University of Cyprus

**DEPARTMENT OF COMPUTER SCIENCE**

# REAL-TIME ADAPTATION TO TIME-VARYING CONSTRAINTS FOR RELIABLE MHEALTH VIDEO COMMUNICATIONS

**DOCTOR OF PHILOSOPHY DISSERTATION**

**ZINONAS C. ANTONIOU**

**2017**

University of Cyprus

**DEPARTMENT OF COMPUTER SCIENCE**

# REAL-TIME ADAPTATION TO TIME-VARYING CONSTRAINTS FOR RELIABLE MHEALTH VIDEO COMMUNICATIONS

ZINONAS C. ANTONIOU

A DissertationSubmitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

December, 2017

2017

# APPROVAL PAGE

Doctor of Philosophy Dissertation

**REAL-TIME ADAPTATION TO TIME-VARYING CONSTRAINTS FOR RELIABLE MHEALTH VIDEO COMMUNICATIONS**

Presented by

Zinonas C. Antoniou,

Research Supervisor: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Constantinos S. Pattichis

Research Co-supervisor: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Andreas S. Panayides

Committee Member: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Andreas Pitsillides

Committee Member: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Vasos Vassiliou

Committee Member: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Anthony G. Constantinides

Committee Member: ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ
Nicolas Tsapatsoulis

University of Cyprus

December, 2017

# DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Zinonas C. Antoniou

# ΠΕΡΙΛΗΨΗ

Ο σημαντικός αριθμός των ιατρικών συστημάτων και υπηρεσιών επικοινωνίας βίντεο σε κινητές συσκευές που αναπτύχθηκαν κατά την τελευταία δεκαετία, αντικατοπτρίζει το ευρύ πεδίο εφαρμογής αυτών των συστημάτων και υπηρεσιών στην καθιερωμένη κλινική πρακτική. Όσον αφορά τις υπηρεσίες επικοινωνίας βίντεο, τα σενάρια εφαρμογών κυμαίνονται, αλλά δεν περιορίζονται, από την απομακρυσμένη διάγνωση και φροντίδα, την αντιμετώπιση έκτακτης ανάγκης, την παροχή δεύτερης γνώμης και την παρακολούθηση στο σπίτι. Η ενσωμάτωση των συστημάτων επικοινωνίας ιατρικών βίντεο στην παροχή ιατροφαρμακευτικής περίθαλψης, μπορεί να βελτιώσει σημαντικά την ποιότητα της περίθαλψης, μειώνοντας ταυτόχρονα τους χρόνους νοσηλείας και τα συναφή έξοδα ιατροφαρμακευτικής περίθαλψης.

Η πρόκληση έγκειται στην παροχή επαρκώς υψηλών ψηφιακών αναλύσεων βίντεο και ρυθμών καρέ (frame rate) απαιτώντας χαμηλή καθυστέρηση και χαμηλά ποσοστά απώλειας πακέτων, παρέχοντας ένα ισοδύναμο επίπεδο κλινικής εμπειρίας με εκείνο των ενδό-νοσοκομειακών εξετάσεων. Το τελευταίο απαιτεί έλεγχο της διαδικασίας ροής βίντεο σε πραγματικό χρόνο, ώστε να διευκολυνθούν τα επαρκή επίπεδα κλινικής ποιότητας βίντεο που απαιτούνται για την υποστήριξη της αξιόπιστης διάγνωσης. Σκοπός της παρούσας διδακτορικής διατριβής είναι η προσαρμογή της επικοινωνίας του ιατρικού βίντεο στη χρόνο-μεταβλητότητα των αλληλοεπηρεαζόμενων παραγόντων που αναφέρονται στην κλινική του ποιότητα, στον χρόνο κωδικοποίησής του και στην αποστολή του μέσω του διαθέσιμου ασύρματου δικτύου. Μία από τις σημαντικότερες συνεισφορές της μελέτης αυτής είναι ότι αυτή η προσαρμογή επιτυγχάνεται σε πραγματικό χρόνο.

Πιο συγκεκριμένα, η προσέγγιση βασίζεται στην βελτιστοποίηση πολλαπλών στόχων, η οποία μεγιστοποιεί ταυτόχρονα την ποιότητα και την ταχύτητα κωδικοποίησης του κωδικοποιημένου βίντεο, ενώ ελαχιστοποιεί τις απαιτήσεις του ρυθμού μετάδοσης bit (bitrate) μέσα από το διαθέσιμο ασύρματο δίκτυο. Για το σκοπό αυτό, κατασκευάζεται ένας πυκνός χώρος κωδικοποίησης και χρησιμοποιείται η λογαριθμική γραμμική παλινδρόμηση για την εκτίμηση μοντέλων πρόβλεψης για την ποιότητα, τον ρυθμό μετάδοσης bit και την υπολογιστική πολυπλοκότητα. Τα μοντέλα πρόβλεψης χρησιμοποιούνται στη συνέχεια από το προτεινόμενο προσαρμοστικό σύστημα ελέγχου, το οποίο μπορεί να τελειοποιήσει την κωδικοποίηση βίντεο με βάση τους περιορισμούς σε πραγματικό χρόνο. Το προτεινόμενο σύστημα επικυρώνεται χρησιμοποιώντας έναν αλγόριθμο όπου χρησιμοποιούμε εκπαίδευσής και αξιολόγησης, που εφαρμόζεται σε δέκα βίντεο υπερήχων της κοινής καρωτιδικής αρτηρίας. Δηλαδή, κάθε φορά χρησιμοποιούμε τα 9 βίντεο για την εκτίμηση μοντέλων πρόβλεψης και το 1 βίντεο για την αξιολόγησή του. Αυτή η διαδικασία ακολουθείται 10 φορές. Τα μοντέλα πρόβλεψης μπορούν να υπολογίσουν την ποιότητα που μετριέται με τον δείκτη Structural SIMilarity (SSIM) με ένα μέσο σφάλμα ακρίβειας μικρότερο από 1%, απαιτήσεις ρυθμού μετάδοσης bit με σφάλμα απόκλισης 10% ή λιγότερο και κωδικοποίηση ρυθμού καρέ εντός περιθωρίου 6%. Η προσαρμογή σε πραγματικό χρόνο σε επίπεδο ομάδας φωτογραφιών (Group of Pictures - GOP) καταδεικνύεται χρησιμοποιώντας το πρότυπο κωδικοποίησης βίντεο υψηλής απόδοσης (High Efficiency Video Coding - HEVC) . Η αποτελεσματικότητα του προτεινόμενου συστήματος σε σύγκριση με τις στατικές μη προσαρμοστικές προσεγγίσεις, αποδεικνύεται για διαφορετικούς τρόπους λειτουργίας, επιτυγχάνοντας σημαντικά κέρδη ποιότητας, μειώσεις της ζήτησης ρυθμού μετάδοσης bit και βελτιώσεις απόδοσης, σε σενάρια πραγματικής ζωής που υπόκεινται σε χρόνο-μεταβλητούς περιορισμούς.

# ABSTRACT

The significant number of mobile health (mHealth) systems developed over the past decade reflect the broad applicability spectrum of such systems and services in standard clinical practice. In terms of medical video communications, application scenarios range, but are not limited, from remote diagnosis and care, to emergency response, second opinion provision, and home monitoring. Integration of medical video communication systems in the healthcare provision pathway can significantly enhance the quality of care, while reducing hospitalization times and associated healthcare costs.

The challenge lies in delivering sufficiently high video resolutions and frame rates with the low-delay and low packet loss rates requirements that will accommodate an equivalent level of clinical experience to that of in-hospital examinations. The latter requires real-time control of the video streaming process so as to facilitate the adequate levels of clinical video quality required to support reliable diagnosis. The objective of this PhD thesis is to provide a scalable, video modality, encoder, and wireless network agnostic framework, that will support real-time adaptation to time-varying wireless networks' state while guaranteeing diagnostically lossless video communications and conforming to end-user device constraints for real-time performance.

More specifically, the approach is based on multi-objective optimization, that jointly maximizes the encoded video's quality and encoding rate, while minimizing bitrate demands. For this purpose, a dense encoding space is constructed, and logarithmic linear regression is used to estimate forward prediction models for quality, bitrate, and computational complexity. The prediction models are then used by the proposed adaptive control framework that can fine-tune video encoding based on real-time constraints. The proposed framework is validated using a leave-one-out

algorithm applied to ten ultrasound videos of the common carotid artery. The prediction models can estimate Structural SIMilarity (SSIM) quality with a median accuracy error of less than 1%, bitrate demands with deviation error of 10% or less, and encoding frame rate within a 6% margin. Real-time adaptation at a Group of Pictures (GOP) level is demonstrated using the High Efficiency Video Coding (HEVC) standard. The effectiveness of the proposed framework compared to static, non-adaptive approaches is demonstrated for different modes of operation, achieving significant quality gains, bitrate demands reductions, and performance improvements, in real-life scenarios imposing time-varying constraints.

Zinonas C. Antoniou - University of Cyprus, 2017

# ACKNOWLEDGEMENTS

Reaching the completion of my Ph.D. dissertation, at the Department of Computer Science of the University of Cyprus, I would like to express my sincere gratitude to all those who guided and supported me to achieve my goals during this long, difficult, but fruitful journey.

Foremost, I would like to deeply thank my research supervisor, Professor Constantinos Pattichis. I'm thankful to him for his sincerity, encouragement, patience and support in different levels. This thesis would not have been possible without his unconditional support. I owe him lots of gratitude for having shown me how to be an ideal scientist, a great supervisor and a really kind person. Also, I would like to thank Dr. Andreas Panayides, who undertook the co-supervision of my Ph.D. Heartfelt thanks go to him who given me so much energy for my research study. His expertise in the field of video compression and mHealth video frameworks was quite valuable in developing the proposed approaches. It gives me great pleasure to acknowledge his guidance, his valuable suggestions, his constructive criticism, his incredible patience and his selfless attitude. The intellectual and ethical abilities of both Professor Constantinos Pattichis and Dr. Andreas Panayides have contributed substantially throughout this Ph.D. study.

Besides my advisers, I would like to take the opportunity to thank the rest of my examining committee: Professor Andreas Pitsillides, Assistant Professor and Vasos Vassiliou, both members of the faculty of the Computer Science Department of the University of Cyprus, for serving as readers of my thesis. Also, I would like to thank Professor Anthony G. Constantinides, faculty of Engineering of Department of Electrical and Electronic Engineering at the Imperial College

London and Associate Professor Nicolas Tsapatsoulis, faculty of the Department of Communication and Internet Studies of the Cyprus University of Technology. I really thank them all for their feedback, direction and assistance along with their significant and valuable comments that gave me the opportunity to wider my research horizons.

Gratitude is extended to Professor Marios Pattichis, faculty of the Department of Electrical and Computer Engineering of the University of New Mexico, for his persistent support and motivation. With his inspiration, his enthusiasm, and his great efforts to explain things clearly and simply, he helped me to clarify and highlight different aspects of the adaptive mHealth framework.

Moreover, I would like to express my sincere appreciation to the academic and technical support of the Department of Computer Science which provided my the support and equipment I needed to complete my thesis. Yet, I would like to thank Dr. Yiannos Mylonas and Dr. George Matheou for their significant help whenever I asked. Additionally, special thanks go to the members of the eHealth Laboratory, for providing an enjoyable working environment as well as for stimulating casual discussions on research topics and helping me develop my ideas.

Special recognition goes out to parents, my father Charalambos and my mother Maria, for proving a loving and caring environment for me from the very beginning of this study until the end. Also, I express my sincere thanks to my siblings, my brother Pavlos and my sister-in-law Christina, my sister Anna and my brother-in-law Spyros, for their support, encouragement and patience while pursuing of my thesis. I could never forget to thank my parents-in-law, Petros and Eleftheria for their unconditional support and love.

Last, but never least, I would like to thank my lovely wife Christina, who provided persistent patience and encouragement during the entire process. Her generous support and tireless efforts behind the scene everyday, made me overcome many difficulties and made everything possible. Also, special recognition goes to my two little kids, Charalambos and Marina, who missed out

on a lot of daddy time, especially the last couple of months. I thank all three of them for their patience and love them more than they will ever know.

I could not forget that I owe a heartfelt thanks to my spiritual father, father Vasileios, who supported me unselfishly during this journey of knowledge, and he continue to support every single step of my life. Finally, I thank our God and Even-Virgin Mary from the bottom of my heart for answering my prayers and for all Their blessings in every aspect of my life.

"Εάν είσαι αγωνιστής, μπορείς και κάθε ώρα να βάζεις αρχή."

Αββάς Σιλουανός

***Dedicated to my beloved parents Charalambos and Maria,***
***my lovely wife Christina,***
***and my children Charalambos and Marina.***

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ACRONYMS

$1G$ First generation mobile telecommunications

$2G$ Second generation mobile telecommunications

$3G$ Third generation mobile telecommunications

$3GGP$ Third Generation Partnership Project

$4G$ Fourth generation mobile telecommunications

$5G$ Fifth generation mobile telecommunications

$AAP$ Alternative Approval Process

$ACT$ Adaptive Colour Transform

$AES$ Advanced Encryption System

$ALF$ Adaptive Loop Filtering

$AMC$ Adaptive Modulation and Coding

$AMVP$ Advanced Motion Vector Prediction

$AQMS$ Adaptive Quantization Matrix Selection

$ARQ$ Automatic Repeat Request

$ASO$ Arbitrary Slice Ordering

$AVC$ Advanced Video Coding

$B2$ 2 consecutive B frames (IPBBP)

$B4$ 4 consecutive B frames (IPBBBBP)

$BA$ Block-based Adaptation

$BALF$ Block-based Adaptive Loop Filter

$BE$ Best-Effort service

$BL$ Base Layer

$BLIINDS-II$ Blind Image Integrity Notator DCT Statistics

$BO$ Band Offset

$BS$ Base Station

$CABAC$ Context-based Adaptive Binary Arithmetic Coding

$CAVLC$ Context-Adaptive Variable-Length Coding

$CBP$ Intra-coded Block Pattern

$CBR$ Constant-Bit-Rate

$CCP$ Cross-Component Prediction

$CDMA$ Code Division Multiple Access

$CID$ Connection Identifier

$CME$ Continued Medical Education

$CoMP$ Coordinated Multi-point Transmission and Reception

$CPB$ Coded Picture Buffer

$CSF$ Contrast Sensitivity Function

$CTU$ Coding Tree Unit

$CU$ Coding unit

$DBF$ Deblocking Filter

$DFT$ Discrete Fourier transform

$DIF$ Directional Interpolation Filtering

$DL$ Downlink

$DLM$ Detail Loss Metric

$DL-MAP$ Downlink Map Messages

$DP$ Data Partition

$DPB$ Decoded Picture Buffer

$DST$ Discrete Sine Transform

$EAP$ Extensible Authentication Protocol

$eHealth$ Electronic Health

$EL$ Enhancement Layer

$EO$ Edge Offset

$ETSI$ European Telecommunications Standards Institute

$FDD$ Frequency-Division Duplex

$FDIS$ Final Draft International Standard

$FEC$ Forward Error Correction

$FMO$ Flexible Macroblock Ordering

$GPSS$ Grant Per Subscriber Station

$GSM$ Global System for Mobile communications

$HARQ$ Hybrid Automatic Repeat Request

$HD$ High-Refinition

$HDTV$ High-Definition Television

$HEB$ High-Efficiency Binarization

$HEVC$ High Efficiency Video Coding

$HP$ High Profile

$HSPA$ High-Speed Packet Access

$HTB$ High-Throughput Binarization

$HVS$ Human Visual System

$IBC$ Intra Block Copy

$IDR$ Instantaneous Decoder Refresh

$IEEE$ Institute of Electrical and Electronics Engineers

$IF$ Interpolation Filtering

$IFC$ Information Fidelity Criterion

$IMT$ International Mobile Telephony

$IMT - Advanced$ International Mobile Telecommunications-Advanced

$ITU$ International Telecommunication Union

$ITU - R$ ITU Radiocommunication Sector

$ITU - T$ ITU Telecommunication Standardization Sector

$JVT$ Joint Video Team

$LCU$ Largest Coding Unit

$LOS$ Line Of Sight

$LTE$ Long-Term Evolution

$MAC$ Medium Access Control layer

$MB$ Macroblock

$MBAmap$ Macroblock Allocation map

$MBMS$ Multimedia Broadcast Multicast Service

$MDIS$ Mode-Dependent Intra Smoothing

$mHealth$ Mobile Health

$MIMO$ Multiple Input Multiple Output

$MOMS$ Maximal-Order-Minimum-Support

$MOS$ Mean Opinion Score

$MOVIE$ Motion-based Video Integrity Evaluation

$MPEG$ Motion Pictures Experts Group

$MSE$ Mean Square Error

$MV$ Motion Vector

$MV-HEVC$ [Multi-View High Efficiency Video Coding]

$NAL$ Network Abstraction Layer

$NLOS$ Non-Line Of Sight

$NQM$ Noise Quality Metric

$nrtPS$ Non real-time Polling Service

$NSS$ Natural Scene Statistics

$NTIA$ National Telecommunication and Information Administration

$OFDM$ Orthogonal Frequency Division Multiplexing

$OFDMA$ Orthogonal Frequency-Division Multiple Access

$PDU$ Protocol Data Units

$PHR$ Personal Health Record

$PHY$ Physical layer

$PIPE$ Probability Interval Partitioning Entropy

$PKM$ Privacy Key Management

$PLR$ Packet Loss Rates

$PS$ Polling Service

$PSNR$ Peak Signal-to-Noise Ratio

$PU$ Prediction unit

$QALF$ Quadtree-based Adaptive Loop Filter

$QoE$ Quality of Experience

$QoS$ Quality of Service

$QP$ Quantization Parameter

$QPSK$ Quadrature Phase Shift Keying

$RA$ Region-based Adaptation

$RD$ Rate-Distortion

$RDO$ Rate Distortion Optimization

$RDOQ$ Rate Distortion Optimized Quantization

$RDPCM$ Residual Differential Pulse Code Modulation

$RExt - HEVC$ Range Extensions High Efficiency Video Coding

$RF$ Radio Frequency

$RHIN$ Refugee Health Information Network

$RLC$ Radio Link Control

$ROI$ Regions of Interest

$RQT$ Residual Quadtree

$RS$ Redundant Slices

$rtPS$ Real-time Polling Service

$SAO$ Sample Adaptive Offset

$SC$ Single Carrier

$SCC$ Screen Content Coding

$SD-FDMA$ Single-carrier Frequency-Division Multiple Access

$SF$ Separable Filters

$SHVC$ Scalable High efficiency Coding

$SNR$ Signal-to-Noise Ratio

$SOFDMA$ Scalable Orthogonal Frequency Division Multiple Access

$SP/SI$ Switching-Predictive/Switching Intra

$SS$ Subscriber Stations

$SSIM$ Structural SIMilarity

$SVC$ Scalable Video Coding

$SVT$ Spatially Varying Transform

$TDD$ Time-Division Duplex

$TDMA$ Time-Division Multiple Access

$TD-SCDMA$ Time Division Synchronous Code Division Multiple Access

$TU$ Transform Unit

$UEP$ Unequal Error Protection

$UGS$ Unsolicited Grant Service

$UHD$ Ultra High Definition

$UL$ Uplink

$UL-MAP$ Uplink Map Messages

$UVLC$ Universal Variable Length Coding

$VCEG$ Video Quality Experts Group

$VD$ Visual Distortion

$VIF$ Visual Information Fidelity

$VIIDEO$ Video Intrinsic Integrity and Distortion Evaluation Oracle

$VLC$ Video Coding layer

$VMAF$ Video Multi-Method Assessment Fusion

$VQA$ Video-Quality Assessment

$VSNR$ Visual Signal to Noise Ratio

$WCDMA$ Wide-band Code Division Multiple Access

$WiMAX$ World Wide Interoperability for Microwave Access

$WMAN$ Wireless Metropolitan Area Network

$WPP$ Wavefront Parallel Processing

$WSNR$ Weighted Signal to Noise Ratio

$ZL$ Zero latency

# Chapter 1

# Introduction

## 1.1 Introduction

The term mobile health (mHealth) was first introduced implicitly as ”Unwired e-med” in the first special issue on wireless telemedicine systems of the IEEE Transactions on Information Technology in Biomedicine journal [1]. Since then, there have been significant advances in wireless communications and network technologies. These advances have already made significant impact on current eHealth and telemedicine systems [2, 3, 4].

In historical order of introduction and development of such terms, the information and technology health domains (or classes) include telemedicine (1905 and 1969), eHealth (1999), and mHealth (2003). Telemedicine can be defined as the delivery of health care and sharing of medical knowledge over a distance using telecommunication means. It aims at providing distribute medical expertise and services anywhere that health care is needed. It was firstly introduced as a concept a few decades ago, when telephone and fax machines were the first telecommunication means used. Telemedicine applications, including those based on wireless technologies, cover large areas of emergency health care such as telecardiology, teleneurology, teleleoncology, and many more [5].

eHealth is a relatively recent healthcare practice supported by electronic processes and communication and can be defined as the cost-effective and secure use of information and communication technologies in support of the health and health-related fields including healthcare, health surveillance and health education, knowledge and research [6]. The term mHealth was introduced into the literature in 2003 [7] in response to the vast expansion of mobile communication technology and its perceived usefulness in facilitating access to healthcare, especially in the developing world. An important feature of mHealth is its person-centered nature and ubiquity enabled by mobile phone technology and connection to the Internet.

The significant number of mHealth systems developed over the past decade reflect the broad spectrum of usage, and highlight the necessity of such systems and services [5, 4]. A timely and effective way of handling emergency cases can prove essential for a patient's recovery or even for a patient's survival. Especially in cases of serious injuries the way of transporting and, generally the way of providing care are crucial for the quality of life the patient. Today with the assistance of the wireless technology, health applications include remote diagnosis and care, home monitoring of patients with chronic diseases and the elderly, and assistive technologies. Furthermore, services offered by telemedicine are designed to help in improving healthcare access and information services while reducing the isolation between healthcare providers and residents, especially in rural areas. Wireless telemedicine reduces time and costs of patient's transportation from rural areas, while it is used in distinct areas such as consultations in neurology, cardiology, and general medicine. Telemedicine is also used for Continued Medical Education (CME), research and development.

However, despite the rapid growth of telemedicine systems, wireless channels remain error prone, while the continuous bitrate and compression efficiency increases are soon met by the rising expectations of the amount of data to be transmitted. Any increase in available data rates

is soon met by continuously increasing demands for medical video bandwidth. Different from conventional video-streaming quality requirements, the quality of transmission systems in eHealth is measured in terms of their diagnostic yield. Any compromise of clinically sensitive video regions can lead to deterioration of the system's objective of remote diagnosis and care [8].

In terms of wireless mobile infrastructure, the Global System for Mobile communications [9] was the most popular standard. GSM signified the transition from analog first-generation (1G) to digital second-generation (2G) technology. It had been designed for voice communication (circuit switched), but it can also carry data at rates up to 9.6 kbps. At such low rates, GSM can only be used for still images but not for the transmission of medical video. The evolution of mobile telecommunication systems from 2G to 2.5G and subsequently to 3G, 3.5G, mobile WiMAX, and LTE systems has facilitated both an always-on model (as compared with the circuit-switched mode of GSM), as well as the provision of faster data transfer rates and lower delays, thus enabling the development of more-responsive telemedicine systems [4].

In terms of bandwidth, both 2.5G and 3G in most cases provide sufficient rates for medical image and biosignal transmission. For medical video transmission, 3G rates are sufficient for QCIF (176 x 144) resolution medical video, as well as specific regions of interests (ROIs). Furthermore High-speed packet access High-Speed Packet Access ($HSPA$) and HSPA+ 3.5G technologies enable the transmission of high quality CIF (352 x 288) as well as with some limitations up to 4CIF (704 x 576) resolution video and beyond [10].

WiMAX release 2.0 and LTE-advanced networks conforming to the International Mobile Telecommunications-Advanced (IMT-advanced) requirements [11] constitute the next-generation family of technologies, namely 4G, whilst 5G is expected towards 2020. Low latency, high mobility, high bandwidths (targeting 100 Mbps for high mobility and 1 Gbps for low mobility, in the downlink), and quality-of-service (QoS) provisions provided by 4G wireless networks, are

expected to significantly boost the development of mHealth systems and services. Moreover, 5G planning aims at higher capacity and better coverage compared to current 4G, allowing a higher density of mobile broadband users, and supporting device-to-device, more reliable, and massive machine communications. 5G research and development also aims at lower latency ($<$1ms) than 4G equipment and lower battery consumption [12]. 5G is expected to boost mHealth and eHealth applications in a major way to enable the introduction of additional services such as personalized or precision medicine initiatives with distributed, patient-centric approaches [13]. As a result, the wider availability of wireless network infrastructure and the emerging high efficiency video compression methods will significantly advance existing mHealth applications and support higher reliability of communication.

In addition to new wireless networks' standards, efficient video compression systems can be build using the video coding standards in order to provide for both an efficient (size wise) and timely (real time) encoding. The latter will enable the transmission of high definition quality.

The advance of digital video compression in the last three decades has produced numerous fruitful results. Several international image and video coding scenarios have been standardized, for example, ISO/IEC JPEG for still images, ITU H.261/H263 for video telephony and, ISO/IEC MPEG-1 and MPEG-2 for video CD and digital TV [14]. The next video coding standard, MPEG-4 part 2 was released in 1999. The most significant addition to the video coding standard was H.264/MPEG-4 Advanced Video Coding (AVC) standard, which was finalized in 2003 [15]. H.264/AVC video format has a very broad application range that covers all forms of digital compressed video from low bitrate Internet streaming applications to High-definition TV (HDTV) broadcast and digital cinema applications, with nearly lossless coding. Many devices, including mobile devices, already support this video format.The current video coding standard is the High Efficiency Video Coding (HEVC) [16] standard, that provides additional coding efficiency and

network reliability, that can strongly benefit the development of mHealth video communication systems [17].

## 1.2 Motivation

Wider adoption of mHealth systems in routine clinical practice necessitates the delivery of sufficiently high video resolutions and frame rates, subject to low-delay and low packet loss rates requirements, for reliable diagnosis. However, this approach requires real-time control to provide for adequate levels of clinical video quality. Currently, this is limited by the inability of current technologies to support real-time medical video transmission at the acquired resolution and frame rate, throughout a video streaming session that suffers from time-varying wireless networks' constraints. In addition, the reduction in video resolution and frame rates, and the need to critically assess compression-induced artifacts, creates further obstacles to wider adoption. Thus, broader adoption of real-time mHealth systems can only be achieved using adaptive video encoding approaches. These should support real-time adaptation based on the time-varying wireless networks state, in order to guarantee clinically acceptable performance throughout the streaming session, whilst conforming to device capabilities for supporting real-time encoding.

The goal of this thesis was to develop an adaptive medical video communication framework that facilitates increased quality of care while reducing hospitalization times and associated healthcare costs, and at the same time to match the clinical experience levels of in hospital examinations. This framework is expected to transform healthcare delivery by fostering responsive emergency systems, in unrestricted and cross border settings that will significantly improve patient's quality of care and hence quality of life.

## 1.3 Thesis Original Objectives

This thesis relies on two studies. The first one proposes an effective video communication framework for the wireless transmission of H.264/AVC medical ultrasound video over mobile WiMAX networks.

The second study, proposes a scalable, video modality, encoder, and wireless network agnostic framework, that supports real-time adaptation to time-varying wireless networks' state while guaranteeing diagnostically lossless video communications and conforming to end-user device constraints for real-time performance.

The objectives of the first study were:

- **Relationship between spatial resolution (QCIF, CIF, 4CIF) and clinical diagnosis**: This relationship between video resolution and clinical quality was investigated via multiple clinical evaluation sessions. Experiments involved medical experts evaluating a video's diagnostic quality based on the actual clinical protocol that is followed during common carotid artery (CCA) ultrasound examination, displayed at different resolutions. The medical experts were asked to validate the clinical content of these three resolutions. The findings verified the hypothesis that higher resolution accommodates a larger amount of clinical information. The use of higher 4CIF ($704 \times 576$) resolution underpins new clinical quality standards that are closer to standards used for in hospital exams.

- **Medical video communications over Mobile WiMAX networks**: Within the context of the present thesis, the case study of medical video communication over Mobile WiMAX networks was evaluated using the OPNET modeler and video traces of real CCA ultrasound videos. Recommendations for mobile WiMAX network's parameters setup and utilization towards maximizing the communicated video's clinical capacity are hence proposed while

findings motivated the proposed adaptive video communication framework. Key parameters that are highlighted in the thesis involve three different channel modulation and coding schemes, various distances from the base station (BS), and diverse mobility patterns. To demonstrate the effectiveness of the proposed framework, we successfully transmitted clinically acceptable ultrasound video of 4CIF resolution over Mobile WiMAX networks.

The objectives of the second study were:

- **Multi-objective optimization for real-time operation**: The thesis proposes and evaluates a multi-objective optimization framework for adaptive video delivery that leverages video quality (application-modality level adaptation), bitrate demands (wireless network adaptation), and encoding frame rate (device adaptation for real-time operation). Thus, the proposed solution considers encodings that simultaneously maximize video quality and encoding rate, while minimizing the required bitrate. The required space of solutions forms a Pareto front that is used for solving constrained optimization problems.

- **Robust prediction models**: Robust prediction models were generated for each optimization objective, that allow real-time encoding adaptation for real-time operation. More specifically, forward prediction models were developed for SSIM (video) quality, bitrate demands, and encoding frame rate, for three different group of pictures (GOP) encoding structures, namely zero latency (ZL), B2 and B4 respectively. Overall, prediction accuracy results were promising, especially for video quality and bitrate demands, since the depicted errors were within reasonable bounds for all investigated models.

- **Real-Time adaptation using pareto-optimal encoding configurations**: Real-time adaptation at a GOP level was demonstrated using the HEVC standard. The effectiveness of the

proposed framework compared to static, non-adaptive approaches was demonstrated for different modes of operation, achieving significant quality gains, bitrate demands reductions, and performance improvements, in real-life scenarios imposing time-varying constraints.

- **Video modality, encoder, and wireless network agnostic framework**: The proposed approach is generic enough and should be applicable to other medical video modalities and for different applications, provided that the appropriate training described in the thesis is performed.

## 1.4 Publications List

This section provides a list of all publications stemming out of this thesis. Major contributions of this thesis have been published in:

**Journal papers**

1. Andreas Panayides, Zinonas Antoniou, Yiannos Mylonas, Marios Pattichis, Andreas Pitsillides, and Constantinos Pattichis, "High-resolution, low-delay, and error-resilient medical ultrasound video communication using H. 264/AVC over mobile WiMAX networks," in IEEE journal of Biomedical and Health Informatics 17.3 (2013): 619-628.

2. Zinonas Antoniou, Andreas Panayides, Marios Pantziaris, Anthony Constantinides, Constantinos Pattichis, and Marios Pattichis, "Real-Time Adaptation to Time-Varying Constraints for Medical Video Communications," in IEEE journal of Biomedical and Health Informatics (2017), Early Access.

   Zinonas Antoniou, Andreas Panayides, Marios Pattichis, and Constantinos Pattichis, "A Comparative Evaluation of Scalable-HEVC and MPEG-DASH for Adaptive Video Delivery Telemedicine Applications," To be submitted in February 2018.

**Book chapters**

1. Andreas Panayides, Zinonas Antoniou, and Anthony Constantinides, "An overview of mHealth medical video communication systems," in Mobile Health. Springer International Publishing, 2015, pp. 609-633.

**Conference and Workshop papers**

1. Andreas Panayides, Zinonas Antoniou, Marios Pattichis, Constantinos Pattichis and Anthony Constantinides, "High efficiency video coding for ultrasound video communication in mHealth systems," in Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pp. 2170-2173, 2012.

2. Zinonas Antoniou, Andreas Panayides, Marios Pattichis, Stavros Stavrou, Efthyvoulos Kyriacou, Andreas Spanias, Anthony Constantinides, and Constantinos Pattichis, "Adaptive emergency scenery video communications using HEVC for responsive decision support in disaster incidents." Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015.

3. Zinonas Antoniou, Andreas Panayides, Marios Pattichis, Stavros Stavrou, Efthyvoulos Kyriacou, Andreas Spanias, Anthony Constantinides, and Constantinos Pattichis, "A unifying framework for m-Health video communication systems using HEVC," in 8th Cyprus Workshop on Signal Processing and Informatics, July 03 2015, University of Cyprus, Nicosia, Cyprus.

4. Zinonas Antoniou, Andreas Panayides, Marios Pantzaris, Anthony Constantinides, Constantinos Pattichis, and Marios Pattichis, "Dynamic Network Adaptation for Real-Time Medical Video Communication," in XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016. Springer International Publishing, 2016.

5. Zinonas Antoniou, Andreas Panayides, Marios Pantzaris, Anthony Constantinides, Constantinos Pattichis and Marios Pattichis, "Adaptive Video Communications for Real-Time mHealth Applications," in Proceedings of the 39th Annual Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, July 11-15, 2017.

6. Zinonas Antoniou, "Adaptive Video Encoding Framework based on Multi-objective Optimization for Real-time mHealth Applications,"in 10th Cyprus Workshop on Signal Processing and Informatics, July 19 2017, University of Cyprus, Nicosia, Cyprus.

### 1.5 Guide to Thesis Contents

Chapter 2 provides an overview of video coding standards development along with a detailed review of the latest video coding standards, namely H.264/AVC and HEVC, summarizing their most important characteristics and capabilities.

Chapter 3 incorporates the characteristics of wireless technologies with an emphasis to 4G, that are expected to allow real-time medical video transmission of higher resolution and frame rate. Communication protocols necessary for establishing a connection between the transmitting/receiving parties and responsible for conveying clinical video data are also highlighted.

Chapter 4 describes an effective video communication framework for the wireless transmission of H.264/AVC medical ultrasound video over mobile WiMAX networks. Medical ultrasound video is encoded using diagnostically driven, error resilient encoding, where quantization levels are varied as a function of the diagnostic significance of each image region. This chapter demonstrates how our proposed system allows for the transmission of high-resolution clinical video that is encoded at the clinical acquisition resolution and can then be decoded with low delay. To validate performance, we perform OPNET simulations of mobile WiMAX medium access control

and physical layers characteristics that include service prioritization classes, different modulation and coding schemes, fading channel's conditions, and mobility. We encode the medical ultrasound videos at the 4CIF (704×576) resolution that can accommodate clinical acquisition that is typically performed at lower resolutions. Video quality assessment is based on both clinical (subjective) and objective evaluations. Comprehensive experimentation showed that low-delay high resolution 4CIF ultrasound video transmission is possible over mobile WiMAX networks, even at speeds of 100 km/h and distances of 1 km from the BS. The investigated channel modulation and coding schemes verified that QPSK 1 /2 is the most robust scheme, especially when transmitting from locations with low SNR. On the other hand, 16-QAM 3/4 and 64-QAM 3/4 provide higher network capacities and are preferable when the transmitting station is closer to the BS. The performance of the system in terms of transmitted video's quality was evaluated using both objective and subjective evaluations. Clinical validation verified the capacity of mobile WiMAX networks to provide robust, clinically acceptable 4CIF ultrasound video transmission, thus enabling the transmission of ultrasound video at resolutions close to the original's video acquired resolution. The scope of this study is to identify the need for adaptive, real-time mHealth video communications framework.

Chapter 5 discusses a scalable, medical modality and technology independent approach for adaptive, real-time mHealth video communications. An adaptive video encoding framework is designed and developed, based on multi-objective optimization that jointly maximizes the encoded video's quality and encoding rate (in frames per second) while minimizing bitrate demands. For this purpose, we construct a dense encoding space and use linear regression to estimate forward prediction models for quality, bitrate, and computational complexity. The prediction models are then used in an adaptive control framework that can fine-tune video encoding based on real-time constraints. The system is developed using a leave-one-out algorithm applied to ten ultrasound

videos of the common carotid artery. The prediction models can estimate SSIM quality with a median accuracy error of less than 1%, bitrate demands with a deviation error of 10% or less, and encoding frame rate within a 6% margin. Real-time adaptation at a Group of Pictures (GOP) level is demonstrated using the High Efficiency Video Coding (HEVC) standard. Results demonstrate that efficient adaptation at a GOP level is possible using our proposed models, while the effectiveness of the proposed framework compared to static, non-adaptive approaches is demonstrated for different modes of operation, achieving significant quality gains, bitrate demands reductions, and performance improvements, in real-life scenarios imposing time-varying constraints. Our approach is generic and should be applicable to other medical video modalities with different applications.

Chapter 6 provides the conclusions and future work. Main achievements are summarized both in terms of technical and clinical views. Results demonstrate that efficient adaptation at a GOP level is possible using our proposed real-time methodology, significantly outperforming static approaches and demonstrates higher flexibility and precision compared to currently dominant pre-computed states found in the literature [18, 19]. Future work incorporates further exploitation of HEVC features and Medical Video Quality Assessment linked with knowledge emanating from video transmission simulations and find threshold values (correlation) for which clinical quality is maintained with respect to video quality assessment algorithms.

# Chapter 2

## Review of Video Standards and Video Quality

### 2.1 Introduction

Standardization of video compression algorithms benefited by the success of recent multimedia systems. Several international image and video coding scenarios have been standardized for video telephony, video CD and digital TV as well. The advance of digital video compression in the last three decades resulted numerous fruitful accomplishments in this field, culminating in standardization of H.264/MPEG-4 Advanced Video Coding (AVC) in 2003.

H.264/AVC met the growing demand of multimedia and video services by providing enhanced compression efficiency significantly outperforming all prior standards. H.264/AVC represents a number of advances in standard video coding technology, in terms of coding efficiency improvement, error/loss robustness enhancement, and flexibility for effective use over a broad variety of network types and application domains. Its video coding layer (VCL) design is based on conventional block-based motion-compensated hybrid video coding concepts, but with some important differences relative to prior standards, which include enhanced entropy coding methods, use of a small block-size exact-match transform, adaptive in-loop deblocking filter and enhanced motion prediction capability. H.264/AVC can provide for bitrate reductions of up to 50% for equivalent

perceptual quality compared to its predecessors. H.264/AVC offers a range of error resilience techniques for a wide variety of applications. To this end, H.264/AVC defines different profiles and levels. Each profile and level specify restrictions on bitstreams, hence limits on the capabilities needed to decode this bitstreams [15].

The current state-of-the-art video coding standard is the High Efficiency Video Coding (HEVC) [16] standard that will provide additional coding efficiency and network reliability that can strongly benefit mHealth video communication systems [17]. HEVC is designed for video resolutions ranging from 416x240 (WVGA) to 2560x1600 [20, 21], and aims to reduce bitrate demands by as much as 50% for equivalent perceptual quality, compared to the current standard.

mHealth video communication systems rely on video coding standards and the underlying wireless network. In order to evaluate the quality of the transmitted compressed video resulting from such system, a diagnostic validation method has to be adopted. Diagnostic validation is the most significant requirement for emerging medical video transmission systems. Diagnostic validation requires an accurate assessment of the diagnostic capacity of the transmitted medical video.

