separate RC Model | True |
| $T_{thresh}$ | 5% of frame allocation |

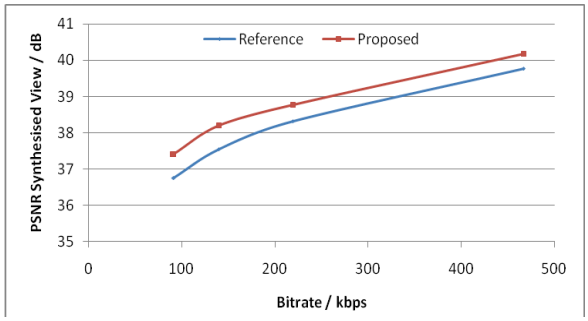



Fig. 7: R-D plot - *Balloons* Synthesized View 3

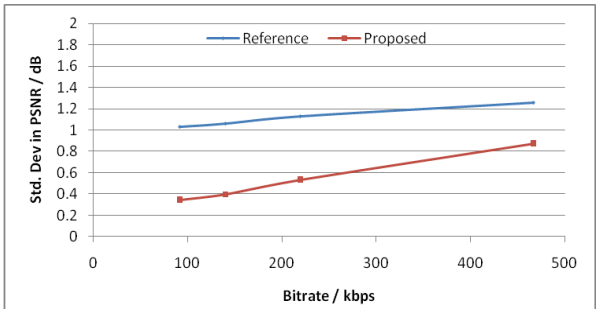



Fig. 8: Std. Deviation plot - *Balloons* Synthesized View 3

As shown in Fig. 7, the proposed rate control algorithm imparts a better R-D performance when compared to the reference algorithm used in MV-HEVC. Moreover, the standard deviation of the synthesized view quality is also lower owing to smaller bit-budget estimation errors, which result in smoother buffer fullness variation and hence more consistent video quality.

**Table 3.** Rate Control Performance
(synthesized view is indicated by the camera number shown in **bold**)

| Sequence / Resolution | Target bitrate (kbps) | % Δ PSNR | % Δ bitrate error | BD-PSNR (dB) |
|---|---|---|---|---|
| *Balloons (1-**3**-5) 1024x768* | 466.7 | 1.01 | -0.61 | 0.55 |
| | 219.3 | 1.23 | -0.79 | |
| | 140.1 | 1.28 | -0.00 | |
| | 91.4 | 1.88 | -0.03 | |
| *Book (8-**9**-10) 1024x768* | 295.7 | 1.20 | -0.27 | 0.41 |
| | 152.1 | 0.91 | -0.18 | |
| | 101.5 | 1.29 | -0.13 | |
| | 69 | 1.13 | -0.09 | |
| *Kendo (1-**3**-5) 1024x768* | 597.5 | 0.76 | -0.22 | 0.3 |
| | 314.7 | 0.72 | -0.15 | |
| | 213.9 | 0.66 | -0.10 | |
| | 147.5 | 0.60 | -0.08 | |
| *Newspaper (2-**4**-6) 1024x768* | 370.1 | 1.47 | -0.65 | 0.66 |
| | 181.5 | 1.79 | -0.07 | |
| | 115.9 | 2.16 | -0.05 | |
| | 77.2 | 2.70 | -0.04 | |
| *PoznanHall (7-**6**-5) 1920x1088* | 205.8 | 1.01 | -0.28 | 0.39 |
| | 108.3 | 1.06 | -0.16 | |
| | 77.8 | 0.71 | -0.07 | |
| | 57 | 0.32 | -0.06 | |
| *PoznanStreet (5-**4**-3) 1920x1088* | 332.7 | 0.69 | -0.21 | 0.39 |
| | 150.2 | 1.12 | -0.34 | |
| | 96.7 | 0.82 | -0.11 | |
| | 66.4 | 1.04 | -0.02 | |
| **Average** | | **1.15** | **-0.20** | **0.45** |

## 5. CONCLUSION

In this paper we have presented a depth map specific rate control algorithm for MV-HEVC MVD representation based on a bit allocation refinement technique for key frames together with a depth-map centric R-λ model and an adaptive clipping function to exploit the depth map characteristics. The texture rate control algorithm was not modified in this work. Experimental results using standard test sequences suggested in the CTC [18] have shown that the proposed rate control scheme improves the synthesized view PSNR and BD-PSNR by 1.15% and 0.45dB respectively when compared to the reference algorithm used in MV-HEVC. The improvement in the synthesized view visual quality is achieved whilst improving the rate control accuracy by 0.2% and reducing the PSNR standard deviation, thereby ensuring a more consistent video quality and making this scheme attractive for MVD transmission over bandwidth limited channels.

## 6. REFERENCES


[1] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in Proc. IEEE Int. Conf. Image Process., Sep. 2007, pp. 201-204.

[2] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI, 2004, vol. 5291.

[3] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," IEEE Trans. Circuits & Syst. for Video Technol., vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[4] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," IEEE Trans. on Broadcast., vol. 55, no. 4, pp. 761–766, Dec. 2009.

[5] A. Vetro, S. Uea, and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays," in Proc. of SPIE Conference on Applications of Digital Image Processing XXXI, vol. 7073, Sep. 2008.

[6] H. Choi, J. Nam, J. Yoo and D. Sim, "Rate control based on unified RQ model for HEVC," in JCT-VC H0213, 8th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC29/WG 11, San Jose, CA, USA, 2012.

[7] B. Li, H. Li, L. Li and J. Zhang, "Rate control by R-lambda model for HEVC," in JCT-VC K0103, 11th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Shanghai, CN, 2012.

[8] L. Sun, O. Au and W. Dai, "An adaptive frame complexity based rate quantization model for intra-frame rate control of high efficiency video coding (HEVC)," APSIPA ASC, 2012, 1–6.

[9] J. Si, S. Ma and W. Gao, "Adaptive rate control for HEVC," in JCT-VC I0433, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Geneva, CH, 2012.

[10] B. Li, D. Zhang, H. Li, and J. Xu, "QP determination by lambda value", in JCTVC-I0426, Geneva, CH, 27 April - 7 May 2012.

[11] X. Wang and M. Karczewicz, "Intra frame rate control based on SATD," in JCT-VC M0257, Incheon, 2013.

[12] M. Wang, K. Ngi Ngan and H. Li, "An efficient frame-content based Intra frame rate control for high efficiency video coding," in IEEE Signal Processing Letters, vol. 22, no. 7, pp. 896-900, July 2015.

[13] J. Si, S. Ma and W. Gao, "Efficient bit allocation and CTU level rate control for high efficiency video coding," Proc. of Picture Coding Symposium, pp.89-92, Dec. 2013.

[14] J. Si, S. Ma and W. Gao, "Laplace distribution based CTU level rate control for HEVC," Proc. of Visual Communications and Image Processing, pp. 1-6, Nov. 2013.

[15] W. Lim, D. Sim and I. V. Bajić, "JCT3V – Improvement of the rate control for 3D multi-view video coding," ISO/IEC JTC1/SC29/WG11, JCT3V-C0090, Geneva, Switzerland, Jan. 2013.

[16] Tan, S., Si, J., Ma, S., Wang, S., Gao, W., "Adaptive Frame Level Rate Control in 3D-HEVC," Visual Communication and Image Processing, Malta, Dec. 2014.

[17] A.Vetro, Y. Chen, K. Müller, "HEVC-Compatible Extensions for Advanced Coding of 3D and Multiview Video", Data Compression Conference (DCC), Apr. 2015.

[18] Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) of ITU-T VCEG and ISO/IEC MPEG, "Common test conditions of 3DV core experiments," JCT3V-G1100, San Jose, USA, Tech. Rep., Jan. 2014.

[19] M. Cordina, C.J. Debono, "An adaptive lagrange multiplier technique for multi-view video plus depth coding," Proc. of Picture Coding Symposium, pp. 249-252, Dec. 2013.

[20] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679-697, 1986.

[21] "HEVC reference software," [Online]. Available: http:// hevc.kw.bbc.co.uk/trac/browser/jctvc-htm/tags/HM-11.0

[22] G. Bjøntegaard, "Calculation of average PSNR difference between RD-curves," Doc. VCEG-M33, Apr. 2001

# An Adaptive Lagrange Multiplier Technique for Multi-view Video plus Depth Coding

Mario Cordina[1], Carl J. Debono[2]
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
{[1]mario.cordina, [2]c.debono}@ieee.org

***Abstract*—The characteristics of the depth map video are different from those of the texture and thus the empirical function of the Lagrange multiplier λ normally used in the rate-distortion optimization (RDO) of texture views might not be suitable for depth map coding. In this paper, we propose a technique whereby the Lagrange multiplier used to select the macroblock (MB) mode is adapted based on the discontinuity region of the depth map and the frame type. This technique, which was tested on various standard test sequences, preserves the depth discontinuities leading to an average depth map bit rate saving of 12.5% without incurring significant quality degradation in the synthesized views compared to using the fixed Lagrange multiplier used in the respective multi-view texture videos.**

## I. INTRODUCTION

Three Dimension (3D) technologies have reached commercialization in a wide spectrum of products and applications ranging from smartphones, TV, gaming, and medical imaging. This widespread proliferation of 3D technology is driving the market which is expected to reach $227.27 billion in 2016, an estimated year-on-year increase of 15.8% [1]. An application of Multi-view Video (MVV), which is increasingly gaining commercial and research interest, is Free-View Video (FVV) [2] whereby the user can interractively select an arbitrary viewing position in relation to the 3D scene being watched. This requires the use of a large number of cameras to capture a scene and thus the multi-view video plus depth (MVD) representation [3] is beneficial, as it allows the flexible rendering of a continuum of views at the receiving end through the use of depth image-based rendering techniques [4]. In MVD, a small set of texture views together with a corresponding set of per-pixel depth map views are independently compressed using the Multi-view Video Coding (MVC) standard by exploiting the redundancy in space, time and in between views [5, 6]. As the characteristics of depth map views is significantly different from those of texture views, the value of the Lagrange multiplier λ used to encode texture views might not be the best option for the depth map views.

In this paper, we propose a technique to adapt the Lagrange multiplier, used during depth map coding, at a macroblock (MB) level based on the discontinuity region of the depth map and the frame type. The rest of the paper is organized as follows: Section 2 gives a summary of the rate-distortion optimization (RDO) mode selection process used in H264/MVC; Section 3 describes the proposed technique to adapt the Lagrange multiplier λ for the encoding of depth map views; Section 4 gives a description of the simulation environment and results, while Section 5 provides some comments and conclusions.

## II. RATE DISTORTION OPTIMIZED MB MODE SELECTION

Similar to H264/AVC, the MVC encoder uses a RDO mode selection technique to choose the coding mode of a MB based on the Rate-Distortion (R-D) cost of each coding mode. The supported MB prediction modes include SKIP, DIRECT, Inter16×16, Inter16×8, Inter8×16, Inter8×8, Inter8×8Frext, Intra16×16, Intra8×8 and Intra4×4. The prediction mode with the lowest R-D cost, given by (1), is chosen to encode the MB [7].

$$\begin{aligned} &J(s,c,MODE \mid QP) \\ &= SSD(s,c,MODE \mid QP) + \lambda_{MODE} R(s,c,MODE \mid QP) \end{aligned} \quad (1)$$

where $s$ and $c$ denote the source and reconstructed MB respectively, *MODE* is the candidate MB mode, *QP* is the MB quantization parameter, $R(s, c, MODE \mid QP)$ is the total bits of the MB header, motion vectors and discrete cosine transform (DCT) coefficients, $\lambda_{MODE}$ is the Lagrange multiplier for the mode decision, and $SSD(s, c, MODE \mid QP)$ is the sum of square differences which reflects the distortion between the original and reconstructed MB given by:

$$SSD(s,c,MODE \mid QP) = \sum_{x,y} \left( s[x,y] - c[x,y] \right)^2 \quad (2)$$

The Lagrange multiplier $\lambda_{MODE}$ controls the trade-off between the rate and distortion. As $\lambda_{MODE}$ decreases, the R-D cost function is biased towards minimizing the distortion whilst allowing a higher rate and vice-versa. Hence, selecting the optimal value of $\lambda_{MODE}$ for a specific sequence is not trivial. Several studies on texture views were carried out to find empirical solutions to this problem [8]-[10], however the most commonly used solution is to calculate $\lambda_{MODE}$ as a function of QP as shown in (3).

$$\lambda_{MODE} = 0.85 \times 2^{\frac{(QP-12)}{3}} \quad (3)$$

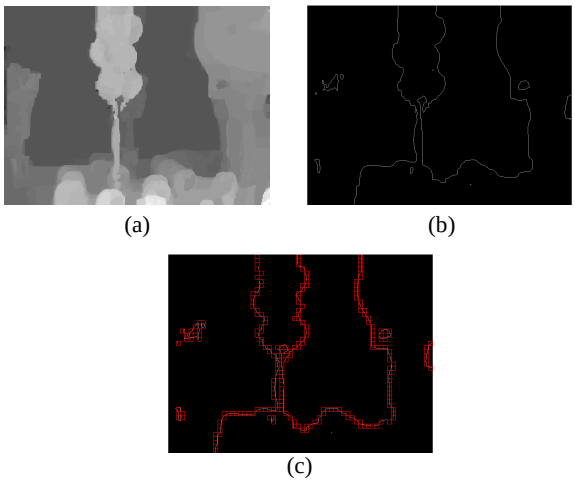
Depth map views represent the distance between an object and a camera as a gray scale image. Thus, they have limited texture and prominent contours giving rise to depth discontinuities which are critical for the correct synthesis of intermediate views. These characteristics are significantly different from those of texture views however, the same empirical function shown in (3), is usually used for the depth map view.

## III. PROPOSED TECHNIQUE

As discussed in Section II, the preservation of depth discontinuities is critical for the correct synthesis of intermediate views. Thus, an MB edge mask indicating MBs containing edges, needs to be generated.

### A. *Edge Detection*

Various edge detection techniques have been proposed in recent years, amongst which, Laplacian edge detection techniques [11] are commonly used due to their high noise immunity. These detect edges by searching for zero crossings in the second derivative of the image whilst the high noise immunity is achieved by Gaussian smoothing the image prior to applying the Laplacian filter, yielding the Laplacian of Gaussian (LoG) detection. This forms the basis of Canny edge detection [12] which is used in this work to generate a binary edge map from the video depth map. Considering the original video depth map of the *Balloons* sequence shown in Fig. 1(a), the binary edge map at the output of the Canny edge filter is shown in Fig. 1(b). The algorithm shown in Fig. 2 is used to create an improved MB edge mask by reducing false edges created by the Canny filter. This results in the MB edge mask shown in Fig. 1(c).



(a) (b)

(c)

**Figure 1.** MB edge mask generation for *Balloons* – Camera 1

### B. *Adaptation of Lagrange multiplier λ*

Using the MB edge mask generated earlier, the constant term in (3) (usually fixed to 0.85) is modified such that MBs located at edge boundaries will have their Lagrange multiplier lowered, thereby giving more weight to the distortion term in

```
Algorithm 1 Macroblock edge mask algorithm
  for all Frames in Video Depth Map do

    for Each Macroblock in Video Depth Frame do
      Split MB into 16 4x4 Sub-Macroblocks

      for Each Sub-Macroblock do
        if Sub-Macroblock contains an Edge then
          Mark Sub-Macroblock as containing an Edge
        end if
      end for

      for Each Sub-Macroblock containing an Edge do
        Check whether the neigbouring 4 Sub-Macroblocks contain an edge

        if Numb. of Sub-Macroblocks containing an Edge greater than 3 then
          Flag MB as containing an Edge
        end if

      end for
    end for
  end for
```

**Figure 2.** MB edge mask algorithm

the RDO model. Conversely, MBs located in relatively flat areas will have their constant term modified such that the Lagrange multiplier is increased giving more weight to the rate term in the RDO model. Moreover, as the frame type has a significant effect on the coding efficiency, this adaptation is carried out for all MBs in a predicted (P) frame. The MBs in an Intra (I) frame still have a fixed constant term, set to give more weight to the distortion term in the RDO model. This adaptation process is summarized in Fig. 3, where $\alpha_{edge}$ and $\alpha_{non\text{-}edge}$ are the parameters used to modify the constant term in the Lagrange multiplier of MBs at edge boundaries and flat areas respectively.

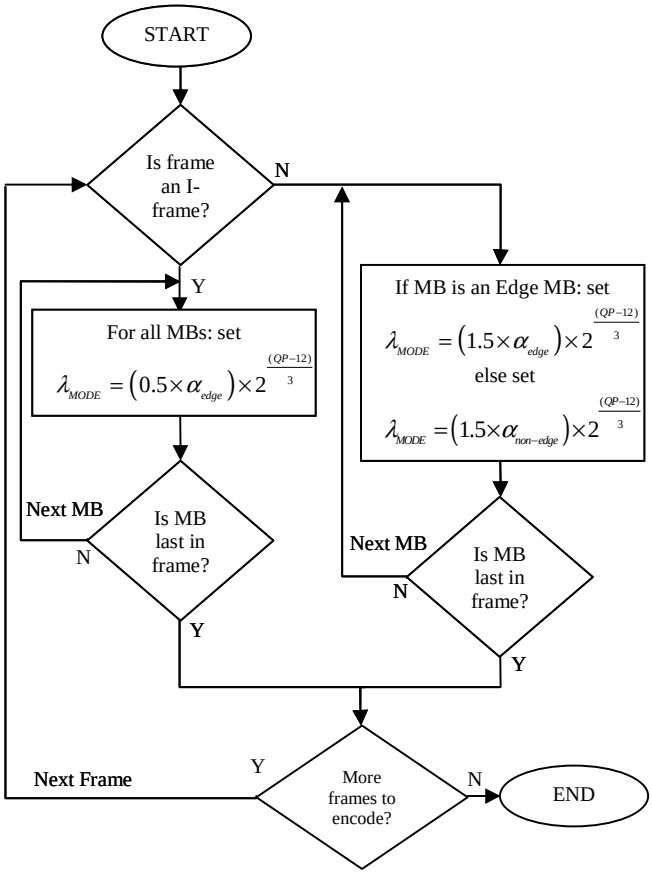



**Figure 3.** Flowchart of proposed technique

## IV. Simulation Environment and Results

A series of experiments were conducted to test the efficacy of the proposed technique and study the relationship between the newly defined parameters ($\alpha_{edge}$, $\alpha_{non\text{-}edge}$) and the quantization parameter for the depth map views ($QP_{Depth}$) for various test sequences. The test setup used for these experiments is shown in Figure 4, while the encoding parameters together with the test sequences used and their characteristics are shown in Tables 1 and 2, respectively.

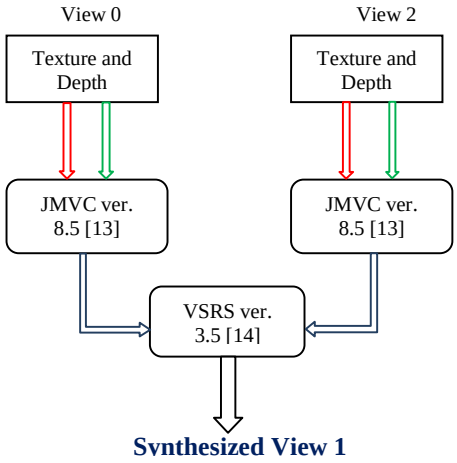



**Figure 4.** Test setup used during experiments

**Table 1.** Encoding parameters

| | |
|---|---|
| GOP Size | 1 |
| Intra Period | 12 |
| $QP_{Texture}$ | 32 |
| $QP_{Depth}$ | 24, 28, 32, 38 and 42 |
| Prediction Structure | I-P-P-P in time-domain [15] |
| Encoded Frames | 60 |

**Table 2**. Test sequence parameters and characteristics

| Sequence | Image Size | Cameras (L-C-R) | Characteristics (video features, frame rate, baseline) |
|---|---|---|---|
| *Balloons* | 1024 x 768 | 1-2-3 | Complex object motion; 30 fps; 5cm in 1D plane |
| *Champ. Tower* | 1280 x 960 | 37-38-39 | Slow but complex object motion; 30 fps; 5cm in 1D plane |
| *Pantomine* | 1280 x 960 | 37-38-39 | Medium complex object motion; 30 fps; 5cm in 1D plane |
| *Book Arrival* | 1024 x 768 | 10-9-8 | Moderate object motion; 15 fps; 6.5cm in 1D plane |

For each test sequence and $QP_{Depth}$, the parameters $\alpha_{edge}$ and $\alpha_{non\text{-}edge}$ were varied between [0.6-2.5] and [0.7-2.2] respectively. The parameter settings which yielded the best compromise between visual quality degradation and depth bit rate savings, when compared to the fixed constant case (0.85), were determined experimentally. With reference to Figure 5 and Figure 6, the parameters $\alpha_{edge}$ and $\alpha_{non\text{-}edge}$ generally follow a downward linear trend with $QP_{Depth}$ except for the case of the *Champ. Tower* sequence. In this case, $\alpha_{non\text{-}edge}$ increases with $QP_{Depth}$ due to the slow object motion and relatively simple background of this specific test sequence, allowing a higher Lagrange multiplier in the background areas. Moreover, it is evident that although the gradient is approximately the same across all tested sequences, the intercept point is sequence dependent. Hence further studies need to be carried out to identify the relationship between the sequence characteristics and parameter set ($\alpha_{edge}$, $\alpha_{non\text{-}edge}$) to allow for the real-time estimation of these parameters using sequence statistics and $QP_{Depth}$.