## 2.2 Video Standards

### 2.2.1 Overview of Video Coding Standards

Since the early 1990s, ITU-T video coding standards, namely H.261, H.263 and H.264, have been the engines behind the commercial success of digital video compression. They have played pivotal roles in spreading the technology by providing the power of interoperability among products developed by different manufacturers, while at the same time allowing enough flexibility for ingenuity in optimizing the technology to fit a given application and making the cost-performance

trade-offs best suited to particular requirements [22]. All ITU-T standards since H.261 [23] have in common that they are based on the so-called block-based hybrid video coding approach. The basic source coding algorithm is a hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the prediction error signal to reduce spatial redundancy. Each picture of a video signal is partitioned into fixed-size macroblocks of 16x16 samples, which can be transmitted in one of several coding modes depending on the picture or slice coding type. Common to all standards is the definition of INTRA coded pictures or I-pictures. In I pictures, all macroblocks are coded without referring to other pictures in the video sequence. Also common is the definition of predictive-coded pictures, so-called P-pictures and B-pictures. In predictive-coded pictures, typically one of a variety of INTER coding modes can be chosen to encode each macroblock. In order to manage the large number of coding tools included in standards and the broad range of formats and bit-rates supported, the concept of profiles and levels is typically employed to define a set of conformance points, each targeting a specific class of applications. These conformance points are designed to facilitate interoperability between various applications of the standard that have similar functional requirements. A profile defines a set of coding tools or algorithms that can be used in generating a compliant bit-stream, whereas a level places constraints on certain key parameters of the bit-stream, such as the picture resolution and bit-rate.

H.261 developed in 1988 to 1990 by ITU-T, former CCITT. H.261 compression was designed for videotelephony and videoconferencing applications and it was part of the H.320 group of standards which describes the different components of a video conferencing system and define a narrow-band multimedia terminal. Its compression algorithm took advantage of both the spatial and the temporal redundancy of video sequences to achieve high compression ratios. H.261 design goals were low delay and Constant Bit Rate (CBR) but also had some limitations, including limited rubustness to bit errors, no recording/playback and no backward or forward seek due to any frame

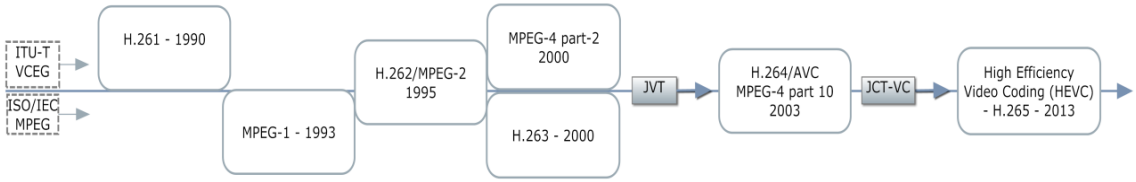



Figure 1: Timeline of video coding standard's development [25]

encode with entirely intra-coded block. H.261 descendant was H.263 [24] which was standardized in 1996 and replaced H.261 in many applications. Its basic design was very similar to H.261 but had numerous optional improvements to improve compression, robustness, and flexibility of use. While H.261 only supported two resolutions (QCIF (176x144) and CIF (352x288)), H.263 also supported SQCIF (128x96), 4CIF (704x576) and 16CIF (1408x1152) and instead using 16x16 macroblocks for prediction it used four 8x8 blocks for better detail. The primary goal of H.263 was to improve quality at lower bit rates.

### 2.2.2 H.264/Advanced Video Coding (AVC)

H.264/AVC was jointly developed by the ISO/IEC motion pictures experts group (MPEG) and ITU-T video quality experts group (VCEG) who formed the Joint Video Team (JVT). H.264/AVC met the growing demand of multimedia and video services by providing enhanced compression efficiency significantly outperforming all prior standards (MPEG-x and H.26x, see Figure 1).

H.264/AVC can provide up to 50% for bitrate reductions for equivalent perceptual quality compared to its predecessors [26]. Its design enables transportation over heterogeneous networks to be carried out in a friendly-manner. To achieve this, H.264/AVC defines a VCL and a network abstraction layer (NAL). VCL is responsible for video coding and is a unit already known from prior standards, maintaining its block-oriented coding functionality [27]. Its enrichment and refinement resulted in the provided compression efficiency. Figure 2 depicts the basic encoding structure of H.264/AVC. On the other hand, NAL is a novel concept aiming at a network-friendly

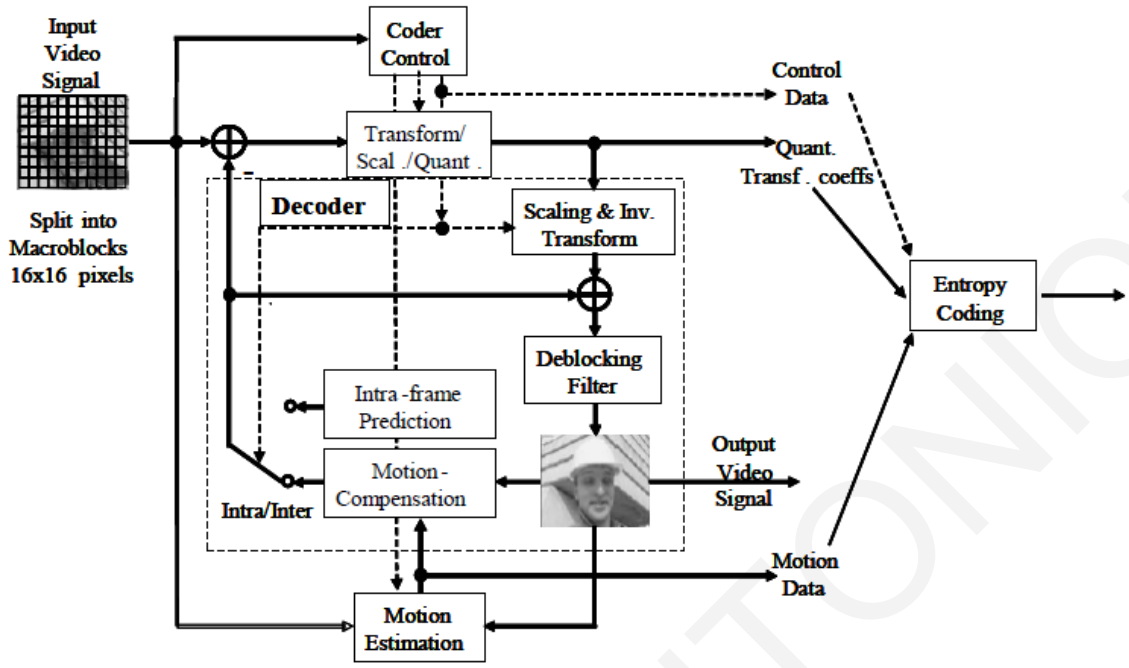



Figure 2: H.264/AVC coding structure (example based on [26])

adaptation of VCL content to candidate heterogeneous networks (or storage devices). NAL functionality is a substantial improvement constituting H.264/AVC coding and transmission network-independent. An example of VCL and NAL functionality is illustrated in Figure 3. The scope of the standard is centered on the decoder, due to that only the decoder is standardized, allowing great flexibility to the encoder.

H.264/AVC offers a range of error resilience techniques for a wide variety of applications. For this purpose, H.264/AVC defines different profiles and levels. Each profile and level specify restrictions on bitstreams, hence limits on the capabilities needed to decode the corresponding bitstreams [15]. Baseline, main, extended and high profiles assume different processing devices tailored for different applications and offer incremental level capabilities (and therefore complexity), that is alleviating constraints on bitstreams [28]. Figure 4 demonstrates unique features of each profile, while Table 1 summarizes some of the different capabilities of each level, including

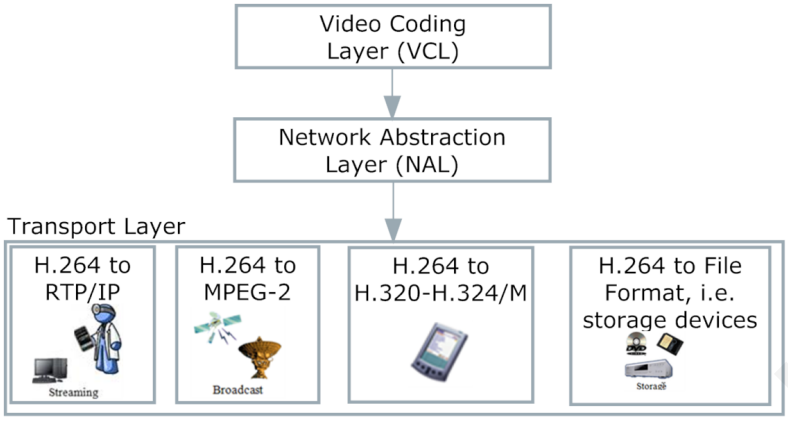



Figure 3: H.264/AVC network abstraction layer (NAL) and video coding layer (VCL) [25]

resolution, frame rate, maximum allowed bandwidth, and maximum coded picture buffer (CPB) and decoded picture buffer (DPB).

A level is a specified set of constraints that indicate a degree of required decoder performance for a profile. For example, a level of support within a profile specifies the maximum picture resolution, frame rate, and bit rate that a decoder may use. A decoder that conforms to a given level must be able to decode all bitstreams encoded for that level and all lower levels. Lower levels mean lower resolutions, lower allowed max bitrates and lower memory to store reference frames. A level primarily used for device compatibility. For example, the iPhone supports H.264 Level 3, which means that a video's peak bitrate can't exceed 10,000 kbps. Also, video coding layer of H.264 as depicted in Figure 3 can support a maximum video bit rate for each level for each profile. Referring to the previous example about iPhone, video's peak bitrate 10.000 kbps means that iPhone supports Baseline, Extended and Main profiles. Additionally, as can be seen in Table 1, level 3 can support up to four high resolutions in a specific frame rate, for example 720x526 at 25.0 frames per second. Generally, the maximum bit rate for High Profile is 1.25 times that of the Base/Extended/Main Profiles, 3 times for Hi10P, and 4 times for Hi422P/Hi444PP. The number

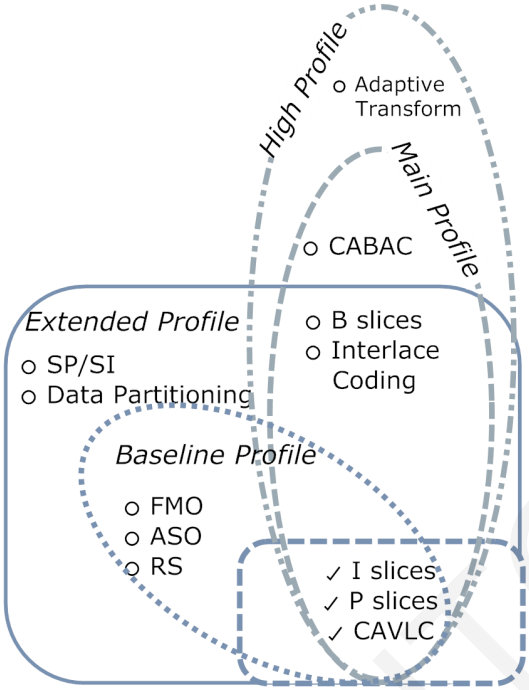



Figure 4: H.264/AVC baseline, main, extended, and high profiles features [25]. Context-adaptive variable-length coding(CAVLC), flexible macroblock ordering (FMO), arbitrary slice ordering (ASO), redundant slice (RS), switching-predictives/switch (intra SP/SI), context-adaptive binary arithmetic coding (CABAC)

of luma samples is 16x16=256 times the number of macroblocks and the number of luma samples

per second is 256 times the number of macroblocks per second.

Table 1: H.264/AVC Level maximum supported parameters. Levels address the problem of bandwidth, max resolution, and memory issues on the decoder side. Any given device supports a max profile depending by the memory and the max resoultion available. Lower levels mean lower resolutions, lower allowed max bitrates and lower memory to store reference frames.

| Level id. | Max. Video Bitrate | Max Frame (MB) | Max MB per second | Max CPB[a] Size (MB) | Max Delay at Max Bitrate | Max DPB[b] size (bytes) | Resolution, Frame Rate, Max Buffer Picture (max stored frames) |
|---|---|---|---|---|---|---|---|
| 1 | 64 kbit/s | 99 | 1485 | 175 | 2.7 | 152 064 | **SQCIF** 128x96 @ 30fps (8)<br>**QCIF** 176x144 @ 15fps (4) |
| 1b | 128 kbit/s | 99 | 1485 | 350 | 2.7 | 152 064 | **QCIF** 176x144 @ 15fps (4) |
| 1.1 | 192 kbit/s | 396 | 3000 | 500 | 2.6 | 345 600 | **QCIF** @ 30fps (9) |
| 1.2 | 384 kbit/s | 396 | 6000 | 1 | 2.6 | 912 384 | **CIF** 352x288 @ 15fps (6) |
| 1.3 | 768 kbit/s | 396 | 11880 | 2 | 2.6 | 912 384 | **CIF** x 30fps (6) |
| 2 | 2 Mbit/s | 396 | 11880 | 2 | 1 | 912 384 | **CIF** x 30fps (6) |
| 2.1 | 4 Mbit/s | 792 | 19800 | 4 | 1 | 1 824 768 | **525 HHR** 352x480 @ 30fps (7)<br>**625 HHR** 352x576 @ 25fps (6) |
| 2.2 | 4 Mbit/s | 1620 | 20250 | 4 | 1 | 3 110 400 | **525 SD** 720x480 @ 15fps (6)<br>**625 SD** 720x576@12.5fps(5) |
| 3 | 10 Mbit/s | 1620 | 40500 | 10 | 1 | 3 110 400 | **525 SD** x 30fps (6)<br>**625 SD** x 25fps (5)<br>**VGA** 640x480@30fps (6) |
| 3.1 | 14 Mbit/s | 3600 | 108000 | 14 | 1 | 6 912 000 | **720p HD**<br>**1280x720** @30fps(5) |
| 3.2 | 20 Mbit/s | 5120 | 216000 | 20 | 1 | 7 864 320 | **720p HD** x 60fps (5) |
| 4 | 20 Mbit/s | 8192 | 245760 | 25 | 1.25 | 12 582 912 | **720p** HD x 60fps (9)<br>**1080 HD** (1920x1088) $\times$ 30fps (4)<br>**2k x 1k** (2048x1024) $\times$ 30fps (4) |
| 4.1 | 50 Mbit/s | 8192 | 245760 | 62.5 | 1.25 | 12 582 912 | as above |
| 4.2 | 50 Mbit/s | 8704 | 522240 | 62.5 | 1.25 | 13 369 344 | **1080 HD** x 60fps (4)<br>**2k x 1k** x 60fps (4) |
| 5 | 135 Mbit/s | 22080 | 589824 | 135 | 1 | 42 393 600 | **2k x 1k** x 72fps (13) |
| 5.1 | 240 Mbit/s | 36864 | 983040 | 240 | 1 | 70 778 880 | **2k x 1k** @120fps (16)<br>**4k x 2k** 4096x2048@30fps (5) |

Macroblock (MB), [a]Coded picture buffer (CPB),[b]Decoded picture buffer (DPB)

#### 2.2.2.1 Encoding Modes and Frame Types

Frame encoding modes can have a significant impact on both error propagation and video compression performance. A summary of the different modes is provided next:

- **Intra-mode:** Intra-mode is the procedure where intra-prediction is used for coding a video frame (I-frame) [26] as a part of the basic encoding structure of H.264/AVC as depicted in Figure 2 and as can be seen in Figure 4, it is supported by all the profiles of H.264/AVC. All the information used for encoding is restricted within the frame. No prediction from previous or future frames is allowed. As a result, intra-mode encoded frames require higher bitrates than inter-mode encoded frames. On the other hand, the use of intra-mode coding significantly limits error-propagation in wireless video transmission networks [25].

- **Inter-mode:** Inter-mode is the procedure where inter-prediction is used for coding a video frame.

  · **P-mode:** P-mode uses prediction from previously decoded frames. In inter-mode, the encoder's side provides all the necessary information for accurate motion estimation of the spatial displacement between the decoder's reference picture and the current picture in the sequence at the encoder. With respect to Figure 2, this procedure is described as motion compensation [29]. Clearly, decoding errors in the reference picture will be propagated to the predicted frame. As can be seen in Figure 4, P slices are supported by all the profiles of H.264/AVC.

  · **B-mode:** Whereas in P-mode at most one motion compensated signal is employed, B-mode provides the ability to make use of two motion compensated signals for the prediction of a picture. B-mode is also referred to as bi-prediction as not only it allows the utilization of previously decoded pictures but also the utilization of forthcoming

ones. Again, errors from previously decoded pictures propagate to the predicted frame. On the other hand, B-mode pictures require less bandwidth than both P-mode and I-mode pictures [22, 25]. As depicted in Figure 4, B slices are supported by all the profiles of H.264/AVC except the main.

The use of predictive coding, like P-frames or B-frames, or not, like I-frames, depends on the application. Depending on time and quality constraints imposed, one mode may be preferred over the other and the other way around. The ratio between P-frames and B-frames used is also application specific. A ratio of 2:1, or IBBPBBP coding structure has proved to be a good balance between single-directional and bi-directional prediction, widely used for internet video streaming [26, 25, 29] .

#### 2.2.2.2 Intra Updating

In broadcasting applications it is essential to insert an Intra coded frame in a sequence of every GOP frames, being utilized for random access to the transmitted bitstream such as joining an ongoing session. However, the transmission of videos using completely intra coded frames in non pre-encoded (real-time) applications, is rather limited. This is because intra coding requires increased encoding time and involves considerable bandwidth (size) compared to predictive coded frames, usually unacceptable in limited bandwidth, strict time delay applications. Pre-defined, as well as random intra-macroblock refresh is used instead to limit error propagation in error prone wireless environments [30, 31]. Intra-macroblock refreshing can prove particularly efficient in the presence of a low end-to-end delay feedback channel. A feedback channel can provide information as to which part of the picture is affected by losses and needs to be intra-coded in order to limit error propagation. Referring to Figure 4, Intra Updating is supported by all the profiles of H.264/AVC except the main.

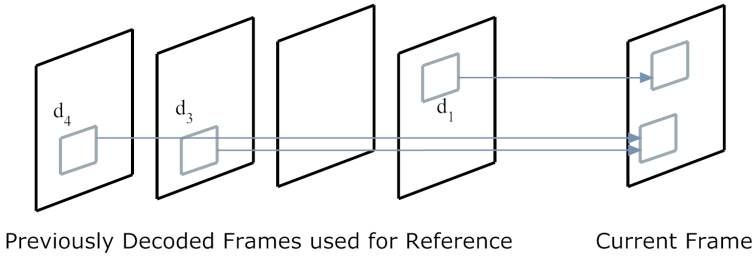


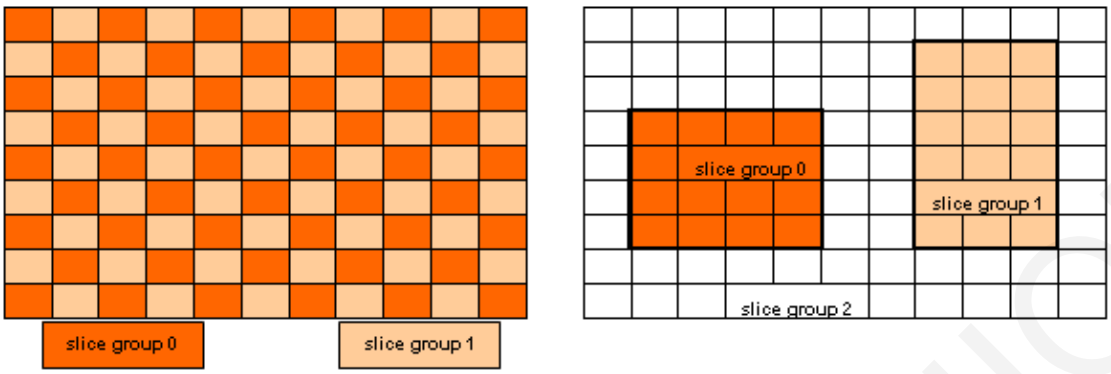

Figure 5: Multiple reference frame prediction. In addition to the motion vector, also an image reference parameter $d_t$ is transmitted, where $d_4$=4, $d_3$=3, and $d_1$=1. Example based on [26]

#### 2.2.2.3 Multiple Reference

H.264/AVC allows the utilization of up to 16 reference frames for prediction during encoding. Employing a certain number of previous or future frames which all contribute to the encoding of the current frame, rather than just a single reference frame (past or future) enhances predictive coding accuracy. In this concept, the linear combination of prediction signals is also made regardless of the temporal direction. Given the computational intensity of using multiple reference frames though and increased memory usage at the decoder, the number of reference frames should be selected wisely, especially for real time applications. Studies have shown multiple reference frames to work better in the presence of a feedback channel, notifying the encoder to avoid erroneously received frames at the decoder for motion estimation purposes [31]. Furthermore, using H.264/AVC it is possible to use images containing B-slices as reference frames for further predictions which was not possible in any former standard [29]. Multiple Reference is supported by all profiles of H.264/AVC.

#### 2.2.2.4 Flexible Macroblock Ordering, Redundant Slices and Arbitrary Slice Ordering

An innovative error resilient feature introduced by H.264/AVC is flexible macroblock ordering [32] (FMO), which is essentially a slice structuring approach, where a frame is partitioned into independently transmitted and decoded slices. Each frame may be partitioned in up to eight




(a) (b)

Figure 6: Flexible macroblock ordering. Subdivision of a QCIF frame into slices. (a)Slice Groups 0 and 1 are transmitted separately useful for error concealment in video conferencing applications and (b) Dispersed macroblock allocation where Slice Groups 0 and 1 are the foreground and slice Group 2 is the leftover for use in region-of-interest type of coding applications Example based on [26]

different slices and a frame may still be decoded even if not all slices are present at the decoder. A slice contains a number of macroblocks (MBs), the basic block coding unit of H.264/AVC. Using FMO, a picture can be split into many macroblock scanning patterns such as interleaved slices, a dispersed macroblock allocation, one or more ”foreground” slice groups and a ”leftover” slice group, or a checker-board type of mapping. The latter two are illustrated in Figure 5. The left-hand side macroblock to slice group mapping has been demonstrated for use in region-of-interest type of coding applications. The right-hand side macroblock to slice group mapping has been demonstrated useful for concealment in video conferencing applications where slice group 0 and slice group 1 are transmitted in separate packets and one of them is lost. In this manner and in conjunction with proper utilization of the spatial relationships between error free slices and MBs therein, concealment of errors becomes much more efficient.

Seven different types of FMO are defined (i.e. patterns for MB to slice allocation). A macroblock allocation map (MBAmap) is used to keep track of MBs assigned to slices.

The most interesting case is FMO type 2, designed for defining rectangular slices as foreground(s) and background. These slices can be used to define regions of interest for encoding and transmission. Slices may overlap. However, the MBs can only belong to one slice. In the event of a packet carrying a whole slice gets dropped, H.264/AVC allows the transmission of redundant slices (RS). A RS can be encoded both differently as well as with the same encoding setting as the corresponding primary slice. The decoder is responsible for replacing a corrupted primary slice with its equivalent redundant representation. The latter error resilience technique is highly efficient for communication in noisy environments in the absence of a back channel. More details on FMO can be found in [32, 33].

Arbitrary Slice Ordering (ASO) [26] enables slices to be essentially transmitted independently of their order within a picture. As a result, they can be also decoded out of sequence, thus reducing the decoding delay at the decoder. ASO is particularly effective in environments where out-of-order delivery of a packet is possible such as the internet or wireless networks, or packet based networks in general.

As depicted in Figure 4, FMO, RS and ASO are only supported by the baseline and extended profiles of H.264/AVC.

#### 2.2.2.5 Data Partitioning, SP/SI Slices

The basic idea in data partitioning lies in the observation that not all bits in a bitstream carry equal information. On the contrary, data bits can be categorized according to their importance, with certain bits being more important than others. Data partitioning in H.264/AVC allows the partitioning of a normal slice in up to three parts. Each part can be paired accordingly with unequal error protection (UEP) during transmission. Data partition A (DP) contains the most important slice information such as MB types and motion vectors (MVs), and possible loss or corruption

of DP A, constitutes the remaining two partitions of no use. Second in importance comes DP B, which consists of intra-coded block patterns (CBPs) and I-block transform coefficients, while DP C incorporates inter CBPs and P-block coefficients. More detailed description can be found in [26, 34] along with recommended actions when partition loss is detected [26].

Switching-predictive/switching-intra (SP/SI) [35] are two new picture types introduced in the H.264/AVC design that allow the decoder to switch between two or more pre-encoded bit-streams. The decoder of H.264/AVC can be seen in the same figure with the encoder, in Figure 2. Switching-predictive (SP) pictures efficiently exploit motion-compensated prediction whereas switching-intra (SI) pictures can exactly reconstruct SP pictures. The switching between two bit streams using SI and SP pictures is illustrated in Figure7. Pre-encoded bitstreams are constructed from the same source sequence, but are of different bitrate and quality. Besides the obvious benefit of channel adaptation, this dual nature feature proves particularly efficient in terms of error resilience, especially in the presence of a feedback channel which enables the decoder to trigger the encoder to perform a bitstream switch, regaining in that way lost synchronization resulting from data losses or errors. Nevertheless, valuable bandwidth is preserved, since recovering from an error does not incorporate the transmission of an I-frame. The SP/SI scheme can be further used for operations such as fast-forward, reverse, etc.

As can be seen in Figure 4, SP/ SI slices and data partitioning are only supported by the extended profile of H.264/AVC.

### 2.2.3 High Efficiency Video Coding (HEVC)

In February 2011, a total of 27 proposals were submitted to JCT-VC and the subjective image evaluation was done in March. The evaluation results discussed in the April JCT-VC meeting at Dresden, Germany [36] showed that some proposals could reach the same visual quality as AVC

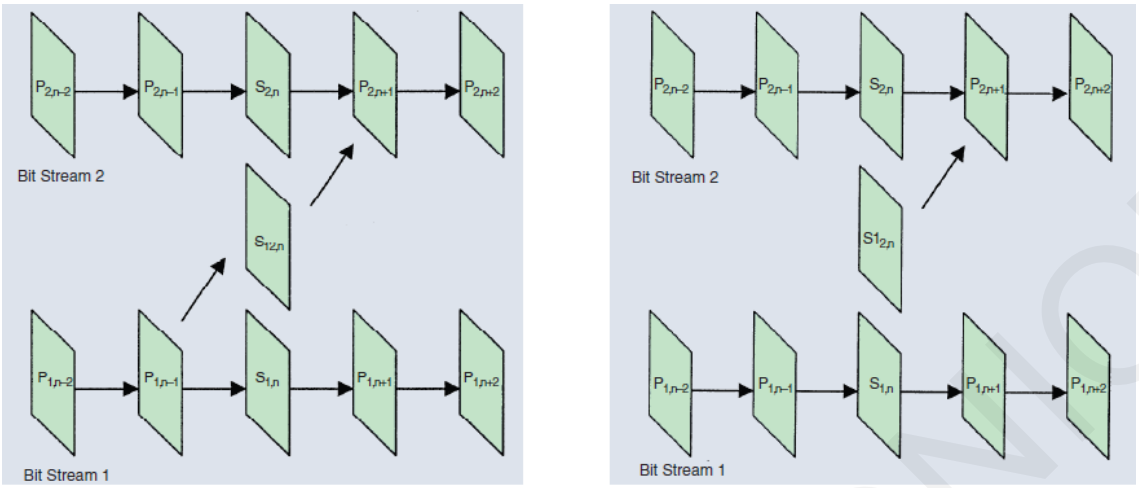



Figure 7: SP/SI Slices [29]. (a) SP slice is coded such that efficient switching between different precoded pictures becomes possible. (b) SI slice allows an exact match of a macroblock in an SP slice for random access and error recovery purpose.

at only half the bit rate in many of the test cases, at the cost of 2x to 10x increase in computational complexity; and some proposals achieved good subjective quality and bit rate results with lower computational complexity than the reference AVC High profile encodings. On January 2013, the ITU announced that HEVC had received first stage approval (consent) in the ITU-T Alternative Approval Process (AAP) [37]. Also, MPEG announced that HEVC had been promoted to Final Draft International Standard (FDIS) status in the MPEG standardization process [38]. The first version of HEVC was published in June 2013. The second version was completed and approved in 2014 and published in early 2015, while the third vesrion approved in April 2015 and adds the 3D Main profile. Currently, we have the fourth version that was approved in December 2016 and adds seven screen content coding extensions profiles, three high throughput extensions profiles, and four scalable extensions profiles.

HEVC is also known as H.265 and MPEG-H Part 2 [39, 40, 41]. It is expected to satisfy the ever increasing requirements for cost effective video encoding process by optimizing the video

quality, the compression efficiency, the spatial and temporal resolution, and finally the computational complexity.

HEVC development was motivated by the need of increased coding efficiency compared to state-of-the-art hybrid video codecs like H.264/AVC. This increased coding efficiency is achieved by introducing new coding tools as well by improving components already known from H.264/AVC. Figure 8 shows the block diagram of the basic HEVC design — as it is implemented in the HM 5.0 software codec. As can be observed, the main structure of the HEVC encoder resembles that of the H.264/AVC. On contrast, Figure 9 shows a general diagram of the main stages of the HEVC decoder.

New tools introduced in HEVC are variable size block partitioning using quadtrees for the purpose of prediction and transformation and an additional in-loop filter, namely, sample adaptive offset (SAO). Improvements include more intra-prediction angles, advanced motion vector prediction (AMVP), a new block merging mode that enables neighboring blocks to share the same motion information, larger transform sizes and a more efficient transform coefficient coding. HEVC incorporates only one entropy coder, which is basically CABAC from H.264/AVC. While objective test results of the latest HEVC reference software show bit rate reductions up to 35% compared to H.264/AVC high profile (HP), preliminary subjective test results indicate that a 50% bit rate reduction for a comparable visual quality can be achieved [42]. A preliminary subjective test was conducted by JCT-VC members to quantify the rate-distortion (RD) gain of the HEVC draft codec HM5.0 against a similarly configured H.264/AVC JM 18.3 codec. The results showed that an average RD-gain of 57.1% is achieved based on the subjective test data in the form of Mean Opinion Scores (MOSs). A more detailed objective and subjective evaluation of HM5.0 was reported in [43], which again suggested that HM5.0 has achieved the target of 50% RD gain over H.264/AVC and the actual savings can be even higher. Although these subjective tests and

evaluations were on random access coding configuration only and more comprehensive tests are still yet to be conducted, it is speculated that similar improvement may be achieved in other test conditions, and thus HEVC is very promising at achieving its initial RD performance target.

The HEVC standard version 1 defines three profiles, Main, Main 10 and Main Still Picture and it also contains provisions for additional profiles [44]. The Main profile allows for a bit depth of 8-bits per colour with 4:2:0 chroma sub-sampling which is the most common type of video used with consumer devices. The Main 10 profile allows for a bit depth of 8-bits to 10-bits per colour with 4:2:0 chroma sub-sampling [44]. A higher bit depth allows for a greater number of colours. 8-bits per colour allows for 256 shades per primary colour (a total of 16.78 million colours) while 10-bits per colour allows for 1024 shades per primary colour (a total of 1.07 billion colours) [45]. A higher bit depth allows for a smoother transition of colour which resolves the problem known as colour banding [45]. The Main 10 profile allows for improved video quality since it can support video with a higher bit depth than what is supported by the Main profile. Also, in the Main 10 profile 8-bit video can be coded with a higher bit depth of 10-bits, which allows improved coding efficiency compared to the Main profile [46]. The Main Still Picture profile allows for a single still picture to be encoded with the same constraints as the Main profile [44] As a subset of the Main profile the Main Still Picture profile allows for a bit depth of 8-bits per colour with 4:2:0 chroma sub-sampling [44, 45]. An objective performance comparison was done in April 2012 in which HEVC reduced the average bit rate for images by 56% compared to JPEG [47].

The new tool features are summarized in tabular form in [36] as a comparison with AVC and the main key features of the latest version of HEVC can be reviewed as follows:

· Wide-range variable block-size prediction, with block sizes ranging from 64x64 down to 8x8 pixels and the size of prediction blocks can be adaptively chosen by using recursive quad-tree partitioning [50].

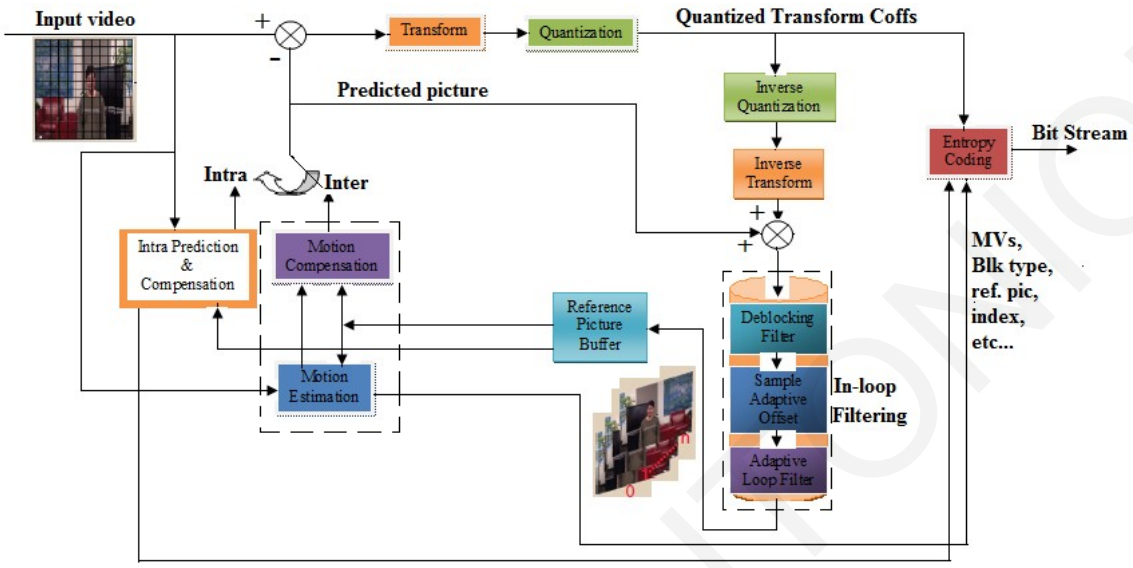



Figure 8: Block diagram of an HM-5.0 encoder. Example based on [48]

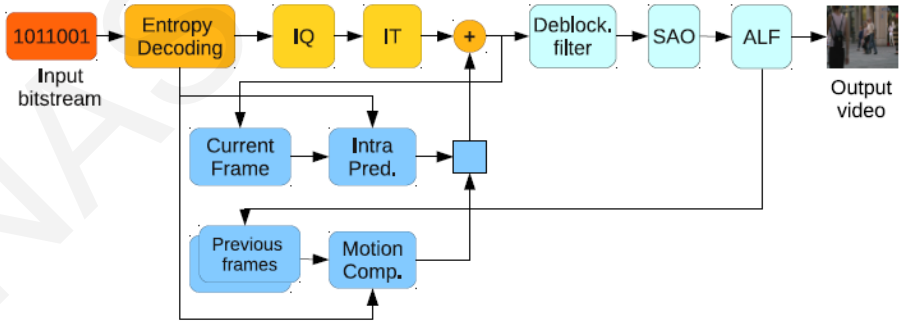



Figure 9: General block diagram of HEVC decoder. Inverse Quantization (IQ), Inverse Transformation (IT), Adaptive Loop Filter (ALF), Sample Adaptive Offset (SAO). Example based on [49].

· The block sizes used for DCT based residual coding is adapted to the characteristics of the residual signal by using a nested, so-called residual quadtree (RQT) for partitioning of a given prediction block [51]. It supports several integer transforms, ranging from 32x32 down to 4x4 pixels, as well as nonsquare transforms [52].

· Neighboring blocks can be merged into one region, such that motion information has to be transmitted only once for a whole region [53].

· Interpolation of fractional-sample positions for motion- compensated prediction is based on a fixed-point implementation of the Maximal-Order-Minimum-Support (MOMS) algorithm using an IIR/FIR filter [54].

· Improved mechanisms to support parallel encoding and decoding, including tiles and wavefront parallel processing (WPP) [52].

· Extensive in-loop processing on reconstructed pictures: In addition to the deblocking filter, sample adaptive offset (SAO) and adaptive loop filtering (ALF), a separable Wiener filter is applied within the coding loop [55].

· More intraprediction modes (most of which are directional), which can be done at several block sizes [52].

· The novel probability interval partitioning entropy (PIPE) coding scheme provides the coding efficiency and probability modeling capability of context based adaptive binary arithmetic coding (CABAC) [52]. In addition to the novel algorithms mentioned above, the HEVC proposals also includes the following features [45]; Simple quad-tree structure supporting large macroblock sizes of 32x32 and 64x64, Low complexity B pictures that only use integer

motion vectors for SKIP and DIRECT modes, and Spatially Varying Transform (SVT) that allows the position of the transform change arbitrarily within the macroblock.

The main aspects of the proposed coding tools are described in detail in the subsequent subsections.

#### 2.2.3.1 Quadtree-Based block partitioning

An important difference of HEVC compared to H.264/AVC is the frame coding structure. In HEVC each frame is divided into the basic processing unit scheme called Largest Coding Units (LCUs) [50]. LCUs can be recursively split into smaller Coding tree Units (CUs) using a generic quadtree segmentation structure (a nested quadtree structure) that indicates the subdivision of the CU for the purpose of prediction and residual coding. CUs can be further split into Prediction Units (PUs) used for intra- and inter-prediction and Transform Units (TUs) defined for transform and quantization as can be seen in Figure 10.

However, in H.264/AVC, each picture is partitioned into 16x16 macroblocks, and each macroblock can be further split into smaller blocks (as small as 4x4) for prediction [52]. As the picture resolution of videos increases from standard definition to HD and beyond, the chances are that the picture will contain larger smooth regions, which can be encoded more effectively when large block sizes are used.

This is the reason that HEVC supports larger encoding blocks than H.264/AVC, while it also has a more flexible partitioning structure to allow smaller blocks to be used for more textured and in general uneven regions. Hence, it has been designed to target ultra high resolution with higher frame rates compared to H.264/AVC. Taking this into consideration, HEVC has introduced a new partitioning image scheme concept based on a quadtree structure with a larger block size – of 64x64 Coding Unit (CU) that can be recursively further split into 4 CUs (Quadtree), which are

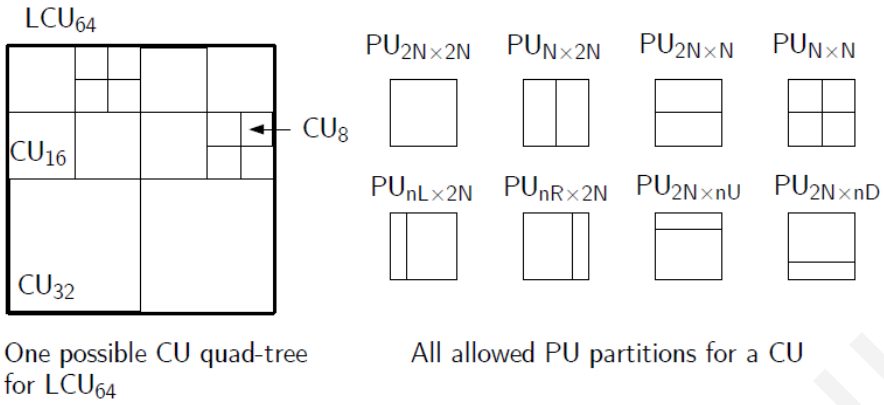



Figure 10: Block Partitioning in HEVC. Example based on [57]. Largest Coding Unit (LCU), Coding tree Unit (CU), Prediction Unit (PU).

used as the basic unit for intra- and inter-coding. The size of CUs can be as large as that of LCUs or as small as 8x8, depending on the picture content. Because of recursive quarter-size splitting, a content-adaptive coding tree structure comprised of CUs is created in HEVC [56].