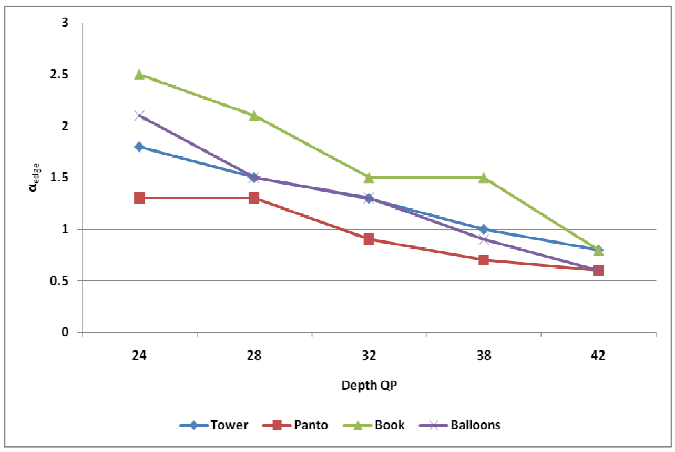



**Figure 5.** $\alpha_{edge}$ with $QP_{Depth}$

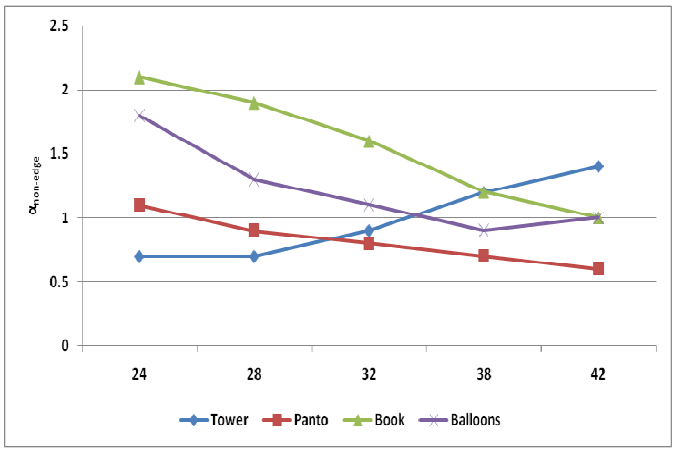



**Figure 6.** $\alpha_{non\text{-}edge}$ with $QP_{Depth}$

Through the adaptation of the Lagrange multiplier, the proposed technique is able to preserve the depth discontinuities. This leads to significant depth bit-rate savings over all the test sequences and $QP_{Depth}$ without significantly degrading the synthesized view quality, when compared to the fixed Lagrange multiplier case, as shown in Figure 7 and 8 respectively. As expected, as $QP_{Depth}$ increases, the bit rate

savings diminishes as we start being limited by the amount of Peak Signal-to-Noise Ratio (PSNR) degradation that we can allow. Nevertheless, the proposed technique achieves an average depth bit-rate saving of 12.5% over all test sequences and $QP_{Depth}$. This result is obtained with a PSNR loss which is kept below 0.1dB compared to the fixed $\lambda$ scenario.

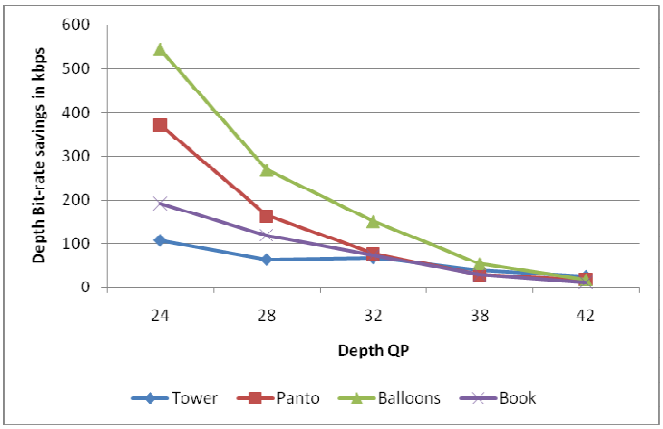



**Figure 7.** Depth Bit-Rate savings with $QP_{Depth}$

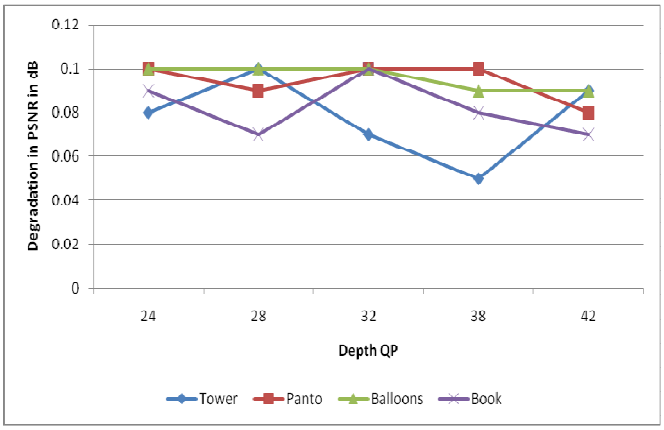



**Figure 8.** Degradation in PSNR from reference case with $QP_{Depth}$

## V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an adaptive Lagrange multiplier technique for the encoding of depth map videos in MVD systems. The MB edge mask, obtained from the depth map sequence, is exploited to selectively modify the Lagrange multiplier of MBs depending on their location within the depth map and frame type. Simulation results on test sequences exhibiting a variety of video characteristics and encoded using different $QP_{Depth}$ have shown that the proposed technique is able to preserve the depth discontinuities and leads to an average depth bit-rate saving of 12.5%, without any significant degradation in PSNR. In our future work, we shall study the relationship between the sequence characteristics and parameter set ($\alpha_{edge}$, $\alpha_{non\text{-}edge}$) to find a general solution that allows real-time estimation of these parameters using sequence statistics and $QP_{Depth}$.


## ACKNOWLEDGEMENT

The authors would like to thank Nagoya University for providing *Balloons*, *Champagne Tower*, *Pantomine* test sequences and Fraunhofer-HHI for providing *Book Arrival* test sequence.


## REFERENCES


[1] 3D/4D Technology Market by Products, Applications & Technology. Internet:http://www.marketsandmarkets.com/ Market-Reports/3d-4d-technology-market-646.html

[2] A. Smolic, and P. Kauff, "Interactive 3-D Video representation and coding technologies," in *Proc. of the IEEE*, vol. 93, no.1, pp. 98-110, Jan 2005.

[3] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Process.*, pp. 201-204, Sep. 2007.

[4] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291.

[5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. Circuits & Syst. for Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[6] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. on Broadcast.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[7] Y. S. Ho, and K. J. Oh, "Overview of multiview video coding," in *Proc. of International Workshop on Systems, Signals and Image Processing*, pp. 5-12, Jun. 2007.

[8] S. Kim, Y. Choe, "Rate-Distortion optimization for mode decision with sequence statistics in H264/AVC," in *Proc. of the Fourth International Conference Signal Processing, Pattern Recognition and Applications*, pp. 140-145, Feb. 2007.

[9] T. Wiegand, and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. of Int. Cong. on Image Processing*, pp. 542-545, Oct. 2001.

[10] K. Takagi, "Lagrange multiplier and RD-characteristics (JVT-C084)," in JVT Meeting (Joint Video Team of ISO/IEC MPEG & ITU-T VCEG), Fairfax, USA, May 6-10, 2002.

[11] M. Juneja, and P. Singh Sandhy, "Performance evaluation of edge detection techniques for images in spatial domain", *Int. Journal of Comp. Theory and Eng.*, vol. 1, no. 5, pp. 614-621, Dec. 2009.

[12] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679-697, 1986.

[13] JMVC 8.5, Available: garcon.ient.rwthaachen.de, Sep. 2011.

[14] "View Synthesis Reference Software (VSRS) 3.5," wg11.sc29.org, March 2010.

[15] M. Cordina, C.J. Debono, "A Novel view-level target bit rate distribution estimation technique for real-time multi-view video plus depth", in *Proc. of IEEE Int. Conf. on Multimedia and Expo.*, pp. 878-883, July 2012.

# AN ADAPTIVE TEXTURE-DEPTH RATE ALLOCATION ESTIMATION TECHNIQUE FOR LOW LATENCY MULTI-VIEW VIDEO PLUS DEPTH TRANSMISSION

Mario Cordina[1], Carl J. Debono[2]
Department of Communications and Computer Engineering
University of Malta
Msida, Malta
{[1]mario.cordina, [2]c.debono}@ieee.org

**ABSTRACT**

This paper presents an adaptive texture-depth target bit rate allocation estimation technique for low latency multi-view video plus depth transmission using a multi-regression model. The proposed technique employ the prediction mode distribution of the macroblocks at the discontinuity regions of the depth map video to estimate the optimal texture-depth target bit rate allocation considering the total available bit rate. This technique was tested using various standard test sequences and has shown efficacy as the model is able to estimate, in real-time, the optimal texture-depth rate allocation with an absolute mean estimation error of 2.5% and a standard deviation of 2.2%. Moreover, it allows the texture-depth rate allocation to be adapted to the video sequence with good tracking performance, allowing the correct handling of scene changes.

***Index Terms*—**Multi-view video, online estimation, multi-regression models, texture-depth rate allocation

## 1. INTRODUCTION

Immersive 3D technology has seen significant improvements over the past few years and is estimated to increase household penetration from 10% in 2011 to more than 50% by 2019 worldwide [1]. One of the main drivers behind the proliferation of such devices is 3D TV [2] and associated 3D content, whereby customers enjoy an immersive multimedia experience. Another application of Multi-view Video (MVV), which is increasingly gaining commercial and research interest, is Free-View Video (FVV) [3]. In FVV, the user interactively selects an arbitrary viewing position in relation to the 3D scene being watched, thereby requiring a large number of cameras to capture a scene. The multiview video plus depth (MVD) representation [4] is attractive as it allows the flexible rendering of a continuum of views at the receiving end through the use of depth image-based rendering techniques [5]. Generally, in MVD, a small set of texture views together with a corresponding set of per-pixel depth map views are independently compressed using the Multi-view Video Coding (MVC) standard by exploiting the redundancy in space, time and in between views [6,7]. As the compressed MVD stream is typically transported over a bandwidth limited network, a rate control scheme operating across the view, texture-depth, and frame layer is essential. This, in turn, requires the correct estimation of the target bit rate allocation at each layer for optimum operation.

Although the depth information is used only for virtual view rendering, recent studies have shown that multi-view texture and depth compression has a significant effect on the virtual view rendering quality [4]. Similar to the texture data, any artefacts introduced by coding the depth map at a low bit rate can have a significant impact on the quality of the synthesized views [8]. Thus under the constraint of the total bit rate, the correct balance between texture and depth rate allocation is critical. In this paper, we propose a technique to estimate, online, the optimal texture-depth rate allocation by exploiting the prediction mode distribution of the macroblocks at the discontinuity regions of the depth map video. The proposed technique was tested using test sequences, which were not used during the development of the model, exhibiting widely differing video characteristics and capture techniques to ensure that the model can be generically applied whilst exhibiting good tracking performance when subjected to video scene changes.

The rest of the paper is organized as follows: Section 2 gives a summary of related work; Section 3 describes the proposed online texture-depth allocation estimation technique; Section 4 gives a description of the general linear model formulation and its performance while Section 5 provides some comments and conclusions.

## 2. RELATED WORK

The disparate bit rate allocation between the texture data and associated depth map can have an impact on the quality of the virtual views when the total bit rate is constrained. This problem has attracted research attention as it forms the basis of rate control for MVD applications. The authors of [5] adopted a generic approach whereby the texture-depth rate ratio is fixed (5:1). Although simplistic and suitable for real-time applications, this approach cannot guarantee an optimal virtual view rendering quality. In [9], the authors proposed a joint texture-depth bit allocation algorithm to combine the texture and depth rate-distortion (R-D) curves yielding a single R-D surface. The optimal texture-depth rate ratio is

then found using a fast hierarchical optimization algorithm employing an orthogonal search pattern which exploits the smooth monotonic properties of the R-D surface. The main drawback of this algorithm is that the authors assume that a real view is available at the synthesis position allowing a full reference assessment [10] of the synthesised view. Moreover, it requires offline analysis to generate the R-D surface making it unsuitable for real-time applications.

In [11], Liu *et al.* attempt to resolve the no-reference assessment problem, in texture-depth rate allocation, by proposing a distortion model to characterize the synthesized view quality without requiring the original reference image. Here the authors propose an additive distortion model accounting for the video coding, depth quantization and the inherent geometry induced distortion. Using this model, a joint rate allocation method searches for the correct balance between the texture and depth bit rate to maximize the view synthesis quality. Although the proposed algorithm is able to find an optimal balance between the texture and depth bit rate accurately, its application is restricted to non-real time applications due to the offline processing required and the relatively high complexity. In [12], Cheung et al. demonstrates that the optimal quantization levels for the texture and depth map is equivalent to the shortest path in a specially constructed 3-D trellis. Another solution to the texture-depth rate allocation problem was developed by Yuan *et al.* [13]. In their work, the authors propose a concise distortion model for the synthesized virtual view based on the average distortion of the left and the right view and the average distortion of the compressed depth maps. Using this distortion model and the fractional rate-quantisation model [14], the joint bit allocation problem is formulated as a constrained optimization problem. This is solved using the Lagrangian multiplier method yielding the optimal quantization step parameter for both texture, $Q_t$, and depth $Q_d$ as given by:

$$Q_t = \frac{a_t + \sqrt{\frac{\nu a_t a_d}{\mu}}}{R_c - b_t - b_d} \qquad Q_d = \sqrt{\frac{\mu}{\nu}\frac{a_d}{a_t}}Q_t \qquad (1)$$

where $R_c$ is the total bit rate contraint and $a_t$, $a_d$, $\mu$, $\nu$, $b_t$ and $b_d$ are model parameters which need to be estimated by pre-encoding the texture and depth map video at different quantization steps, limiting this solution to non-real time applications.

In comparison with other works, this solution makes use of hierarchical B coding order thereby constricting itself to non-real-time video sequences, as future frames are not known, and precluding scene change handling.

## 3. ONLINE ESTIMATION OF THE TEXTURE-DEPTH RATE ALLOCATION

### 3.1. Low latency prediction structure

Generally, video sequences are sequentially encoded using the H.264/MVC hierarchical B coding structure as this yields a better compression ratio. However, as discussed in Section 2, such coding order uses future frames making it unsuitable for low latency applications, such as videoconferencing. This can be overcome by using the I-P-P-P structure in the time domain by setting the Group of Pictures (GOP) size to 1. With this GOP setting, a real-time prediction structure, as shown in Fig. 1, can be implemented by setting the appropriate reference frames for each view.

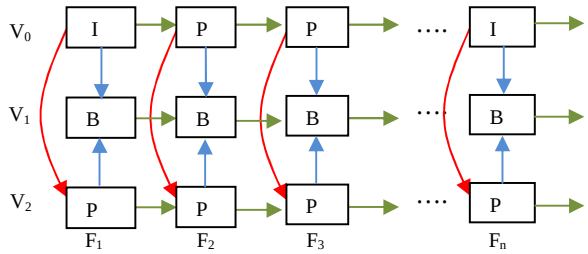



**Fig. 1.** Low latency prediction structure

### 3.2. Optimal Texture-Depth Ratio

Similar to the study carried out by Bosc *et al.* [15], a series of experiments were first conducted to determine the optimal texture-depth ratio for various total bit rates and test sequences. However, unlike [15], the low latency prediction coding structure was used during the experiments. The texture-depth ratio and total bit rate were adjusted by varying the Quantization Parameter (QP) for both the texture and the depth map video. The test setup used for these experiments is shown in Fig. 2 whilst the encoding parameters together with the test sequences used and their respective characteristics are shown in Table 1 and 2, respectively.

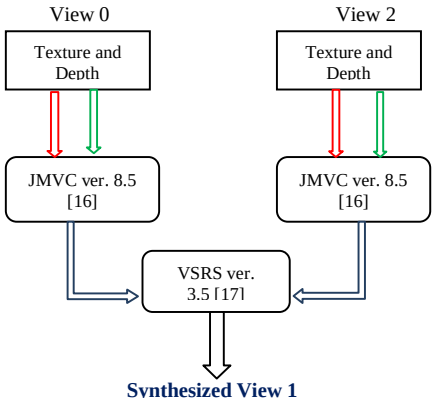



**Fig. 2.** Test setup used during experiments

**Table 1.** Encoding parameters

| | |
|---|---|
| GOP Size | 1 |
| Intra Period | 12 |
| $QP_{Texture}$ | 20 to 44 |
| $QP_{Depth}$ | 12 to 50 |
| Prediction Structure | Low latency |

**Table 2**. Test sequence parameters and characteristics
([V] indicates the sequence used for model verification only)

| Sequence | Image Size | Cameras (L-C-R) | Characteristics (video features, frame rate, baseline) |
|---|---|---|---|
| *Balloons* | 1024 x 768 | 1-2-3 | Complex object motion; 30 fps; 5cm in 1D plane |
| *Champ. Tower* | 1280 x 960 | 37-38-39 | Slow but complex object motion; 30 fps; 5cm in 1D plane |
| *Pantomine* | 1280 x 960 | 37-38-39 | Medium complex object motion; 30 fps; 5cm in 1D plane |
| *Kendo* | 1024 x 768 | 1-2-3 | Complex object motion with moving camera;30 fps; 5cm in 1D plane |
| *Ballet* | 1024 x 768 | 0-1-2 | Moderate object motion; 15 fps; 20cm in 1D arc |
| *Book Arrival*[V] | 1024 x 768 | 10-9-8 | Moderate object motion; 15 fps; 6.5cm in 1D plane |
| *Breakdancer*[V] | 1024 x 768 | 0-1-2 | Complex object motion; 15 fps; 20cm in 1D arc |

A plot of the synthesized view Peak Signal-to-Noise Ratio (PSNR) with the depth-texture ratio of the *Champagne Tower* video sequence is shown in Fig. 3. For a given total bit rate, the PSNR follows a downward concave parabola yielding an optimal depth-texture ratio of 40.7%. The PSNR of the synthesised view remains largely constant over a small range of depth-texture ratio (+/-5% in this example) however if we consider a depth-texture ratio of 30% or 60%, the PSNR degrades by 0.4dB and 1.1dB, respectively. Other sequences exhibit a narrower concave parabola yielding a larger degradation in PSNR when deviating away from the optimal value, highlighting the importance of a model which is able to estimate the optimal depth-texture ratio accurately.

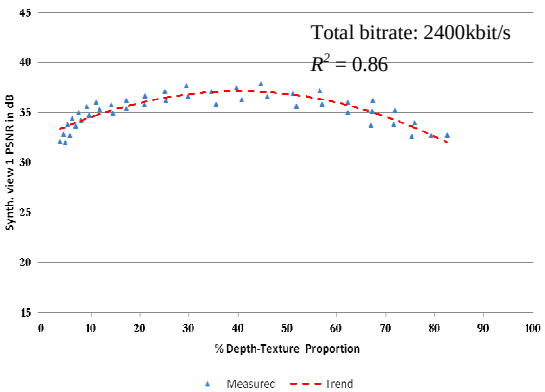



**Fig. 3.** *Champagne Tower* - PSNR vs depth-texture ratio

Plotting the optimal depth-texture ratio for increasing values of total bit rate (lowering $QP_{Texture}$ and $QP_{Depth}$) and test sequences, imparts different linear trends between the test sequences as shown in Fig. 4. This implies that the optimal depth-texture ratio is to a certain extent dependent upon the test sequence characteristics and the total bit rate.

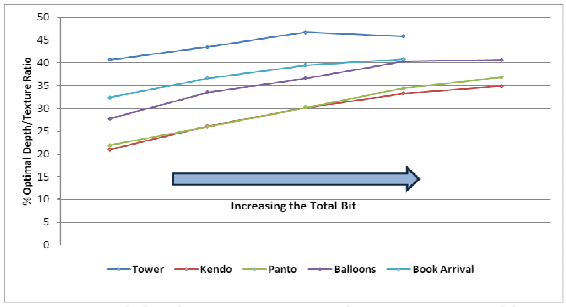



**Fig. 4.** Optimal depth-texture ratio with increasing total bit rate

In this study we explore the correlation between the depth map characteristics, specifically the macroblock (MB) mode type distribution at the discontinuity regions of the video depth map, $QP_{Depth}$, and the optimal depth-texture ratio. Prior to determining the MB mode type distribution at the discontinuity regions of the video depth map, an MB edge mask indicating MBs containing edges, needs to be generated for each view.

### 3.3. Edge detection and Macroblock mode distribution

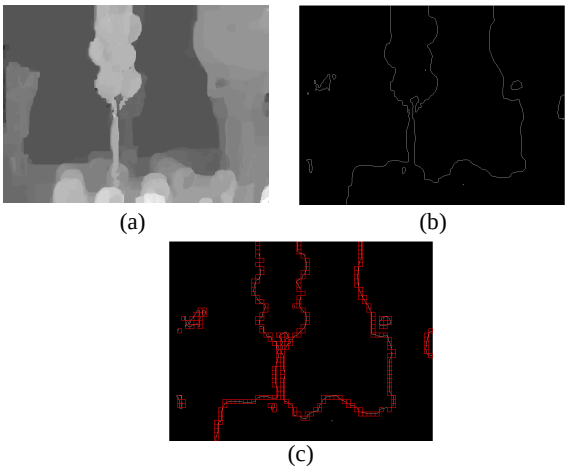
Edge detection is an important research topic in image processing and various edge detection techniques have been proposed in recent years. However, the majority of the techniques may be grouped into two categories namely:

*Gradient edge detection [18]:*These detect edges by searching for the extrema in the first derivative of the image, such as the Prewitt and Sobel filter. Although they are easy to compute they exhibit a high noise sensitivity.

*Laplacian edge detection [18]:*These detect edges by searching for zero crossings in the second derivative of the image. To improve the noise immunity, the image is Gaussian smoothed prior to applying the Laplacian filter yielding the Laplacian of Gaussian (LoG) detection which forms the basis of Canny edge detection [19]. Canny edge detection is used in this work to generate a binary edge map from the video depth map. Considering the original video depth map shown in Fig. 5(a), the binary edge map at the output of the Canny edge filter is shown in Fig. 5(b). To create an MB edge mask whilst reducing false edges, an algorithm shown in Fig. 6 is used to generate an MB edge mask as shown in Fig. 5(c).

The supported MB prediction modes in the Joint Multi-View Model (JMVM) include SKIP, DIRECT, Inter16x16, Inter16x8, Inter8x16, Inter8x8, Inter8x8Frext, Intra 16x16, Intra8x8 and Intra4x4, where the prediction mode with the lowest R-D cost is chosen to encode the MB [2]. JMVC ver. 8.5 was modified to use the MB edge mask of each

frame and extract the cumulative prediction mode distribution, at the discontinuous regions of the depth map video, for various $QP_{Depth}$ settings. This prediction mode distribution captures the characteristics of the video depth map sequence and has a direct impact on the optimal depth-texture ratio. Thus, by observing the optimal depth-texture ratio at various $QP_{Depth}$ settings and prediction mode distribution, a statistical regression model can be formulated to estimate the optimal depth-texture ratio.