Each CU may contain one or more PUs, and each PU can be as large as their root CU or as small as 4x4 in luma block sizes [53]. While an LCU can be recursively split into smaller and smaller CUs. The splitting of a CU into PUs is non recursive (it can be done only once). PUs can be symmetric or asymmetric. Symmetric PUs can be square or rectangular (non square) and are used in both intra prediction (uses only square PUs) and inter prediction. In particular, a CU of size 2Nx2N can be split into two symmetric PUs of size Nx2N or 2NxN or four PUs of size NxN [56] as shown in Figure 2b . Asymmetric PUs is used only for inter prediction. This allows partitioning, which matches the boundaries of the objects in the picture [56].

Finally, since HEVC applies a DCT-like transformation to the residuals to decorrelated data, Transform Unit (TU) is the basic unit for transform and quantization, which may exceed the size of PU, but not that of the CU. Only two TU modes are considered [58], signaled by transform unit size flag: i) If the Transform unit size flag = 0 $\rightarrow$ 2Nx2N (i.e., the same as the CU size), ii) Else if the Transform unit size flag = 1 $\rightarrow$ Square units of smaller size are considered: NxN if PU

splitting is symmetric or N/2xN/2 if PU splitting is asymmetric. So, the size and the shape of the TU both depends on the size of the PU. The size of square-shape TUs can be as small as 4x4 or as large as 32x32. Nonsquare TUs can have sizes of 32x8, 8x32, 16x4, or 4x16 luma samples. Each CU may contain one or more TUs and each square CU may split into smaller TUs in a quad-tree segmentation structure. Figure 2(a-c) shows some examples of partitioning in HEVC.

For transform sizes larger than 8x8, the proposal [59] utilizes truncated transforms where only the 8x8 low frequency coefficients are calculated. This results in a significant computational complexity saving, and implies that only 4x4 and 8x8 quantization kernels are used. In some cases, this may improve visual quality by avoiding ringing artifacts. Other features like a novel technique called Spatially Varying Transform (SVT), where the position of the transform block within the macroblock is not fixed but can be varied is proposed [59, 60].

HEVC introduced tiles as a means to support parallel processing, with more flexibility than normal slices in H.264/AVC but of considerably lower complexity than flexible macroblock ordering (FMO). Tiles are specified by vertical and horizontal boundaries with intersections that partition a picture into rectangular regions [52]. To support parallel processing, each slice in HEVC can be subdivided into smaller slices called entropy slices. Each entropy slice can be independently entropy decoded without reference to other entropy slices. Therefore, each core of a CPU can handle an entropy-decoding process in parallel. However, they both come with a performance penalty since prediction dependencies are broken across boundaries and the statistics used in entropy coding have to be initialized for every slice/tile. To avoid these problems, Wavefront Parallel Processing (WPP) is supported in HEVC [50, 56]. Wavefront processing is a way to achieve parallel encoding and decoding without breaking prediction dependencies and using as much context as possible in entropy encoding.

#### 2.2.3.2 Inter and Intra-Prediction Coding

Predictive coding is the primary tool employed in current video compression technologies, and it is very efficient for removing the correlation between pixels in both spatial and temporal domain. Pixel values to be coded are predicted from already coded and reconstructed adjacent pixel values, and only small prediction errors are coded and the bit rate reduction depends on the correlation coefficient. In HEVC, for generating the prediction signal at subsample positions, one of the proposals [50] uses an interpolation method based on families of so-called maximal-order-minimal-support (MOMS) basis functions. This requires, however, an additional pre-filtering step on the reference picture before the actual interpolation. More details on the subsample interpolation scheme can be found in [50, 54].

**Inter-Prediction Coding**

Inter-prediction explores temporal redundancy between frames to save coding bits. By using motion compensated prediction, the best matching position of current block is found within the reference picture so that only prediction difference needs to be coded.

The enhancements of inter-prediction introduced in HEVC, compared with H.264/AVC, are described as variable PU size motion compensation, improved subpixel interpolation, and motion parameter encoding and improved skip mode [52].

Each PU coded using inter prediction, has a set of motion parameters, which consists of a motion vector, a reference picture index, and a reference list flag. Intercoded CUs can use symmetric and asymmetric motion partitions (AMPs). AMPs allow for asymmetrical splitting of a CU into smaller PUs. AMP can be used on CUs of size 64x64 down to 16x16 and improves the coding efficiency since it allows PUs to more accurately conform to the shape of objects in the picture without requiring further splitting [61].

The existing sub-pel interpolation method has been improved by replacing the fixed filters by the adaptive ones or by redesigning the filter coefficients [36]. Several proposals adaptively update interpolation filters by the least squares method in order to minimize the prediction errors of each video frame.

The accuracy of motion compensation in HEVC is 1/4 pel for luma samples. To obtain the non integer luma samples, separable one-dimensional eight-tap and seven-tap interpolation filters are applied horizontally and vertically to generate luma half-pel and quarter-pel samples, respectively [48].

To further improve inter prediction efficiency, finer fractional motion prediction and better motion vector prediction were proposed. Increasing the resolution of the displacement vector from 1/4-pel to 1/8-pel to obtain higher efficiency of the motion compensated prediction is also suggested in [50]. Furthermore, HEVC introduces a technique called motion merge. For every inter-coded PU, the encoder can choose between using explicit encoding of motion parameters, motion merge mode, or the improved SKIP mode [50].

**Intra-Prediction Coding**

Like H.264/AVC, HEVC uses block-based intra-prediction to take advantage of spatial correlation within a picture. HEVC follows the basic idea of H.264/AVC intra prediction but makes it far more flexible. HEVC has 35 luma intra prediction modes compared with nine in H.264/AVC. Furthermore, intra prediction can be done at different block sizes, ranging from 4x4 to 64x64 (whatever size the PU has) with 33 different directional orientations . Figure 11 shows the luma intra prediction modes of HEVC and H.264/AVC. The number of supported prediction modes

varies based on the PU size, (see Table 2) [56]. As mention before, intra-prediction coding operates according to the PU size, and also previously decoded boundary samples from spatially neighboring PUs are used to form the prediction signal.

HEVC also includes a planar intra prediction mode, which is useful for predicting smooth picture regions. In planar mode, the prediction is generated from the average of two linear interpolations (horizontal and vertical) [62]. Also, a DC prediction (a flat surface with a value matching the mean value of the boundary samples) can also be used. For chroma, the horizontal, vertical, planar, and DC prediction modes can be explicitly signaled, or the chroma prediction mode can be indicated to be the same as the luma prediction mode and, as a special case to avoid redundant signaling, when one of the first four choices is indicated and is the same as the luma prediction mode, the Intra-Angular [63] mode is applied instead.

Table 2: Luma intra prediction modes supported for different PU sizes

| **PU Size** | **Intra prediction Modes** |
|---|---|
| 4x4 | 0-16, 34 |
| 8x8 | 0-34 |
| 16x16 | 0-34 |
| 32x32 | 0-34 |
| 64x64 | 0-2, 34 |

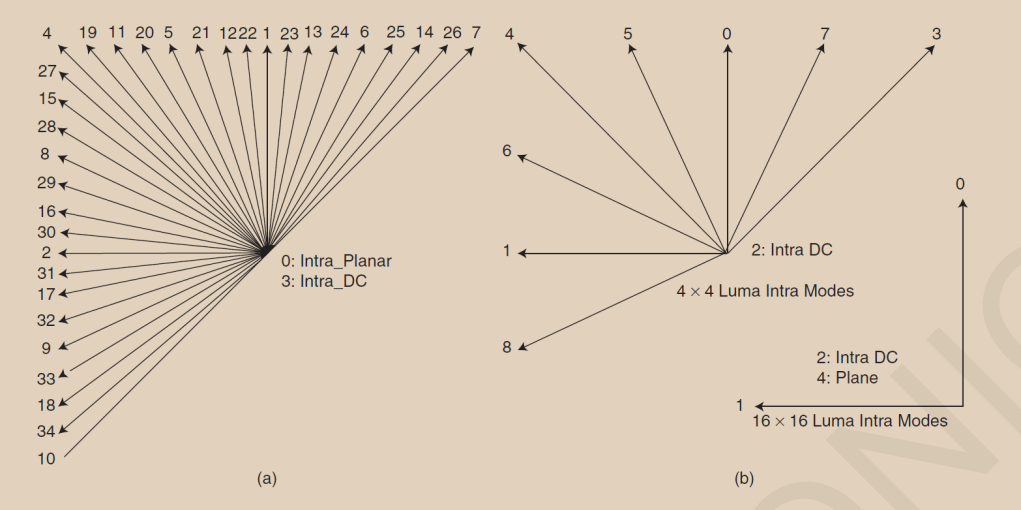



Figure 11: Luma intra prediction modes of (a) HEVC and (b) H.264/AVC. Example based on [52]

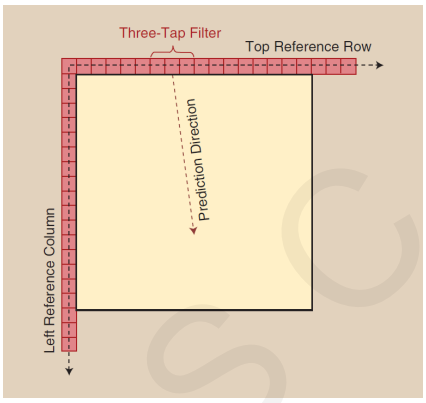



Figure 12: An example of using the MDIS filter for intra-prediction. Example based on [52]

To improve the performance of intra prediction, mode-dependent intra smoothing (MDIS) is used for some intra modes. MDIS involves applying a simple low-pass finite impulse response filter with coefficients (1, 2, 1)/4 to the samples being used for prediction. This smoothing of the reference signal improves the prediction performance, especially for large PUs. MDIS is enabled based on the PU size and intra mode. In general, MDIS is used for large PU sizes and directional modes, except in the horizontal and vertical modes. An example of an MDIS applications in shown in Figure 12.

Generally, MDIS filters the source pixels used for intra prediction with different degrees of smoothing, with the degree determined by the intra prediction direction [64]. Mode dependent

| Index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | … | 33 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mode | VER | HOR | DC | VER-8 | VER-4 | VER+4 | VER+8 | HOR-4 | HOR+4 | HOR+8 | VER-6 | VER-2 | VER+2 | VER+6 | HOR-6 | HOR-2 | HOR+2 | HOR+6 | … | |
| 4x4 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8x8 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16x16 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 32x32 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

Figure 13: Table of MDIS Filter Index. Taken from [64]

intra smoothing is a technique to selectively apply intra smoothing to the neighbor pixels of intra prediction. For each prediction block, the codec then select either filter 0 (no filter) or filter 1 or filter 2 according to the LUT shown in Figure 13. The filter index selected depends on the block size and prediction mode.

#### 2.2.3.3 Transform Coding

Similar to H.264/AVC, HEVC applies a DCT-like integer transform on the prediction residual. As can be seen in the block diagram of HEVC structure in Figure 8, HEVC includes transforms that can be applied to blocks of sizes ranging from 4x4 to 32x32 pixels [65]. HEVC also supports transforms on rectangular (non-square) blocks where the row transform and column transforms have different sizes. The integer transforms used in HEVC are better approximations of the DCT than the transforms of H.264/AVC. The basis vectors of the HEVC transforms have equal energy. So as a result, there is no need to compensate for the different norms, as in H.264/AVC.

HEVC also incorporates a 4x4 discrete sine transform (DST), which is used for blocks coded with some directional intra prediction modes. When using intra prediction, the pixels close to the ones used for prediction (i.e. near the top or left boundaries) will usually be predicted more accurately than the pixels further away. Therefore, the residuals tend to be larger for pixels away from the boundaries. The DST will usually be better at encoding these kind of residuals, because the DST basis functions start low and increase, compared with the DCT basis functions that start high and decrease [66].

#### 2.2.3.4 Quantization

In the MC/DCT hybrid video coding schemes, uniform scalar quantization schemes are usually utilized to quantize the transform coefficients, and the quantization step size, which determines the quantization strength, is transmitted to the receiver. To achieve better quantization, optimized quantization decision at the macroblock level and at different coefficient positions are proposed. More recently, $HEVC$ [48] gives an improved, more efficient Rate Distortion Optimized Quantization ($RDOQ$) implementation. Adaptive Quantization Matrix Selection ($AQMS$) [67], a method deciding the best quantization matrix index, where different coefficient positions can have different quantization steps, is proposed to optimize the quantization matrix at a macroblock level. The quantization weighting matrix, which is a controlling element can be either uniquely defined and sent to the decoder as coding parameters, or substituted by a default one. To match the statistics of the transform coefficient distribution, adaptive selection of the quantization weighting matrix is proposed [36].

#### 2.2.3.5 Coding

After transformation and quantization, entropy coding is applied to code all the syntax elements and quantized transform coefficients in video coding techniques. In H.264/AVC, context-adaptive variable-length coding (CAVLC) is the base entropy coder, and context-adaptive binary arithmetic coding (CABAC) is optionally used in the main and high profiles. CABAC can provide better coding efficiency than CAVLC because of its arithmetic coding engine and more sophisticated context modeling. While CABAC improves the coding efficiency, it increases coding complexity. This is more pronounced at higher bit rates [small quantization parameters (QPs)], where the transform coefficient data have a dominant role in encoded bit streams. In HEVC, to improve the worst-case throughput, the codec uses a higher-throughput alternative mode for coding

transform coefficient data. Figure 12 illustrates the block diagram of HEVC entropy coding. As can be seen, there are two modes of HEVC entropy coding: high-efficiency binarization (HEB) and high-throughput binarization (HTB) [68]. The HEB mode is entirely CABAC based while the HTB mode is partially based on the well-known CAVLC residual coding module. HTB is intended to serve as the high-throughput mode of HEVC, and its use is signaled at slice level (one bit identifier indicating whether HTB is used). In HTB mode, all syntax elements except the residual coefficients are coded using CABAC while the residual coefficients are coded using CAVLC. Using this harmonized design, HEVC entropy coding uses the best features of both CABAC and CAVLC coding (i.e., high efficiency and low complexity, respectively).

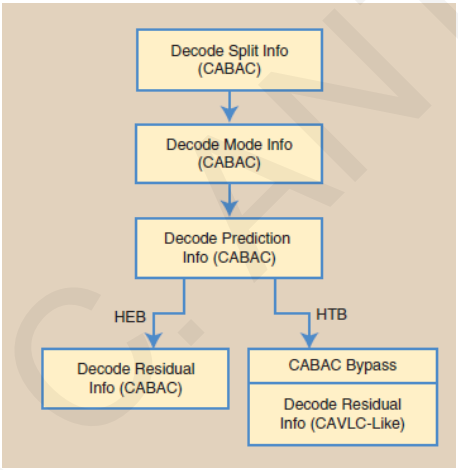



Figure 14: HEVC entropy coding. Example based on [52]. Context-based adaptive binary arithmetic coding (CABAC), high-efficiency binarization (HEB),high-throughput binarization (HTB), context-adaptive variable-length coding (CAVLC)

#### 2.2.3.6 In-loop Filtering

Referring to Figure 8, loop filtering is applied after inverse quantization and transform, but before the reconstructed picture is used for predicting other pictures through motion compensation. The name loop filtering reflects the fact that filtering is done as part of the prediction loop rather

than post processing. HEVC employs a deblocking filter similar to the one used in H.264/AVC but also expands an in-loop processing by introducing two new tools: Sample Adaptive Offset filter (SAO) and Adaptive Loop Filter (ALF). These techniques are intended to undo the distortion introduced in the main steps of the encoding process (prediction, transform, and quantization). By including filtering as part of the prediction loop, the pictures will serve as better references for motion-compensated prediction since they have less encoding distortion.

**Interpolation Filtering (IF)**

Similar to H.264/AVC, the proposal for HEVC utilizes a translational motion model with motion vectors having quarter pixel accuracy. Here to obtain the samples at fractional pixels, two sets of interpolation filters are utilized. The first set of interpolation filters is referred to as the Directional Interpolation Filter (DIF) and is used for all the quarter-pixel positions. Directional filters use either 6-tap or 12-tap filter for each sub-pixel, and has significantly less complexity compared to H.264/AVC interpolation, where 6x6 separable filters are utilized [67]. The second set of interpolation filters consists of Separable Filters (SF) where interpolated samples are calculated by first applying a 6-tap filter horizontally and then vertically.

**Deblocking Filter (DBF)**

Blocking is known as one of the most visible and objectionable artifacts of block-based compression methods. For this reason, in H.264/AVC, low-pass filters are adaptively applied to block boundaries according to the boundary strength. This improves the subjective and objective quality of the video. HEVC uses an in-loop deblocking filter similar to the one used in H.264/AVC. In HEVC, there are several kinds of block boundaries, such as CUs, PUs, and TUs). The set of boundaries that may be filtered in HEVC is the union of all of these boundaries (except for 4x4

blocks, which are not filtered to reduce complexity). For each boundary, a decision is made to turn the deblocking on or off and whether to apply strong or weak filtering. This decision is based on the pixel gradients across the boundary and thresholds derived based on the QP in the blocks.

**Sample Adaptive Offset filter (SAO)**

The sample adaptive offset (SAO) filter is applied in between the deblocking filter and the ALF. SAO is a new coding tool introduced in HEVC, which involves classifying pixels into different categories based on either intensity or edge properties and adding a simple offset value either band offset (BO) or edge offset (EO), to the pixels in each category in a region to reduce distortion.

BO classifies all pixels of a region into multiple bands, with each band containing pixels in the same intensity interval [59]. The intensity range is divided into 32 equal intervals from zero to the maximum intensity. The 32 bands are divided into two groups and only offsets in one group are transmitted. The encoder decides which group of bands to apply SAO, so 16 offsets will be encoded in the bit stream [69].

EO uses four 1-D 3-pixel patterns for pixel classification with consideration of edge directional information, as illustrated in Figure 15. Each region of a picture can select one pattern to classify pixels into multiple categories by comparing each pixel with its two neighboring pixels. Each pixel can be classified as a peak (if it is greater than two neighbors), valley (if it is less than the two neighbors), edge (if it is equal to one neighbor, categories 2 and 3), or none of these. Four offset values will be calculated for these four categories [7]. The encoder can choose to apply either BO or EO to different regions of a picture and also signal that neither BO nor EO is used for a region.

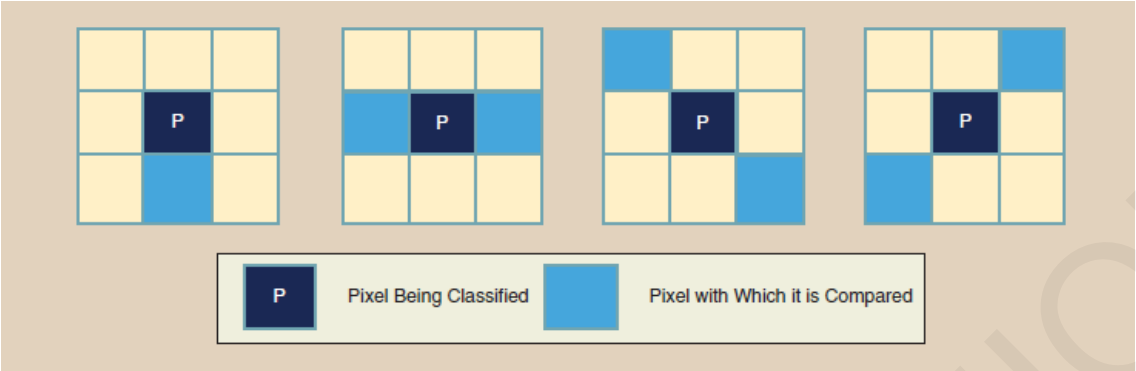



Figure 15: Patterns used in EO mode. Each region of a picture can select one pattern to classify pixels into multiple categories by comparing each pixel with its two neighboring pixels. Example based on [52]

**Adaptive Loop Filter (ALF)**

An ALF is applied to the reconstructed signal after the deblocking filter and SAO. This filter is adaptive in the sense that the coefficients are signalled in the bit stream and can therefore be designed based on image content and distortion of the reconstructed picture.

Two different ALF techniques were adopted so far: Quadtree-based Adaptive Loop Filter (QALF) and Block-based Adaptive Loop Filter (BALF) [67]. The main idea of Quad- tree ALF (QALF) is to signal the on/off decision of filtering through a quad-tree partition process. QALF is adopted, and improved by providing multiple filters for adaptation.

There are two modes that can be used for applying different filters to different pixels within each picture: region based adaptation (RA) and block-based adaptation (BA). In RA mode, the picture is divided into 16 regions of equal size. These regions can be merged, and each region remaining after merging will have its own filter (with a unique set of coefficients). In BA mode, 4x4 blocks are classified into 1 of 16 categories based on edge activity and direction. These categories can be merged, and in the end, one filter will be designed for each of the categories left

after merging. The filter coefficients for each region can be calculated based on the autocorrelation and cross-correlation of the original pixels and the reconstruction pixels in the region (using Wiener–Hopf equations) [70].

The ALF can be enabled or disabled for different picture areas based on the partitioning of LCUs into CUs (in a quad-tree segmentation structure).

### 2.2.4 HEVC released versions

Additionally to $HEVC$ version 1 components that were described earlier, Version 2 of HEVC was approved in October 2014. This version adds 21 range extensions ($RExt$) profiles, two scalable extensions profiles ($SHVC$), and one multi-view ($MV-HEVC$) profile [71, 72]. All of the inter frame range extension profiles have an Intra profile [71]. $HEVC-RExt$ was developed to support non-4:2:0 colour formats, e.g. 4:4:4 and 4:2:2, and high bit-depth video, e.g. up to 16-bit. Even though most screen content is captured in the 4:4:4 colour format, $HEVC$ version 1 does not support it. Thus, in this version more attention was given to the coding of screen content. The coding tools that improved the coding efficiency for screen content in $HEVC-RExt$ compared with $HEVC$ version 1 include improvements to transform skip mode, Residual Differential Pulse Code Modulation ($RDPCM$) and Cross-Component Prediction ($CCP$). The new version of transform skip mode supports block sizes up to 32x32 (versus only 4$times$4 support in version 1). Transform skip rotation, allows the encoder to indicate a rotation of residual data for 4x4 transform skip blocks while Transform skip context enabling, uses a separate context for entropy coding the indication of which blocks are coded using transform skipping [73]. $RDPCM$ allows a vertical or horizontal spatial-predictive coding of residual data in transform skip and transform-quantization bypass blocks which can be selected for use in intra blocks, inter blocks, or both [73].

$CPP$ improves coding efficiency using prediction between the chroma/luma components. In bit rate it can be up to 7% for YCbCr 4:4:4 video [73].

Version 3 of $HEVC$ approved in April 2015 adds one 3D profile: 3D Main [73] while the February 2016 draft of the screen content coding extensions added seven Screen Content Coding ($SCC$) extension profiles, three high throughput extensions profiles, and four scalable extension profiles [71]. The new coding tools of the 3rd version are adaptive color transform, adaptive motion vector resolution, intra block copying and palette mode [74].

In Adaptive Colour Transform ($ACT$) a CU-level adaptation is used to convert residual to different colour spaces. More precisely, an image block in the RGB colour space can be coded directly or it can be converted adaptively to the *YCoCg* colour space during coding. Due to the fact that much screen content uses the RGB colour space, removing inter-colour component redundancy is important for efficient coding.

Using adaptive motion vector resolution, a slice-level control is enabled to switch the motion vectors between full-pel and fractional resolutions.

$HEVC\ SCC$ introduces a new CU mode in addition to the conventional intra and inter modes, referred to as intra block copy ($IBC$). When a CU is coded in $IBC$ mode, the PUs of this CU find similar reconstructed blocks within the same picture.

Palette mode enumerates different colour values that may exist in many blocks in screen contents. Then for each sample, sends an index to indicate to which colour it belongs. In some cases, palette mode can be more efficient than the prediction-then-transform representation.

The current version 4 approved in December 2016 includes seven $SCC$ extension profiles, three high throughput extension profiles, and four scalable extension profiles [75].

An overview of $HEVC$ extension can be found in [76].

#### 2.2.4.1 Scalability Extension

As mentioned in the previous subsection, Scalable Video Coding ($SVC$) is included in the second version of $HEVC$, which has been finalized in October 2014. The Scalabilty extension of HEVC ($SHVC$) has been designed as a high-level syntax only extension to allow reuse of existing decoder components. It's important to note that SHVC minimizes implementation complexity by enabling repositioning of multiple single-layer HEVC cores to achieve efficient scalable coding. Thus, SHVC achieves high scalable coding efficiency without requiring any block-level coding logic changes to the single-layers HEVC cores [76], [77].

The objective of the $SVC$ standardization is to enable the encoding of a high-quality video bit stream that contains one or more subset bitstreams. These bitstreams can themselves be decoded with a complexity and reconstruction quality similar to that achieved using the existing Scalable High efficiency Video Coding $SHVC$ design, having the same quantity of data as in the subset bit stream. More specifically, $SVC$ provides a mechanism for coding video in multiple layers, where each layer represents a different quality representation of the same video scene. The base layer ($BL$) is the lowest quality representation while one or more enhancement layers ($ELs$) may be coded by referencing lower layers and provide enhanced video quality. Having more than one layers to encode, encoding of $SVC$ typically costs more bits to achieve the same video quality compared to non-scalable video coding [76].

An $HEVC$ decoder offers the option to select which layer of the encoded bitstream will decode.

$SHVC$ offers three levels of scalability like $H.264/AVC$: (a) spatial, (b) temporal, and (c) quality ($SNR$). Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) or frame

rate (temporal resolution), respectively. In quality scalability, the substream provides the same spatio-temporal resolution as the complete bit stream, but with a lower fidelity – where fidelity is often informally referred to as signal-to-noise ratio (SNR) [78].

Another feature of the $SVC$ is error resiliency [76]. In cases where network errors will lead to loss of $EL$ data, the resulted video will have much less quality degradation compared to non-scalable data. In cases where the encoder has the capability to switch from a lower layer to a higher layer in the bitstream, a video decoder can fully recover from an earlier loss in the higher layer.

Applications that use $SVC$ can benefit from the ability to adapt the video bitstream according to the requirements of the decoders and conditions of the networks.

### 2.2.5 Performance Comparison of HEVC vs H.264

The most widely used video coding format in 2016 was H.264 while the current state-of-the-art, HEVC, is expected to dominate the market in the near future. Tables 3 and 4 summarize the average bitrate savings for equivalent PSNR to validate the performance gains of HEVC compared to H.264/AVC, for both entertainment videos and common carotid ultrasound videos. In both studies, the Bjøntegaard measurement method [79] for calculating objective differences between rate-distortion curves was used as evaluation criterion. In [80], the test sequence included five classes of different video and video resolution ranged between 416×240 to 2560×1600, while the frame rate range is 24 Hz to 60 Hz. The video duration were 5s to 10s. As can be seen in Table 3, HEVC employing Main Profile achieved up to 35.4% average bitrate savings compared to H.264/AVC employing High Profile, for equivalent PSNR for entertainment videos. It's important to note that applications achieved up to 49.3% average bitrate savings for entertainment application scenario based on subjective assessment results. Additionally, in [81], the test sequence included

Table 3: Average Bit Rate Savings for equivalent PSNR for Entertainment Videos

| ***Encoding*** | ***Bitrate Savings Relative to*** | | | |
|---|---|---|---|---|
| | *H.264/MPEG-4 AVC HP* | *H.263 CHC* | *MPEG-4 ASP* | *MPEG-2/ H.262 MP* |
| ***HEVC MP*** | *35.4%* | *63.7%* | *65.1%* | *70.8%* |
| ***H.264/MPEG-4 AVC HP*** | - | *44.5%* | *46.6%* | *55.4%* |
| ***H.263 CHC*** | - | - | *3.9%* | *19.7%* |
| ***MPEG-4 ASP*** | - | - | - | *16.2%* |

Table 4: Average Bit Rate Savings for equivalent PSNR for CCA Ultrasound Videos

| ***Encoding*** | ***Bitrate Savings Relative to*** | | | |
|---|---|---|---|---|
| | *H.264/MPEG-4 AVC HP* | *H.263 CHC* | *MPEG-4 ASP* | *MPEG-2/ H.262 MP* |
| ***HEVC MP*** | *33.2%* | *54.6%* | *58.3%* | *71%* |
| ***H.264/MPEG-4 AVC HP*** | - | *32.3%* | *37.7%* | *56.8%* |
| ***H.263 CHC*** | - | - | *7.5%* | *32.4%* |
| ***MPEG-4 ASP*** | - | - | - | *27.4%* |

ten ultrasound videos of the common carotid artery with video resolution 560x$\times$416 at frame rate equal to 50 Hz. The video duration 8s to 10s. As depicted in Table 4, HEVC employing Main Profile achieved up to 33.2% average bitrate savings compared to H.264/AVC employing High Profile, for equivalent PSNR for common carotid ultrasound videos. Consequently, the trend is that using HEVC encoder we have important bitrate gains for equivalent PSNR in any video modality.

### 2.3 Video Quality Assessment

The transmission of digital video on a daily base necessitate a ubiquitous service and a broad range of applications rely on it. Globally, IP video traffic was up to 73% in 2016 and will be up to 82% of all consumer internet traffic by 2021 as Cisco Visual Networking Index forecasts [82].

As a video passes through various stages, like acquisition, digitization, encoding, transmission, and decoding until its consumption by a viewer, many distortions may be introduced affecting the perceptual quality of the video. This necessitates the evaluation of the visual quality of a video.

The ability to quantify the quality of a video is the most significant requirement for emerging medical video transmission systems. Diagnostic validation requires an accurate assessment of the diagnostic capacity of the transmitted medical video [25]. Different from conventional video-streaming quality requirements, the quality of transmission systems is measured in terms of their diagnostic yield, thus depending on their ability to provide for a reliable and dependable diagnosis.

To evaluate the diagnostic yield of a streamed medical video both objective and subjective (clinical) approaches exist.

### 2.3.1 Objective Medical Video Quality Assessment

Objective video quality assessment ($VQA$) algorithms seek to make predictions of the visual quality in the absence of any human subjective judgment. They are divided in three categories (a) full reference, (b) blind or no reference and (c) reduced reference base on the amount of the available knowledge that they use.

#### 2.3.1.1 Full Reference Algorithms

Image quality assessment has seen significant growth over the last years and as a result today's most established VQAs are extensions of algorithms originally designed for image quality assessment. Such examples are the Peak Signal-to-Noise Ratio (PSNR) and the average Structural SIMilarity Index (SSIM) [83].

These are full reference algorithms since they require a reference video with respect to which the quality of the decoded video is assessed. Such algorithms reach a satisfactory level of reliable performance since they demonstrate high correlations with human subjective judgments.

**PSNR**

PSNR utilizes the Mean Square Error (MSE) between the original and transmitted videos on a frame basis. PSNR is one of the most established metrics for both image and video quality assessment and is often used as the benchmark metric for measuring the objective performance of transmitted video. Given a reference, uncompressed video **I** and a distorted (i.e. following transmission) video **K**, with mxn spatial dimensions, the PSNR is computed as follows:

$$\mathrm{MSE} = \frac{1}{\mathrm{m\,n}} \sum_{\mathrm{i}=0}^{\mathrm{m}-1} \sum_{\mathrm{j}=0}^{\mathrm{n}-1} [\mathrm{I}(\mathrm{i},\mathrm{j}) - \mathrm{K}(\mathrm{i},\mathrm{j})]^2 \tag{1}$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{2}$$

**SSIM**

Another popular image quality assessment metric that has been extended to address video quality assessment is Structural SIMilarity (SSIM) [83]. Similarly to PSNR and all image quality assessment algorithms used for *VQA*, the index is computed separately for each frame and the average yields the video quality assessment index. The following equations are abstracted from [83].

$$S(\mathrm{x},\mathrm{y}) = \mathrm{l}(\mathrm{x},\mathrm{y}) \,.\mathrm{c}(\mathrm{x},\mathrm{y}) \,.\mathrm{s}(\mathrm{x},\mathrm{y}) = \frac{4\mu_\chi \mu_\mathrm{y} \sigma_{\chi\mathrm{y}}}{\left(\mu_\chi^2 + \mu_\mathrm{y}^2\right)\left(\sigma_\chi^2 + \sigma_\mathrm{y}^2\right)} \tag{3}$$

where $l(x,y)$, $c(x,y)$, and $s(x,y)$ are the luminance, the contrast and the structure comparison measures respectively.

$$\mu_\chi = \overline{\chi} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \chi_\mathrm{i}, \qquad \mu_\mathrm{y} = \overline{\mathrm{y}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{y}_\mathrm{i} \tag{4}$$

$$\sigma_\chi^2 = \frac{1}{\mathrm{N}-1} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x}_\mathrm{i} - \overline{\mathrm{x}})^2, \qquad \sigma_\mathrm{y}^2 = \frac{1}{\mathrm{N}-1} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y}_\mathrm{i} - \overline{\mathrm{y}})^2 \tag{5}$$

$$\sigma_{\chi y}=\frac{1}{\mathrm{N}-1}\sum_{\mathrm{i}-1}^{\mathrm{N}}\left(\chi_{\mathrm{i}}-\overline{\chi}\right)\left(\mathrm{y}_{\mathrm{i}}-\overline{\mathrm{y}}\right) \tag{6}$$

To overcome the computation stability problem observed when $\left(\mu_{\chi}^{2}+\mu_{\mathrm{y}}^{2}\right)$ or $\left(\sigma_{\chi}^{2}+\sigma_{\mathrm{y}}^{2}\right)$ is close to 0, especially for flat image regions, equation 3 is modified so that it now reads:

$$\mathrm{SSIM}\left(\chi,\mathrm{y}\right)=\frac{\left(2\mu_{\chi}\mu_{\mathrm{y}}+\mathrm{C}_{1}\right)\left(2\sigma_{\chi\mathrm{y}}+\mathrm{C}_{2}\right)}{\left(\mu_{\chi}^{2}+\mu_{\mathrm{y}}^{2}+\mathrm{C}_{1}\right)\left(\sigma_{\chi}^{2}+\sigma_{\mathrm{y}}^{2}+\mathrm{C}_{1}\right)} \tag{7}$$

where $C_1$ and $C_2$ are two constants which are given by:

$$\mathrm{C}_{1}=\left(\mathrm{K}_{1}\mathrm{L}\right)^{2}\text{ and }\mathrm{C}_{2}=\left(\mathrm{K}_{2}\mathrm{L}\right)^{2} \tag{8}$$

For the two new constants added, $\mathrm{L}=255$ for 8 bits/pixel gray scale images and $\mathrm{K}_1$ and $\mathrm{K}_2$ are set to $\mathrm{K}_1=0.01,\ \text{and}\ \mathrm{K}_2=0.03$ respectively in [83], values which are also adopted here.

**MOVIE**

Motion-based video integrity evaluation ($MOVIE$) is a full reference VQA algorithm where both spatial and temporal (and spatio-temporal) aspects of distortion assessment are taken into consideration [84]. That means that video quality is evaluated not only in space and time, but also in space-time, by evaluating motion quality along computed motion trajectories. The performance of the resulting algorithm shows that such distortions contribute significantly to the perception of video quality.

The performance of MOVIE algorithm, proves that temporal and spatial distortions contribute significantly to the perception of video quality, and is in agreement with physiological findings, showing that it has high correlation with subjective mean opinion scores.

**VMAF**

Video Multi-Method Assessment Fusion ($VMAF$) is a new VQA proposed by Netflix [85]. VMAF predicts subjective quality by combining multiple elementary quality metrics and seeks

to reflect the viewer's perception of their streaming quality. This metric is focused on quality degradation due to compression an rescaling. The basic rationale is that each elementary metric may have its own strengths and weaknesses with respect to the source content characteristics, type of artifacts, and degree of distortion. VMAF is using a Support Vector Machine ($SVM$) regressor which assigns weights to each elementary metric. The final metric can preserve all the strengths of the individual metrics, and deliver a more accurate final score. The machine-learning model is trained and tested using the opinion scores obtained through a subjective experiment taken from NFLX video dataset.

The current version of the VMAF algorithm incorporates the following elementary metrics:

•**Visual Information Fidelity:** $VIF$ is an image quality metric that integrates natural scene statistics ($NSS$) modeling together with image/video degradation and human visual system ($HVS$) models in their algorithmic design. Natural images are modeled in the wavelet domain using Gaussian scale mixtures ($GSM$) [86]. In VMAF, they adopt a modified version of VIF where the loss of fidelity in each scale is included as an elementary metric.

•**Detail Loss Metric:** $DLM$ is an image quality metric based on the rationale of separately measuring the loss of details which affects the content visibility, and the redundant impairment which distracts viewer attention. The original metric combines both DLM and additive impairment measure ($AIM$) to yield a final score [87]. However, in VMAF, they only adopt the DLM as an elementary metric.

•**Motion:** This is a simple measure of the temporal difference between adjacent frames. This is accomplished by calculating the average absolute pixel difference for the luminance component.

#### 2.3.1.2 No Reference Algorithms

A reliable no reference/blind VQA is needed in cases where no reference signal is available. Most of the blind VQA algorithms generally follow one of the three following approaches:

•**Distortion-specific:** Algorithms quantify one or more distortions such as blur and/or block-effect distortion and score the image accordingly.

•**Training-based:** Algorithms train a model to predict the image quality score on a number of features extracted from the image.

•**Natural scene statistics:** $NSS$ algorithms rely on the hypothesis that natural world images belong to a subspace of the space of all possible images and seek to find a distance value between the examined image and the subspace of natural images.

**BLIINDS-II**

Blind image integrity notator DCT statistics ($BLIINDS-II$) algorithm is based on a non-distortion-specific approach that inherits the advantages of the $NSS$ aproach to VQA [88]. After a set of features is extracted from an image, the algorithm is trained using these features on a simple Bayesian approach, and the learning model then predicts perceptual image quality scores. For feature extraction, a generalized NSS based model of local DCT coefficients is estimated. The model parameters are used to design features suitable for perceptual image quality score prediction. The NSS features are used by the Bayesian probabilistic inference model to quantify the visual quality. The algorithm requires minimal training and its results are correlated highly with human visual perception of quality and yields highly competitive performance.

**VIIDEO**

Video Intrinsic Integrity and Distortion Evaluation Oracle ($VIIDEO$) algorithm is very promising since it is a new no reference VQA approach that does not require the use of any additional information other than the video being quality evaluated [89]. VIIDEO does not require in advance any distortion knowledge like a training video dataset containing anticipated distortions, or human judgments of video quality, distorted or no. Its benefit is that it only models the statistical 'naturalness' or lack thereof of an examined video. This is a significant advantage, having in mind that it's really difficult to create VQA databases containing distorted videos that are all accompanied with human mean opinion scores.

Comparing VIIDEO algorithm with the full reference MSE metric, VIIDEO achieves better correlation of the predictions' quality with human subjective opinion. Currently, the only limitation of this algorithm is that it consumes high computation resources due to the high complexity of filtering divisive normalization operations.

### 2.3.2 Subjective Medical Video Quality Assessment - Clinical Protocol

There are unique challenges associated with both objective and subjective clinical VQA. In addition to the motion and QoS aspects of conventional VQA, unique clinical criteria, often different for each medical modality, need to be properly assessed. These clinical criteria often correspond to specific video portions that are of diagnostic interest. These regions of diagnostic interest are much more sensitive to compression and error impairments given their significant clinical contribution to the diagnostic yield of the particular medical video. On the other hand, diagnostic ROI encoding may lead to diagnostically lossless medical videos. Clearly, both schemes are not adequately assessed by current objective VQA algorithms.