(a) (b)

(c)

**Fig. 5.** MB edge mask generation for *Balloons* – Camera 1

**Algorithm 1** Macroblock edge mask generation

```
for Each MB in Video depth frame do
    Divide MB into 16 4x4 subMB

    for Each subMB do
        if subMB contains an edge then
            Flag subMB as containing an edge
        end if
    end for

    for Each subMB containing an edge do
        Check whether the neighbouring 4 subMB contain an edge
        if 3 or more subMB are in cluster formation then
            Flag MB as containing an edge
        end if
    end for
end for
```

**Fig. 6.** MB edge mask algorithm

### 3.4. General multiple linear regression models

A number of regression models can be found in literature however multi-linear regression models are the most commonly used due to their well-known statistical properties. Multi-regression models are an extension of simple regression models in which several predictors ($QP_{Depth}$ and prediction mode distribution) are used to model a single dependent variable (optimal depth-texture ratio). Given that the dependent variable $y$ having a normal distribution and uncorrelated predictors, the multi-regression model for $n$ observed cases that expresses the dependent variable $y_n$ as a function of $p$ predictors $\beta_p$ is given by [20]:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2)$$

$$\mathbf{y} = \mathbf{x}\beta + \varepsilon \quad (3)$$

where $\beta$ is a $p$-vector of unknown regression parameters, $\varepsilon$ is an $n$-vector of unknown random error terms, $\mathbf{y}$ is an $n$-vector of responses, and $\mathbf{x}$ is an $n$ x $p$ matrix of known constants.

As long as the estimation errors have a normal distribution, the least square estimation method can be employed to estimate the regression coefficients. The least squares estimate $b$, given by:

$$b = \left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y} \quad (4)$$

is chosen to minimize the residual sum of squares given by :

$$\varepsilon'\varepsilon = \sum_{i=1}^{n} \varepsilon_i^2 \quad (5)$$

## 4. MULTI-REGRESSION MODEL FORMULATION AND RESULTS

The prediction mode distribution at the discontinuous region of the video depth map (extracted from the modified JMVC ver. 8.5), for all test sequences shown in Table 2, together with the optimal depth-texture ratio at various $QP_{Depth}$ settings were parsed and imported into IBM SPSS© statistical package for analysis. Before formulating a multi-regression model, the dependent variable was checked for normality using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.905 validating the normality criteria. SPSS© was used to formulate the multi-regression model which uses the least square estimation method described in Section 3.4 to estimate the parameter values. Not all parameters were included in the model due to the fact that some parameters had their p-value greater than 0.05, implying these were not correlated with the variation in the optimal depth-texture ratio. Moreover, other parameters were found to be correlated with each other and thus were eliminated to reduce multi-collinearity effects. The parameters and confidence levels of the parsimonious model are shown in Table 3 whereas the coefficient of determination,$R^2$, of the formulated model is 0.906.

**Table 3.** Parsimonous model

| Parameter | Coefficient | p-value |
|---|---|---|
| Constant Term | 76.76 | 0.000 |
| $QP_{Depth}$ | -1.275 | 0.000 |
| Intra 16x16 | -66.897 | 0.000 |
| Skip | 11.26 | 0.001 |
| Mode 8x8 | -31.264 | 0.001 |

Thus, the parsimonous multi-regression model, given by (6), is composed of 5 significant predictors and an offset parameter which depends upon the camera baseline.

$$OptimalDTR = 76.76 - 1.275QP_{Depth} - 66.897Intra16x16 + 11.26Skip - 31.264Mode8x8 + 20CM \quad (6)$$

where *OptimalDTR* is the estimated optimal depth-texture ratio, $QP_{Depth}$ is the quantization parameter for the video

depth map whilst *Intra*16*x*16, *Mode*8*x*8 and *Skip* are ratios based on the MB prediction mode distribution at the discontinuous regions of the depth map video and calculated over a number of frames and *CM* is a camera baseline dependent parameter which is set to 1 when cameras are arranged in 1D arc and 0 when cameras are arranged in 1D plane.

After fitting the multi-regression model, a number of diagnostic checks need to be carried out to detect any model uncertainty:

*Standardized Residual Distribution [21]*:The standardized residuals are checked for normality using the Kolmogorov-Smirnov test at the 0.05 level of significance yielding a p-value of 0.67, validating the normality criteria.

*Studentized Residual plot [21]:* To check the fitness of the model and detect any outliers, the studentized residual, shown in Fig. 7, was plotted against the unstandardized predicted optimal depth-texture ratio. As the points in the plot are distributed across a horizontal band, indicating homoscedasticity, and the majority of the points lie within the standard $\pm 2$ limit, the model is adequate for such data fit with a very small number of outliers (shown in red).

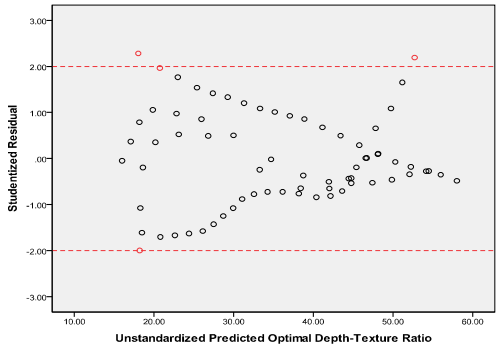



**Fig. 7.** Studentized residual plot

In order to verify the efficacy of the proposed solution, the optimal depth-texture ratio for various $QP_{Depth}$ settings was calculated online using the multi-regression model integrated within the JMVC framework. The plots of the estimated and actual optimal depth-texture ratio for the test sequences used during model formulation and verification are shown in Fig. 8, and Fig. 9, 10, respectively.

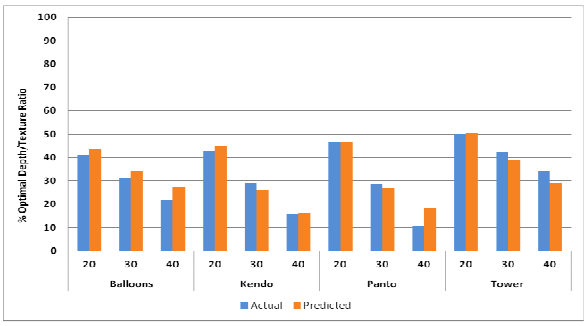



**Fig. 8.** Optimal depth-texture ratio – model formulation

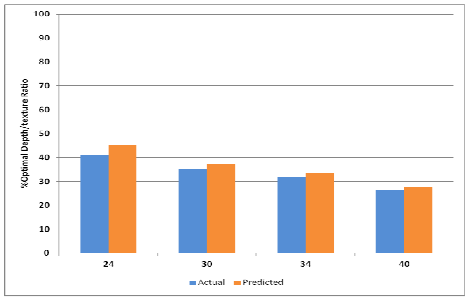



**Fig. 9.** Optimal depth-texture ratio – *Book Arrival* sequence

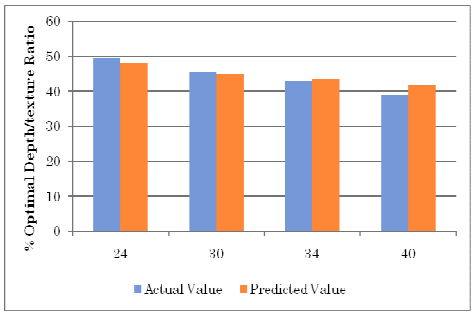



**Fig. 10.** Optimal depth-texture ratio – *Breakdancers* sequence

As shown in Table 4, the formulated multi-regression model was able to estimate the optimal depth-texture ratio with an overall estimation error of 2.8% and standard deviation of 2.1% for the test sequences used during the model formulation. Moreover, with reference to Fig. 10 and Fig. 11, the model is able to estimate the optimal depth-texture ratio for the *Book Arrival* and *Breakdancers* test sequence, which were not used during model formulation, with an estimation error of 2.5% and standard devivation of 2.2%.

**Table 4.** Model estimation error

| Test Sequence | % Absolute Mean Estimation error | % Standard Deviation of the Estimation error |
|---|---|---|
| Model Formulation | 2.8 | 2.1 |
| Model Verification | 2.5 | 2.2 |

The effect of the estimation error on the PSNR of the synthesized view was also evaluated for the *Book Arrival* sequence and this was found to be less than 0.1dB. Furthermore, as shown in Fig. 11, the PSNR of the synthesized view using the proposed technique is on average 0.8dB better than that obtained using a fixed depth-texture ratio of 20%, as proposed in [5], for various total bit rate and $QP_{Depth}$. The adaptation of the optimal depth-texture ratio was tested by combining 40 frames from 3 different video sequences and encoding the combined sequence using the modified JMVC framework with the proposed technique. As shown in Fig. 12, the optimal depth-texture ratio was adapted every 4 frames ensuring good tracking performance and correct handling of scene changes.

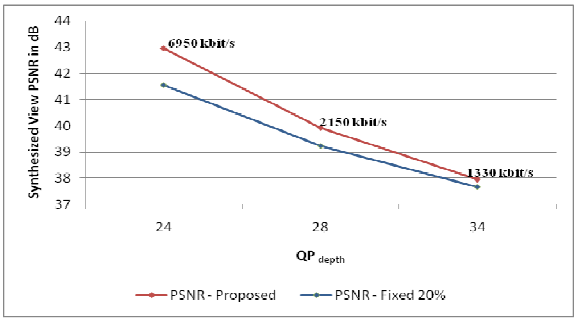



**Fig. 11.** Synthesized view PSNR – proposed vs fixed at 20%

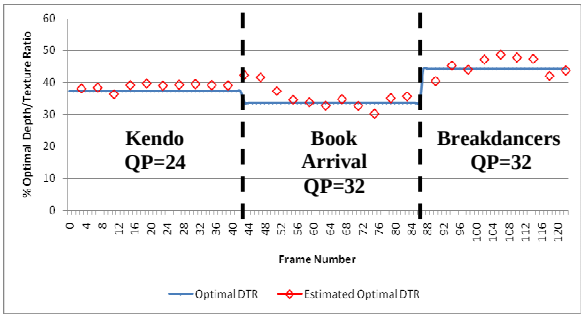



**Fig. 12.** Scene change handling of the proposed technique

## 5. CONCLUSION

In this paper, we have presented a texture-depth rate allocation estimation technique for low-latency MVD systems based on a statistical multi-regression model. Simulation results on test sequences exhibiting a variety of video characteristics have shown that the proposed technique is able to estimate the optimal depth-texture ratio with an absolute mean estimation error of 2.5% for the test sequences not used during model formulation. Moreover, the PSNR of the synthesized view using the proposed technique is on average 0.8dB better when compared to a fixed depth-texture ratio of 20% whilst allowing scene change handling capability.