Table 5: Clinical evaluation rating

| Rate | Description |
|---|---|
| *1* | Lowest score. Video is not suitable for diagnosis. The video is of unacceptably low quality. |
| *2* | Poor video quality. The medical expert cannot evaluate its content. |
| *3* | Diagnostically unacceptable video quality. Video is not suitable for a confident diagnosis. |
| *4* | Diagnostically acceptable threshold. The medical expert can provide a confident diagnosis. There is a diagnostically acceptable loss of minor details. |
| *5* | Highest score. Diagnostically lossless video. The evaluated video carries equivalent clinical information as the original video. |

For subjective clinical evaluation, while we do expect that the basic features of subjective quality assessment criteria described in [90] will play a role in emerging medical video quality assessment standards, unique clinical criteria will also need to be adequately modeled. Diagnostic yield of transmitted medical video is restricted by a number of factors including resolution, frame rate, and end user equipment. For example, diagnostic capacity of a 720x576 resolution medical video at 25 fps displayed on a mobile device largely differs from a 1440x1080 resolution at 40 fps displayed on a laptop. Hence, appropriate clinical rating schemes should also address the aforementioned implications.

Currently, subjective clinical evaluation in this study is accomplished by medical experts which provide a rate for each video. As can be seen in Table 5, the clinical ratings range from one (1) to five (5), which correspond to the lowest and highest possible ratings, respectively. A rating of four (4) translates to diagnostically lossless compression. In other words, despite the inherent loss of some information attributed to compression, the ultrasound video maintains its original clinical capacity.

More information can be seen in Chapter 4, Section 4.4.3, and in Chapter 5, Subsection 5.5.4.

# Chapter 3

## Review of Wireless Networking Standards

### 3.1 Introduction

Each application and/or video modality has unique needs for transmission over wireless channels. Advances in wireless technologies are expected to allow transmission of higher resolution and frame rates for medical video applications. 4G channels promise quality-of-service (QoS) provision through data prioritization classes for low delay-fixed bandwidth allocation. This enables the development of telemedicine systems that will exploit a priori adaptation to channel state and provide for a diagnostically robust system at a required bitrate, irrespective of channels conditions. Based on medical expert's feedback, channel knowledge, and end-user device capabilities, higher bandwidth availability through 4G networks, namely Worldwide Interoperability for Microwave Access and Long Term Evolution, are investigated with respect to the impact on medical video's diagnostic capacity.

### 3.2 Wireless Transmission Technologies

In terms of wireless infrastructure, the Global System for Mobile communications (GSM) [91] is the most widely used wireless technology in the world, available in more than 219 countries and

territories. GSM signified the transition from analog 1st generation (1G) to digital 2nd generation (2G) technology, and despite that originally it was designed almost entirely for voice communication, it is also capable of data transfer at rates from 9.6 kbps up to 14.4 kbps. At such low rates, GSM can only be used for still images and it cannot be used for the transmission of medical video.

Table 6 summarizes access technologies, operating frequency bands, as well as typical delays.

Table 6: Wireless technologies evolution and associated data transfer rates and delays. N/A: not available, TBD: to be determined, DL: downlink, UL: uplink. Information taken from [91, 92, 93, 94, 95, 96, 97, 12, 98]

| **Wireless Technology** | **Frequency Band** | **Theoretical Data Rates** | **Typical Data Rates** | **Delay** |
|---|---|---|---|---|
| 2G-GSM<br>(TDMA) | 850/900/<br>1800/1900 MHz | 9.6 – 115 Kbps. | 10 Kbps | N/A |
| 2.5G-GPRS<br>(TDMA) | as above | DL:UL: 9.6 – 171.2 Kbps | DL:UL: 30–50 Kbps | <700ms |
| 2.5G-EDGE<br>(TDMA) | as above | DL:UL: 9.6–384 Kbps | DL:UL: 75–135 Kbps | <600ms (Rel.99)<br><350 (Rel. 4) |
| Evolved EDGE<br>(TDMA) | as above | DL: 1.89Mbps<br>UL: 947Kbps | DL: 175–350Kbps<br>UL:150–350Kbps | <200ms |
| 3G-UMTS<br>(FDD, W-CDMA) | 800/850/1500/1700/<br>1800/1900/2100 MHz | DL:UL: 144 kbps – 2 Mbps. | DL:UL: 220–384 Kbps | <250ms |
| 3G-UMTS<br>(TDD, TD/CDMA) | 1900–1920/2010–2025/<br>1850–1910/1930–1990/<br>1910–1930/2570–2620 MHz | as above | as above | as above |
| 3.5G-HSPA<br>(HSDPA Rel. 5)<br>(HSUPA Rel. 6) | as above | DL: 14Mbps<br>UL: 5.8 Mbps | DL : 1–4 Mbps<br>UL : 500Kbps–2Mbps | <150ms |
| 3.5G-HSPA+<br>Rel. 8 (Rel. 9) | as above | DL: 42(84) Mbps<br>UL: 11.5(23) Mbps | DL : 1.9–8.8 Mbps<br>UL: 1–4 Mbps | <100ms |
| 3.5G-LTE<br>(OFDMA) | as above + 700/800/1800 MHz<br>2.6 GHz | DL: 326 Mbps<br>UL: 86 Mbps | DL: 6.9–21.5 Mbps<br>UL: 6.0–13.0 Mpbs | <70ms |
| Mobile WiMAX<br>(OFDM)<br>(IEEE 802.16e)<br>4G-LTE-Advanced<br>(OFDMA) | 2.3, 2.5–2.7, 3.5, 5.8 GHz<br>(licensed)<br><br>As in LTE<br><6GHz (IMT-Advanced) | DL: 46 Mbps<br>UL: 5.6 Mbps<br><br>DL: 1Gbps<br>UL: 100 Mbps | as above<br><br>as above | <70ms<br><br><10ms |
| 4G- WiMAX<br>(OFDM)<br>(IEEE 802.16m) | 450-470/698-960/1710-025/<br>2110-2200/2300-2400/2500-<br>690/3400-3600 MHz | DL: 1Gbps<br>UL: 100 Mbps | as above<br>as above | (target<5ms) |
| 5G<br>(by 2020) | 700 MHz, 3.4-3.8 GHz,<br>and 24.25-27.5 GHz | DL: 1-10 Gbps<br>PL: 1-10 Gbps | TBD | TBD<br>(target<1ms) |

The evolution of mobile telecommunication systems from 2G to 2.5G (iDEBN, GPRS, EDGE) and subsequently to 3G (W-CDMA, CDMA2000, TD-CDMA),3.5G (HSPA, HSPA+), mobile WiMAX, and LTE systems has facilitated both an always-on model (as compared with the circuit-switched mode of GSM), as well as the provision of faster data transfer rates and lower delays, thus enabling the development of more-responsive telemedicine systems [4].

In terms of bandwidth, both 2.5G and 3G provide sufficient rates for medical image and bio-signal transmission. For medical video transmission, 3G rates are sufficient for QCIF (176 x 144) resolution medical video (however this is a very low resolution video), as well as specific regions of interests (ROIs). Furthermore High-speed packet access (HSPA) and HSPA+ 3.5G technologies enable the transmission of high quality CIF (352 x 288) as well as up to 4CIF (704 x 576) resolution video and beyond [10].

Satellite systems provide a variety of data transfer rates starting from 2.4 kbps to high-speed data rates of up to nx64 kbps and beyond. Satellite links have also the advantage of to be functional all over the world. Furthermore, satellite links utilization in healthcare benefit from world-wide coverage [99], but require line of sight and comparably higher power for similar bit rates.

WiMAX release 2.0 and LTE-advanced networks conforming to the IMT-advanced requirements [11] will constitute the next-generation family of technologies, namely 4G. Low latency, high mobility, high bandwidths (targeting 100Mbps for high mobility and 1Gbps for low mobility, in the downlink), and QoS provisions, are expected to significantly boost the development of mobile-healthcare systems and services. As a result, the wider availability of wireless network infrastructure and the emerging high efficiency video compression methods will significantly advance existing mHealth applications.

WLAN is a flexible data communications system implemented as an extension to or as an alternative to a wired LAN. WLANs transmit and receive data over the air, minimizing the need

for wired connections. WLANs transmit and receive data over the air using radio frequency (RF) technology. Thus, WLANs combine data connectivity at hundreds of Mbps with, however limited coverage (in the region of tens of meters at the level of an access point or typically few km within an enterprise) and hence user mobility. To extend coverage over larger distances, wireless mesh networks are also being considered. These networks are peer-to-peer multi-hop wireless networks, in which stationary nodes take on the routing functionality thus forming the network's backbone. Basically, they act as a gateway to high-speed wired networks for mobile nodes (clients) which communicate in a peer manner. Clearly, when tens of Mbps are available, we have sufficient bandwidth for transmitting multiple video bit-streams. The problem of coverage still remains.

### 3.2.1 4G Networks

#### 3.2.1.1 Background of 4G Mobile Technologies

In recent years, there have been major advances in mobile communications systems, in particular, the 4G mobile technologies with LTE network that are currently being deployed worldwide. LTE is basically the next generation of cellular networks standardized by the 3rd Generation Partnership Project (3GGP). Furthermore, the IMT-Advanced 4G standards that include LTE-Advanced will provide a global platform on which the next generations of interactive mobile services that will provide faster data access, enhanced roaming capabilities, unified messaging, and broadband multimedia services will be built. On the side, the coexistence of the LTE with high speed packet access, enhanced data rates for global evolution will provide more service choices for mobile healthcare and applications.

The other significant mobile network technology that constitutes part of the 4G evolution is the World Wide Interoperability for Microwave Access (WiMAX) networks. WiMAX aims, in

general, to provide wireless broadband services on the scale of the metropolitan area network and is the commercialization of the IEEE 802.16 standard.

The competitive and effective usage of these two significant 4G constituent technologies for mHealth applications will be one of the key research challenges in the near future.

### 3.2.2 Worldwide Interoperability for Microwave Access (WiMAX)

Worldwide Interoperability for Microwave Access (WiMAX) was firstly standardized for fixed wireless applications in 2004 by the IEEE 802.16-2004 [95] and then for mobile applications in 2005 by the IEEE 802.16e [100] standards. After an initial hype of the WiMAX, lately there has been scepticism as to its successful wide deployment, in favour of the LTE. However, current standardization 802.16m [101], also termed as IEEE WirelessMAN-Advanced met the ITU-R IMT-advanced requirements and succeeded as a 4G technology.

While 3G systems are designed primarily for mobile voice and data users, IEEE 802.16/WiMAX systems are optimized to provide high-rate wireless connectivity for large set of services and applications (e.g., with multimedia traffic) which require QoS guarantee.

The IEEE 802.16 standard, which incorporates several advanced radio transmission technologies such as orthogonal frequency division multiplexing (OFDM), adaptive modulation and coding (AMC), and adaptive forward error correction (FEC), is designed to provide broadband wireless capability using a well defined quality of service (QoS) framework. Therefore, this is a promising technology to provide wireless services requiring high rate transmission (in the range of tens of megabits per second) and strict QoS requirements (e.g., for telemedicine applications) in both indoor and outdoor environments. The potential advantages of using WiMAX and the related design issues are outlined.

IEEE 802.16/WiMAX technology intends to provide broadband connectivity to both fixed and mobile users in a wireless metropolitan area network (WMAN) environment. Key features of physical (PHY) and medium access control (MAC) layers are discussed next.

#### 3.2.2.1 Physical (PHY) Layer

As already mentioned above, WiMAX standards define the air interface and more specifically MAX and PHY layers. The physical layer of the IEEE 802.16 air interface operates at either the 10–66 GHz (i.e., IEEE 802.16) or 2–11 GHz band (i.e., IEEE 802.16a) and supports data rates in the range of 32–130 Mb/s, depending on the operation bandwidth (e.g. 20, 25, or 28 MHz) as well as the modulation and coding schemes used (QPSK, 16-QAM, 64-QAM). The primary features of the physical layer also include hybrid automatic repeat request (HARQ), and fast channel feedback. The IEEE 802.16 standard specifies different air interfaces for different frequency bands. In the 10–66 GHz band the signal propagation between a Base Station (BS) and an Sub-carrier Station (SS) should be line of sight (LOS), and the air interface for this band is wireless single carrier (SC). In the 2–11 GHz band, three different air interfaces supporting non-line-of-sight (NLOS) communication can be used:

· **WirelessMAN-SCa** for single-carrier modulation.

· **WirelessMAN-OFDM** for OFDM-based transmission using 256 subcarriers. For this air interface, the MAC scheme for the SSs is based on time-division multiple access (TDMA).

· **WirelessMAN-OFDMA** for OFDM-based transmission using 2048 subcarriers. The MAC algorithm is based on orthogonal frequency-division multiple access (OFDMA) in which different groups of subcarriers are assigned to different subscriber stations (SS).

WiMAX uses scalable orthogonal frequency division multiple access (SOFDMA) that divides the transmission bandwidth into multiple subcarriers. The number of subcarries ranges from 128 for 1.25 MHz channel bandwidth and extends up to 2048 for 20 MHz channels. In this manner, dynamic QoS tailored to individual application's requirements can be succeeded. In addition, orthogonality among subcarriers allows overlapping leading to flat fading. In other words, multipath interference is addressed by employing OFDM while at the same time available bandwidth can be split and assigned to several requested parallel applications for improved system's efficiency. The latter is true for both downlink (DL) and uplink (UL).

Multiple input multiple output (MIMO) antenna system allows transmitting and receiving multiple signals over the same frequency. Two types of gain are possible, namely spatial diversity and spatial multiplexing. For spatial diversity, unique configuration enables enhanced link quality by combining independent faded signals resulting from simultaneously transmitted duplications of the same information. For spatial multiplexing, increased throughput is achieved via the parallel spatial channels transmission of multiple streams.

#### 3.2.2.2 Medium Access Control (MAC) Layer

IEEE 802.16/WiMAX uses a connection-oriented MAC protocol, which provides a mechanism for the SSs to request bandwidth from the base stations (BS). Although each SS has a standard 48-bit MAC address, the main purpose of this address is for hardware identification. Therefore, a 16-bit connection identifier (CID) is used primarily to identify each connection to the BS. On the downlink, the BS broadcasts data to all SSs in the same network. Each SS processes only the MAC protocol data units (PDUs) containing its own CID and discards the other PDUs. IEEE 802.16 MAC supports grant per SS (GPSS) mode of bandwidth allocation in which a portion

of the available bandwidth is granted to each of the SSs, and each SS is responsible for allocating the bandwidth among the corresponding connections.

In the MAC layer the most important supported features can be summarized in QoS provision through different prioritization classes, direct scheduling for both DL and UL, efficient mobility management, as well as security.

IEEE 802.16/WiMAX supports both frequency-division duplex (FDD) and time-division duplex (TDD) transmission modes. For TDD, a MAC frame is divided into uplink and downlink subframes. The lengths of these subframes are determined dynamically by the BS and broadcast to the SSs through downlink and uplink map messages (UL-MAP and DL-MAP) at the beginning of each frame. Therefore, each SS knows when and how long to receive and transmit data to the BS. In the uplink direction a subframe also contains ranging information to identify an SS, information on the requested bandwidth, and data PDUs for each SS.

As mentioned above, the IEEE 802.16/WiMAX standard defines a QoS framework for different classes of services. The following three major types of services are supported, each of which has different QoS requirements [102].

· **Unsolicited grant service (UGS):** This service type supports constant-bit-rate (CBR) traffic. In this case the BS allocates a fixed amount of bandwidth to each of the connections in a static manner; therefore, delay and jitter can be minimized. UGS service is suitable for traffic with very strict QoS constraints for which delay and loss need to be minimized.

· **Polling service (PS):** This service supports traffic for which some level of QoS guarantee is required. It can be divided into two subtypes: real-time polling service (rtPS) and non-real-time polling service (nrtPS). The difference between these subtypes lies in the tightness

of the QoS requirements (i.e., rtPS is more delay- sensitive than nrtPS). Not only delay-sensitive traffic but also non-real-time Internet traffic can use polling service to achieve a certain through- put guarantee. The amount of bandwidth required for this type of service is determined dynamically based on the required QoS performance and the dynamic traffic arrivals for the corresponding connections.

· **Best-Effort Service (BE):** This is for traffic with no QoS guarantee (e.g., web and e-mail traffic). The amount of bandwidth allocated to BE service depends on the bandwidth allocation policies for the other two types of service. In particular, the bandwidth left after serving UGS and PS traffic is allocated to BE service.

Mobility management is efficient in 802.16e and current 802.16m standards, which was an issue in 802.16d primary standard for fixed connections. With a theoretical support of serving users at 120 km/h in 802.16e, established connections to provide adequate performance for vehicles moving with speeds between 50-100 km/h. More, in 802.16m, mobility support is extended for mobile speeds up to 350 km/h as defined in the evaluation of IMT-advanced requirements.

Finally, enhanced security, especially when compared to competing technologies (like WLANs) is one of the key features in WiMAX networks, protecting the end-user from a variety of threats. Improved security is based on an extensible authentication protocol (EAP) for authentication, while advanced encryption system (AES) is employed for encryption. BS and SS are authenticated via the privacy key management (PKM) algorithm.

### 3.2.3 Long Term Evolution

The long-term evolution (LTE) as defined by the 3rd Generation Partnership Project (3GPP) [96] is a highly flexible radio interface. LTE is standardised in 3GPP Release 8 series, with minor

enhancements described in Release 9. The first release of LTE provides peak rates of 300 Mb/s, a radio-network delay of less than 5 ms, a significant increase in spectrum efficiency compared to previous cellular systems, and a new flat radio-network architecture designed to simplify operation and to reduce cost. LTE supports both frequency-division duplex (FDD) and time-division duplex (TDD), as well as a wide range of system bandwidths in order to operate in a large number of different spectrum allocations.

The goal of LTE is to increase the capacity and speed of wireless data networks using new DSP (digital signal processing) techniques and modulations that were developed around the 2000. Additionally, LTE also aims for a smooth evolution from earlier 3GPP systems such as time division-synchronous code division multiple access (TD-SCDMA) and wide-band code division multiple access/high-speed packet access (WCDMA/HSPA), as well as 3GPP2 systems such as code division multiple access (cdma) 2000. Finally, LTE also constitutes a major step toward international mobile telephony (IMT)-Advanced. In fact, the first release of LTE already includes many of the features originally considered for future fourth-generation systems [103].

The core of the LTE downlink radio transmission is the multiple-carrier multiplexing orthogonal frequency-division multiplexing (OFDM), with data transmitted on a large number of parallel, narrow-band subcarriers. Due to the use of relatively narrowband subcarriers in combination with a cyclic prefix, OFDM transmission is inherently robust to time dispersion on the radio channel without a requirement to resort to advanced and potentially complex receiver-side channel equalization. For the down-link, this is an attractive property because it simplifies the receiver baseband processing with reduced terminal cost and power consumption as consequences. This is especially important considering the wide transmission bandwidths of LTE, and even more so in combination with advanced multi-antenna transmission, such as spatial multiplexing.

On the opposite site for the uplink, where the available transmission power is significantly lower than for the downlink, its design is to enable highly power-efficient transmission. This improves coverage and reduces terminal cost and power consumption at the transmitter. For this reason, single-carrier transmission, based on discrete Fourier transform (DFT)-precoded OFDM, sometimes also referred to as single-carrier frequency-division multiple access (SC-FDMA), is used for the LTE uplink.

The radio link control (RLC) and medium access control (MAC) layers, among other tasks, are responsible for retransmission handling and multiplexing of data flows. In the physical layer, the data that is to be transmitted is turbo coded and modulated using one of the following: quadrature-phase shift keying (QPSK),16-QAM, or 64-QAM, followed by OFDM modulation. The subcarrier spacing is 15 kHz and two cyclic-prefix lengths are supported in both uplink and downlink, a normal cyclic prefix of 4.7 $\mu$s, suitable for most deployments and an extended cyclic prefix of 16.7 $\mu$s for highly dispersive environments. In the downlink, different types of multi-antenna processing, further described below, are applied prior to OFDM modulation. In the uplink, to preserve the single-carrier properties, a DFT precoder is used prior to the OFDM modulator. It's important to mention that the DFT precoder does not compromise orthogonality between subcarriers. In addition, cell-specific reference signals are transmitted in the downlink to support channel estimation for coherent demodulation, as well as for various measurement purposes, including not only measurements for mobility management but also channel quality measurements.

To handle occasional retransmission errors, LTE includes a two-layered retransmission scheme: a fast hybrid-automatic repeat request (ARQ) protocol with low overhead feedback and support for soft combining with incremental redundancy is complemented by a highly reliable selective-repeat ARQ protocol. The use of a two-layered mechanism achieves low latency and low overhead

without sacrificing reliability. Most errors are captured and corrected by the lightweight hybrid-ARQ protocol, which provides feedback to the transmitter for each transmitted subframe; only rarely, in terms of latency and overhead, the more expensive ARQ retransmissions are required. The tight coupling between the two retransmission layers is possible because both mechanisms are terminated in the base station.

To support the LTE features, scheduling decisions, hybrid-ARQ feedback, channel-status reports, and other control information must be communicated between the base station and the terminal. In the downlink, the control signaling is transmitted using (typically) up to three of the first OFDM symbols in each subframe. The code rate of the control signaling for each terminal can be adjusted individually to match the instantaneous channel conditions and to minimize the overhead. Also, the total amount of resources in the downlink used for control signaling can be varied dynamically to minimize the overhead.

LTE allows multi-antenna applications for single and multi users through MIMO technology (up to 4-layers in the downlink and 2-layers in the uplink).

Thus, the the key features provided by LTE systems include enhanced mobility support (up to 350 km/h), efficient multimedia broadcast multicast service (MBMS), QoS provision, security, and cell capacity up to 200 active users.

#### 3.2.3.1 LTE-Advanced

LTE-Advanced is the evolutionary path from LTE Release 8 satisfying the IMT-Advanced requirements. IMT-Advanced is the term used by the ITU for radio-access technologies beyond IMT-2000 and an invitation to submit candidate technologies for IMT-Advanced has been issued by ITU [104]. Such systems provide access to a wide range of telecommunication services including advanced mobile services, supported by mobile and fixed networks, which are increasingly

Table 7: ITU and 3GGP requirements

| Quantity | | IMT-Advanced | LTE-Advanced |
|---|---|---|---|
| Peak data rate | UL | | 1 Gbit/s |
| | DL | | 500 Mbit/s |
| Spectrum allocation | | Up to 40 MHz | Up to 100 Mhz |
| Latency | User plane | 10 ms | 10 ms |
| | Control plane | 100 ms | 50 ms |
| Spectrum efficiency (4 ant BS, 2 ant terminal) | Peak | 15 bit/s/Hz DL<br>6.75 bit/s/Hz UL | 30 bit/s/Hz DL<br>15 bit/s/Hz UL |
| | Average | 2.2 bit/s/Hz DL<br>1.4 bit/s/Hz UL | 2.6 bit/s/Hz DL<br>2.0 bit/s/Hz UL |
| | Cell-edge | 0.06 bit/s/Hz DL<br>0.03 bit/s/Hz UL | 0.09 bit/s/Hz DL<br>0.07 bit/s/Hz UL |

packet-based. All the requirements of the IMT-Advanced cannot be fulfilled by LTE Release 8 and require technology beyond Release 8.

Being an evolution of LTE, LTE-Advanced should be backwards compatible in the sense that it should be possible to deploy LTE-Advanced in spectrum already occupied by the first release of LTE with no impact on existing LTE terminals. A direct consequence of this requirement is that, for an LTE terminal, an LTE-Advanced-capable network should appear as an LTE network. Such spectrum compatibility is of critical importance for a smooth, low-cost transition to LTE-Advanced capabilities within the network. As can be seen in Table 4, requirements are set not only on the peak spectral efficiency, but also on the average and cell-edge spectral efficiency. The latter are, in most practical deployments, more important than the peak rates and [104].

LTE-Advanced has adopted many new technology components [105]. These components include:

***Carrier aggregation*** enables wider bandwidth transmission up to 100 MHz utilizing a combination of frequency blocks, thus increasing system's peak data rates. Carrier aggregation is attractive because it allows operators to deploy a system with extended bandwidth by aggregating several smaller component carriers while providing backward compatibility to legacy users. Three possible aggregation scenarios are possible: contiguous aggregation of component carriers

in a single band, non-contiguous aggregation of component carriers in a single band, and non-contiguous aggregation of component carriers over multiple bands. Up to five component carriers may be aggregated together, providing a maximum bandwidth of 100 MHz.

***Enhanced Multi-antenna transmission*** in LTE-advanced systems include eight-layer transmission in the downlink that can be transmitted using an 8x8 antenna configuration, allowing for a peak spectral efficiency exceeding the requirement of 30 bit/s/Hz and implying a possibility for data rates beyond 1 Gbit/s in a 40 MHz bandwidth and even higher data rates with wider bandwidth. Furthermore, LTE-Advanced systems include spatial multiplexing of up to four layers also for the uplink. With four-layer transmission in the uplink, a peak uplink spectral efficiency exceeding 15 bit/s/Hz can be achieved.

***Coordinated Multi-point Transmission and Reception (CoMP)*** is another novel technique defined in the standard which provides for increased throughput on the cell edge. The key idea is that multiple e-Node Bs cooperate to coordinate transmission relevant aspects that provide for reduced interference and increased throughput for UEs located near the cell edge.

***Relaying*** variation solutions can be envisioned to improve coverage and reduce deployment cost, depending on the scheme applied, although they all share the basic property of relaying the communication between the donor cell and the terminal. The donor cell may, in addition to serving one or several relays, also communicate directly with other terminals. The simplest form of relay is a repeater, which simply amplify and forward the received analog signals. Repeaters simply amplify and forward the received analog signals and are used already today for handling coverage holes. Such repeaters are invisible to both the terminal and the base station. However, more advanced repeater structures can be considered. The intermediate node may also decode and re-encode any received data prior to forwarding it to the served users. This is often referred to

as decode-and-forward relaying. As the intermediate node decodes and re-encodes received data blocks a significant delay is introduced, longer than the LTE subframe duration of 1 ms.

Concluding, the evolution of LTE, also called LTE-Advanced will incorporate additional technology components to further enhance the performance beyond the IMT-Advanced requirements while maintaining backwards compatibility with earlier releases of LTE. These technology components include extended spectrum flexibility, multi-antenna solutions, coordinated multi-point transmission/reception, and the use of advanced repeaters/relaying.

### 3.2.4 5G Standardization Efforts

Throughout the development of 5G, there are already plethora of reports discussing on the potential benefits that 5G technology will enable. It's documented that several industry associations have been researching and supporting the vision of 5G since 2013. Pre-standards research efforts include dozens of 5G research projects across the globe [106], like METIS [107] and 5GNOW [108].

This early research enables 5G to be quickly standardized in organizations like the 3rd Generation Partnership Project (3GPP), the ITU Radiocommunication Sector (ITU-R), the ITU Telecommunication Standardization Sector (ITU-T), the European Telecommunications Standards Institute (ETSI), the Internet Engineering Task Force (IETF) and more. However, it's important to note that 5G standardization is primarily driven through 3GPP [109]. 3GPP has been intensifying its effort to IMT-2020 (5G) development since September 2015, since it held a conference in September 2015 to plan development of the new standard.

Thenceforth, the organization released Release 14 and Release 15 that both represent the first phase of work towards 5G standards with Release 14 focused on the study items towards 5G and Release 15 on the first phase of normative specifications for 5G, to wit what can be achieved.

5G work will involve both the radio access network and the system architecture. However, the architecture today is based upon a set of assumptions that have been around the mobile system for some time. Therefore, 3GPP is trying to compound the diversity of views and projects in order to be proposed for 5G specification. Some tough decisions about what must be taken on in Release 16 are now being made.

# Chapter 4

# High-Resolution, Low-Delay, and Error-Resilient Medical Ultrasound Video Communication Using H.264/AVC Over Mobile WiMAX Networks

## 4.1 Introduction

Continuous advances in medical video coding, together with wider availability of current and emerging wireless network infrastructure, provide the key technologies that are needed to support mHealth video communication technologies in standard clinical practice. Over the past decade, demand for mobile health systems has been growing [5, 10, 4]. Demand is driven by the need for responsive emergency telematics, remote diagnosis and care, medical education, as well as for mass population screening and emergency crisis management. Advancements in mobile health systems are expected to bring greater socioeconomic benefits, improving the quality of life of patients with mobility problems, the elderly, and people residing in remote areas, by enhancing their access to specialized care. Moreover, they will provide a critical time advantage that can prove life saving in life-threatening emergency incidents.

Current research in mHealth video communications systems include modality-aware (m-aware) diagnostically driven systems, which adapt to the underlying wireless transmission medium [10], [110]. Diagnostically driven systems often rely on the use of diagnostic regions of interest (ROIs)[25, 111, 8]. Adaptation to the wireless network's characteristics includes diagnostically relevant selection of the source encoding parameters and error control for addressing inevitable transmission errors. Clinical video quality assessment (VQA) methods are vital for the systems' objective of communicating reliable medical video to the medical expert [25, 112, 113].

In terms of wireless infrastructure, thus far, mHealth video systems have been primarily based on 3G wireless networks [10, 111]. Given the limited upload data rates supported by these channels (up to 384 kb/s), the associated source encoding parameters were bounded to CIF resolution video size. As documented in [25] and [8], medical video resolution directly impacts the clinical capacity of the transmitted video. For atherosclerotic plaque ultrasound video, shifting from QCIF ($176\times144$) to CIF ($352\times288$) resolution enables the assessment of plaque type [8], providing critical clinical information to the medical expert for assessing the possibility of a plaque rupture, leading to stroke. Some recent studies that have briefly highlighted the benefits associated with streaming higher resolutions can be found in [25] and [33, 114, 115]. However, these studies are based on a limited number of cases, while the clinical aspect has not been extensively addressed. Moreover, these previous studies did not address individual network parameters' issues associated with clinical capacity of high-resolution video transmission.

Table 8: Literature Review - Medical Video Communication Systems

| **Author** | **Year** | **Resolution, Frame Rate, Bitrate** | **Encoding Standard** | **Wireless Network** | **Medical Video Modality** | **Video Quality Assessment** |
|---|---|---|---|---|---|---|
| Chu et *al.* [116] | 4 | 320x240 <5fps<br>50-80 Kbps | M-JPEG | 3G-CDMA | Trauma video | No |
| Garawi et *al.* [117] | 6 | 176x144 @ 5fps<br>18.5-60 Kbps | H.263 | 3G-UMTS | Cardiac Ultrasound | No |
| Rao et *al.* [18] | 9 | 360x240 @ 30 fps<br>500 Kbps | MPEG-2 | 3G and beyond | Pediatric videos | Clinical (subjective) evaluation |
| Martini et *al.* [118] | 10 | 480x256@15fps<br>300 Kbps | H.264/AVC | Mobile WiMAX | Cardiac Ultrasound | PSNR/SSIM |
| Panayides et *al.* [8] | 11 | 352x288 @ 15fps<br>197-421 Kbps | H.264/AVC | 3G and beyond | Carotid Artery Ultrasound (CCA) | Clinical (subjective) and objective evaluation |
| Debono et *al.* [114] | 12 | 640x480 @ 25fps | H.264/AVC | Mobile WiMAX | Cardiac Ultrasound | PSNR |
| Khire et *al.* [119] | 12 | 720x480 @ 30fps,<br>125 – 200 Kbps | H.264/AVC | 3G and beyond | Maxillofacial Surgery Clips | PSNR |
| Cavero et *al.* [120] | 13 | 720x576 @ 25fps,<br>200 Kbps | SPIHT | HSUPA, mobile WiMAX | Cardiac Ultrasound | Clinical evaluation |
| Panayides and Antoniou et *al.* [121] | 13 | 704x576 @ 15fps<br>768 Kb/s -1.5 Mbps | H.264/AVC | Mobile WiMAX | CCA Ultrasound | Clinical (subjective) and objective evaluation |

A summary of the afore-described studies appears in Table 8. As evident, early medical video communication systems relied on earlier video coding standards and wireless networks. Consequently, the limited compression capabilities and available data rates resulted in the transmission of low resolution video that in turn translated to limited clinical quality. The next generation of such systems however, employing new video encoding standards and exploiting higher data rates facilitated by new wireless channels, supported higher video resolutions, and were hence of increased clinical capacity. At the same time, the diversity of the investigated medical video modalities, highlights the necessity of adapting medical video communication systems in standard clinical practice, across the range of different application scenarios. For that reason, video quality assessment metrics were further introduced to quantify the communicated video's clinical capacity, towards increasing the reliability and hence impact of such systems.

The motivation of this work was based on the observation that there was no study that thoroughly investigated the implications of wireless networks varying state on clinical video quality. As a result, there is a strong demand to investigate new 3.5G and 4G wireless technologies [122] that can facilitate medical video communication at the clinically acquired video resolution. Ultimately, the goal is to deliver sufficiently high resolutions and video frame rates with the low-delay and low packet loss rates (PLR) that can approach the experience of in-hospital examinations.

In this study, we investigate the added clinical value of high-resolution (4CIF - $704 \times 576$) medical video communications over mobile worldwide interoperability for microwave access (WiMAX) networks for emergency telemedicine. The efficacy of the proposed end-to-end ultrasound video communication scheme is validated based on scalable clinical criteria. For this purpose, the clinically validated approach introduced [122] is extended from the CIF resolution to the higher resolution of 4CIF and the lower resolution of QCIF. In [8], diagnostically relevant

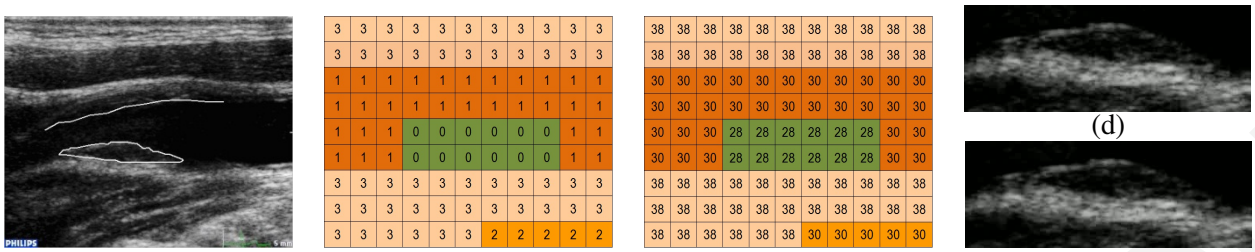



Figure 16: Atherosclerotic plaque ultrasound variable quality slice encoding. (a) Pixel-based segmentation of diagnostically important regions (based on [123]), (b) FMO type 2 Macroblock Allocation Map (MBAmap), (c) The corresponding Quantization Parameter Allocation Map (QPAmap) for variable quality slice encoding (introduced in [8]), (d) Plaque ROI segmented from original 4CIF video, (e) Plaque ROI with QP:28 segmented from 4CIF resolution video with QPs: 38/30/28.

selection of encoding parameters based on video region's clinical importance was used (see Figure 16). Here, we extensively validate different medium access control (MAC) and Physical layer features of mobile WiMAX channels that can support efficient emergency telemedicine mHealth systems. Most importantly, we clinically evaluate ultrasound videos transmitted using different network parameters configurations. The goal of the network study is to provide recommendations for resilient network parameter selection that will accommodate different emergency scenarios and varying network state.

We summarize the primary contributions of this study over previously published work (see, e.g., [8]) in three different areas.

1. Robust video encoding at three different resolutions: We consider QCIF, CIF, and 4CIF resolutions and carefully discuss the clinically validated criteria associated with each spatial resolution. For each case, we measure improvement in terms of the reduction in bitrate, which can be used for increasing the peak signal-to-noise ratio (PSNR) of the reconstructed video as compared to standard H.264/AVC encoding. For this purpose, we employ the BD-PSNR algorithm, which estimates the average bitrate gains for equivalent PSNR levels for a total of 240 cases of QCIF,

CIF, and 4CIF resolutions. Beyond the cases considered in [8], we consider an additional 1500 4CIF transmission cases over Mobile WiMaX.

2. Relationship between spatial resolution (QCIF, CIF, 4CIF) and clinical diagnosis: We summarize the clinical criteria that can be addressed at different resolutions. This relationship is validated through separate evaluation of the different criteria. For example, the use of 4CIF (704 $\times$ 576) resolution allows the use of new clinical criteria that are closer to standards used for in-hospital exams. For the current application, the additional resolution allows us to visualize atherosclerotic plaque morphology, not just the plaque type.

3. Medical video communications over Mobile WiMAX networks: We propose an emulation framework for mobile WiMAX medical video communication based on OPNET modeler. For this purpose, real ultrasound video encodings are used to generate video trace files imported to OPNET to model wireless video transmission. Following transmission, the communicated video packets are mapped back to the original files for decoding. The latter method allows us to realistically measure the effect of each investigated configuration setting both objectively and most importantly, subjectively (clinical evaluation).

Based on the previously described method, we provide recommendations for mobile WIMAX network parameter utilization for maximizing the communicated video's clinical capacity. Typical emergency telemedicine scenarios are constructed as a function of three different channel modulation and coding schemes, hybrid signal attenuation model, various distances from the base station (BS), and diverse mobility patterns. Mobile WIMAX network's performance is validated using quality of service (QoS) measurements such as PLR, packet delay, and PSNR of reconstructed video bitstreams, for an overwhelming number of 1500 video cases.

## 4.2 Mobile WiMAX for Video Communications

WiMAX was first standardized for fixed wireless applications in 2004 by the IEEE 802.16-2004 and then for mobile applications in 2005 by the IEEE 802.16e standard [124]. The current WiMAX standard 802.16m, also termed as IEEE WirelessMAN-Advanced, met the ITU-R IMT-advanced requirements and is considered to be a 4G technology. In what follows, we describe the WiMAX features provided at the physical layer (PHY) and the MAC layer.

A. Physical Layer Features

The primary features of the physical layer include adaptive modulation and coding (QPSK, 16-QAM, 64-QAM), hybrid automatic repeat request (hARQ), and fast channel feedback. WiMAX uses scalable orthogonal frequency division multiple access that divides the transmission bandwidth into multiple subcarriers. The number of subcarriers ranges from 128 for 1.25 MHz channel bandwidth and extends up to 2048 for 20-MHz channels. In this manner, dynamic QoS can be tailored to an individual application's requirements. In addition, orthogonality among subcarriers allows overlapping leading to flat fading. In other words, multipath interference is addressed by employing OFDM, while available bandwidth can be split and assigned to several requested parallel applications for improved system's efficiency. The latter is true for both downlink (DL) and uplink (UL). A multiple-input multiple-output antenna system improves communication performance, including significant increases in data throughput and link range, without additional bandwidth or increased transmit power.

B. MAC Layer Features The most important features of the MAC layer include QoS provision through different prioritization classes, direct scheduling for DL and UL, efficient mobility management, and security. The five QoS categories are described in [124] and [125]. Based on each application's requirements, we have an appropriate QoS class with its corresponding UL burst and

data rate. For real-time video streaming, the best option is to use the real-time polling service (rtPS) QoS class. The rtPS class specifies the minimum sustained data rate, the maximum traffic burst, the maximum tolerated latency, and a traffic priority, which the WiMAX air interface scheduler is designed to accommodate [125].

Mobility management is well addressed in 802.16e and current 802.16m standards, which was an issue in 802.16d primary standard for fixed connections. With a theoretical support of serving users at 120 km/h in 802.16e, established connections provide adequate performance for vehicles moving with speeds between 50 and 100 km/h.

### 4.3 Methodology

We investigate high-resolution medical video communication performance over mobile WiMAX networks based on realistic clinical scenarios. The aim is to model realistic scenarios that can be used to evaluate the challenges associated with developing mHealth video systems for emergency telemedicine. Such a system is illustrated in Figure 17. The key concept is to communicate the patient's video (trauma or ultrasound) to the hospital premises, for remote diagnosis and assistance with in-ambulance care, moreover for better triage and hospital admission related tasks (e.g., surgery chamber preparation).

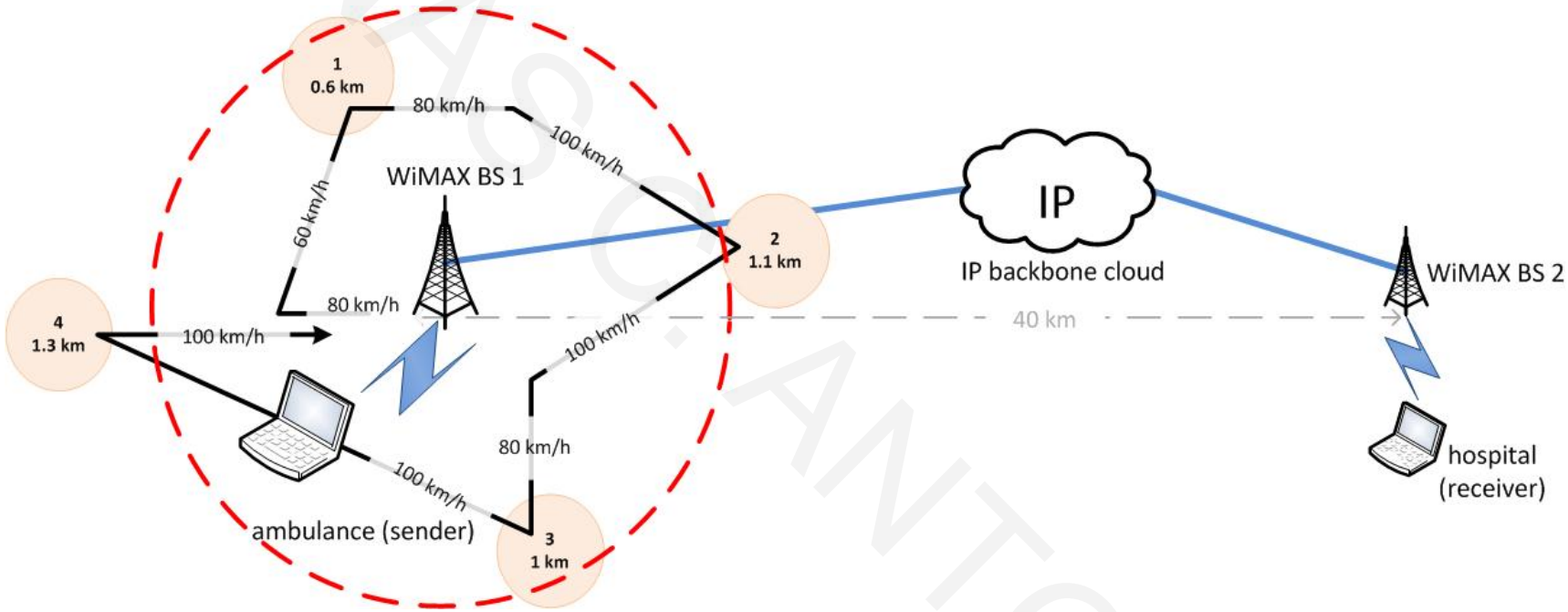



Figure 17: A typical topology for simulating medical video transmission over mobile WiMAX. The ambulance travels with speeds ranging from 60-100 km/h following the course delineated by the black line. Vertices '1'-'4' depict the locations close to the BSs effective coverage zone used during Scenario 2. Example from [121]

For the scenario depicted in Figure 17, the medical ultrasound video transmission is launched once the paramedics have stabilized the patient, utilizing equipment residing in the ambulance. The simulated scenario models a typical route from the emergency incident to the hospital premises and highlights the technological challenges associated with the wireless communication of ultrasound video of adequate diagnostic quality.

In what follows, we provide more detailed descriptions of each block component of the proposed medical video communication framework in the context of the scenario depicted in Figure 17.

### 4.3.1 Pre-processing

This step typically involves video resolution and frame rate adjustments to match the available channel bandwidth (upload data rate) and end-user device capabilities. In this study, high bandwidth mobile WiMAX networks allow the investigation of the transmission of 4CIF video resolution at 15 fps.

### 4.3.2 Diagnostically Relevant Encoding

The proposed system uses a diagnostically relevant (m-aware) and resilient encoding scheme that has been described in [25, 8]. The key idea was to associate video regions of interest with clinical criteria. Each video slice was then assigned a quality level based on its diagnostic significance. These quality levels were implemented by adjusting the values of the quantization parameter as demonstrated in Figure 16. In this manner, significant bitrate requirements can be preserved by compressing the background (non-diagnostically important region). The basic ROI-approach can be extended to different medical imaging modalities and is already gaining ground in the literature [10, 25, 118].

For atherosclerotic plaque ultrasound videos, the correspondence between the ROIs and the clinical significance was as follows (see Figure 16 ):

1. *Plaque region for visualizing plaque type, morphology, and motion*: This was the primary region of interest. By visualizing the plaque type and morphology, we can assess the stability of the plaque (e.g., darker plaques turn to be more dangerous). Plaque motion patterns can also help in assessing plaque stability. The use of 4CIF resolution over WiMAX networks is particularly critical for this region (see Figure 16 and Table 14)

2. *Surrounding plaque region for visualizing stenosis*: A high degree of stenosis is used as a strong predictor of the risk of stroke.

3. *Near and far wall regions for visualizing wall motion*: The interest in visualizing the near and far walls comes from the need to compare motion patterns with the plaque. Unstable plaques can have different motion patterns than the whole plaque.

4. *ECG region for visualizing ECG waveform*: The ECG was used to help visualize plaque and stenosis changes through different parts of the cardiac cycle (e.g., during systole and diastole).

In this study, we considered three different video resolutions, namely QCIF (176x144), CIF (352x288), and 4CIF (704x576), for the encoding setup depicted in the left column of Table 10. The objective was to include scalable screen resolutions that are widely used in the literature today, in addition to investigating high-resolution encodings over mobile WiMAX networks. In the latter case only 4CIF resolution with the recommended, diagnostically acceptable QPs setting of 38/30/28 (see also [8]) was used.

Table 9: Total number of processed videos in this study. We used FMO and FMO ROI RS for constant QP FMO encoding and variable QP FMO with RS respectively.[a]38:background/ 30:wall ROI/ 28:plaque ROI. Example from [121]

| | | **Diagnostically Relevant Encoding** | | **Mobile WiMAX Transmission** | |
|---|---|---|---|---|---|
| *Instances* | | | *Cases* | *Instances* | *Cases* |
| Resolution | | QCIF, CIF, 4CIF | ×3 | 4CIF | ×1 |
| Method | | FMO, FMO ROI RS | ×2 | FMO ROI RS | ×1 |
| QP | FMO | 36/36/36, 32/32/32, 28/28/28, 24/24/24 | ×4 | N/A | ×1 |
| | FMO ROI RS | 42/38/36, 40/34/32, 38/30/28, 36/26/24 | | 38/30/28[a] | |
| Data Set | | 10 ultrasound videos | ×10 | 10 ultrasound videos | ×10 |
| Modulation & Coding | | N/A | N/A | QPSK 1/2, 16-QAM 3/4, 64-QAM 3/4 | ×3 |
| Video StreamingLocation | | N/A | N/A | Scenario 1 (video looped over entire route) | ×1 |
| | | | | Scenario 2 (0.6 km, 1.1 km, 1 km, 1.3 km) | ×4 |
| OPNET | | N/A | N/A | 10 simulations runs | ×10 |
| Total number of processed videos | | | 240 | *Scenario 1* | 300 |
| | | | | *Scenario 2* | 1200 |

A series of ten videos encoded at 15 frames per second (fps) was used to evaluate the proposed concept. H.264/AVC error resilient tool, Flexible Macroblock Ordering (FMO) type 2, was used to implement variable quality slice encoding. Baseline profile, universal variable length coding (UVLC) entropy coding, IPPP encoding structure, with an Intraupdate frame interval of 15 frames, and a total of 100 frames per video summarize the encoding parameters. The JM H.264/AVC reference software [126] had been used for encoding. For the mobile WiMAX video transmission experiments, the obtained results were averaged over 10 simulations runs for each scenario. Redundant slices (RS) at the encoder (one every four coded frames), and simple frame copy error concealment at the decoder, were used to recover from packet losses.

### 4.3.3 Mobile WiMAX Video Transmission

We investigated high-resolution video communication performance based on the scenario illustrated in Figure 17. Our aim was to realistically model the varying state of wireless channels that contributes to ultrasound video degradation when transmitting from the ambulance to the hospital. For this typical scenario, we investigated the use of different channel modulation and coding schemes, signal attenuation due to different signal propagation models, mobility, distance from the base station (BS), bandwidth availability through subcarriers scalability, and QoS prioritization classes.

A synopsis of the parameters associated with the scenario of Figure 17 appears in Table 10, while the total number of processed videos is illustrated in the right column of Table 9. The ambulance travels with speeds ranging from 60-100 km/h and traverses through locations situated near the effective coverage zone of the BS (distance range: 150 m -1.3 km). The multipath channel model is set to ITU Vehicular A, while a vehicular path loss model is also considered; with a shadow fading correction of 12 dBs (OPNET [127] implements differently path loss and

Table 10: Mobile WiMAX Network Configuration Parameters. OFDMA: Orthogonal Frequency Division Multiple Access, TDD: Time Division Duplexing, ARQ: Automatic Repeat reQuest, hARQ: Hybrid Automatic Repeat request, DL: Downlink, UL: Uplink, Msps: Mega Symbols per Second, BS: Base Station, MS: Mobile Station. [1]16-QAM 3/4: 3/b/symbol/Hz * Msps (512 subcarriers) = 8.64/ 1.728 Mbps, 64-QAM 3/4: 4.5 b/symbol/Hz * Msps (512 subcarriers) = 12.96/ 2.592 Mbps, [2]QPSK 1/2: 1 b/symbol/Hz * Msps (2048 subcarriers) = 9.216 /2.6112 Mbps, [3]For the 10 atherosclerotic plaque ultrasound videos of the dataset. Example from [121]

| **Parameter** | **Value** | **Parameter** | **Value** |
|---|---|---|---|
| Access Technology | OFDMA 20MHz | Frame Duration | 5ms |
| Base Frequency | 5.8 GHz | Symbol Duration | 100.8 |
| Subcarrier Frequency Spacing | 10.9375 KHz | Total Capacity DL/UL | 2.88/0.576 Msps (512 subcarriers)[1] 9.216/2.6112 Msps (2048 subcarriers)[2] |
| ARQ/hARQ | Disabled | Duplexing Technique | TDD |
| Number Transmit Receive Antennas | BS: 1, MS: 1 | Multipath Channel Model/ Pathloss Model/ Shadow Fading Additive Correction in dBs | ITU Vehicular A/ Vehicular Environment/ 12 dB |
| Antenna Gain | BS: 15 dBi, MS: -1dBi | MAC Layer QoS Class | Real time polling service (rtps) |
| Maximum Transmission Power | BS: 2 W, MS: 0.5W | Minimum Sustained Data Rate[3] | 768 kbps, 1 Mbps, 1.3 Mbps, 1.5 Mbps |
| Modulation and Coding | QPSK 1/2, 16-QAM 3/4, 64-QAM 3/4 | Mobility | 60-100 Km/h |

multipath fading, and allows shadow fading correction for increasing possible signal attenuation combinations during simulations). Three different channel modulation and coding schemes were investigated, namely QPSK 1/2, 16-QAM 3/4, and 64-QAM 3/4, with the number of subcarriers set to 512, besides QPSK 1/2 which is set to 2014.

### 4.3.4 Scenario 1

For a more realistic evaluation, the ultrasound video traffic sent through the network is modeled via trace files generated using real ultrasound video encodings. In *scenario 1*, the ultrasound

videos were looped over the entire route to examine the wireless channel's performance by measuring the average QoS parameters such as packet loss rates (PLR), end-to-end delay, and delay jitter.

### 4.3.5 Scenario 2

In *scenario 2*, ultrasound video traffic was initiated at four different locations as can be seen in Figure 17, selected to highlight the wireless channel's ability to provide reliable medical video communications at different distances from the BS. Following the wireless transmission and the QoS measurements, the successfully streamed packets were mapped back to the original RTP files, decoded, and evaluated by the relevant medical expert. In addition to the wireless network's QoS measurements, VQA ratings described below summarize the evaluation setup.

### 4.3.6 Video Quality Assessment

VQA includes objective and subjective evaluations. Objective VQA is given in terms of the video quality metric (e.g., PSNR) computed over the specified video slices using the clinical criteria. Clinical evaluation is performed by the relevant medical expert for the clinical criteria provided in Table 14. Ratings are given in the range of 1–5. A rating of 5 is the highest possible and it signifies that the diagnostic information in the decoded video is of essentially the same quality as the original video. A rating of 4 indicates that there is a diagnostically acceptable loss of minor details. At the lowest scale, a rating of 1 signifies that the decoded video is of unacceptably low quality.

Table 11: Average bit rate gains of diagnostically relevant encoding while maintaining the same diagnostic quality for FMO ROI RS. FMO ROI RS stands for variable quality slice encoding with redundant slices. FMO stands for standard H.264/AVC FMO type 2 encoding. Bitrate gains account for the reductions in bitrate requirements for equivalent objective quality (PSNR), estimated using the BD-PSNR algorithm for QP values: 36, 32, 28, and 24. Example from [121].

| | QCIF | CIF | 4CIF |
|---|---|---|---|
| Bit Rate Gain (%) | 34.7 | 39.8 | 42.3 |

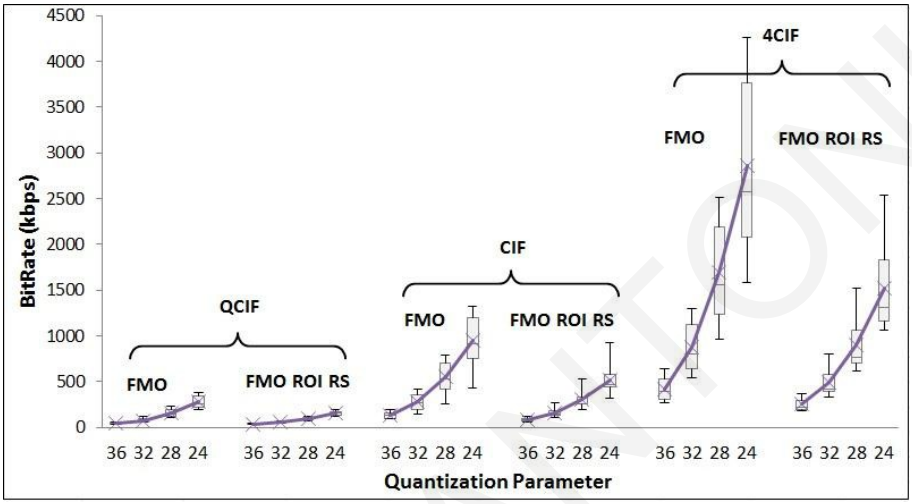



Figure 18: Boxplots depicting bitrate requirements for equivalent perceptual quality of the two investigated encodings schemes, using four QPs and QCIF, CIF, and 4CIF video resolutions. Example from [121].

### 4.4 Results and Discussion

In this section, we discuss the experimental evaluation of the proposed medical ultrasound video transmission framework. We present results in terms of video encoding, medical video transmission over mobile WiMAX channels, and clinical evaluation.

#### 4.4.1 Diagnostically Relevant Encoding

To demonstrate the efficiency of the proposed diagnostically relevant encoding scheme, we provided a comparative evaluation of: (i) FMO with constant QP video slices and (ii) FMO with variable QPs and RS for communications in noisy environments. For each method, we had four sets of quantization levels for the video slices as depicted in Table 9.

Table 12: QoS measurements for Scenario 1. Example from [121].

| Channel Modulation & Coding Schemes | QPSK 1/2 | | | 16-QAM 3/4 | | | 64-QAM 3/4 | | |
|---|---|---|---|---|---|---|---|---|---|
| QoS Parameters | PLR % $\sigma$[1] | Delay (ms) | Jitter (ms) | PLR % ($\sigma$) | Delay (ms) | Jitter (ms) | PLR % ($\sigma$) | Delay (ms) | Jitter (ms) |
| Total Avg. | 1.0 (0.1) | 21.38 | $<1$ | 3.8 (3.85) | 20.34 | $<1$ | 4.38 (5.19) | 20.62 | $<1$ |

[1] $\sigma$: standard deviation.

Table 11 depicts the associated bitrate gains of the proposed variable quality slice encoding scheme when compared to the conventional, uniformly encoded medical video. Figure 18 uses boxplots to illustrate the bitrate requirements of the medical ultrasound video data set, for the two investigated encoding schemes. Bitrate gains for equivalent perceptual quality were computed using the BD-PSNR algorithm [128], based on the four rate points shown in Figure 17. The average bitrate demands reductions are 42.3% for 4CIF, 39.8% for CIF, and 34.7% for QCIF resolution videos. Bitrate gains are functions of the area occupied by the diagnostic ROIs, and most of the savings come from compressing the background (see Figure 16). Here, the medical video data set comprised of videos with diagnostic ROIs ranging between 44%-72% of the entire video [25].

### 4.4.2 Mobile WiMAX Medical Video Transmission

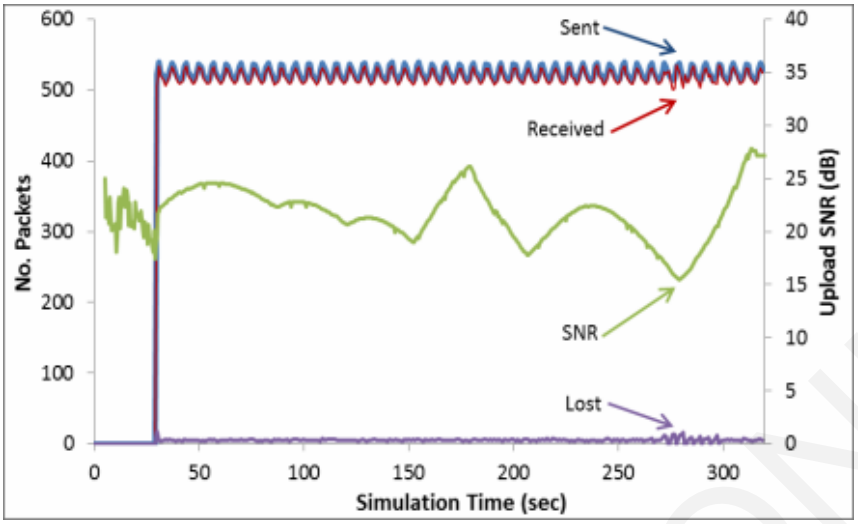
*1. Scenario 1*

Table 12 records the averaged QoS measurements of all video transmission simulations. QPSK 1/2 channel modulation scheme provides for a more robust performance as the packet loss rates measured are in the order of 1%. This is also highlighted in Figure 19(a) for the video shown in Figure 16(a). 16-QAM 3/4 and 64-QAM 3/4 depict comparable performance with PLR extending up to 5%. At these PLR, the reconstructed ultrasound videos still yield acceptable diagnostic performance, due to the use of RS and FMO error-resilience features. By examining the PLR standard deviation in Table12 however, we observe that PLR for 16- QAM 3/4 and 64-QAM 3/4 vary significantly and can reach unacceptably high rates. This is more clearly visualized in

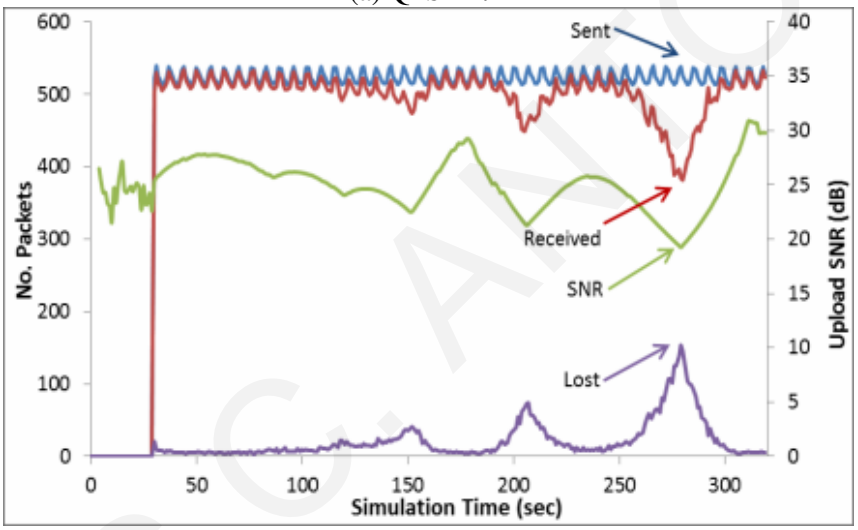
Figure 19(b)-(c), where it is obvious that significant packet losses occur at large distances to the Base Station. This is not the case for QPSK 1/2 scheme, which exhibits a robust performance throughout the simulation irrespective of the varying channel conditions. The reasoning is that QPSK $\frac{1}{2}$ requires lower signal-to-noise (SNR) compared to 16-QAM 3/4 and 64-QAM 3/4 [102] to maintain a quality connection. This is clearly depicted in Figure 19, where the upload SNR (measured each second) fluctuations experienced by packets traversing from the mobile station to the base station do not result in packet losses for QPSK 1/2 as in the rival channel modulation and coding schemes. This benefit comes at the expense of the channel's capacity as depicted in Table 10. QPSK 1/2 conveys information at 1 bit/symbol/Hz, as compared to 16-QAM 3/4 and 64-QAM 3/4, which provide 3 bits/symbol/Hz and 4.5 bits/symbol/Hz, respectively (mobile WiMAX capacity is given at mega symbols per second-Msps). As a result QPSK 1/2 utilizes 2048 subcarriers at 20MHz to meet the channel capacity required to transmit 4CIF resolution medical video, whereas 16-QAM 3/4 and 64-QAM 3/4 only require 512 subcarriers.

Average end-to-end delay of transmitted packets for all three channel modulations and coding schemes examined is less than 22 ms, which is well within the acceptable bounds for medical video streaming applications [102] (300 ms but preferably less than 100 ms). Similarly, delay jitter is negligible for the presented scenario. This is partly due to the fact that in this particular scenario no background traffic was modeled while the RTP packets do not traverse through multiple nodes to reach their destination. However, even in the above-described circumstances, the use of service prioritization classes, allow mobile WiMAX networks to meet the individual QoS requirements of each service.

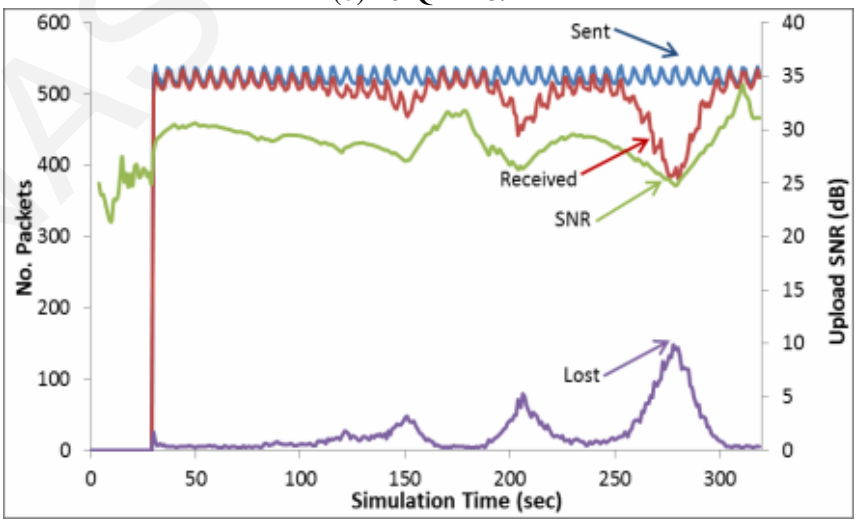
*2. Scenario 2*




(a) QPSK 1/2




(b) 16-QAM 3/4




(c) 64-QAM 3/4

Figure 19: Packet losses as a function of the distance from the Base Station (illustrated via simulation time) and Signal-to-Noise Ratio (SNR) for the investigated channel modulation and coding schemes (for a typical video looped over the entire route of Figure 17, video duratio n=7s, Scenario 1): (a) Packet losses for QPSK 1/2 . (b) Packet losses for 16-QAM 3/4. (c) Packet losses for 64-QAM 3/4. Example from [121].

To better examine the performance of the system for transmission near the boundaries of the BSss effective coverage zone, a 2nd scenario was implemented where the mobile station transmitted only one video loop at each of the 4 different distances from the BS depicted in Figure 17 (in *scenario 1* each video was looped until the end of each simulation). In this way, the actual quality of the transmitted video can be computed, both objectively and subjectively, as well as the specific QoS measurements at these locations, allowing accurate assumptions as to the extent (as a function of the distance from the BS and mobility) that the investigated channel modulation and coding schemes can be used.

Table 13: QoS measurements for Scenario 2. Example from [121].

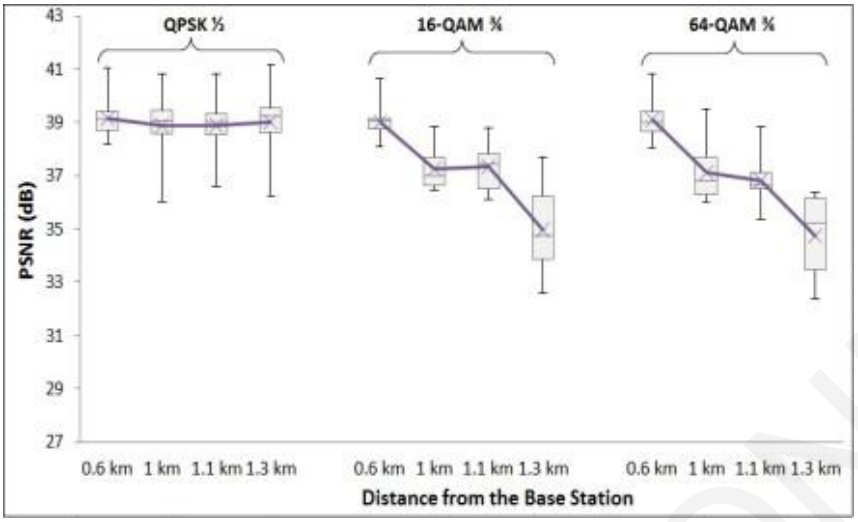
| Channel Modulation & Coding Schemes | QPSK 1/2 | | | 16-QAM 3/4 | | | 64-QAM 3/4 | | |
|---|---|---|---|---|---|---|---|---|---|
| QoS Parameters | PLR % ($\sigma$)[1] | Delay (ms) | PSNR[2](dB) | PLR % ($\sigma$) | Delay (ms) | PSNR (dB) | PLR % ($\sigma$) | Delay (ms) | PSNR (dB) |
| **1** (0.6 km from BS) | 1.42 (0.22) | 20.76 | 39.17 | 1.82 (0.77) | 20.04 | 39.01 | 1.68 (0.98) | 20.24 | 39.11 |
| **2** (1.1 km from BS) | 1.24 (0.32) | 22.65 | 38.87 | 6.07 (6.77) | 19.95 | 37.34 | 7.22 (8.98) | 20.19 | 36.83 |
| **3** (1 km from BS) | 1.25 (0.21) | 21.41 | 38.91 | 5.28 (5.67) | 19.99 | 37.23 | 5.82 (7.39) | 20.20 | 37.11 |
| **4** (1.3 km from BS) | 1.36 (0.50) | 24.07 | 39.01 | 16.27 (17.64) | 19.82 | 34.98 | 16.5 (20.47) | 19.99 | 34.74 |

[1]$\sigma$: standard deviation, [2]:PSNR is given for the atherosclerotic plaque ROI extracted from the transmitted ultrasound video.

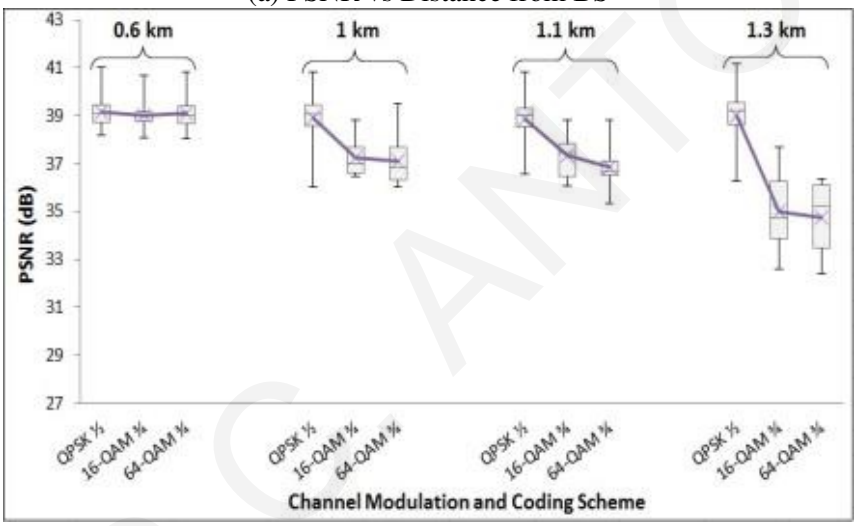
Results are shown in Table 13 and Figure 20. As expected, QPSK 1/2, attains diagnostically acceptable QoS measurements in all 4 locations, with consistently low PLR and high PSNR scores around 39 dBs (see leftmost boxplots of Figure 20(a)). Here, we used the term diagnostically acceptable to refer to the fact that the attained plaque ROI PSNR values are above 35 dB that were shown to qualify for clinical practice in [8].

16-QAM 3/4 and 64-QAM 3/4 deliver diagnostically acceptable performance comparable to QPSK 1/2 only for the first location, which is situated closer to the BS as can be seen in location "1" in Figure 17 and leftmost boxplots of Figure 20(b). As the mobile station (ambulance) was moving away from the BS and signal attenuation was increasing, the quality of the video were significantly degraded. At 1 km from the BS, diagnostically acceptable average PSNR ratings were still obtained. The experienced PLR and the associated PLR standard deviation however, suggest that ultrasound video of unacceptable clinical quality transmitted at some occasions. This is more obvious at location "2" (1.1 km from BS) and 64-QAM 3/4, as shown in Figure 20(c). Here, it is important to note that the depicted results in Figure 20(a)-(b) are boxplots based on PSNR averages of 10 simulation runs of the ten videos parting the ultrasound video data set. On the other hand, Figure 20(c) depicts boxplots reporting the PSNR ratings for each of the ten simulation ran, for the video depicted in Figure 16(a). The latter case demonstrated the extreme channel conditions a mobile station was likely to experience when transmitting at the effective coverage zone of the BS. As shown in Figure 20(c), increased number of packet losses and high PLR standard deviation, as already discussed above, often result in diagnostically unacceptable PSNR ratings.

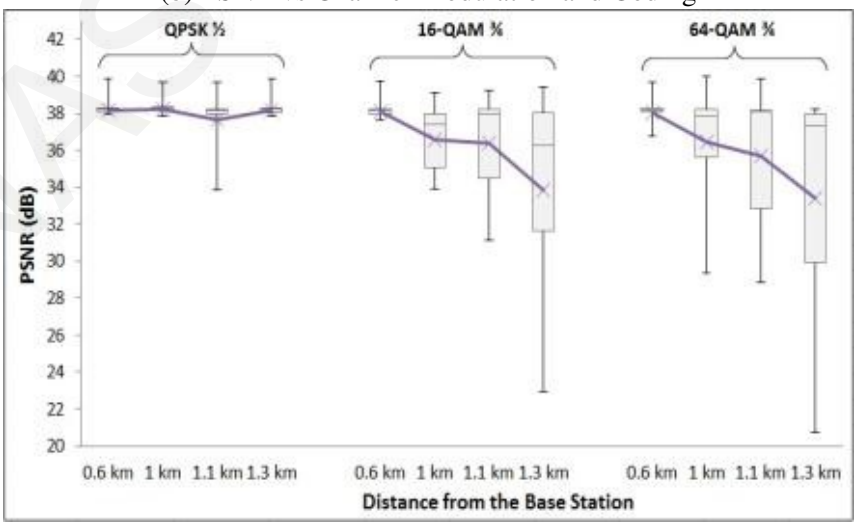
The latter observation was verified during the clinical evaluation where different video instances were clinically validated. At the furthest location, high PLR made the transmitted video




(a) PSNR vs Distance from BS




(b) PSNR vs Channel Modulation and Coding




(c) PSNR vs Distance from BS (video of Figure 19)

Figure 20: Scenario 2 QoS Evaluation: (a) Boxplots depicting the average PSNR ratings for the whole data set and for each channel modulation and coding scheme, as a function of the distance from the BS. (b) Boxplots depicting the average PSNR ratings for the whole data set and for each distance from the BS, as a function of the investigated channel modulation and coding schemes. (c) Boxplots depicting PSNR ratings for the ten random emulations of a single video (video shown in Figure 16 and used in Figure 19) for each channel modulation and coding scheme, as a function of the distance from the BS. Example from [121].

of limited clinical interest. Consequently, distances greater than 1km from the BS provide for a boundary case for this scenario.

The results from the objective evaluation significantly extended the findings of previous studies in the literature [25, 118, 129, 114]. Here, additional experimentation allowed investigating higher mobile speeds up to 100 km/h (compared to 50 km/h [25, 118]), distances between 150 m - 1.3 km from the BS (compared to 500m [25, 118]), hybrid vehicular multipath and path loss propagation models (compared to free space [129, 114] and vehicular models [118]), subcarriers scalability up to 2048 (compared to 512 [118] and 1024 [25]), for a data set composed of 10 ultrasound videos and an overwhelming number of investigated cases. Most importantly, the clinical capacity of the communicated ultrasound videos was supported by the clinical evaluation provided below.

### 4.4.3 Clinical Evaluation

*1. High Resolution Encoding*

Table 14 depicts the added clinical value linked with higher resolution medical video communication. A total of ten videos were displayed at both CIF and 4CIF resolutions. The medical expert was asked to comment on the clinical content of these two resolutions. Ratings were given in the range of 1 to 5 as can be seen in Table 5. A rating of 5 is the highest possible and it signifies that the diagnostic information in the decoded video is of essentially the same quality as the original video. A rating of 4 indicates that there is a diagnostically acceptable loss of minor details. At the lowest scale, a rating of 1 signifies that the decoded video is of unacceptably low quality. Based on previous knowledge, CIF resolution provided for evaluating the clinical criterion of plaque type, something which was not feasible with lower QCIF resolution. The findings verified

Table 14: The relationship between clinical criteria and Video Resolution. Example from [121].

| | **QCIF** | **CIF** | **4CIF** |
|---|---|---|---|
| *Plaque Detection* | √ | √ | √ |
| *Artery Stenosis* | √ | √ | √ |
| *Plaque type* | | √ | √ |
| *Plaque Morphology* | | | √ |

Table 15: Clinical evaluation for the investigated channel modulation and coding schemes as a function of the distance from the BS. Example from [121].

| Resolution: **4CIF**, Frame Rate: **15fps**, QP: **38/30/28**[a], BitRate: **1.5Mbps** | | | |
|---|---|---|---|
| | **QPSK 1/2** | **16-QAM 3/4** | **64-QAM 3/4** |
| **Location**[b] | **1/ 2/ 3/ 4** | **1/ 2/ 3/ 4** | **1/ 2/ 3/ 4** |
| *Plaque Detection* | 5/ 5/ 5/ 5 | 5/4.3/4.5/3.4 | 5/ 4.2/ 4.4/3.6 |
| *Artery Stenosis* | 5/ 5/ 5/ 5 | 5/4.1/4.2/3.4 | 5/ 4/ 4.3/3.6 |
| *Plaque Type & Morphology* | 5/ 5/ 5/ 5 | 5/4.1/4.2/3.3 | 5/ 4/ 4.3/3.5 |

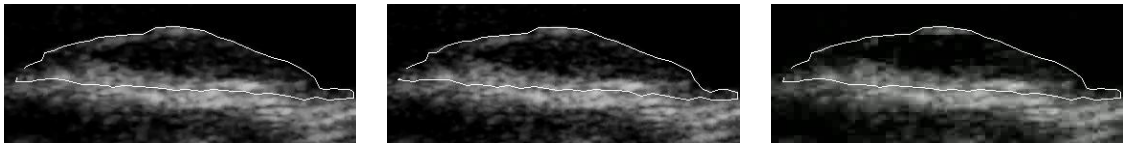
1: Lowest Score, 5: Highest Score

[a]QP are given in the order of background/wall ROI/plaque ROI, i.e. 38:background/ 30:wall ROI/ 28:plaque ROI.

[b]Distance from the BS: 1: 0.6 km, 2: 1.1 km, 3: 1 km, 4: 1.3 km.

the hypothesis that higher resolution is associated with communicating a larger amount of clinical information. Detailed assessment of plaque morphology is made possible for 4CIF resolution medical video. This was not always the case with lower, CIF resolution.

However, the key finding in these experiments was that 4CIF resolution closely matches the clinical capacity of the original video. Similar to [112, 113] the medical expert was also asked to rate whether the encoded video contained the same amount of clinical information as the original video. The medical expert concluded that the clinical information precision found in 4CIF resolution ultrasound video is comparable to that of the ultrasound device's monitor. It was associated with better assessment of the plaque motion and plaque components movement, which leads to confident assessment of plaque type and plaque morphology, aligned with diagnosing possibility of plaque rupture. Moreover, it facilitated better visualization of the intima of the near and far walls, of the plaque components, and the fibrous cap where this was applicable. In general, it was expected to reduce inter-observer variability. On the other hand, a higher frame rate than 15 fps may be required to rival in-hospital examination.



(a) (b) (c)

Figure 21: Video image examples of scenario 2 (pixel segmentation based on [123]). (a) Plaque ROI PSNR: 38.2 dB, channel modulation coding: QPSK $\frac{1}{2}$, distance from BS: 1.3 km, (b) Plaque ROI PSNR: 37.36 dB, channel modulation coding:16 -QAM 3/4, distance from BS: 1.1 km, (c) Plaque ROI PSNR: 29.34 dB, channel modulation coding: 16-QAM 3/4, distance from BS: 1.3 km. Example from [121].

*2. Scenario 2*

Table 15 summarizes the clinical evaluation of investigated locations of the afore-described scenario 2. It evaluates the capability of mobile WiMAX networks to communicate high-resolution medical ultrasound video. Here, we present mean opinion scores of a representative sample of the 120 instances (3 channel modulation and coding schemes x 4 locations x 10 simulation runs = 120 instances) of the video shown in Figure 16(a).

Videos decoded after transmission using QPSK 1/2 attained the highest clinical ratings. This is aligned with the objective assessment depicted in Table 13, and the associated high PSNR scores. 16-QAM 3/4 and 64-QAM 3/4 attained diagnostically acceptable ratings (marginal at 1.1 km from the BS) in all but the most distant (last) location, where they failed to qualify for clinical practice. Clearly, when the distance from the BS exceeds 1 km, a switch to a more robust channel modulation scheme will prevent clinical quality to fall below of what is acceptable. Figure 21 depicts video image examples of the investigated schemes. The artifacts in Figue 21(c) demonstrate the limits in trying to visualize plaque morphology at this larger distance. Even at 4CIF resolution, the plaque morphology cannot be visualized with 16-QAM 3/4 at a distance of 1.3km.