## 6. ACKNOWLEDGMENT

The authors would like to thank Nagoya University for providing *Balloons*, *Kendo*, *Champagne Tower*, *Pantomine* test sequences, Fraunhofer-HHI for providing *Book Arrival* test sequence, Microsoft Research for providing *Ballet* and *Breakdancers* test sequences.

## 7. REFERENCES

[1] 3D Display Technology and Market Forecast Report. Internet:http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/120926_tv_to_drive_global_demand_for_3d_ready_devices.asp.

[2] Y. S. Ho, and K. J. Oh, "Overview of multiview video coding," in *Proc. of International Workshop on Systems, Signals and Image Processing*, pp. 5-12, Jun. 2007.

[3] A. Smolic, and P. Kauff, "Interactive 3-D Video representation and coding technologies," in *Proc. of the IEEE*, vol. 93, no.1, pp. 98-110, Jan 2005.

[4] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, pp. 201-204.

[5] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," in *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291.

[6] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. Circuits & Syst. for Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[7] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. on Broadcast.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[8] A. Vetro, S. Uea, and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays", in *Proc. of SPIE Conference on Applications of Digital Image Processing XXXI*, vol. 7073, Sep. 2008.

[9] Y. Morvan, D. Farin, and P. De With, "Joint Depth/Texture bit-allocation for multi-view video compression", in *Proc. of Picture coding Symposium*, Nov. 2007.

[10] ITU-R BT.1683, Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference. Geneva, June 2004.

[11] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model," *Signal Process.: Image Commun.*, vol. 24, no. 8, pp. 666-681, Sep. 2009.

[12] G. Cheung, V. Velisavljevic and A. Ortega, "On Dependent Bit Allocation for Multiview Image Coding with Depth-Image-based Rendering," *IEEE Trans. on Image Proc.*, vol.20, no.11, Nov. 2011.

[13] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding", *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 21, no. 4, Apr. 2011.

[14] S. Ma, W. Gao, and Y, Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Cicuits Syst. for Video Technol.*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.

[15] E. Bosc, V. Jantet, M. Pressigout, L. Morin, and C. Guillemot, "Bit-rate allocation for multi-view video plus depth", in *Proc. of 3DTV Conference*, May 2011.

[16] JMVC 8.5, Available: garcon.ient.rwthaachen.de, Sep. 2011.

[17] "View Synthesis Reference Software (VSRS) 3.5," wg11.sc29.org, March 2010.

[18] M. Juneja, and P. Singh Sandhy, "Performance Evaluation of edge detection techniques for Images in Spatial Domain", *Int. Journal of Comp. Theory and Eng.*, vol. 1, no. 5, pp. 614-621, Dec. 2009.

[19] J. F. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679-697, 1986.

[20] K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate Analysis (Probability and Mathematical Statistics), Academic Press, Edition 1, Feb. 1980.

[21] M. Norusis, IBM SPSS Statistics 19 Advanced Statistical Procedures Companion, Pearson, Feb. 2011.

# A NOVEL VIEW-LEVEL TARGET BIT RATE DISTRIBUTION ESTIMATION TECHNIQUE FOR REAL-TIME MULTI-VIEW VIDEO PLUS DEPTH

Mario Cordina
Department of Communication and Computer Engineering
University of Malta
Msida MSD2080, Malta
E-mail: mario.cordina@ieee.org

Carl J. Debono
Department of Communication and Computer Engineering
University of Malta
Msida MSD2080, Malta
E-mail: c.debono@ieee.org

***Abstract*—This paper presents a novel view-level target bit rate distribution estimation technique for real-time Multi-view video plus depth using a statistical model that is based on the prediction mode distribution. Experiments using various standard test sequences show the efficacy of the technique, as the model manages to estimate online the view-level target bit rate distribution with an absolute mean estimation error of 2% and a standard deviation of 0.9%. Moreover, this technique provides adaptation of the view-level bit rate distribution providing scene change handling capability.**

**Keywords- online estimation, multi-view video, view-level rate control, generalised linear models**

## I. INTRODUCTION

The latest advances in high speed networking together with improvements in video capturing devices, have given rise to a plethora of multimedia services. Moreover, customers are increasingly demanding a more realistic, immersive, multimedia experience. Multi-view Video (MVV) is appealing as it provides viewers with a 3D scene representation together with the ability to select and control the viewpoint. Such features allow MVV to be used in various applications such as tele-presence, tele-medicine, Free-View Video (FVV) and 3D TV [1, 2]. With progress in multimedia capture and compression technologies together with the launch of the first commercially available full size auto-stereoscopic 3D display by Toshiba and Phillips [3], MVV technology is gaining ground in the entertainment and consumer arena. In addition, 3D video content is gradually permeating the consumer market through the proliferation of 3D Blu-ray disks, 3D broadcasts and the Internet [4].

There are several 3D video representations in literature, such as model-based representations [5] and point sample-based representations [6]. However, the Multi-view Video plus Depth (MVD) data format is being proposed as the main 3D representation format as it provides good rendering quality and flexible processing capabilities [7, 8]. Such data format fulfils the 3D video system's requirements whilst supporting wide angle 3D displays and auto-stereoscopic displays [4]. Moreover, it allows the rendering of a continuum of output views with high image quality and low complexity [8] through the use of depth image-based rendering (DIBR) techniques. MVD needs the deployment of several video cameras to simultaneously capture a scene from different angles (views) and locations, leading to a significant amount of data. For this reason, MVD is generally jointly compressed using the Multi-view Video Coding (MVC) standard by exploiting the redundancy in space, time and in between views [9, 10]. Additionally, the depth map is often treated as the luminance component of color video, thereby allowing it to be compressed using either H264/AVC or H264/MVC. The resulting MVD stream is then transported over a bandwidth limited network and thus efficient rate control in the encoder is necessary to ensure that the 3D video coding satisfies the channel bandwidth and the decoder buffer constraints.

Rate control in MVC is still in its infancy and no rate control solution has yet been defined in the standard. Moreover, current work focuses on either extending the H264/AVC rate control solutions or the implementation of offline strategies which are not suitable for real-time video. MVD rate control is generally divided into three levels namely view-level, video/depth-level and frame/macro block-level. Correct estimation of the target bit rate at each level is essential for optimum operation. In this paper, we propose a view-level target bit rate distribution estimation technique for real-time MVD based on statistical analysis of the prediction modes used in the different view types.

The rest of the paper is organized as follows: Section 2 gives a summary of related work; Section 3 describes the proposed online view-level target bit rate distribution estimation technique together with an overview on statistical regression models; Section 4 gives a detailed account of the statistical model formulation and its performance while Section 5 provides some comments and conclusions.

## II. RELATED WORK

Rate control in 2D video coding has garnered significant attention in the past decade and a number of algorithms have been proposed, including MPEG-2 TM5 [11], H.263 TMN8 [12] and the H264 rate control algorithm [13,14,15]. However, to the best knowledge of the authors, limited work exists on rate control for MVC and MVD video coding.

A 3D Multi-view rate control algorithm based on the human visual system was first proposed by Lim et al. [16], however this was not based on the new H264/MVC standard. Yan et al. proposed a Multi-view video rate control algorithm for H264/MVC [17] based on [13]. In their work, the authors first propose an improved quadratic rate-quantization (R-Q) model to cater for the additional prediction modes and frames in MVC. Then, using the

978-0-7695-4711-4/12 $26.00 © 2012 IEEE
DOI 10.1109/ICME.2012.5

fluid-flow traffic model, hypothetical reference decoder (HRD) and the improved R-Q model, they estimate the target bit rates for the group of group of pictures (GGOP), group of pictures (GOP) and frame levels.

Work on rate control for video plus depth representation was proposed in [18]. In this paper, the authors develop a rate control technique for 3D video based on Evolution Strategy (ES) for offline H264/AVC 3D video encoding using subjective quality assessment. The main drawback of this algorithm lies in the fact that it is not based on the new H264/MVC standard and is not suitable for real-time applications. The authors of [19] propose a similar approach using Lagrangian optimisation techniques.

Another aspect of rate control is the bit allocation problem. In [20], Morvan et al. studied the joint depth/texture bit allocation problem for multi-view video compression. In order to ensure efficient transmission of 3D video, the compression of both texture and depth images is required. They proposed a joint depth/texture bit allocation algorithm for the compression of MVD by combining the depth and texture rate-distortion (R-D) curves yielding a single R-D surface. A fast hierarchical optimisation algorithm employing an orthogonal search pattern exploits the smooth monotonic properties of the R-D surface, thereby allowing the optimisation of the joint bit-allocation problem in relation to the rendering quality. Although this joint model could be readily integrated into H264/MVC, it requires offline analysis to generate the R-D surface.

Another solution for MVD was developed by Liu et al. [21]. In their study, the authors propose a rate control technique for MVD based 3D video coding by using an image-stitching method to simultaneously encode video and depth. This is followed by a joint 3-level rate control algorithm made up of:

*View level rate allocation:* In MVC, different inter-view predictions can lead to a different R-D performance [21]. For instance, the I-view typically requires a higher bit rate than a P-view or a B-view at the same visual quality. From experiments conducted by the authors, they found that the average rate proportion among the three views (I-view, B-view and P-view) for different quantisation parameter (QP) settings is approximately equal to 6:4:5 for the *Breakdancers* sequence and 4.3:3.1:3.9 for the *Ballet* sequence [21]. The statistical rate allocation for a given sequence is pre-calculated, by offline encoding several frames from each view, and used to assign a different bit rate proportion to the different views depending upon the view type.

*Video and depth allocation:* Since the depth map is strictly used for the rendering of virtual views, the depth sequence can be significantly compressed to attain the required channel bandwidth constraint. In their work, the authors define the bit rate for the depth sequence $R_d$ as a fraction of the bit rate for the video sequence $R_v$. Using a linear R-Q model, the relationship between QP of the depth sequence, $Q_d$, and QP of the video sequence, $Q_v$, is given by:

$$Q_d = Q_v + \left\lceil 6\log_2\left(\frac{K_d}{\alpha K_v + (\alpha C_v - C_d) * 2^{\frac{(Q_v-4)}{6}}}\right)\right\rceil \quad (1)$$

where $K_v$, $K_d$, $C_v$, and $C_d$ are sequence dependent constants which are initialised after pre-encoding several frames from each view during the view-level rate allocation stage.

*Frame-level rate control:* The authors used a hierarchical rate allocation method together with coding complexity and buffer constraints to regulate the target bits for each frame [21]. Using the estimated target bits, the quantisation parameter is computed using the quadratic R-Q model [22]. Furthermore, according to the special characteristics of the Multi-view HRD, the buffer-related rate control is also considered to prevent the decoder buffer from overflow or underflow even when outputting multiple views.

Similar to other works, the main drawback of this technique is its dependency on offline processing. Moreover, the use of hierarchical B coding order is inherently restricting its use to non-real-time video sequences, as future frames are not known. Furthermore, it also precludes the handling of scene changes unless the complete video sequence is analysed *apriori* leading to a two-pass rate control technique.

## III. Online View-Level Target Bit Rate Distribution Estimation

Consider an advanced MVD-based 3D video system as proposed in [4]. Further assume that the MVC encoder is able to simultaneously encode multiple views. Through the use of DIBR techniques, views can be synthesised at the receiver side, thereby allowing us to encode and transmit a subset of the views. In this work, we are considering three views, namely the view 0, view 1 and view 2.

In general, video sequences are sequentially encoded using the H264/MVC hierarchical B coding structure. Although this yields a better compression ratio, it comes at the expense of a higher coding delay as by the time the reference frames are available, the current frame being decoded should have been already displayed. Real-time applications such as video conferencing are very sensitive to delay and industry standards limit the end-to- end delay to a maximum of 300ms making such prediction structure unsuitable to real-time communications. By default, the current Joint Multi-view Video Coding (JMVC) software model uses the hierarchical B coding structure however, as from JMVC ver. 2.3, the model supports a GOP size of 1 which allows the use of an I-P-P-P structure in the time domain. When using this GOP setting, the coding structure needs to be defined in the configuration file, whereby for each view, the reference frames that may be used must be

specified. By setting the appropriate references, a real-time prediction structure, utilising inter-view B frames as shown in Fig. 1, can be implemented without additional time delay as these can be catered for by parallel hardware.

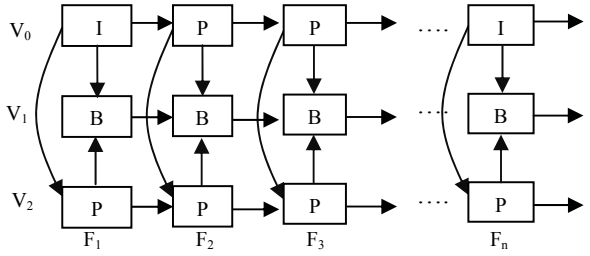



Figure 1. Real-time prediction structure.

A key aspect in such a rate control framework is the estimation of the allocated bit rate in proportion to the different views, depending on the view type. This view-level target bit rate allocation defines the target bit rate bounds of the video/depth and the frame layer rate control algorithm. Ideally, this estimation is carried out periodically and in real-time to adapt to the changing video sequence. Moreover, it cannot make use of the controlled bit rate values at the output of the rate controller as this directly depends on the correct estimate at the input. However, as we have seen in section II, the view-layer target bit rate distribution is usually estimated using offline processing techniques with no adaptation. By exploiting the distribution of the prediction modes in the different views, which apart from other factors depends on the video sequence characteristics and prediction structure, we can infer the bit rate allocation for the different views.

To apply adaptation, we propose the use of a statistical regression model, which has been trained offline, to estimate online the view-level target bit rate distribution, as shown in Figure 2. This estimation is carried out periodically (say at every anchor period) in order to provide scene change handling. The distribution of the prediction modes has a direct impact on the bit rate. Therefore, by observing the statistical distribution of the Macroblock (MB) prediction modes, view type (I, P or B) and the view-level bit rate for various video sequences and QPs, a statistical regression model can be formulated.

In JMVM, the supported MB prediction modes include SKIP, DIRECT, Inter16x16, Inter16x8, Inter8X16, Inter8x8, Inter8x8Frext, Intra16x16, Intra8x8 and Intra4x4. The prediction mode with the lowest R-D cost [3], given by (2), is chosen to encode the MB.

$$J(S_k, I_k \mid \lambda) = SSD(S_k, I_k) + \lambda R(S_k, I_k) \tag{2}$$

where $S_k$ and $I_k$ denote the $k^{th}$ MB and the corresponding MB mode respectively, $\lambda$ is the Lagrange multiplier for mode decision, $SSD(S_k,I_k)$ is the sum of squared difference between the reconstructed MB and the original MB, and $R(S_k,I_k)$ is the rate after entropy coding.

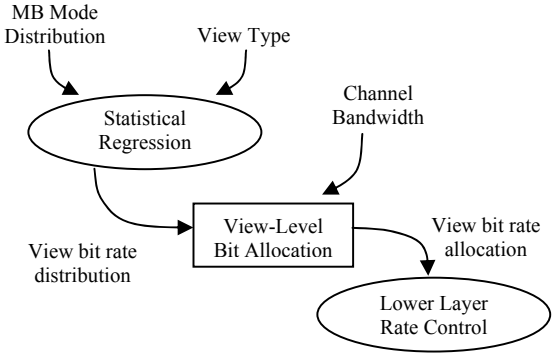



Figure 2. Overview of bit rate distribution estimation.

There are several regression models in literature, however linear regression models are the most commonly used due to their well-known statistical properties and are typically composed of a single dependent variable (view-level bit rate) and several predictors (QP, view type and prediction modes). A well known general linear model that accommodates predictors which are a combination of factors and covariates, as in our case, is the Analysis of Covariance (ANCOVA). For $n$ observed cases, the ANCOVA model that expresses the dependent variable $y_n$ as a function of $p$ predictors $\beta_p$ is given by [23]:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{3}$$

$$\mathbf{y} = \mathbf{x}\beta + \varepsilon \tag{4}$$

where $\beta$ is a $p$-vector of unknown regression parameters estimated using least squares estimation, $\varepsilon$ is an $n$-vector of unknown random error terms, $\mathbf{y}$ is an $n$-vector of responses which apart from having a normal distribution, it is also linearly related and highly correlated with each quantitative predictor, and $\mathbf{x}$ is an $n$ x $p$ matrix whose elements are a mixture of real and dummy (0 or 1 indicators) values and are uncorrelated with each other.

Given that the errors have a normal distribution, the least square estimate, which is equivalent to the maximum likelihood estimator $b$, is given by:

$$b = \left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y} \tag{5}$$

Although regression models, such as the ANCOVA model, have been used extensively by researchers for prediction, inference and modelling, these models rely heavily on the underlying assumptions. In particular, regression models assume that the dependent variable has a normal distribution and that the predictors are independent. However, these assumptions are not always valid and other modelling techniques have to be explored. One of the most significant contributions in statistical modelling is the

concept of generalised linear models (GLM) [24]. These models relate the dependent variable to the linear predictors through any invertible link function and accommodate any error distribution present in the exponential family. Although GLM can be viewed as a generalisation of the general linear model, they still rely on the assumption that the responses are independent. GLM are characterised by three components:

*1)* For $n$ observed cases, the dependent variable $y_n$ is assumed to be independent and follows a distribution that is a member of the exponential family.

*2)* The predictors influence the distribution of the dependent variable $y_n$ through a single linear function known as the linear predictor and given by:

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + .... + \beta_p x_{pi} = \mathbf{x}\beta \quad (6)$$

*3)* The mean $\mu = E(\mathbf{y})$ is related to the linear predictor through an invertible link function $g(\mu)$.

$$g(\mu_i) = \eta_i = \mathbf{x}\beta \quad (7)$$

Thus, the GLM is given by $E(\mathbf{y}) = \mu = g^{-1}(\mathbf{x}\beta)$ where $\mathbf{y}$ is an $n$-vector of responses, $\beta$ is a $p$-vector of unknown regression parameters estimated using maximum likelihood techniques, and $\mathbf{x}$ is an $n$ x $p$ matrix whose elements are a mixture of real and dummy (0 or 1 indicators) values.

## IV. Statistical Model Formulation

The proposed online view-level target bit rate distribution estimation technique was implemented in two parts: First the statistical model was formulated, and then the statistical model integrated in JMVC ver. 8.3.1. The latter also included the verification of its effectiveness in estimating the view-level target bit rate distribution. Prior to the statistical model formulation, the JMVC model was modified to extract information such as the prediction mode distribution, view type, and average bit rate for various QP settings, parse it and import it into IBM SPSS© statistical analysis package for analysis. The encoding parameters together with the test sequences used and their respective characteristics are shown in Table I and II respectively.

TABLE I. Encoding Parameters

| | |
|---|---|
| GOP Size | 1 |
| Intra Period | 12 |
| QP | 26 to 50 |
| Prediction Structure | Real-time |
| Encoded Frames | 60 |

Before selecting a statistical regression model, the dependent variable was checked for normality using the One-Sample Kolmogorov-Smirnov test at the 0.05 level of significance. The result of this test, indicates that the data is not normally distributed, as the $p$-value is lower than 0.05, thereby eliminating the use of a general linear model, such as the ANCOVA model. This was further confirmed after plotting the histogram of the dependent variable, as shown in Fig. 3.

TABLE II. Test Sequence Parameters and Characteristics ([V] indicates the sequence is used for verification only)

| Sequence | Image Size | Frame Rate | Features |
|---|---|---|---|
| Kendo[V] | 1024 x 768 | 30 | Complex object motion; Moving camera |
| Break dancers | 1024 x 768 | 15 | Slow and very fast motion; No camera motion |
| Balloons | 1024 x 768 | 30 | Complex object motion; Moving camera |
| Champ. Tower | 1280 x 960 | 30 | Slow but complex object motion; No camera motion |
| Book Arrival | 1024 x 768 | 15 | Moderate object motion; No camera motion |

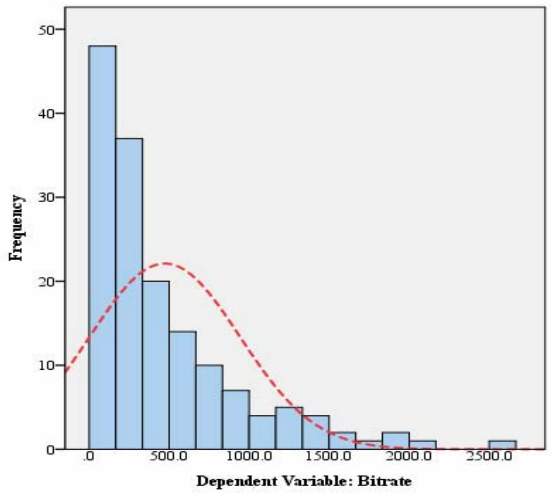


Figure 3. Dependent Variable Histogram.

This shows that the dependent variable follows a distribution which is a member of the exponential family. Therefore, a generalised linear model using the identity link function was chosen to fit the data. The initial GLM parameter estimates together with the Wald-Chi square and confidence levels are shown in Table III. As shown in this table, the Mode 16x16 and Mode16x8 have a $p$-value which is higher than the 0.05 level of significance. This implies that their contribution in explaining the variation in the bit rate was not found to be significant and thus these two predictors need to be removed from the model fitting system. Excluding these two predictors and re-calculating the parameter estimates we obtain the parsimonious GLM model shown in Table IV.

Thus, the parsimonious GLM model is composed of 10 significant predictors and is given by (8).

$$B = 1945.5 - 179.2V_1 + 161.9V_2 + 1903.7DirectB - 13472.9Mode8x16 + 2021.7Intra4x4 + 61861.8Mode8x8 - 3073.5Intra16x16 - 1349.4Skip - 9.9QP \quad (8)$$

where $B$ is the estimated view-level bit rate; DirectB, Mode8x16, Intra4x4, Mode8x8, Intra16x16 and Skip are ratios based on the MB prediction mode distribution over a number of frames or anchor period, QP is the quantisation parameter whereas $V_1$ and $V_2$ are flags used to indicate view type. $V_1$ and $V_2$ are set to 0 and 1 for an I-view, 1 and 0 for a B-view, 1 and 1 for a P-View respectively.