## 4.5 Conclusions

This study proposes an H.264/AVC-based framework for the wireless transmission of atherosclerotic plaque ultrasound video over mobile WiMAX networks. The depicted diagnostically driven encoding scheme shows that equivalent clinical quality can be obtained at significantly reduced bitrate demands. When combined with recent postprocessing error concealment techniques [114], it can provide for additional diagnostic resilience. Comprehensive experimentation showed that low-delay high-resolution 4CIF ultrasound video transmission is possible over mobile WiMAX networks, even at speeds of 100 km/h and distances of 1 km from the BS. The investigated channel modulation and coding schemes verified that QPSK 1/2 is the most robust scheme, especially when transmitting from locations with low SNR. On the other hand, 16-QAM 3/4 and 64-QAM 3/4 provide higher network capacities and are preferable when the transmitting station is closer to the BS. Based on these findings, and driven by the greater need to provide reliable and confident diagnosis throughout the video streaming session while optimally exploiting available resources, the adaptive video communication framework proposed in Section 5.4 was motivated. The performance of the system in terms of transmitted video's quality was evaluated using both objective and subjective evaluations. Clinical validation verified the capacity of mobile WiMAX networks to provide robust, clinically acceptable 4CIF ultrasound video transmission, thus enabling the transmission of ultrasound video at resolutions close to the original's video acquired resolution.

# Chapter 5

# Real-Time Adaptation to Time-Varying Constraints for Medical Video Communications

## 5.1 Introduction

Advancement and wider adoption of mobile health (mHealth) systems and services has been identified as a key priority both at a global and at the European level [2], [130]. The goal is to facilitate increased quality of care while reducing hospitalization times and associated healthcare costs. In terms of mHealth medical video communication systems, application scenarios range from remote patient monitoring, diagnosis, and care, to emergency incident response and medical education [3]. Once adopted in standard clinical practice, these systems are expected to transform healthcare delivery by fostering responsive emergency systems, in unrestricted and cross border settings that will significantly improve patient's quality of care and ultimately life expectancy.

Over the past decade, there has been increased interest in the development of low-delay and high-quality medical video communications systems [131], [132]. The development of the new systems has benefited from advances in video compression and wireless network technologies,

such as the emergence of the high efficiency video coding (H.265/ HEVC) standard [133], the deployment of 4G LTE-Advanced and ongoing research towards 5G systems [134].

Despite these advancements, mHealth video communication systems have not experienced wider clinical usage in spite of successes of proof-of-concept and pilot studies [121, 8, 18, 135, 19, 136, 115, 119, 137, 120, 81, 138]. This is partly due to the fact that most studies are modality-specific, rely on specific video compression and wireless technologies, do not address computational complexity, and more importantly, do not facilitate efficient adaptation to real-time video communication constraints. Furthermore, beyond constraints on the communications channels, there is also a strong need to provide real-time or faster video encoding while guaranteeing an acceptable level of clinical video quality.

The goal of the work in this chapter is to provide methods for real-time end-to-end systems that dynamically adapt to time-varying channel state while maximizing clinical video quality or guaranteeing a minimum, acceptable level of clinical video quality. The latter is expected to expedite the adoption of mHealth medical video communications in standard clinical practice by preserving the quality thresholds for remote diagnosis and decision making.

The requirement for real-time communications implies that the video encoding rate needs to be higher than the communicated frame rate. Otherwise, the proposed system will not be useful for interactive video communications. Furthermore, by maintaining a minimum level of video quality [139], we guarantee that the communicated video will always meet or exceed diagnostic quality requirements. On the other hand, when the available bandwidth is low, we still require that the system delivers the maximum possible clinical quality, since the proposed system will adapt to the available data rate. In the opposite end, non-adaptive systems would simply fail, causing a video transmission interruption until the available bitrate matches the rate to which they transmit. Low-bandwidth scenarios are particularly important for emergency medical video communications.

Related research focuses on switching between pre-encoded video content in response to available bandwidth [18, 135, 19, 115, 113, 138]. These approaches cannot guarantee real-time encoding, neither minimum clinical video quality levels can be secured against bandwidth fluctuations for which a pre-encoded instance has not been foreseen. In other words, except for bandwidth constraints, there are no constraints considered on encoding time and clinical video quality throughout a remote ultrasound video streaming session. More importantly, encoding time is never considered as an optimization objective. Instead, in this thesis we consider approaches that are jointly optimal in the performance-bitrate-quality space and effectively adapt to real-time objectives' constraints.

Our approach is based on multi-objective optimization of video quality (application-modality level adaptation), bitrate demands (wireless network adaptation), and encoding frame rate (device adaptation for real-time operation). Thus, we only consider encodings that simultaneously maximize video quality and encoding rate, while minimizing the required bitrate. The required space of solutions forms a Pareto front that is used for solving constrained multi-objective optimization problems.

To describe our approach, let $VQ$ denote video quality (currently expressed in terms of SSIM), $B_r$ denotes the bitrate demands, and $FPS$ denotes the encoding rate in terms of frames per second. Our approach seeks optimal solutions subject to realistic constraints on video quality $VQ \geq VQ_{min}$, bitrate demands $B_r \leq B_{r,max}$, and encoding frame rate $FPS \geq FPS_{min}$. We thus seek the set of encoding configurations $E_n$ that are optimal in the multi-objective sense given by:

$$\max(\mathrm{VQ}(\mathrm{E_n}), -\mathrm{B_r}(\mathrm{E_n}), \mathrm{FPS}(\mathrm{E_n})) \tag{9}$$

The set of solutions of equation 9 identifies the set of Pareto optimal encoding configurations. Here, we note that: (i) quality directly affects the clinical capacity of the video, (ii) bitrate determines candidate wireless networks that support data rates able to accommodate the encoded video

for real-time transmission, and (iii) encoding rate measures the device's capabilities of supporting real-time encoding and also provides an indication of the required power/energy consumption.

In general, the relationship between video encoding and energy consumption can be very complex (e.g., see [140]). In the simplest model, we have that energy is given by $E = P_{avg} \cdot t$ where $P_{avg}$ denotes the average power consumption and $t$ denotes the total encoding time. Now, assuming that we have the same $P_{avg}$ for the different encodings, for encoding N video frames, we have that $t = N/FPS$ which gives that $E = N \cdot P_{avg}/FPS$. Thus, we have an inverse relationship between the total energy and number of encoded frames per second. Beyond the energy required for video encoding, it is important to note that we also have significant savings in transmitting videos at lower bitrates.

### 5.2 Semi-Adaptive Medical Video Communication Systems

A limited number of studies investigated the robustness in terms of clinical video quality facilitated by mHealth video communication systems that are capable of adapting to time varying constraints. As depicted in Table 16, these studies mostly relied on a Markov decision process triggering a switch to a pre-encoded video instance based on cross-layer information. Pre-encoded video instances or precomputed states refer to videos encoded offline with different configurations that correspond to different quality (e.g. Peak Signal-to-Noise Ratio (PSNR) or SSIM) and bitrate demands. This concept is similar to scalable video encoding (SVC) which provides quality scalability by encoding videos at different rates [76]. SVC has been widely studied in the literature for adaptive video streaming of general purpose and social videos [141].

Table 16: Literature Review - Semi-Adaptive Medical Video Communication Systems

| **Author** | **Year** | **Resolution, Frame Rate, Bitrate** | **Encoding Standard** | **Wireless Network** | **Adaptation Methodology** | **Adaptation Approach** | **Medical Video Modality** | **Video Quality Assessment** |
|---|---|---|---|---|---|---|---|---|
| Istepanian et *al.* [135] | 9 | 176x144 @ 8-10fps | H.264/ AVC | 3.5G and beyond | Markov Decision Process | Available bandwidth | Abdomen Ultrasound | PSNR |
| Martini et *al.* [136] | 10 | 480x256 @ 15fps 300 Kbps | H.264/AVC | Mobile WiMAX | Pre-encoded, diagnostic ROI (d-ROI) | Available bandwidth | Cardiac Ultrasound | PSNR/SSIM |
| Debono et *al.* [114] | 12 | 640x480 @ 25fps | H.264/AVC | Mobile WiMAX | Pre-encoded, d-ROI | Wireless SNR | Cardiac Ultrasound | PSNR |
| Khire et *al.* [119] | 12 | 720x480 @ 30fps, 125 – 200 Kbps | H.264/AVC | 3G and beyond | Pre-endoded, d-ROI | A mentor changes quality of ROI | Maxillofacial Surgery Clips | PSNR |
| Alinejad et *al.* [115] | 12 | 704x480 @ 20fps | Windows Media Video2 | HSUPA, mobile WiMAX | Markov Decision Process | Environment's state | Ultrasound | PSNR |
| Ognenoski et *al.* [142] | 13 | 636x420 400 Kbps - 1Mbps | Not defined | 3.5G and beyond | MPEG-DASH | Several technical parameters | Ultrasound | PSNR |
| Cicalò et *al.* [138] | 16 | 720x576 @ 25fps | H.264/ AVC | LTE | Scalable Video Coding | Available bandwidth | Disaster and Emergency | PSNR/SSIM |
| Balaji et *al.* [143] | 17 | 640x480 @ 30 fps | H.264/AVC | LTE-Advanced | Scalable Video Coding | Available bandwidth | Emergency | PSNR |
| Hosseini et *al.* [144] | 17 | Not defined | Not defined | 2G to 4G | MPEG-DASH | Available bandwidth | Emergency | Own metric |
| Antoniou et *al.* [145] | 17 | 560x448 @ 40fps | HEVC | Adaptive | Adaptive | Time-varying constraints | CCA Ultrasound | Clinical (subjective) and objective evaluation |

In [146], [147], encoder control in resource constraint contexts have been studied, employing rate-distortion optimization (RDO) to select the optimal coding mode while reducing the computational complexity. The current thesis goes significantly beyond such prior studies by considering diagnostic quality, image and video quality metrics beyond PSNR, and joint optimization and control in the quality-rate-performance space.

In cross-layer architectures [129], systems typically abstract information from the medium access control (MAC), physical (PHY), and application layers. The former two characterize the wireless medium state, while the latter defines requirements that relate to the underlying medical modality. Then, during real-time communications the objective is to trigger a switch to a precomputed state that meets the available bandwidth while conforming to a clinical quality threshold.

The emerging MPEG-DASH standard [148] responds to changes in the wireless networks' state by triggering a receiver initiated switch to a different video instance, demonstrating promising results [149]. In contrast to the proposed approach, there is no guarantee that any of these instances correspond to optimal video encodings. Also, video packet retransmission using adaptive HTTP streaming can only be performed within very short time periods.

In [18], for pediatric respiratory distress videos, the authors consider switching between six precomputed states in response to changes of available bandwidth. To both enhance video's diagnostic capacity and preserve network resources, a diagnostically relevant encoding approach was employed, where the region of diagnostic interest (d-ROI) was encoded in higher quality than the background. Diagnostically relevant encoding based on d-ROIs [131] was also investigated for atherosclerotic plaques [8], [121] and cardiac [136] ultrasound videos, as well as trauma videos [119].

Pre-encoded states were also considered in [150]. The five employed encoding configurations, besides video quality, also controlled the degree of the error resilience. In [129], the authors

proposed a reinforcement Q-learning algorithm based on a finite discrete-time Markov decision process. A total of 216 states were considered with 27 possible actions (switches) per state. The optimal switch was selected based on minimizing a cost function which took into account video quality and communication network's available bandwidth (deduced from round-trip time (RTT)). In [135], using the same reinforcement algorithm, the authors considered frame rate adaptation that maintained a predefined clinical quality threshold. Both studies considered a limited sample of cardiac ultrasound videos. A Markov process was also considered in [113], where the objective was to determine the wireless channel's state in terms of packet loss rates, so as to trigger retransmissions that would enhance the ultrasound video's robustness to wireless transmission errors.

Data prioritization for jointly streaming multiple videos (i.e., ambient and ultrasound videos in emergency incidents) was considered in [138]. The objective was to meet the available throughput by imposing priority weights on bandwidth utilization of the different video feeds, based on the application scenario. For each video, four different rate points (encodings) were considered using SVC.

Based on the afore-described approaches, adapting to a time-varying wireless channels' state becomes a problem of conforming to the available bandwidth, while application level adaptation refers to meeting clinical quality criteria. Thus, there is a great need to consider the problem as a multi-objective optimization problem. Our proposed approach provides a multi-objective optimization framework with specific constraints on quality, bitrate, and encoding time. Multi-objective optimization allows us to reject sub-optimal encodings that could incur substantial overhead with no measurable benefits. Moreover, the proposed approach supports real-time adaptation to time-varying constraints on quality, bitrate, and encoding time as opposed to the current literature focus on bandwidth constraints. Conforming to time-varying constraints in a real-time fashion is essential for real-life applications and a decisive factor in clinical practice adoption.

## 5.3 An Introduction to Multi-Objective Optimization

Multi-objective optimization is a field of optimization that involves more than one objective functions to be optimized simultaneously. Many real-world optimization problems are ideally suited to be modeled using multiple conflicting objectives. For instance, in engineering by minimizing the error in both the position and the orientation of a robotic arm [151]. Consequently, such approach is of significant practical importance.

In mathematical terms, a multi-objective optimization problem can be formulated as:

$$\mathrm{P}: \ \min \mathrm{f_m(x)} \text{ subject to } \mathrm{x} \in \mathrm{S} \tag{10}$$

where $P$ is the name of the problem, the integer $m \geq 2$ is the number of objectives and the set $S$ is the feasible solution space, which typically contains the constraint functions [152], [153], [154]. The ideal vector-valued feasible solution $z^*$ is given as:

$$\mathrm{z}^* \ = \ \mathrm{f}^* \ = \ (\mathrm{f}_1^*, \mathrm{f}_2^*, ..., \mathrm{f}_\mathrm{M}^*)^\mathrm{T} \tag{11}$$

However, in multi-objective optimization, there is not a feasible solution that minimizes all objective functions at the same time. For this reason, the Pareto-optimality concept [152] has been intensively used in the literature and has significantly contributed in the elaboration of a large set of works. Thus, a set of solutions exists, namely Pareto-optimal set, where all points are non-dominated solutions. This means that a solution is the best that could be achieved for one objective without disadvantaging at least one other objective. More specifically, a point $x_0$ is called Pareto optimal for the problem $P$ if $x_0 \in S$ and there is no $x_0 \neq x \in S$ with $f_m(x) \leq f_m(x_0)$ for $m = 1, 2, ..., M$, with a strict inequality for at least one $m$, $1 \leq m \leq M$.

In there literature, there are numerous available multi-objective optimization techniques based on different approaches, like biology inspired algorithms [155]. This category includes evolutionary algorithms which emulate the survival of the fittest process of natural ecosystems, and

swarm based algorithms, that mimic the collective behavior of populations. For example, behavior of honey bees like Artificial Bee Colony Algorithm (ABC) [156], and of ant colonies like Ant Colony Optimization Algorithm (ACO) [157]. Concerning evolutionary algorithms, many classical genetic algorithms and differential evolution algorithms were developed, including Non-dominated Sorting Genetic Algorithm (NGSA) [158], Niched Pareto Genetic Algorithm (NPGA) [159], Pareto Archived Evolution Strategy (PAES) [160], Pareto Envelope-Based Selection Algorithm (PESA) [161], NPGA2 [162] and NSGA-II [163]. Some other multi-objective optimization techniques, are based on physics [155], geography or social culture approaches.

To the best of our knowledge, there are no other related studies on adaptive medical video encoding communication frameworks. A brief literature review of semi-Adaptive Medical Video Communication Systems is depicted in Table 16.

### 5.4 Methodology

We present a top-level description of the proposed method in Figures 23 and 24. As described in pseudo-codes in Figure 23(a) and (b) and shown in Figure 23(c), a training set is used to estimate the objective models for quality, bitrate, and encoding time as functions of the video encoding configuration. Then, as depicted in Figure 24, for each GOP, we retrieve the constraint optimization mode and compute the possible configurations that satisfy the current constraints. We then use the estimated objective models to determine Pareto-optimal configurations that are used to adaptively encode the video.

The proposed system architecture is depicted in Figure 22.

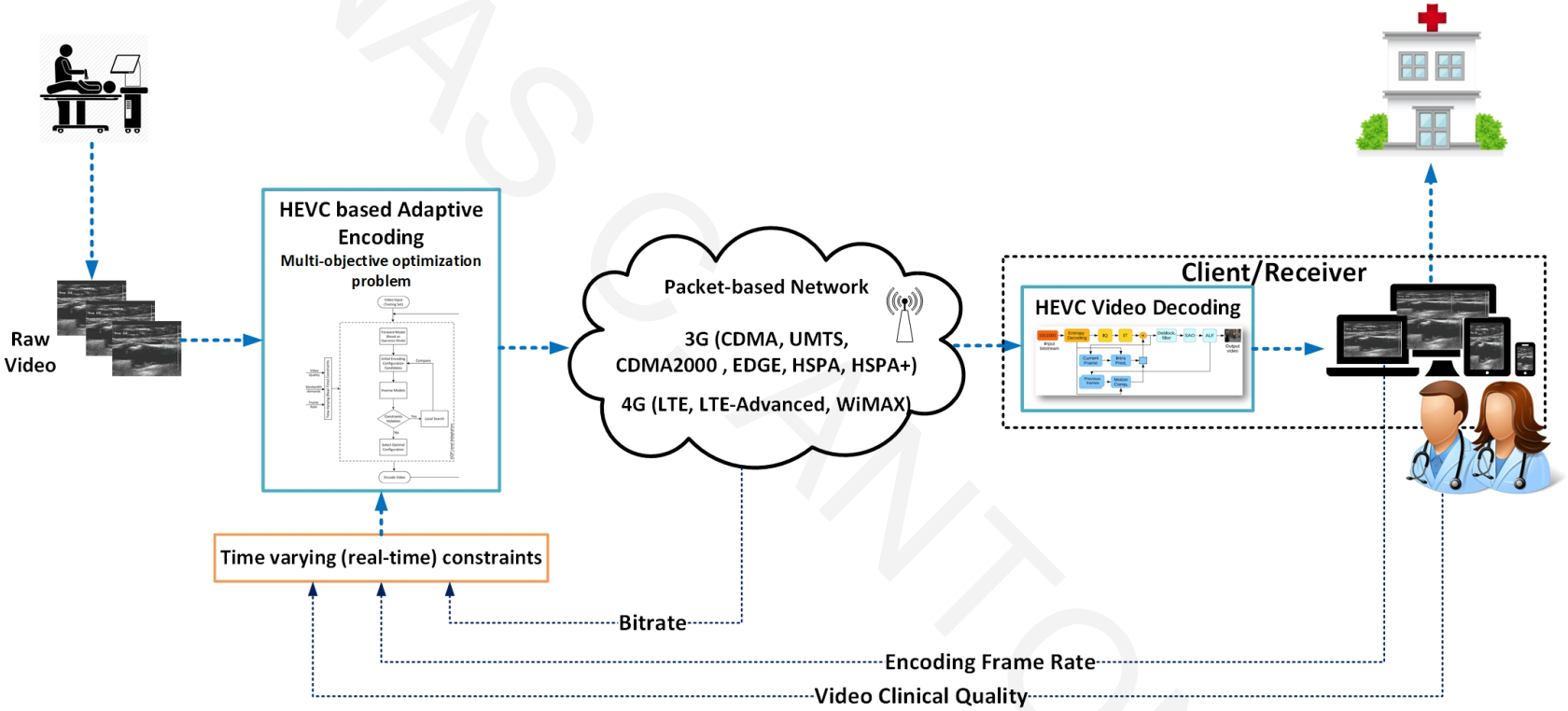



Figure 22: Proposed system architecture.

### 5.4.1 Adaptive Video Encoding

We begin with a summary of the constraint optimization modes that we seek to solve using adaptive video encoding [164].

#### 5.4.1.1 Maximum video quality mode

We define the maximum video quality mode using:

$$\mathrm{maxVQ} \quad \mathrm{subject} \quad \mathrm{to} \quad (\mathrm{B_r} \leq \mathrm{B_{r,max}}) \quad \mathrm{and} \quad (\mathrm{FPS} \geq \mathrm{FPS_{min}}) \tag{12}$$

where the goal is to encode the video at the maximum possible quality level without exceeding available bandwidth ($B_{r,max}$) while maintaining an acceptable encoding frame rate ($FPS_{min}$).

#### 5.4.1.2 Minimum bitrate demand mode

We define the minimum bitrate mode using:

$$\mathrm{minB_r} \quad \mathrm{subject} \quad \mathrm{to} \quad (\mathrm{VQ} \geq \mathrm{VQ_{min}}) \quad \mathrm{and} \quad (\mathrm{FPS} \geq \mathrm{FPS_{min}}) \tag{13}$$

where the goal is to minimize bandwidth demands as long as the video is of sufficient video quality ($VQ_{min}$) and an acceptable encoding frame rate ($FPS_{min}$) is maintained.

#### 5.4.1.3 Maximum performance mode

We define the maximum performance mode using:

$$\mathrm{maxFPS} \quad \mathrm{subject} \quad \mathrm{to} \quad (\mathrm{VQ} \geq \mathrm{VQ_{min}}) \quad \mathrm{and} \quad (\mathrm{B_r} \leq \mathrm{B_{r,max}}) \tag{14}$$

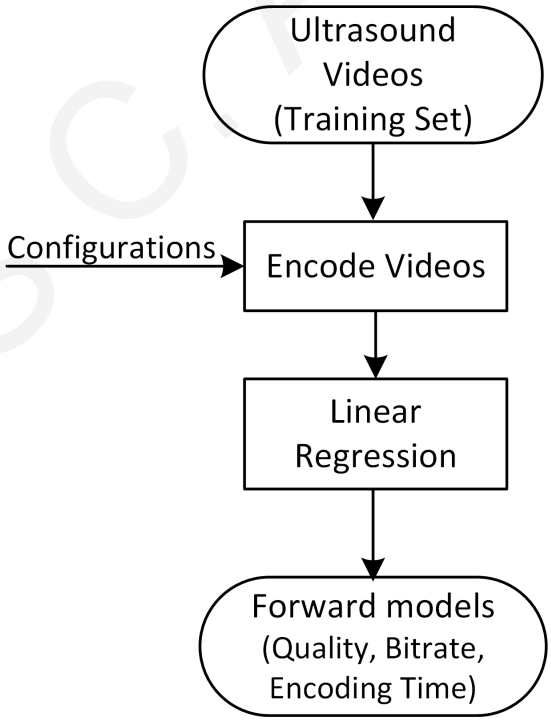
where the goal is to maximize encoding frame rate denoted FPS (thus minimizing encoding time) while conforming to clinically acceptable video quality ($VQ_{min}$) and not exceeding available bandwidth ($B_{r,max}$).

```
function VideoEncoding(Vd, Cnf)
▷ Input: Input Videos V, Video Database Vd,
    Configuration C, Configuration Files Cnf
▷ Output: Parameter Values Pv
    for (∀ V ∃ Vd) do
        for (∀ C ∃ Cnf) do
            Encode V
            Pv ← QP, VQ, B_r, FPS
        end for
    end for
end function
▷ QP refers to Quantization Parameter, VQ to Video Quality,
  B_r to Bandwidth and FPS to Encoding Frame Rate
```

(a): Pseudo code of video encoding

```
function ForwardModels(Pv)
▷ Input: Parameter Values Pv
▷ Output: Forward Models Fm
    for ({VQ, B_r, FPS} in Pv) do
        Fm_VQ ← LR(Pv_VQ)
        Fm_Br ← LR(Pv_Br)
        Fm_FPS ← LR(Pv_FPS)
    end for
end function
▷ QP refers to Quantization Parameter, VQ to Video Quality,
  B_r to Bandwidth, FPS to Encoding Frame Rate and LR
  to Linear Regression
```

(b): Pseudo code of forward models

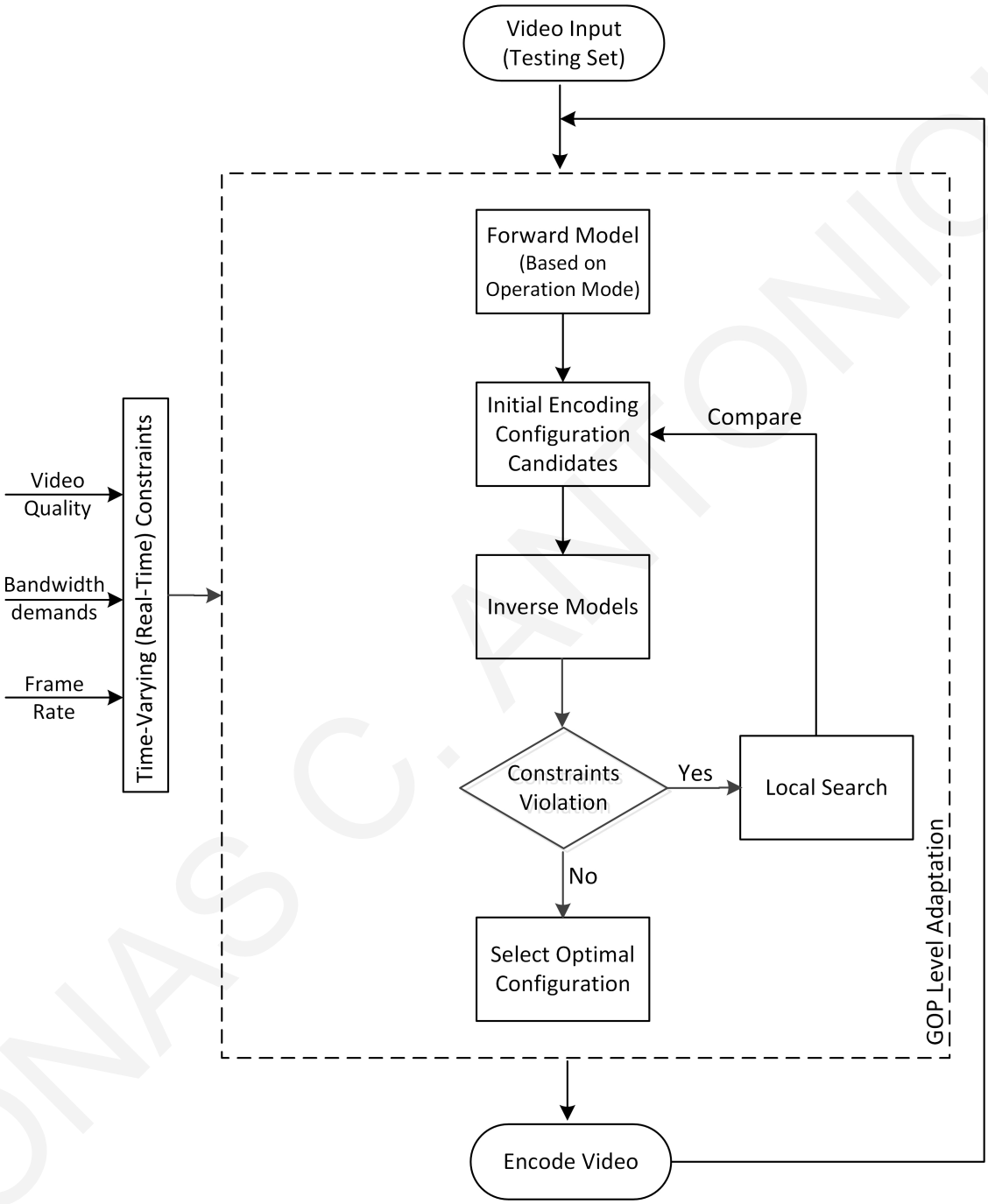



(c): System diagram

Figure 23: Forward models estimation for SSIM quality, bitrate demands, and encoding rate using linear regression.

```
1: function AdaptiveEncoding ( )
2: ▷ This procedure adaptively encodes a video stream
3:     while (more GOP video frames to encode) do
4:         Retrieve constraints and optimization mode
5:         Compute possible configurations sets C1, C2
6:                 and QP range sets QP_r1 and QP_r2
7:                 that satisfy the active constraints within
8:                 certain percentage tolerances.
9:         Combine configuration sets and QP ranges into
10:                candidate sets C_all and QP_all.
11:        Compute predicted objective values:
12:                SSIM_all, FPS_all, B_all
13:                by applying the regression models to
14:                candidate sets C_all and QP_all.
15:        Compute Pareto-front by eliminating
16:                points whose objectives are not
17:                Pareto-optimal.
18:        Solve the constrained optimization problem by
19:                selecting the C_opt and QP_opt that
20:                produce points that lie on the Pareto-front.
21:        Encode the video using C_opt and QP_opt.
22:    end while
23: end function
```

(a) Pseudo code



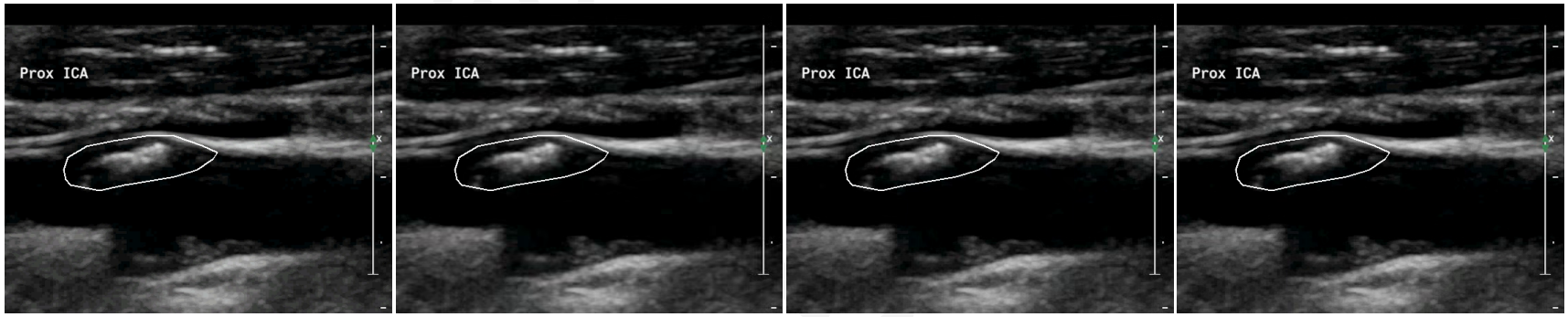

(b) System diagram

Figure 24: Real-time adaptive encoding framework based on time-varying constraints.

### 5.4.2 Video Database Clinical Evaluation

For validating the proposed approach, we used a data set of ten ultrasound videos of atherosclerotic plaques located in the common carotid artery. Here, we note that our proposed methods are developed to be modality independent. The models should be applicable to other medical imaging modalities (e.g., endoscopic surgery videos), provided that we also provide clinical validation.

Extracted video image examples of a typical video appear in Figure 25. The videos record the left or right carotid artery of both symptomatic and asymptomatic subjects. For more information regarding carotid artery disease we refer to [165]. The raw ultrasound videos were acquired using a frame grabber [166] connected to a Philips ATL 5000 ultrasound machine during a routine clinical exam at a video resolution of 560x448@40fps. Clinical evaluation was based on the clinically established protocol defined in [8], [121]. The protocol requires the assessment of the impact of the compression on clinical diagnosis, as we shall document in the results in Section 5.5.




(a) (b) (c) (d)

Figure 25: Ultrasound video image examples of *maximum video quality mode* example described in Section 5.5.4.1. Atherosclerotic plaque segmentation is outlined by white lines [8], [81]. Video Resolution: 560x448, Frame Rate:40 fps. (a) Original (uncompressed) video image, (b) SSIM: 0.921, Bitrate: 290 kbps (constraint: $\leq$ 300 kbps), FPS: 48.29 (constraint: real-time encoding, i.e., $\geq$ 40 fps), QP: 31, Encoding Structure: B4, (c) SSIM: 0.949, Bitrate: 884 kbps (constraint: $\leq$ 1000 kbps), FPS: 46.63 (constraint: $\geq$ 40 fps), QP: 26, Encoding Structure: B2, (d) SSIM: 0.961, Bitrate: 1873 kbps (slight violation of constraint: $\leq$ 1800 kbps), FPS: 43.98 (constraint: $\geq$ 40 fps), QP: 23, Encoding Structure: B2.

More specifically, the clinical protocol requires the visual assessment of (a) *Artery Stenosis*, which assesses the ability to accurately compute the degree of the artery stenosis, (b) *Plaque Morphology*, which involves determining plaque characteristics such as texture [165] that translate to assessing the possibility of a plaque rupture, and (c) *Plaque Motion*, which evaluates within plaque but also near and far wall motions that affect the degree of stenosis but primarily how prone the plaque is to rupture and hence a stroke incident. The clinical ratings range from one (1) to five (5), which correspond to the lowest and highest possible ratings, respectively. A rating of four (4) translates to diagnostically lossless compression. In other words, despite the inherent loss of some information attributed to compression, the ultrasound video maintains its original clinical capacity as it carries equivalent clinical information as the original, uncompressed video, enabling the medical expert to provide a confident diagnosis [139]. On the other hand, a rating of three (3), signals that there is some loss of clinical information in the compressed video, prohibiting its use for clinical practice. The reader is referred to [8] and [121] for more details regarding the employed clinical evaluation protocol.

### 5.4.3 Video Encoding Configurations

To support real-time video communications, we investigated new HEVC encoding instances based on the ultrafast preset of the x265 1.7 encoder, using the HEVC main profile, as depicted in Table 17. Furthermore, different encoding frame structures were considered by varying the number of consecutive B-frames, namely IPPPP (zero latency or $ZL$), IPBBP ($B2$), and IPBBBBP ($B4$). All video sequences were encoded using a group of pictures ($GOP$) size of 96. An instantaneous decoder refresh ($IDR$) picture was inserted at the beginning of each $GOP$ (to eliminate error propagation from erroneously received packets between GOPs and facilitate smooth adaptation during real-time video streaming), followed by 32 P-frames having 2 consecutive B-frames

Table 17: New HEVC Configurations for Ultrafast preset.

| ***Parameter*** | ***Value*** | ***Parameter*** | ***Value*** |
|---|---|---|---|
| Profile | Ultrafast | Frame Threads | 8 |
| Encoding Structure | ZL (IPPP), B2, B4 | SAO | On/Off |
| GOP structure | Open/Close | DBF | On/Off |
| QP | 16-36 | Tuning | zerolatency |
| Max intra period | 96 | CTU | 64 |
| ***Total configurations per video*** | | | ***252*** |

$ZL$ = Zero latency (IPPP), $B2$ = 2 consecutive frames (IPBBP), $B4$ = 4 consecutive frames (IPBBBBP), $SAO$ = Sample Adaptive Offset, $DBF$ = Deblocking Filter, $CTU$ = Coding Tree Unit.

in their midst for B2, 19 P-frames having 4 consecutive B-frames in their midst for B4 and 95 P-frames for ZL structures. All configurations relied on 1 reference frame, while both open and closed $GOP$ structures were considered. Compression levels were varied based on the employed quantization parameter ($QP$), ranging between 16 and 36 (step size of 1).

We set the coding tree unit ($CTU$) that replaced the macroblock ($MB$) unit of previous video coding standards to the maximum of 64 x 64 (width x height) pixels (compared to standard 16 x 16 pixels MB size of H.264/AVC and previous standards). The larger the maximum coding unit ($CU$) size, the more efficiently HEVC encodes flat areas of a picture, providing large reductions in bitrate. However, this comes at a loss of parallelism with fewer rows of CUs that can be encoded in parallel. On the other hand, decoding becomes faster. Thus, increasing the $CTU$ size results in a slight decrease in encoding rate compensated by finer quality and faster decoding time.

$HEVC$ is the first video coding standard to officially introduce parallel encoding tools, despite the fact that H.264/AVC implementations like x264 involve a significant amount of parallelism. Frame parallelism was investigated in previous work by our group, by considering different frame threads per video sequence, namely 1, 2, 4, 6 and 8 [167]. In this study, we selected 8 frame threats for parallel encoding, as this value was found to optimize encoding speed with respect to CPU utilization on our encoding platform (described in Section 5.5). Frame threads value is encoding platform specific.

Also new in $HEVC$ standard, are the sample adaptive offset ($SAO$) and deblocking ($DBF$) filters that were examined jointly, by simultaneously enabling or disabling their operation.

### 5.4.4 Estimating Forward Prediction Models

The objective here is to determine forward models that will predict the encoded video's state, namely video quality, bitrate, and encoding frame rate (see 9). Our approach is to generate these models using the offline process summarized in Figure 23, and then use the models to adapt to time-varying constraints during real-time operation. The latter will alleviate the need of accessing any of the precomputed videos described in 5.4.3.

To determine a suitable model for each of the afore-described encoded video characteristics, we use linear model estimation that also includes logistic regression, as provided by the R Statistical Package [168]. The goal here is to estimate each state (SSIM, bitrate, and encoding frame rate) as a function of the video encoding configurations provided in Table 17. Consequently, for each model, we use a total of 252 different configurations per video.

Next, we employ logistic linear regression to construct the forward models that will predict the quality, the bitrate and the encoding frame rate. In terms of model order, we consider both linear and logistic regression models up to $4^{th}$ order polynomials. While constructing a model, we also consider more than one parameter as predictor variables (e.g. QP and SAO and DFB filters), that may have a strong impact to response variable, here quality, bitrate and encoding frame rate respectively. By default, regression analysis generates a model's equation to describe the statistical relationship between one or more predictor variables and the response variable. Thus, for each model we evaluate the resulting p-values for each predictor variable to measure its significance on the response variable. It's important to note that a small p-value (typically $\leq 0.05$ [169]) suggests that changes in the predictor are strongly associated with changes in the response, while a large

(insignificant) p-value suggests that changes in the predictor are not associated with changes in the response. Based on regression results, QP value has lower p-value than the filters, thus it's the most significant predictor variable that implements different quality levels.

To avoid overfitting with more complex models, we use stepwise regression that minimizes the cross-validated, residual error while accounting for model complexity.

### 5.4.5 Adaptive Video Encoding using Inverse Prediction Models and Local Search

To encode each video stream, we employ the adaptive encoding algorithm described in pseudocode format in Figure 24(a), and schematically depicted in Figure 24(b). Each GOP is adaptively encoded so as to solve one of the constraint optimization modes that was given in Section 5.4.1.

Based on time-varying constraints and selected optimization mode, the corresponding configuration sets and candidate compression levels are computed. For example, for the maximum video quality mode of operation, bitrate and encoding time constraints are retrieved. Then, for each of the candidate forward models for bitrate demands and encoding frame rate, an inverse process is employed to predict the optimal encoding parameters values. For this purpose, we use the Newton's algorithm [170].

Newton' algorithm or Newton-Rapson method is a method for finding successively better approximations to the roots (or zeroes) of a real-valued function. Let's define an $f$ function over real numbers $x$, where:

$$\mathrm{f(x)} = 0 \tag{15}$$

Next, we compute the function's derivative $f'$ and we set an initial guess $x_0$ for a root of the function $f$. In our algorithm, we set $x_0$ equal to 25, since is the median value of QP range. In case the function satisfies the assumptions made in the derivation of the formula and the initial guess is

close, then a better approximation for $x_1$ is:

$$\mathrm{x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}} \tag{16}$$

Then, the process is iterated as:

$$\mathrm{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}} \tag{17}$$

Until we have a tolerance lower than $1 \cdot 10^{-12}$ [171]. Since QP value is a discrete number, we then round the generated real QP value.

In Newton's method, we allow mild violations per input constraint. These violations were introduced so as to address the forward models' prediction error which is analytically described in Section 5.5.2. More precisely, we allow mild violation in the order of 10% for bitrate demands and encoding frame rate, and 0.5% for video quality, respectively. The latter typically generates more than one solutions in terms of QP.

For this example, we set bitrate as the most dominant constraint, provided that exceeding available bandwidth would put video communication at risk of failure, whereas a computationally stronger machine would provide the desired encoding rate. So, we first employ Newton's method to generate QP values that correspond to 90%, 95% and 100% bitrate constraints. Thus, we get maximum three QP values if we don't have bitrate constraint violation. In the next step, we use these QP values in encoding frame rate model as initial values in Newton's method and then we generate three new QP values. We get a maximum set of 27 QP values. Specifically, we have 3 models - bitrate/frame rate/quality - and for each one we have 3 different GOP structures, namely ZL, B2 and B4. Additionally, for each constraint, we allow 3 mild violations. For this example, we considered both filters - Sample Adaptive Offset and Deblocking Filters - off.

We then combine all possible QP solutions and candidate configurations and use the forward models to predict SSIM quality, bitrate demands, and encoding frame rate triplet values per encoding instance. In case we have any constraint(s) violation, we perform a local search to generate new QP value sets. Here, we note that the use of models is computationally very efficient since it eliminates the need to re-encode the video. In what follows, we compute the Pareto front so that all resulting candidate configurations are optimal in the multi-objective sense. It's important to note that we perform the afore-described process for the three different GOP structures, namely ZL, B2 and B4. Thus, we have a set of QP values for each GOP structure. Finally, we solve the constrained optimization problem by selecting the optimal solution from the Pareto front, from the three different GOP structures, that maximizes or minimizes the selected objective while satisfying the time-varying input constraints.

### 5.5 Results and Discussion

The method was implemented using the x265 open source software [172] (most efficient open-source HEVC implementation to date) running on a Ubuntu 14.04.4 LTS/Linux 64-bit platform with 48GB RAM using two AMD Opteron(TM) Processor 6276 CPUs with 16 cores (16 threads) running at 1.4 GHz each.

In what follows, we summarize the benefits of considering different encoding configurations in the proposed adaptive framework, describe the resulting prediction models, highlight the significance of using Pareto optimal solutions, and demonstrate adaptive video encoding efficiency compared to static approaches using real-life examples for input constraints.

### 5.5.1 Average Improvements for ZL, B2, and B4 Encoding Structures with and without HEVC Filtering

To demonstrate the advantages of the proposed approach, we consider 3 $GOP$ encoding configurations ($ZL$, $B2$, and $B4$) that can be used with or without $HEVC$ filtering ($SAO$ and $DBF$ off or on). For each optimization mode, for the ten videos in our dataset, we report the average improvements based on the use of the optimal configurations over the remaining configurations in Table II, for a reduced configuration space (selected $QP$s).

As Table 18 depicts, there is not a significant difference between different encoding configurations for $SSIM$ values for $QP$s lower than 32. The most dominant encoding parameter here, which also holds for bitrate demands and encoding rate is the compression level (different $QP$ values). On the other hand, the optimal encoding configuration can produce significant bitrate reductions that range from 13.5% to 22.5%. This is also the case for encoding rate, documenting a substantial variation between 28%-30.5% throughout the range of examined $QP$s.

Since the $QP$ is held constant for each row of Table 18, it is deduced that all improvements are due to the selection of the optimal encoding structure and filtering option. Clearly, these fixed $QP$ improvements cannot be realized with the standard use of fixed structures and filtering options. On the other hand, along each row, the lack of substantial improvements in video quality suggests that $QP$ is the dominant parameter to use for controlling encoded video's quality. Furthermore, it is clear that adjusting the $QP$ will also result in dramatic changes in bitrate and the encoded rate. We thus need to use suitable forward prediction models to capture these complex relationships.

### 5.5.2 Forward Prediction Models Generation and Error Distribution

For fitting the models, we examined the use of linear, quadratic, and cubic polynomials using polynomial regression and logistic-polynomial regression. To avoid overfitting, we use stepwise

regression with leave-one-out to evaluate the performance of the different models. The best and most robust models were found to be the logistic linear models that will be described next. The logistic linear models improved prediction accuracy by at least 5% over non-logistic models.

Let *i* be used to index the six encoding configurations as a function of the GOP coding structure ($ZL$, $B2$, and $B4$) for $SAO$ and $DBF$ filters on and off. The objective functions are then expressed as functions of $QP$ as given by:

$$\log(\mathrm{SSIM_i}) \quad = \quad \mathrm{a_{i,0}} \cdot \mathrm{QP} + \mathrm{b_{i,0}} \tag{18}$$

$$\log(\mathrm{Bitrate_i}) \quad = \quad \mathrm{a_{i,1}} \cdot \mathrm{QP} + \mathrm{b_{i,1}} \tag{19}$$

$$\log(\mathrm{FPS_i}) \quad = \quad \mathrm{a_{i,2}} \cdot \mathrm{QP} + \mathrm{b_{i,2}} \tag{20}$$

where $a_{i,0}, a_{i,1}, a_{i,2}$ denote the $QP$ coefficients, $b_{i,0}, b_{i,1}, b_{i,2}$ denote constants, and 0, 1, 2 are used to index the $SSIM$, Bitrate, and $FPS$ objectives respectively.

Table 18: Average improvements for Image Quality ($SSIM$), Bitrate reduction ($Kbps$), and Encoding Framerate ($FPS$). We consider 3 $GOP$ configurations ($ZL$, $B2$ and $B4$) with and without filtering ($SAO$ and $DBF$). We have a total of 3x2=6 encoding configurations for each $QP$.

| ***QP***[a] | ***Average best SSIM*** | ***ΔSSIM*** | ***ΔSSIM% Difference*** | ***Average best BITRATE (kbps)*** | ***ΔBITRATE (kbps)*** | ***ΔBITRATE % Difference*** | ***Average best FPS*** | ***ΔFPS*** | ***ΔFPS% Difference*** |
|---|---|---|---|---|---|---|---|---|---|
| 16 | 0.9893 | 0.0027 | 0.27 | 5549 | 749 | 13.50 | 39.55 | 11.10 | 28.07 |
| 20 | 0.9804 | 0.0042 | 0.43 | 2724 | 479 | 17.58 | 40.88 | 11.66 | 28.52 |
| 24 | 0.9685 | 0.0065 | 0.67 | 1187 | 260 | 21.90 | 44.34 | 13.45 | ***30.33*** |
| 28 | 0.9531 | 0.0086 | 0.90 | 501 | 112 | ***22.36*** | 48.65 | 14.33 | 29.46 |
| 32 | 0.9342 | 0.0105 | 1.12 | 223 | 48 | 21.52 | 55.93 | 15.75 | 28.16 |
| 36 | 0.9096 | 0.0123 | ***1.35*** | 112 | 18 | 16.07 | 69.36 | 20.38 | 29.38 |

[a]for selected QPs.

Table 19: Forward Models Coefficients Median Values and Percentage Variation

| | | ***log (SSIM)*** | | | | ***log(Bitrate)*** | | | | ***log (FPS)*** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Filters | Encoding structure | Mediana0 | % Var.[a] a0 | Median bo | % Var. b0 | Mediana1 | % Var. a1 | Medianb1 | % Var. b1 | Mediana2 | %Var. a2 | Medianb2 | %Var.b2 |
| On | ZL | -4.05E-03 | 4.23 | 6.09E-02 | 4.42 | -1.95E-01 | 1.43 | 1.21E+01 | 0.48 | 1.81E-02 | 7.81 | 2.37E+00 | 1.64 |
| | B2 | -3.99E-03 | 3.69 | 5.68E-02 | 4.18 | -2.00E-0 | 0.93 | 1.19E+01 | 0.42 | 1.46E-02 | 5.94 | 3.20E+00 | 0.76 |
| | B4 | -4.02E-03 | 3.60 | 5.61E-02 | 4.41 | -2.01E-01 | 0.96 | 1.19E+01 | 0.44 | 3.47E-02 | 4.39 | 2.67E+00 | 1.23 |
| Off | ZL | -5.64E-03 | 4.22 | 8.58E-02 | 4.53 | -1.99E-01 | 1.79 | 1.23E+01 | 0.60 | 2.36E-02 | 6.32 | 2.58E+00 | 1.58 |
| | B2 | -5.03E-03 | 3.51 | 7.33E-02 | 3.57 | -2.01E-01 | 1.02 | 1.20E+01 | 0.43 | 2.16E-02 | 4.13 | 3.29E+00 | 0.71 |
| | B4 | -4.80E-03 | 3.41 | 6.83E-02 | 3.35 | -2.03E-01 | 0.98 | 1.20E+01 | 0.43 | 3.67E-02 | 4.95 | 2.83E+00 | 1.23 |

[a]Variation.

Table 19 summarizes the median coefficient values over the whole data set for the different encoding configuration options considered. That is, the leave-one-out method is applied ten times, each time leaving a different video out for validation purposes. The maximum percentage variation is defined using:

$$\textit{Percentage variation of } \mathrm{C} = \frac{\max(\mathrm{C}) - \min(\mathrm{C})}{\mathrm{median}(\mathrm{C})} \tag{21}$$

where *C* can be $a_{i,0}, a_{i,1}, a_{i,2}$ or $b_{i,0}, b_{i,1}, b_{i,2}$.

The maximum percentage variation is particularly important as it highlights the robustness of the proposed framework's resulting models. More specifically, $a_{i,0}$ and $b_{i,0}$ coefficients in video quality prediction equation have a maximum variation of less than 5%. Similarly, coefficients' variation in bitrate demands prediction is well below 2% and 1%, for $a_{i,1}$ and $b_{i,1}$ respectively. Encoding frame rate, being the most difficult parameter to accurately predict, shows slightly higher maximum percentage variation extending up to approximately 6% and 8% for $ZL$ with filters enabled and disabled, respectively, while typically lies between 4% to 6% for the rest encoding configurations for $a_{i,2}$ and less than 2% for $b_{i,2}$.

The percentage prediction error is depicted in Figures 26(a), 26(c), and 26(e), while the average QP error is also given in Figs. 26(b), 26(d), and 26(f), for each configuration. The results of Figure 26 indicate that QP error is not significant. To see this, note that after estimating the QP values from the prediction modes, the optimal QP is typically within +/-2 for most configurations.

The percentage median error for SSIM video quality is less than 1% for all investigated models corresponding to a +/-1.5 median QP deviation. The latter finding suggests that the desired objective quality is accurately estimated using SSIM forward models. With respect to bitrate demands prediction, the median percentage error is approximately 10% while associated QP error is in the order of +/-0.5 QP. Error distribution is comparable in all prediction models. Based on

this observation, during real-time operation, we allow mild violations for bitrate prediction in the order of 10%.

The only models that have significant deviations in terms of error distribution, following a similar pattern to models coefficient variation, are the encoding frame rate models. For ZL, approximately 8% median percentage error is experienced, with an approximate +/-3 QP value median error. The error percentage drops significantly to slightly higher than 4% and well below 6% for B2 and B4 models, respectively. QP value error distribution is also reduced to approximately +/-2 for both prediction models. Overall, prediction accuracy results are very promising, especially for video quality and bitrate demands, since the depicted errors are within reasonable bounds for all investigated models. Predicting the encoding frame rate is slightly more difficult. However, real-time adaptive encoding requires solving the multi-objective optimization problem for the employed optimization mode of operation, and as a result Pareto optimal QP values further minimize prediction errors.

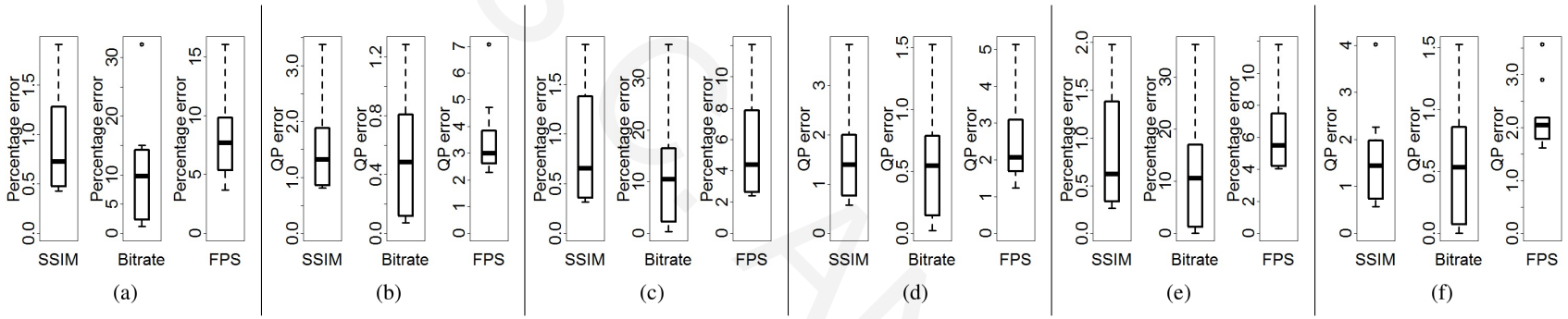



Figure 26: Forward prediction models error distribution. Boxplots depicting the percentage error of SSIM, Bitrate, and Encoding Frame Rate for encoding structures (a) ZL, (c) B2, and (e) B4. Boxplots depicting the QP error of SSIM, Bitrate, and Encoding Frame Rate for encoding structures (b) ZL, (d) B2, and (f) B4. In each plot we display the median, lower, and upper quartiles and confidence interval around the median. The '-' sign indicates the lower and upper whiskers that correspond to the most extreme values. The 'o' sign indicates possible outliers.

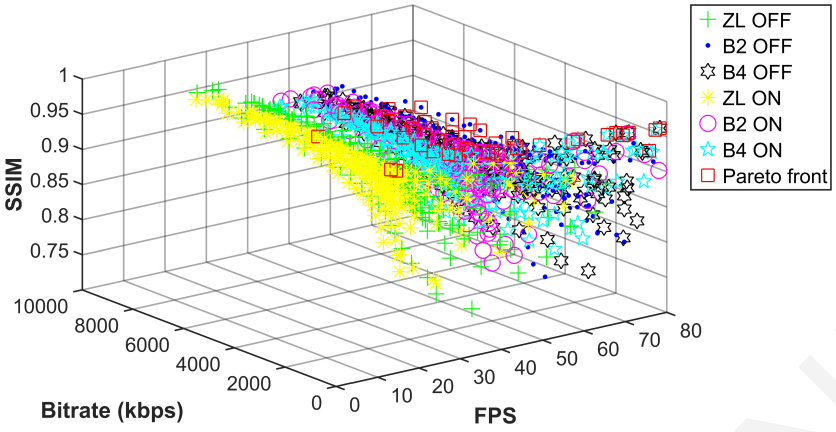
### 5.5.3 Pareto Front Using Prediction Models' Coefficients Median Values

To demonstrate the efficiency of using adaptive encoding and depict why different prediction models (for each of the ZL, B2, and B4 encoding structures with DBF and SAO filters enabled or disabled) are more suitable for this purpose, we construct new prediction models based on the coefficients median values described in Table 19. Then, we use these models to predict SSIM, bitrate, and encoding rate values for each prediction model, using the QP range that appears in Table 17. Based on the resulting values we construct the Pareto Front by selecting the optimal configurations that jointly maximize video quality and encoding frame rate while minimizing bitrate. Figure 27(a) depicts Pareto optimal points along with individual points of each prediction model for the ten videos of our data set. A 3D $\rightarrow$ 2D Pareto front projection of each individual combination pair of the three distinct objectives in our multi-objective optimization framework appears in Figures 27(b)-(d) (without loss of generality and due to space limitations, we only plot points $\leq$ 2.5 Mbps). An important observation is that encoding structure ZL having filters disabled does not contribute any Pareto optimal points. More importantly, Figure 27 clearly demonstrates that all remaining candidate configurations provide non-dominated Pareto front points. Consequently, switching between such configurations is likely to result in a finer and more accurate adaptation during real-time operation. The latter constitutes a significant contribution over the state-of-the-art, provided that adaptation methods in the literature do not consider different encoding configurations, and hence result in suboptimal selection between available, pre-encoded configurations. As a result, our proposed adaptation method initially considers all possible prediction models to compute candidate encoding solutions, before eliminating non Pareto optimal points, and then solving the constraint optimization problem to select the optimal configuration and compression level that satisfies the time-varying constraints per mode of operation (see Figures 23 and 24).

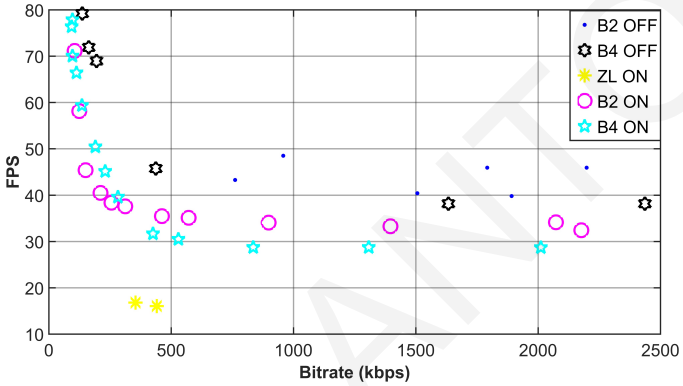
### 5.5.4 Adaptive Encoding Using Different Optimization Modes and Comparison to Static Approaches

The objective here is to validate the effectiveness of the proposed adaptive video encoding framework and highlight its efficiency compared to static approaches, using real-life scenario examples. For this purpose, we demonstrate and discuss how time-varying constraints, typical in real-time video streaming events, trigger an encoding configuration switch, for each mode of operation described in Section 5.4.1.

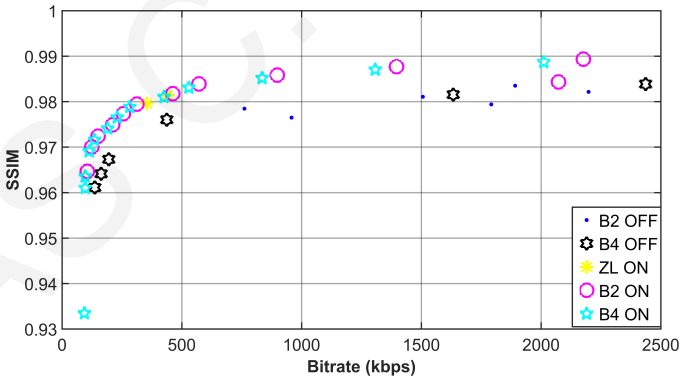
Despite the fact that forward prediction models $B2$ and $B4$ with filters on contribute the most Pareto optimal points as depicted in Figure 27, they are also subject to greater overall prediction errors than the corresponding models with deactivated filters as can be seen in Table 20. In this Table we can see the averages for 10 videos. Thus, for the next optimization examples, adaptation is performed using prediction models with filters off.



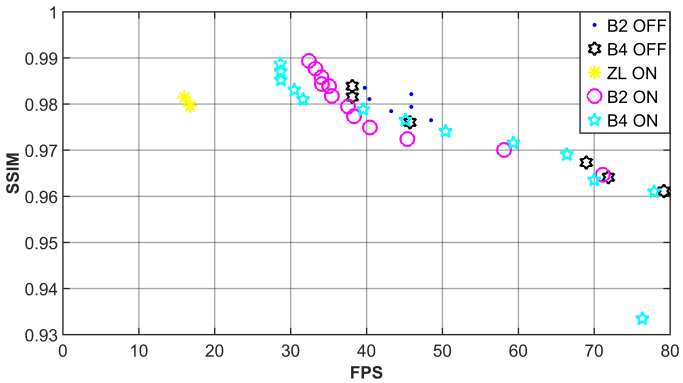

(a) 3D Pareto front points over the entire configuration space




(b) 3D → 2D Pareto front of FPS vs Bitrate




(c) 3D → 2D Pareto front of SSIM vs Bitrate




(d) 3D → 2D Pareto front of SSIM vs FPS



Figure 27: (a) 3D Pareto front plot over the entire configuration space and the whole data set. (b)–(d) 3D → 2D Pareto front plots depicting optimal points that maximize video quality and encoding frame rate while minimizing bitrate demands.

Table 20: Filters ON vs OFF: Error averages for Framerate, Bitrate and SSIM video quality

| | FILTERS OFF | | | FILTERS ON | | |
|---|---|---|---|---|---|---|
| | **ZL** | **B2** | **B4** | **ZL** | **B2** | **B4** |
| | **Averages** | **Averages** | **Averages** | **Averages** | **Averages** | **Averages** |
| **FPS** | | | | | | |
| Average FPS error Percentage | 4.8% | 3.7% | 4.9% | 4.6% | 4.0% | 5.6% |
| Standard Deviation FPS error Percentage | 4.6% | 2.1% | 5.0% | 4.5% | 2.8% | 5.5% |
| Average FPS error per QP percentage | 1.7% | 1.8% | 2.4% | 1.4% | 1.4% | 2.5% |
| Standard Deviation FPS error per QP percentage | 2.0% | 1.1% | 2.3% | 1.7% | 1.1% | 2.3% |
| Average QP error | 2.619 | 1.839 | 1.873 | 3.441 | 2.819 | 2.223 |
| **BITRATE** | | | | | | |
| Average BITRATE error percentage | 10.8% | 11.2% | 11.2% | 10.4% | 11.3% | 11.3% |
| Standard Deviation BITRATE error percentage | 9.0% | 7.4% | 7.0% | 6.4% | 6.7% | 6.1% |
| Average BITRATE error per QP percentage | 10.2% | 10.2% | 10.1% | 9.6% | 10.2% | 10.0% |
| Standard Deviation BITRATE error per QP percentage | 3.2% | 1.7% | 1.6% | 2.1% | 1.7% | 1.5% |
| Average QP error | 0.568 | 0.589 | 0.584 | 0.554 | 0.600 | 0.591 |
| **SSIM** | | | | | | |
| Average SSIM error percentage | 0.7% | 0.6% | 0.6% | 0.6% | 0.6% | 0.6% |
| Standard Deviation SSIM error percentage | 0.6% | 0.5% | 0.5% | 0.4% | 0.4% | 0.4% |
| Average SSIM error per QP percentage | 0.4% | 0.4% | 0.4% | 0.4% | 0.3% | 0.3% |
| Standard Deviation SSIM error per QP percentage | 0.2% | 0.2% | 0.2% | 0.2% | 0.1% | 0.1% |
| Average QP error | 1.256 | 1.315 | 1.360 | 1.593 | 1.545 | 1.540 |

#### 5.5.4.1 Maximum Video Quality Mode

In this optimization mode the objective is to maximize the video quality while conforming to bandwidth constraints in terms of typical upload data rates found in commercially available wireless networks and maintaining an encoding speed greater than the ultrasound video's frame rate ($\geq$40 fps). This scenario reflects real-time, diagnostically lossless ultrasound video streaming in emergency incidents, while the ambulance traverses from the incident scene to the hospital.

Launching the ultrasound video streaming session (e.g., ambulance departing from the incident scene), we select a conservative typical initial upload data rate found in 3G wireless channels ($\leq$300 kbps) [173]. Using the inverse prediction methodology described in Section 5.4.5, the optimal performing encoding configuration is B4 with a QP of 31. As evident in Table 21, the resulting configuration meets the data rate constraint at 290 kbps and is well above the required threshold for real-time encoding performance. As the ambulance traverses towards the hospital premises, the available data rate is increased, switching from a 3G to a 3.5G High Speed Packet Access ($HSPA$) network with typical upload data rate in the order of 1 Mbps [173]. The latter triggers an encoding mode switch to adapt to the new wireless channel's state. Solving the Newton's equation for the considered prediction models provides us with a new encoding structure configuration, namely B2, and a QP of 26. Again, both constraints are met, while a significant increase in video quality is achieved (SSIM value of 0.949 compared to the initial 0.921 value), as per the employed mode of operation. It is important to note here that the adaptation is performed at the GOP level, and throughout the study we consider and allow adaptations every GOP. The third and last example of this scenario, corresponds to further increases in available data rate accommodated in HSPA+ channels ($\leq$1.8 Mbps) [173]. The proposed framework adjusts to this change by considering finer quality encoding employing a QP of 23 while using the same encoding structure,

resulting in an SSIM value of 0.961. Real-time encoding performance is also maintained. Here, there is a mild violation of $\leq$3% in terms of bandwidth demands which is, however, within the acceptable limits. Results for this example are summarized in Table 21 and Figure 28. In Figure 28, the upper and lower bounds of the present example are also plotted. The objective is to depict the benefits from adaptive encoding and video delivery throughout a streaming session subject to time-varying constraints compared to static approaches. The static encoding at the lower end, suffers from low-quality SSIM ratings, which are marginally diagnostically acceptable. On the opposite end, diagnostically lossless $SSIM$ value is achieved using the static approach, however, high bandwidth requirements put video communication at risk of failure when high data rates are not supported. At the same time, bitrate savings of 52% are attained using the adaptive framework for clinically acceptable SSIM value of 0.949 compared to the high static approach.

In the clinical evaluation Tables 22, 25 and 28, the medical expert provides separate ratings for each video segment and for the entire video. Here, we note that there is clearly more information in the entire video as opposed to the individual segments. Thus, clinical ratings for the whole video will be higher than ratings of its constituent video segments. Nevertheless, as shown in the tables, the differences are small since the ultrasound videos are periodic. Furthermore, we note that the whole video ratings essentially reflect an instance of the proposed system, where the encoding bitrate does not change throughout the video streaming session, and consequently there is no need to trigger an encoding mode switch.

As we can see in Table 22, for the lowest SSIM quality of 0.921, where the video is encoded using B4 and a QP of 31, the ability to assess plaque morphology is compromised. This is not the case when assessing the whole video as subsequent frames make available the required clinical information for diagnostically lossless ratings. Here, it is important to note the medical expert's comment, underlining that the first two cardiac cycles of the particular video obscure some clinical

information attributed to patient movement during acquisition. This information is made available in subsequent cardiac cycles. As one would expect, increasing quality results in higher clinical ratings. This is the case for the next two encodings, where diagnostically lossless ratings are provided for all investigated clinical criteria, the clinical information in the compressed video being in some cases identical to the original, uncompressed video (a rating of 5).

Table 21: Maximum Video Quality Optimization Mode example

| ***Encoded Frames*** | ***# of B frames*** | ***QP*** | ***Bitrate (kbps)*** | ***FPS*** | ***SSIM*** |
|---|---|---|---|---|---|
| 0-191 | 4 | 31 | 290 | 48.29 | 0.921 |
| | | constraints | ≤300 | ≥40 | |
| 192-383 | 2 | 26 | 884 | 46.63 | 0.949 |
| | | constraints | ≤1000 | ≥40 | |
| 384-479 | 2 | 23 | 1873 | 43.98 | 0.961 |
| | | constraints | ≤1800 | ≥40 | |

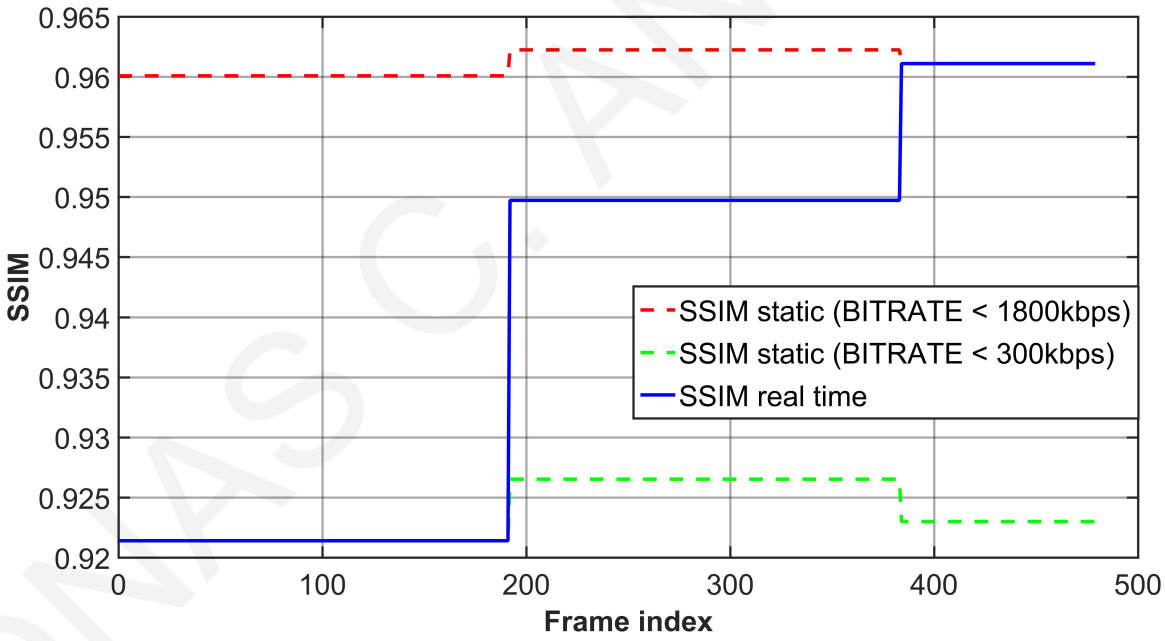



Figure 28: Adaptive encoding based on Maximum Video Quality Optimization Mode.

Table 22: Clinical Assessment[a] for Maximum Video Quality Mode Optimization example

| *QP* | *SSIM* | *Encoding Frames* | *Stenosis* | *Morphology* | *Motion* |
|---|---|---|---|---|---|
| 31 | 0.921 | 0-191 | 4 | 3 | 4 |
| | | whole video | 4 | 4 | 4 |
| 26 | 0.949 | 192-383 | 4 | 4 | 4 |
| | | whole videos | 4 | 4 | 4 |
| 23 | 0.961 | 384-537 | 4 | 4 | 4 |
| | | whole video | 5 | 4 | 4 |

[a]A clinical score from 1 (lowest) to 5 (highest) is assigned for Artery Stenosis, Morphology and Motion as described in Section 5.4.2.

#### 5.5.4.2 Minimum Bitrate Demands Mode

In the minimum bitrate demands mode, the goal is to minimize bandwidth requirements while maintaining clinically acceptable video quality and real-time performance. Such scenarios are likely to occur in disaster incidents and mass population screening cases, especially in developing countries, where wireless networks resources are unstable and shared by many users (e.g., first responders and medical experts). We first seek high quality ultrasound video communication during real-time streaming to the hospital and/or disaster control site. Thus, we set a high video quality constraint SSIM value $\geq$ 0.97 and encoding frame rate higher than 40 fps. Then, the video quality constraint is gradually lowered to $SSIM$ values $\geq$ 0.95 and $\geq$ 0.93 as per the medical expert's recommendation while maintaining its clinical capacity. Indeed, clinically acceptable ratings (greater or equal to 4) are depicted in Table 25, that support the hypothesis that significant bitrate gains can be achieved using adaptive video encoding without compromising the ultrasound video's diagnostic capacity. As commented in the previous optimization example, the low clinical assessment score of 3.5 for artery morphology is primarily attributed to the ultrasound video content rather than the compression effects. Lastly, for experimentation purposes, we employ an $SSIM$ threshold $\geq$ 0.90 to demonstrate that maximum bitrate savings may be achieved,

Table 23: Minimum Bitrate Optimization Mode example

| ***Encoded Frames*** | ***# of B frames*** | ***QP*** | ***Bitrate (kbps)*** | ***FPS*** | ***SSIM*** |
|---|---|---|---|---|---|
| 0-95 | 2 | 30 | 3417 | 41.30 | 0.973 |
| | | constraints | | $\geq$40 | $\geq$0.97 |
| 96-191 | 2 | 24 | 1309 | 48.55 | 0.956 |
| | | constraints | | $\geq$40 | $\geq$0.95 |
| 192-287 | 4 | 29 | 335 | 50.02 | 0.937 |
| | | constraints | | $\geq$40 | $\geq$0.93 |
| 288-479 | 2 | 36 | 132 | 56.4 | 0.887 |
| | | constraints | | $\geq$40 | $\geq$0.90 |

Table 24: Bitrate Gains examples of Adaptive Encoding against static approaches: Minimum Bitrate Optimization Mode

| ***Encoded Frames*** | 96-191 | 192-287 | 288-537 |
|---|---|---|---|
| ***Bitrate Gain (%)*** | 59.7 | 89.7 | 96.1 |

however, at the expense of compromising ultrasound video's diagnostic quality. For the latter encoding instance, clinical ratings fall below of what is diagnostically acceptable (a clinical score of 3 throughout the video for artery morphology). The resulting GOP level encoding adaptation configurations are depicted in Table 23 and Figure 29, while bitrate gains compared to the initial static approach are highlighted in Table 24. As in the previous mode of operation, in addition to selecting a new compression level via different QP values, we observe that there is also an encoding structure switch between B2 and B4. Furthermore, the medical expert found encodings with B4 structure to have superior clinical motion, which is a subject of ongoing investigation by our group. All resulting encoding configuration instances based on the derived forward models meet the imposed constraints besides the last instance, where there is a mild violation of SSIM constraint of approximately 1%. Compared to the initial configuration or the static high bitrate approach, we observe that there is drastic reduction in bitrate demands in the order of 90% (see Table 24) for videos encoded with a QP of 29, while maintaining the video's diagnostic capacity. As a result, multiple users can share the network's infrastructure efficiently, by consuming only the resources necessitated for clinical purposes.

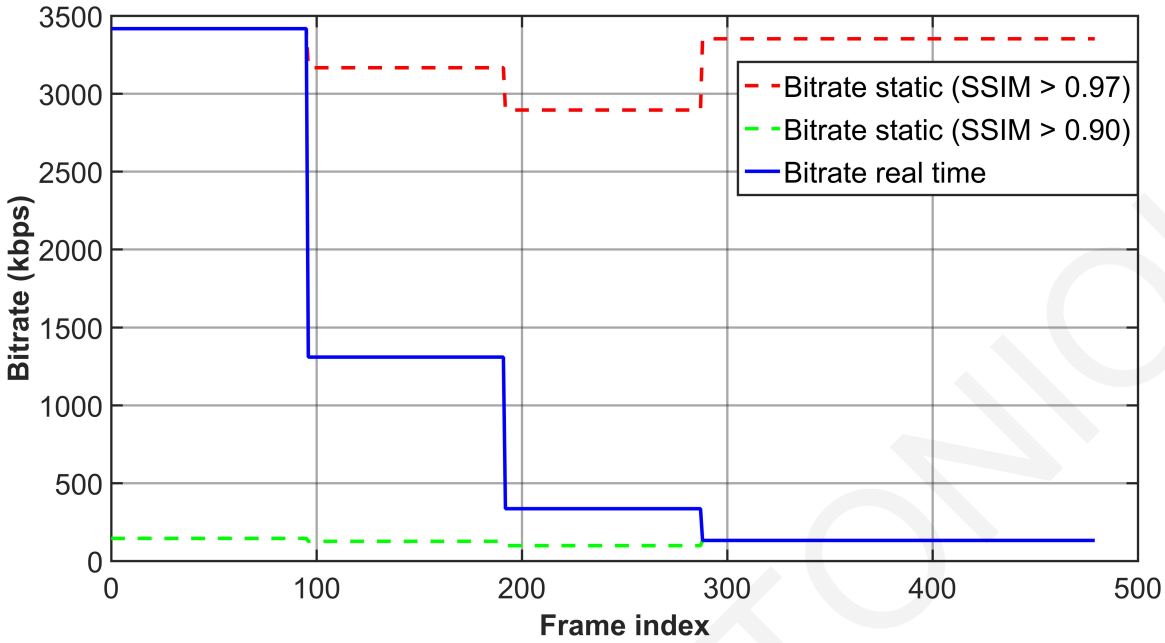



Figure 29: Adaptive encoding based on minimum bitrate optimization mode.

Table 25: Clinical Assessment for Minimum Bitrate Optimization Mode example

| *QP* | *SSIM* | *Encoding Frames* | *Stenosis* | *Morphology* | *Motion* |
|---|---|---|---|---|---|
| 20 | 0.973 | 0-95 | 4 | 3.5 | 4 |
| | | whole video | 4.5 | 4 | 4 |
| 24 | 0.956 | 96-191 | 4 | 4 | 4 |
| | | whole videos | 4.5 | 4 | 4 |
| 29 | 0.937 | 192-287 | 4 | 4 | 4.5 |
| | | whole video | 4.5 | 4 | 4.5 |
| 36 | 0.887 | 288-537 | 4 | 3 | 4 |
| | | whole video | 4 | 3 | 4 |

#### 5.5.4.3 Maximum Performance (FPS) mode

The goal of the maximum performance optimization scenario is to achieve any medical video modality and any-device real-time encoding performance. Conforming to a predefined, diagnostically acceptable SSIM value $\geq 0.95$ (see Table 26), and considering available data rate not an issue, the objective is to increase the encoding performance. We observe that the initial encoding setup does not qualify for real time streaming as the encoding frame rate is lower that the generated video's frame rate. In real-life scenarios, this would cause a video streaming buffer underflow,

Table 26: Maximum Performance Optimization Mode example

| ***Encoded Frames*** | ***# of B frames*** | ***QP*** | ***Bitrate (kbps)*** | ***FPS*** | ***SSIM*** |
|---|---|---|---|---|---|
| 0-191 | 4 | 25 | 995 | 36.58 | 0.952 |
| | | constraints | $\leq$1000 | | $\geq$0.95 |
| 192-383 | 2 | 24 | 1249 | 45.81 | 0.959 |
| | | constraints | $\leq$1400 | | $\geq$0.95 |
| 384-479 | 2 | 23 | 1873 | 43.98 | 0.961 |
| | | constraints | $\leq$1800 | | $\geq$0.95 |

Table 27: Performance Gains examples of Adaptive Encoding against static approaches: Maximum Performance Optimization Mode

| ***Encoded Frames*** | 192-383 | 384-479 |
|---|---|---|
| ***Performance (fps) Gain (%)*** | 18.8 | 14 |

resulting in poor remote ultrasound video based diagnosis experience. Increasing the available bandwidth to 1.4 Mbps triggers an encoding structure switch from B4 to B2 and QP reduction from 25 to 24. Bitrate demands are increased accordingly while SSIM value is also increased from 0.952 to 0.959. Most importantly, the encoding frame rate is significantly increased by approximately 19% to 45.81 fps, well above the real-time streaming threshold. A further increase in bitrate availability to 1.8 Mbps provides for a slightly better performance. Results are depicted in Table 26, while percentage performance increase compared to the static approach is given in Table 27. Figure 30 provides a schematic interpretation of the example scenario. As depicted in Table 28, all encodings maintain clinically acceptable ratings (besides the initial cardiac cycles). Here, it is important to note that to the best of the authors' knowledge, there is no study that investigates encoding rate performance in adaptive medical video streaming applications. Encoding rate translates to power/energy consumption, a crucial factor in mHealth applications supported by mobile devices.

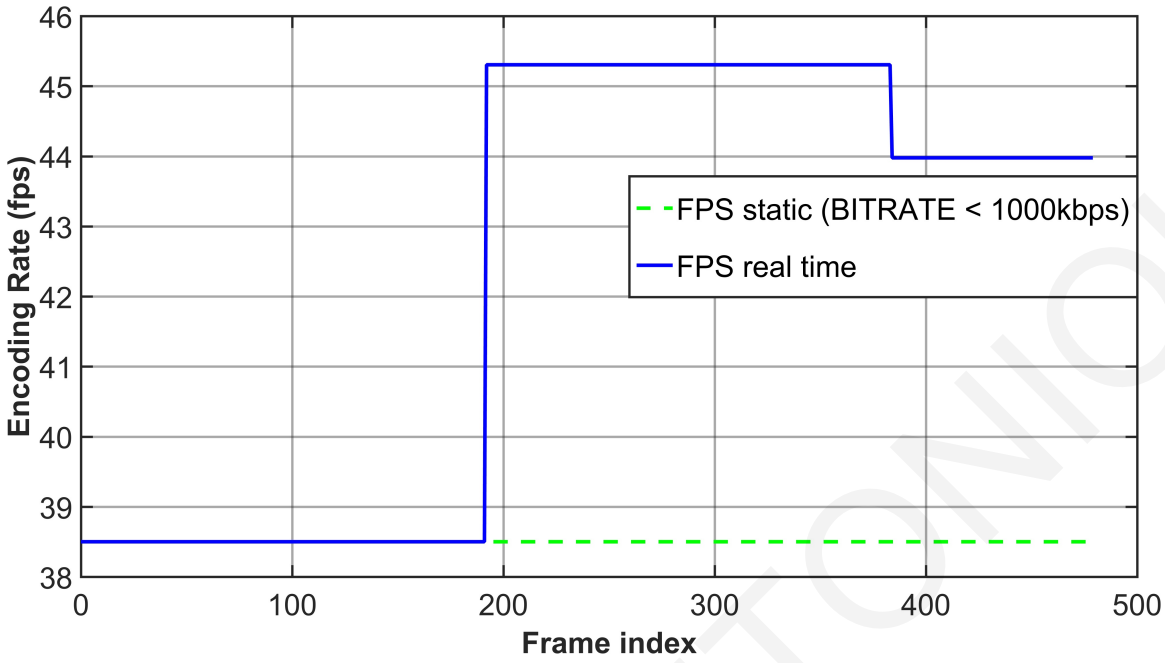



Figure 30: Adaptive encoding based on maximum performance (encoding rate) optimization mode.