TABLE III. INITIAL GLM MODEL

| Parameter | Coefficient | Wald Chi-Square value | *p*-value |
|---|---|---|---|
| Constant Term | -700.3 | 16.137 | 0.000 |
| V1 | -179.71 | 12.975 | 0.000 |
| V2 | 150.31 | 19.163 | 0.000 |
| DirectB | 3192 | 35.903 | 0.000 |
| Mode16x16 | 851.72 | 1.888 | 0.169 |
| Mode16x8 | 14195.62 | 3.304 | 0.069 |
| Mode8x16 | -21459.65 | 9.288 | 0.002 |
| Intra4x4 | 2878.49 | 7.729 | 0.005 |
| Mode8x8 | 56273.25 | 33.154 | 0.000 |
| Intra16x16 | -1725.3 | 17.253 | 0.000 |
| QP | -11.38 | 14.48 | 0.000 |

TABLE IV. PARSIMONOUS GLM MODEL

| Parameter | Coefficient | Wald Chi-Square value | *p*-value |
|---|---|---|---|
| Constant Term | 1945.51 | 16.06 | 0.000 |
| V1 | -179.17 | 12.63 | 0.000 |
| V2 | 161.99 | 23.22 | 0.000 |
| Direct B | 1903.65 | 4.93 | 0.026 |
| Mode 8x16 | -13472.9 | 7.43 | 0.006 |
| Intra 4x4 | 2021.7 | 4.13 | 0.042 |
| Mode8x8 | 61861.78 | 48.24 | 0.000 |
| Intra 16x16 | -3073.47 | 22.49 | 0.000 |
| Skip | -1349.43 | 5.85 | 0.016 |
| QP | -9.927 | 11.95 | 0.001 |

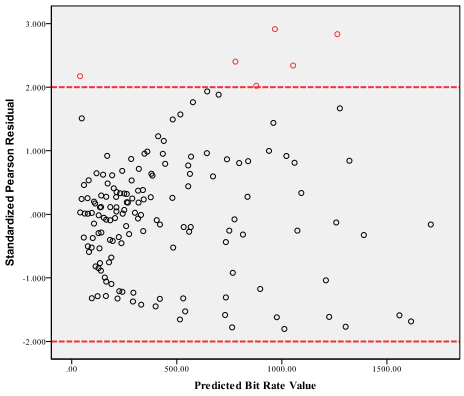



Figure 4. Standardised Pearson Residual.

The Standardised Pearson Residual [25], shown in Fig. 4, is plotted against the predicted bit rate value to check for the efficacy of the fitted GLM model. The points in the Standardised Pearson Residual plot are largely spread across a horizontal band indicating that the residuals have a constant variance (homoscedasticity). An element of heteroscedasticity in the model is present as the variance increases slightly with the predicted value. However, since the majority of the points lie between the standard $\pm 2$ limit, the model is adequate for such data fit with a very small number of outliers (shown in red).

The GLM model was integrated within the JMVC framework and the estimated view-level bit rate distribution was calculated online from the estimation of the view-level bit rate of each view. This was done for all the test sequences and the plots of the estimated and actual view-level bit rate distribution during Model Formulation and Verification are shown in Fig. 5 and 6 respectively.

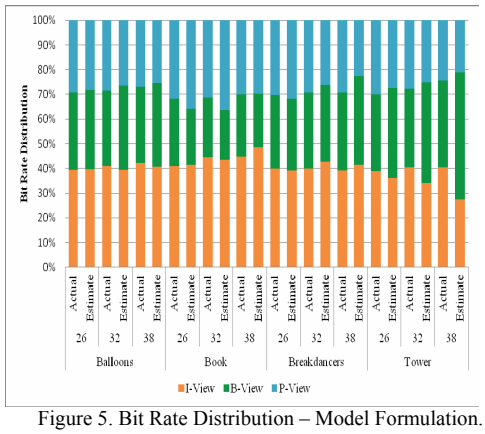



Figure 5. Bit Rate Distribution – Model Formulation.

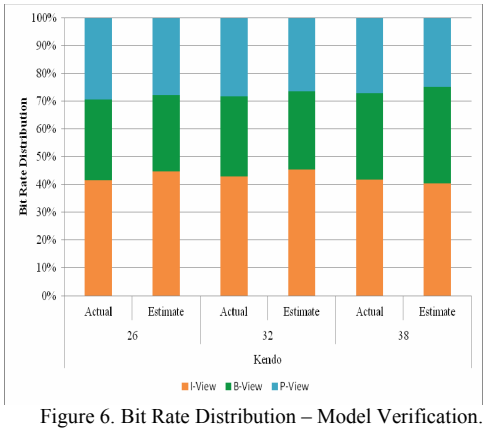



Figure 6. Bit Rate Distribution – Model Verification.

The mean and standard deviation of the estimation error for the test sequences used in the model formulation and verification are shown in Table V. The formulated statistical regression model was able to estimate the view-level bit rate distribution with an overall estimation error of 3.48% and a standard deviation of 3.42% for the test sequences used during the model formulation. With reference to Fig. 6, it is noted that the model manages to accurately estimate the bit rate for sequences which were not used in model

formulation with an estimation error of around 2% and standard deviation of 0.9%. Although this is an acceptable error for target view-level bit rate setting, this performance can be further improved by including other test sequences during model formulation.

TABLE V. MODEL ESTIMATION ERROR

| Test Sequence | % Absolute Mean Estimation error | | | | % Standard Deviation of the Estimation error | | | |
|---|---|---|---|---|---|---|---|---|
| | *I-View* | *B-View* | *P-View* | *Over all* | *I-View* | *B-View* | *P-View* | *Over all* |
| Model Formulation | 3.06 | 4.64 | 2.72 | 3.48 | 3.56 | 4.34 | 1.88 | 3.42 |
| Model Verification | 2.31 | 1.92 | 1.96 | 2.02 | 1.0 | 1.4 | 0.35 | 0.9 |

## V. CONCLUSION

In this paper, we have presented a view-level target bit rate distribution estimation technique that can be applied to real time MVD systems based on a statistical regression model. Simulation results on sequences exhibiting a variety of video characteristics have shown that the technique is able to estimate the view-level target bit rate distribution in real-time with an absolute mean estimation error of 2% for the test sequences not used during model formulation. The model integrated in the JMVC framework allows view-level target bit rate setting for the underlying rate control algorithm whilst allowing scene change handling capability.

## ACKNOWLEDGMENT

The authors would like to thank the Interactive Media Group of Microsoft Research for providing the *Breakdancers* test sequences, Nagoya University for providing *Balloons*, *Kendo* & *Champagne Tower* test sequences and Fraunhofer-HHI for providing *Book Arrival* test sequence.

## REFERENCES

[1] A. Smolic, P. Kauff, "Interactive 3-D Video representation and coding technologies," in Proc. of the IEEE, vol.93, no.1, pp. 98-110, Jan 2005.

[2] Y.S. Ho, K.J. Oh, "Overview of multiview video coding," in Proc. of International Workshop on Systems, Signals and Image Processing, pp. 5-12, Jun 2007.

[3] International Consumer Electronics Show 2011. Internet: http://www.cesweb.org

[4] A. Vetro, S. Yea and A. Smolic, "Towards a 3D video format for auto-stereoscopic displays," in Proc. Of SPIE: Appl. Digital Image Process. XXXI, Vol. 7073, Sep. 2008.

[5] C. Theobalt, G. Ziegler, M. Magnor, and H.P. Seidel, "Model-based free-viewpoint video acquisition, rendering and encoding," in Proc. Picture Coding Symp., San Francisco, pp. 1-6, Dec. 2004.

[6] S. Wurmlin, E. Lamboray, and M. Gross, "3D video fragments: Dynamic point samples for real-time free-viewpoint video," Computers & Graphics, vol. 28, no. 1, pp. 3-14, Feb. 2004.

[7] Y.K. Park, K. Jung, Y. Oh, S. Lee, J.K. Kim, H. Lee, K. Yun, N. Hur and J. Kim, "Depth-image-based rendering for 3DTV service over T-DMB," Signal Process.: Image Commun., vol. 24, no. 1-2, pp. 122-136, Jan. 2009.

[8] P. Kauff, N. Atzpadin, C. Fehn, K. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for Advanced 3DTV services providing interoperability and scalability," Signal Process.: Image Commun., vol. 22, no. 2, pp. 217–234, Feb. 2007.

[9] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," IEEE Trans. Circuits & Syst. For Video Technol., vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[10] L.Q. Shen, Z. Liu, S.X. Liu, Z.Y. Zhang, and P. An, "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," IEEE Trans. on Broadcast., vol. 55, no. 4, pp. 761–766, Dec. 2009.

[11] "MPEG 2 Test Model 5", Rev. 2 ISO/IEC JTC1/SC29WG11, Apr. 1993.

[12] J. Ribas-Corbera and S. Lei, "Rate control for low-delay video communications", ITU Study Group 16, VCEG, Portland, Doc. Q15-A-20, 1997.

[13] Z.G. Li, F. Pan, K.P. Li, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT", Pataya II, Thailand, Doc. JVT-G012e1, 2003.

[14] S. Ma, F. Wu, and Z. Li, "Proposed draft of adaptive rate control", Geneva, Switzerland, Doc. JVT-H017r3, 2003.

[15] Z. Min, T. Ikenaga, and S. Goto, "A Novel Rate Control Algorithm for H.264/AVC", 23rd International Technical Conference on Circuits/Systems, Computers and Communications, 2008.

[16] J.E. Li, J. Ki, K.N. Ngan, and K. Sohn, "Advanced rate control technologies for 3D-HDTV", IEEE Trans. Consumer Electronics, vol. 49, no.4, pp. 1498-1507, Nov. 2003.

[17] T. Yan, P. An, L. Shen, Z. Li, H. Wang, Z. Zhang, "Frame-layer rate control algorithm for multi-view video coding", GEC Summit 2009, pp. 1025-1028, 2009.

[18] S.L.P. Yasakethu, W.A.C. Fernando, and A. Kondoz, "Rate controlling in offline 3D video coding using evolution strategy", IEEE Trans. Consumer Electronics, vol. 55, no. 1, pp. 150-157, Feb. 2009.

[19] B. Kamolrat, W.A.C. Fernando, and M. Mrak, "Rate controlling for colour and depth 3D video coding", in Proc. SPIE: Appl. Digital Image Process. XXXI, San Diego, CA, Aug. 2008.

[20] Y. Morvan, D. Farin and, P. De With, "Joint Depth/Texture bit-allocation for multi-view video compression", Picture coding Symposium, Nov. 2007.

[21] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, S. Ci, and H. Tang, "A Novel Rate Control Technique for Multiview Video Plus Depth Based 3D Video Coding", IEEE Transactions on Broadcasting, pp. 562-571, Jun. 2011.

[22] H.J. Lee, T. Chiang, and Y.Q. Zhang, "Scalable rate control for MPEG-4 video", IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 6, pp. 878-894, Sep. 2000.

[23] K.V. Mardia, J.T. Kent and J.M. Bibby, Multivariate Analysis (Probability and Mathematical Statistics), Academic Press, Edition 1, Feb. 1980.

[24] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman and Hall, 2nd Edition, Aug. 1989.

[25] M. Norusis, IBM SPSS Statistics 19 Advanced Statistical Procedures Companion, Pearson, Feb. 2011.