Table 28: Clinical Assessment for Maximum Performance Optimization Mode example

| *QP* | *SSIM* | *Encoding Frames* | *Stenosis* | *Morphology* | *Motion* |
|---|---|---|---|---|---|
| 25 | 0.952 | 0-191 | 4 | 3.5 | 4 |
| | | whole video | 4 | 4 | 4 |
| 24 | 0.959 | 192-383 | 4 | 4 | 4 |
| | | whole videos | 4.5 | 4 | 4 |
| 23 | 0.961 | 384-537 | 4 | 4 | 4 |
| | | whole video | 5 | 4 | 4 |

### 5.6 Conclusions

The contributions for this study were that we propose a scalable, medical modality and technology independent approach for adaptive, real-time mHealth video communications. Application level requirements in terms of clinical quality, time-varying bandwidth availability, and heterogeneous devices' performance capabilities were modeled as constraints to an optimization problem. The proposed framework determined an encoding configuration that jointly optimizes the afore-described parameters for real-time operation. Results demonstrated that efficient adaptation at a

GOP level was possible using our proposed models, significantly outperforming static approaches and demonstrating higher flexibility and precision compared to approaches relying on precomputed states.

# Chapter 6

## Concluding Remarks and Future Work

Over the past four decades, mobile communication networks have fundamentally changed the way people communicate. From 1G to today's widely used 4G, a significant progress has been allowing the transmission of higher data rates. The evolution from 4G to 5G is still taking shape but there have already been numerous reports on the potential benefits that 5G technology will enable. Contrary to the evolution between previous generations of technology, 5G will offer advances along three fronts simultaneously: data rates, connectivity, and reliability. Additionally, key capabilities of 5G technology [134] include (a) peak data rates that should reach 10 Gbps, (b) very low latency (less than or equal to 1 ms), (c) high mobility (up to 500 km/h) and (d) more energy efficiency and more spectrum efficiency. Consequently, high-speed 5G networks will let physicians see high-resolution video remotely and enable the deployment of healthcare solutions that stream ultra-high-definition content.

However, the current bandwidth restrictions of 3G and 4G networks limit the amount of information that can be sent. So, for mHealth video communication systems, the question often arises as to whether to choose high resolution at high compression (to be transmitted over 4G and beyond), versus low resolution at low compression (to be transmitted over 3G), if both result in the

same bitrate. Since mHealth applications rely on clinical video quality, the answer to this question depends on guaranteeing a minimum, acceptable level of clinical video quality along with the video modality and the screening device of the medical expert. For example, in case the medical expert will use a video device supporting low resolution, it will be wiser to transmit low resolution medical video at low compression taking advantage of the full range of the available bitrate. On the other hand, if the medical expert has a high-resolution screen, it will be better to encode the video in high-resolution considering an upper bound for the compression based on the accepted quality. If this won't be feasible, a better-quality trade-off would be to encode at a lower resolution to eliminate the encoding artifacts.

Nowadays, there is a great need for efficient video compression methods, since original uncompressed HD and Ultra HD (UHD) video resolutions have high bitrate demands. For example, a video having 1280x720 resolution (720 progressive mode), having 24 bits per pixel with 60 frames per second, needs 1.3 Gbps. For UHD resolutions have even higher bitrate demands that 5G networks will not support. In terms of ultrasound video communication purposes, an original uncompressed video would require 93.18 Mbps (560 width resolution $\times$ 416 horizontal resolution $\times$ 50 frames/s $\times$ 8 bits per pixel = 93.18 Mbps). Using HEVC encoding for a QP of 28, which achieves diagnostically lossless ultrasound video quality, the transmission rate is reduced to 340 kbps. In other words, a compression ratio of 274 is achieved [81]. Thus, the necessity for adaptive video compression will still remain.

Additionally, real-time adaptive encoding will still be a necessary mass population screening cases, especially in developing countries, where wireless networks resources are unstable and shared by many users. Also, in disaster incident scenarios, where natural disasters can cause breakdown in communication systems. Thus, in these scenarios the encoding process must adapt

to the available bandwidth in a real-time manner, without compromising the clinical quality of the transmitted video.

Furthermore, multi-view video streaming systems [174] can be benefited by real-time adaptation based on time-varying constraints, because they capture the video content using multiple cameras and then encode it for transmission. Also, such systems can support different video resolutions based on the cameras' capabilities. Consequently, growing bandwidth available in current and emerging mobile networks can't support the transmission of the multi-view video content. Similar approach applies for 3D video transmission [175], [176], where a 3D video provides an added dimension of depth to conventional video. 3D medical video transmission allows improved visual perception and thus better diagnosis.

### 6.1 Concluding Remarks

The motivation and objective of the present thesis is the wider adoption of mHealth video communication systems in standard clinical practice. Toward this direction, technical objectives involve the delivery of (i) sufficiently high video resolutions and frame rates, subject to (ii) low-delay and low packet loss rates requirements, by (iii) adapting to wireless networks' varying state. The latter requires real-time control of the video streaming process so as to facilitate the adequate levels of clinical video quality required to support reliable diagnosis. The goal is hence to match clinical experience levels of in-hospital examinations. The thesis proposes a scalable, video modality, encoder, and wireless network agnostic framework, that will support real-time adaptation to time-varying wireless networks' state while guaranteeing diagnostically lossless video communications and conforming to end-user device constraints for real-time performance.

The approach is based on multi-objective optimization, that jointly maximizes the encoded video's quality and encoding rate while minimizing bitrate demands. For this purpose, a dense

encoding space is constructed, and linear regression is used to estimate forward prediction models for quality, bitrate, and computational complexity. The prediction models are then used by the proposed adaptive control framework that can fine-tune video encoding based on real-time constraints. The proposed framework is validated using a leave-one-out algorithm applied to ten ultrasound videos of the common carotid artery. The prediction models can estimate SSIM quality with a median accuracy error of less than 1%, bitrate demands with deviation error of 10% or less, and encoding frame rate within a 6% margin. Real-time adaptation at a GOP level is demonstrated using HEVC standard. The effectiveness of the proposed framework compared to static, non-adaptive approaches is demonstrated for different modes of operation, achieving significant quality gains, bitrate demands reductions, and performance improvements, in real-life scenarios imposing time-varying constraints.

This proposed framework has the potential to transform healthcare delivery by fostering responsive emergency systems, in unrestricted and cross border settings that will significantly improve patients' quality of care. The clinical benefits of transmitting real-time clinical video of adequate diagnostic quality are an open and active area of research.

The contributions of the first study are:

- **Relationship between spatial resolution (QCIF, CIF, 4CIF) and clinical diagnosis**: This relationship between video resolution and clinical quality was investigated via multiple clinical evaluation sessions. Experiments involved medical experts evaluating a video's diagnostic quality based on the actual clinical protocol that is followed during common carotid artery (CCA) ultrasound examination, displayed at different resolutions. The medical experts were asked to validate the clinical content of these three resolutions. The findings

verified the hypothesis that higher resolution accommodates a larger amount of clinical information. The use of higher 4CIF ($704 \times 576$) resolution underpins new clinical quality standards that are closer to standards used for in hospital exams.

- **Medical video communications over Mobile WiMAX networks**: Within the context of the present thesis, the case study of medical video communication over Mobile WiMAX networks was evaluated using the OPNET modeler and video traces of real CCA ultrasound videos. Recommendations for mobile WIMAX network's parameters setup and utilization towards maximizing the communicated video's clinical capacity are hence proposed while findings motivated the proposed adaptive video communication framework. Key parameters that are highlighted in the thesis involve three different channel modulation and coding schemes, various distances from the base station (BS), and diverse mobility patterns. To demonstrate the effectiveness of the proposed framework, we successfully transmitted clinically acceptable ultrasound video of 4CIF resolution over Mobile WiMAX networks. Here, it's important to note that despite the fact that this study relied on Mobile WiMAX networks (a highly promising technology at the time of investigation but with limited adoption today), the approach is technology agnostic and the same methods can be applied and adopted by any wireless network such as LTE or LTE-Advance.

The contributions of the second study are:

- **Multi-objective optimization for real-time operation**: The thesis proposes and evaluates a multi-objective optimization framework for adaptive video delivery that leverages video quality (application-modality level adaptation), bitrate demands (wireless network adaptation), and encoding frame rate (device adaptation for real-time operation). Thus, the

proposed solution considers encodings that simultaneously maximize video quality and encoding rate, while minimizing the required bitrate. The required space of solutions forms a Pareto front that is used for solving constrained optimization problems.

- **Robust prediction models**: Robust prediction models were generated for each optimization objective, that allow real-time encoding adaptation for real-time operation. More specifically, forward prediction models were developed for SSIM (video) quality, bitrate demands, and encoding frame rate, for three different group of pictures (GOP) encoding structures, namely zero latency (ZL), B2 and B4 respectively. Overall, prediction accuracy results were very promising, especially for video quality and bitrate demands, since the depicted errors were within reasonable bounds for all investigated models.

- **Real-Time adaptation using pareto-optimal encoding configurations**: Real-time adaptation at a GOP level was demonstrated using the HEVC standard. The effectiveness of the proposed framework compared to static, non-adaptive approaches was demonstrated for different modes of operation, achieving significant quality gains, bitrate demands reductions, and performance improvements, in real-life scenarios imposing time-varying constraints.

- **Video modality, encoder, and wireless network agnostic framework**: The proposed approach is generic enough and should be applicable to other medical video modalities and for different applications, provided that the appropriate training described in the thesis is performed.

## 6.2 Future Work

Future work includes investigation of higher quality encodings over 4G and beyond (5G) wireless networks and experimenting with larger and different data sets (including emergency incidents

videos) so as to increase the validity of the proposed methods. Moreover, to extend forward prediction models of video quality, beyond SSIM. The latter involves clinical evaluation (subjective evaluation) of different medical video encodings offline based on a clinically established protocols per investigated modality. The objective will be then to find threshold values (correlation) for which clinical quality is maintained with respect to the corresponding VQA algorithms. Furthermore, future work is directed towards dynamically reconfiguring the generated forward models during real-time operation based on the communicated video characteristics. The latter approach will further minimize prediction errors increasing robustness and reliability.

### 6.2.1 Self-controlled adaptive framework

An ambitious goal is to revisit the adaptive control framework so as to take advantage of the all the resulting optimal configurations in real-time after a constant number of iterations, and use them in order to dynamically reconfigure on-the-fly the generated forward models for quality, bitrate, and computational complexity.

The latter will be triggered when precision (prediction accuracy) is not adequate, based on statistic measurements, signaling that a new forward model tailored to the specific video content (and motions), needs to be computed on the fly. The objective is to maximize system performance and optimize resource utilization. Based on preliminary findings, this approach achieves prediction errors reductions, thus improving the system reliability.

### 6.2.2 Integrate no-reference video quality metrics in the proposed adaptive video communication framework

The ability to quantify the quality of a video is the most significant requirement for emerging medical video transmission systems. Diagnostic validation requires an accurate assessment of the

diagnostic capacity of the transmitted medical video. Since the quality of transmission systems is measured in terms of their diagnostic yield, the efficiency of medical video transmission systems is measured in terms of their ability to provide for a reliable and dependable diagnosis. Hence, there is a strong demand to provide diagnostically robust medical video communications. The absence of efficient video quality assessment algorithms, both objective and subjective (clinical), contributes to the challenges involved in the design of a system of consistent diagnostic quality.

Thus, in addition to the most widely VQA used in the literature, namely PSNR and SSIM full reference algorithms, blind (e.g. no reference) metrics developed at LIVE [177] will be investigated. In particular, the VIIDEO [89], BLIINDS-II [88], VMAF [85], and MOVIE [84] metrics will be investigated. The objective is to generate forward prediction models of adequate clinical video quality that will be based on no reference VQA algorithms. That is, no access to the original uncompressed video will be required to trigger an encoding configuration switch subject to time-varying constraints. Instead, the decision will be made in real-time subject to the transmitted video's quality. The latter approach constitutes a significant advancement of the state-of-the-art.

### 6.2.3 Investigate the proposed adaptive framework for MPEG-DASH streaming

Also, future work includes investigating the integration of the proposed adaptive video communication framework for MPEG-DASH streaming. Towards this direction, we will adopt the concept described in Chapter 5 to select pareto optimal encoding configurations for applications leveraging the Scalable-HEVC and MPEG-DASH standards. The vision is then to generate an MPEG-DASH client-based adaptation mechanism for clinical video scenarios based on multi-objective optimization (video quality, bitrate, decoding rate). Having concluded this, the objective is to perform a comparative evaluation of Scalable-HEVC and MPEG-DASH for adaptive video delivery telemedicine applications.

### 6.2.4 Realistic Wireless Network Modeling

Future work includes video transmission simulations to realistically model wireless network behavior and evaluate in real-time both the responsiveness and effectiveness of the proposed methods. For this purpose, the philosophy of the adaptive video controller will be incorporated in the TruNET [178] and OPNET [127] wireless network simulators. In terms of wireless channel modeling, we will investigate different wireless broadband (802.11x) and cellular (3G, 4G, and beyond (5G)) networks, different signal propagation environments (vehicular, urban, suburban, free space) and multipath fading states (vehicular, indoor to outdoor pedestrian, etc), principal mobility scenarios (low, medium, and high mobility), and effective channel modulation and coding schemes (with respect to distance from the base station and topology).

Capitalizing on the unparalleled capabilities of such simulators, we will further design building structure topologies that are analogous to the Cyprus wireless infrastructure, for the purpose of comparing simulations with real measurements, hence providing a better insight as to bandwidth fluctuation and availability in near real-life settings. The latter will allow simulating the available throughput and bandwidth based on the different signal propagation routes and corresponding signal attenuation based on the specific 3D path losses.

# Bibliography


[1] R. Istepanian, E. Jovanov, and Y. Zhang, "Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, pp. 405—414, Dec. 2004.

[2] WHO, *mHealth: New horizons for health through mobile technologies, Global Observatory for eHealth series - Volume 3*. WHO, 2011.

[3] McKinsey, Company, and GSMA, "mHealth: a new vision for healthcare," 2010.

[4] E. Kyriacou, M. S. Pattichis, C. S. Pattichis, A. Panayides, and A. Pitsillides, "m-health e-emergency systems: Current status and future directions," *IEEE Antennas and Propagation Magazine*, vol. 49, pp. 216–231, Feb. 2007.

[5] C. Pattichis, E. Kyriacou, S. Voskarides, M. Pattichis, and R. Istepanian, "Wireless telemedicine systems: an overview," *IEEE Antennas and Propagation Magazine*, vol. 44, pp. 143–153, Apr. 2002.

[6] WHO, "WHO EMRO - eHealth - Health topics," 2005.

[7] R. S. H. Istepanian and C. S. Pattichis, *M-Health: emerging mobile health systems*. Birkhäuser, 2006.

[8] A. Panayides, M. Pattichis, C. Pattichis, C. Loizou, M. Pantziaris, and A. Pitsillides, "Atherosclerotic plaque ultrasound video encoding, wireless transmission, and quality assessment using h.264," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, pp. 387 –397, May 2011.

[9] GSMA, "Gsma.com." `http://www.gsma.com`.

[10] A. Panayides, M. S. Pattichis, C. S. Pattichis, and A. Pitsillides, "A tutorial for emerging wireless medical video transmission systems [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 53, pp. 202—-213, Apr. 2011.

[11] I. R. M. 1645, "Framework and overall objectives of the future development of IMY-2000 and systems beyonf IMT-2000," Jan. 2003.

[12] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?," *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[13] G. Americas, *5G Services and Use Cases*. Nov. 2017. (Accessed on 12/31/2017).

[14] Y. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards. 2008*. CRC Press, Boca Raton.

[15] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC, "Advanced video coding for generic audiovisual services," May 2003.

[16] ITU-T, "High efficiency video coding (HEVC) - h.265." `http://www.itu.int/en/ITU-T/studygroups/com16/video/Pages/jctvc.aspx`.

[17] R. Istepanian and Y.-T. Zhang, "Guest editorial introduction to the special section: 4G health - the long-term evolution of m-health," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, pp. 1 –5, Jan. 2012.

[18] S. P. Rao, N. S. Jayant, M. E. Stachura, E. Astapova, and A. Pearson-Shaver, "Delivering diagnostic quality video over mobile wireless networks for telemedicine," *Int. J. Telemedicine Appl.*, vol. 2009, pp. 1:1–1:9, Jan. 2009.

[19] M. G. Martini, R. S. Istepanian, M. Mazzotti, and N. Philip, "Robust multilayer control for enhanced wireless telemedical video streaming," *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 5–16, 2010.

[20] I. VCEG and I. MPEG, "Report of subjective testing of responses to joint call for proposals (CfP) on video coding technology for high efficiency video coding (HEVC)," 2010.

[21] V. Baroncini, G. J. Sullivan, and J. R. Ohm, *Report of subjective testing of responses to Joint Call for Proposals (CfP) on video coding technology for High Efficiency Video Coding (HEVC)*. document JCTVC-A204 of JCT-VC, 2010.

[22] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," Aug. 2004.

[23] ITU-T, "Video codec for audiovisual services at px64 kbit/s," *ITU-T Recommendation H.261*, Nov. 1990.

[24] ITU-T, "Video coding for low bitrate communication," *ITU-T Recommendation H.263*, Nov. 1995.

[25] A. Panayides, *Diagnostically resilient encoding, wireless transmission, and quality assessment of medical video*. PhD thesis, Ph. D Dissertation, Department of Computer Science, University of Cyprus, Nicosia, 2011.

[26] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560–576, July 2003.

[27] A. C. Bovik, *The essential guide to image processing (2nd ed.)*. Academic Press, 2009.

[28] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.

[29] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7– 28, 2004.

[30] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, pp. 61–82, July 2000.

[31] W. Zia, T. Afzal, W. Xu, G. Liebl, and T. Stockhammer, "Interactive error control for mobile video telephony," in *IEEE International Conference on Communications, 2007. ICC '07*, pp. 1797–1802, IEEE, June 2007.

[32] S. Wenger, "FMO: flexible macroblock ordering," May 2002.

[33] P. Lambert, W. De Neve, Y. Dhondt, and R. Van de Walle, "Flexible macroblock ordering in H.264/AVC," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 358–375, Apr. 2006.

[34] S. Kumar, L. Xu, M. K. Mandal, and S. Panchanathan, "Error resiliency schemes in H.264/AVC standard," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 425–450, Apr. 2006.

[35] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 637–644, July 2003.

[36] H.-M. Hang, W.-H. Peng, C.-H. Chan, and C.-C. Chen, "Towards the next video standard: High efficiency video coding," in *Second APSIPA Annual Summit and Conference*, (Biopolis, Singapore), pp. 609–618, Dec. 2010.

[37] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, *ITU Recommendation H.265*. ITU-T.

[38] I. S. W. 11, "MPEG HEVC – the next major milestone in MPEG video history is achieved," (Geneva, Switzerland), Jan. 2013.

[39] B. Li, G. J. Sullivan, and J. Xu, "Compression performance of high efficiency video coding (HEVC) working draft 4," *IEEE*, vol. 2012, pp. 886–889.

[40] JCT-VC, *Report of Subjective Test Results of Responses to the Joint Call of Proposals (CfP) on Video Coding (HEVC)*. document JCTVC-A204 of JCT-VC, Apr. 2010.

[41] F. Bossen, *Common test conditions and software reference configurations*. ISO/IEC JTC1/SC29/WG11, JCTVC-B300, July 2010.

[42] D. Gomez-Barquero, *Next Generation Mobile Broadcasting*. CRC Press, Mar. 2013.

[43] T. K. Tan, A. Fujibayashi, Y. Suzuki, and J. Takiue, "[AHG 8] objective and subjective evaluation of HM5.0.," *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 8th Meeting: San José, CA, USA, 1–10 February, 2012*.

[44] ITU, "H.265 : High efficiency video coding." `http://www.itu.int/rec/T-REC-H.265`, June 2013.

[45] P. Bordes, G. Clare, F. Henry, M. Raulet, and J. Vieron, "An overview of the emerging HEVC standard," 2010.

[46] C. Furgusson, "Focus on...HEVC: the background behind the game-changing standard-ericsson," tech. rep., Ericsson, June 2013.

[47] J. Lainema, *On HEVC still picture coding performance*. JCT-VC, Apr. 2013.

[48] E. A. Ayele and S. B. Dhok, "Review of proposed high efficiency video coding (HEVC) standard," *International Journal of Computer Applications*, vol. 59, Dec. 2012.

[49] M. A. Mesa, C. C. Chi, T. Schierl, and B. Juurlink, "Evaluation of parallelization strategies for the emerging hevc standard," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012.

[50] D. Marpe, H. Schwarz, T. Wiegand, S. Bobe, B. Bross, H. Lakshman, T. Nguyer, S. Oudin, M. Siekmann, S. Karsten, and M. Winken, "Improved video compression technology and the emerging high efficiency video coding standard," (Berlin), IEEE, 2011.

[51] M. Winken, P. Helle, D. Marpe, H. Schwarz, and T. Wiegand, "Transform coding in the HEVC test model," in *IEEE International Conference on image processing (ICIP)*, Sept. 2011.

[52] M. T. Pourazad, C. Doutre, M. Azimi, and P. Nasiopoulos, "HEVC: the new gold standard for video compression," *IEEE consumer electronics magazine*, vol. 1, no. 3, pp. 36–46, 2012.

[53] S. Oudin, P. Helle, J. Stegemann, C. Bartnik, B. Bross, D. Marpe, H. Schwarz, and T. Wiegand, "Block merging for quadtree-based video coding," pp. 1–6, IEEE, July 2011.

[54] H. Lakshman, B. Bross, H. Schwarz, and T. Wiegand, "Fractional-sample motion compensation using generalized interpolation," in *Picture Coding Symposium (PCS)*, Dec. 2010.

[55] M. Siekmann, S. Bobe, H. Schwarz, and T. Wiegand, "Separable wiener filter based adaptive in-loop filter for video coding," in *Picture Coding Symposium (PCS)*, Dec. 2010.

[56] B. Bross, J. Han, J. R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 6," Apr. 2013.

[57] M. M. D. Tikekar, *Circuit implementations for high-efficiency video coding tools*. PhD thesis, Massachusetts Institute of Technology, 2012.

[58] H. Koumaras, M.-A. Kourtis, and D. Martakos, *Benchmarking the Encoding Efficiency of H.265/HEVC and H.264/AVC*. IIMC International Information Management Corporation, 2012.

[59] K. Ugur, K. Andersson, A. Fuldseth, G. Bjontegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Rusert, J. Samuelsson, R. Sjoberg, and Z. Wu, "Low complexity video coding and the emerging HEVC standard," in *28th Picture Coding Symposium, PCS2010*, Dec. 2010.

[60] C. Zhang, K. Ugur, J. Lainema, and M. Gabbouj, "Video coding using variable block-size spatially varying transforms," pp. 905–908, IEEE, Apr. 2009.

[61] I. K. Kim, W. J. Han, J. H. Park, and X. Zheng, "CE2: test results of assymetric motion partition (AMP)," July 2011.

[62] J. Chen and T. Lee, *Planar intra prediction improvement*. JCT-VC-F483, July 2011.

[63] J. Lainema and K. Ugur, "Angular intra prediction in high efficiency video coding (HEVC)," in *Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on*, pp. 1–5, 2011.

[64] J. Zhao and A. Segall, *On intra prediction and MDIS*. JCTVC-E437, 2011.

[65] A. Fuldseth, G. Bjontegaard, and M. Budagavi, "CE10: core transform design for HEVC," Nov. 2011.

[66] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," pp. 726–729, IEEE, 2010.

[67] L. Liu, R. Cohen, H. Sun, A. Vetro, and X. Zhuang, *New Techniques for Next Generation Video Coding*. TR2010-058, Apr. 2010.

[68] J. Lainema, K. Ugur, and A. Hallapuro, *Single entropy coder for HEVC with a high throughput binarization mode*. JCTVC-G569, Nov. 2011.

[69] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, "Sample adaptive offset in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1755–1764, Dec. 2012.

[70] C.-Y. Tsai, C.-Y. Chen, C.-M. Fu, Y.-W. Huang, and S.-M. Lei, "One-pass encoding algorithm for adaptive loop filter in high-efficiency video coding," in *Visual Communications and Image Processing (VCIP)*, Nov. 2011.

[71] ITU-T, "H.265 : high efficiency video coding (version 3)." `http://www.itu.int/rec/T-REC-H.265`, 10 2014. (Accessed on 02/09/2017).

[72] J. Chen, J. Boyce, Y. Ye, M. Hannuksela, G. J. Sullivan, and Y. Wang, "Hevc scalable extensions (shvc) draft text 7 (separated text)," *JCT-VC, Retrieved*, pp. 07–13, 2014.

[73] ITU-T, "Itu-t recommendation database." `http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=12455`, 4 2015. (Accessed on 02/09/2017).

[74] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging hevc screen content coding extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 50–62, 2016.

[75] ITU-T, "H.265 : high efficiency video coding (version 4)." `http://www.itu.int/rec/T-REC-H.265-201612-P`, 12 2016. (Accessed on 02/09/2017).

[76] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.

[77] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (hevc)," *IEEE Journal of selected topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013.

[78] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, pp. 1103 –1120, Sept. 2007.

[79] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *ITU-T Q. 6/SG16 VCEG, 15th Meeting, Austin, Texas, USA, April, 2001*, 2001.

[80] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

[81] A. S. Panayides, M. S. Pattichis, C. P. Loizou, M. Pantziaris, A. G. Constantinides, and C. S. Pattichis, "An effective ultrasound video communication system using despeckle filtering and HEVC," *IEEE journal of biomedical and health informatics*, vol. 19, no. 2, pp. 668–676, 2015.

[82] *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*. (Accessed on 10/27/2017).

[83] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, pp. 121–132, Feb. 2004.

[84] K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on*, vol. 19, pp. 335 –350, Feb. 2010.

[85] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, 2016.

[86] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[87] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, 2011.

[88] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[89] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 289–300, 2016.

[90] I. R. BT.500-11, *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, Geneva, Switzerland, 2002.

[91] *4G Americas*.

[92] G. TS, *High Speed Downlink Packet Access (HSDPA) Stage 2 – Release 5*.

[93] 3GPP, *Overview of the 3GPP Release 6, V0.1.1*. 2010.

[94] G. T. 25.913, *Requirements for Evolved (E-UTRA) and Evolved UTRAN (E-UTRAN)*. Mar. 2006.

[95] "IEEE std 802.16-2004: IEEE standard for local and metropolitan area networks part 16: Air interface for fixed broadband wireless access systems," June 2004.

[96] G. T. 36.300, "Evolved universal terrestrial radio access (e-UTRA) and evolved universal terrestrial radio access network (e-UTRAN); overall description; stage 2."

[97] I. . B. W. A. W. Group, "The draft IEEE 802.16m system description document (SSD)," July 2008.

[98] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[99] R. Perez, *Wireless Communications Design Handbook: Interference Into Circuits: Aspects of Noise, Interference, and Environmental Concerns*. Academic Press, Oct. 1998.

[100] "802.16e-2005 and IEEE std 802.16-2004/Cor1-2005: IEEE standard for local and metropolitan area networks part 16 : Air interface for fixed and mobile broadband wireless access systems amendment 2 : Physical and medium a c c e s s control layers for combined fixed and mobile operation in licensed bands," corrigendum 1, Dec. 2005.

[101] "IEEE std 802.16m-2011, standard for local and metropolitan area networks - part 16: Air interface for broadband wireless access systems - amendment 3: Advanced air interface," 2011.

[102] D. Niyato, E. Hossain, and J. Diamond, "IEEE 802.16/WiMAX-based broadband wireless access and its application for telemedicine/e-health services," *Wireless Communications, IEEE*, vol. 14, pp. 72 –83, Feb. 2007.

[103] D. Astely, E. Dahlman, P. Frenger, R. Ludwig, M. Meyer, S. Parkvall, P. Skillermark, and N. Wiberg, "A future radio-access framework," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 693–706, Mar. 2006.

[104] G. T. 36.913, *Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)*. Mar. 2009.

[105] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, pp. 44–52, Apr. 2009.

[106] G. Americas, *Summary of Global 5G Initiatives*. Apr. 2014. (Accessed on 12/31/2017).

[107] H. Tullberg, Z. Li, A. Hoglund, P. Fertl, D. Gozalvez-Serrano, K. Pawlak, P. Popovski, G. Mange, and O. Bulakci, "Towards the metis 5g concept: First view on horizontal topics concepts," in *Networks and Communications (EuCNC), 2014 European Conference on*, pp. 1–5, IEEE, 2014.

[108] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Ten Brink, I. Gaspar, N. Michailow, A. Festag, *et al.*, "5gnow: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, 2014.

[109] J. K. R. M. B. T. G. L. J. N. Paul Nikolich, Chih-Lin I and S. Zhang, "Standards for 5g and beyond: Their use cases and applications," *IEEE 5G Tech Focus*, vol. 1, June 2017. (Accessed on 01/04/2018).

[110] A. Panayides, M. Pattichis, A. Constantinides, and C. Pattichis, "M-health medical video communication systems: an overview of design approaches and recent advances," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 7253–7256, IEEE, 2013.

[111] A. Panayides, M. Pattichis, C. Pattichis, C. Schizas, A. Spanias, and E. Kyriacou, "An overview of recent end-to-end wireless medical video telemedicine systems using 3G," in

*Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 1045–1048, 2010.

[112] A. Alesanco, C. Hernández, A. Portolés, L. Ramos, C. Aured, M. García, P. Serrano, and J. García, "A clinical distortion index for compressed echocardiogram evaluation: recommendations for xvid codec," *Physiological Measurement*, vol. 30, pp. 429–440, May 2009.

[113] E. Cavero, A. Alesanco, and J. Garcia, "Enhanced protocol for real time transmission of echocardiograms over wireless channels," 2012.

[114] C. J. Debono, B. W. Micallef, N. Y. Philip, A. Alinejad, R. S. Istepanian, and N. N. Amso, "Cross-layer design for optimized region of interest of ultrasound video data over mobile wimax," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1007–1014, 2012.

[115] A. Alinejad, N. Y. Philip, and R. S. Istepanian, "Cross-layer ultrasound video streaming over mobile wimax and hsupa networks," *IEEE transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 31–39, 2012.

[116] Y. Chu and A. Ganz, "A mobile teletrauma system using 3G networks," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 8, pp. 456 –462, Dec. 2004.

[117] S. Garawi, R. Istepanian, and M. Abu-Rgheff, "3G wireless communications for mobile robotic tele-ultrasonography systems," *Communications Magazine, IEEE*, vol. 44, pp. 91 – 96, Apr. 2006.

[118] M. G. Martini and C. T. E. R. Hewage, "Flexible macroblock ordering for context-aware ultrasound video transmission over mobile WiMAX," *Int. J. Telemedicine Appl.*, vol. 2010, pp. 6:1–6:7, Jan. 2010.

[119] S. Khire, S. Robertson, N. Jayant, E. A. Wood, M. E. Stachura, and T. Goksel, "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios," in *MILCOM 2012-2012 IEEE Military Communications Conference*, pp. 1–6, IEEE, 2012.

[120] E. Cavero, A. Alesanco, L. Castro, J. Montoya, I. Lacambra, and J. Garcia, "Spiht-based echocardiogram compression: clinical evaluation and recommendations of use," *IEEE journal of biomedical and health informatics*, vol. 17, no. 1, pp. 103–112, 2013.

[121] A. Panayides, Z. Antoniou, Y. Mylonas, M. Pattichis, A. Pitsillides, and C. Pattichis, "High-resolution, low-delay, and error-resilient medical ultrasound video communication using h. 264/AVC over mobile WiMAX networks," *IEEE Biomedical and Health Informatics*, vol. 17, pp. 619–618, May 2013.

[122] R. Research and LCC, "4G mobile broadband evolution: 3GGP release 10 and beyond," Jan. 2012.

[123] C. Loizou, C. Pattichis, M. Pantziaris, and A. Nicolaides, "An integrated system for the segmentation of atherosclerotic carotid plaque," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 11, pp. 661 –667, Nov. 2007.

[124] I. L. S. Committee *et al.*, "Ieee standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1," *IEEE Std 802.16-2004/Cor 1-2005*, 2006.

[125] M. Alasti, B. Neekzad, J. Hui, and R. Vannithamby, "Quality of service in WiMAX and LTE networks [topics in wireless communications]," *Communications Magazine, IEEE*, vol. 48, pp. 104 –111, May 2010.

[126] I. M. . I.-T. VCEG, "H.264/AVC JM 15.1 reference software." http://iphome.hhi.de/suehring/tml/.

[127] *OPNET University Program*.

[128] G. Bjontegaard, "Improvements of the BD-PSNR model," July 2008.

[129] A. Alinejad, N. Y. Philip, and R. S. Istepanian, "c," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 1, pp. 31–39, 2012.

[130] D. eHealth Action Plan 2012-2020 (2012, "Innovative healthcare for the 21st century." http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0736. (Accessed on 08/08/2017).

[131] A. S. Panayides, M. S. Pattichis, and C. S. Pattichis, "Mobile-health systems use diagnostically driven medical video technologies," *IEEE Signal Proc. Mag*, vol. 30, no. 6, pp. 163–172, 2013.

[132] A. S. Panayides, Z. C. Antoniou, and A. G. Constantinides, "An overview of mhealth medical video communication systems," in *Mobile Health*, pp. 609–633, Springer, 2015.

[133] "Itu-t rec. h.265 (04/2013) high efficiency video coding." (Accessed on 12/06/2016).

[134] R. I.-R. M.2083, *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond*. 2015. (Accessed on 12/06/2016).

[135] R. S. Istepanian, N. Y. Philip, and M. G. Martini, "Medical qos provision based on reinforcement learning in ultrasound streaming over 3.5 g wireless systems," *IEEE Journal on Selected areas in Communications*, vol. 27, no. 4, pp. 566–574, 2009.

[136] M. G. Martini and C. T. Hewage, "Flexible macroblock ordering for context-aware ultrasound video transmission over mobile WiMAX," *International journal of telemedicine and applications*, vol. 2010, p. 6, 2010.

[137] A. Panayides, I. Eleftheriou, and M. Pantziaris, "Open-source telemedicine platform for wireless medical video communication," *International journal of telemedicine and applications*, vol. 2013, 2013.

[138] S. Cicalò, M. Mazzotti, S. Moretti, V. Tralli, and M. Chiani, "Multiple video delivery in m-health emergency applications," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 1988–2001, 2016.

[139] M. Razaak, M. G. Martini, and K. Savino, "A study on quality assessment for medical ultrasound video compressed via hevc," *IEEE Journal of biomedical and health informatics*, vol. 18, no. 5, pp. 1552–1559, 2014.

[140] C. Herglotz, D. Springer, and A. Kaup, "Modeling the energy consumption of hevc p-and b-frame decoding," in *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 3661–3665, IEEE, 2014.

[141] X. Wang, M. Chen, T. T. Kwon, L. Yang, and V. C. Leung, "Ames-cloud: a framework of adaptive mobile video streaming and efficient social video sharing in the clouds," *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 811–820, 2013.

[142] O. Ognenoski, M. Razaak, M. G. Martini, and P. Amon, "Medical video streaming utilizing MPEG-DASH," in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 54–59, IEEE, 2013.

[143] L. Balaji and K. Thyagharajan, "A pixel orientation and adaptive search range based complexity reduction in h. 264 scalable video coding," in *Advanced Computing and Communication Systems (ICACCS), 2017 4th International Conference on*, pp. 1–5, IEEE, 2017.

[144] M. Hosseini, Y. Jiang, R. R. Berlin, L. Sha, and H. Song, "Toward physiology-aware dash: Bandwidth-compliant prioritized clinical multimedia communication in ambulances," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2307–2321, 2017.

[145] Z. C. Antoniou, A. S. Panayides, M. Pantziaris, A. G. Constantinides, C. S. Pattichis, and M. S. Pattichis, "Real-time adaptation to time-varying constraints for medical video communications," *IEEE journal of biomedical and health informatics*, 2017.

[146] Z. Sheng, D. Zhou, H. Sun, and S. Goto, "Low-complexity rate-distortion optimization algorithms for hevc intra prediction," in *International Conference on Multimedia Modeling*, pp. 541–552, Springer, 2014.

[147] X. Li, M. Wien, and J.-R. Ohm, "Rate-complexity-distortion optimization for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 957–970, 2011.

[148] "Iso/iec 23009-1:2014 - information technology – dynamic adaptive streaming over http (dash) – part 1: Media presentation description and segment formats." `http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=65274`. (Accessed on 12/06/2016).

[149] C. Zhou, C.-W. Lin, and Z. Guo, "m DASH: A Markov decision-based rate adaptation approach for dynamic HTTP streaming," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 738–751, 2016.

[150] M. Martini, R. Istepanian, M. Mazzotti, and N. Philip, "Robust multilayer control for enhanced wireless telemedical video streaming," *Mobile Computing, IEEE Transactions on*, vol. 9, pp. 5 –16, Jan. 2010.

[151] C. Castejón, G. Carbone, J. García Prada, and M. Ceccarelli, "A multi-objective optimization of a robotic arm for service tasks.," *Strojniski Vestnik/Journal of Mechanical Engineering*, vol. 56, no. 5, pp. 316–329, 2010.

[152] Y. Censor, "Pareto optimality in multiobjective problems," *Applied Mathematics & Optimization*, vol. 4, no. 1, pp. 41–59, 1977.

[153] K. Deb, *Multi-objective optimization*. Springer, 2014.

[154] C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, *et al.*, *Evolutionary algorithms for solving multi-objective problems*, vol. 5. Springer, 2007.

[155] G. Z.-Z. Q. Cui, Yunfei and Y. Han, "Multi-objective optimization methods and application in energy saving," *Energy*, vol. 125, pp. 681–704, 2017.

[156] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Applied soft computing*, vol. 8, no. 1, pp. 687–697, 2008.

[157] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on evolutionary computation*, vol. 1, no. 1, pp. 53–66, 1997.

[158] N. Srinivas and K. Deb, "Muiltiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary computation*, vol. 2, no. 3, pp. 221–248, 1994.

[159] J. Horn, N. Nafpliotis, and D. E. Goldberg, "A niched pareto genetic algorithm for multiobjective optimization," in *Evolutionary Computation, 1994. IEEE World Congress on Computational Intelligence., Proceedings of the First IEEE Conference on*, pp. 82–87, Ieee, 1994.

[160] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," *Evolutionary computation*, vol. 8, no. 2, pp. 149–172, 2000.

[161] D. W. Corne, J. D. Knowles, and M. J. Oates, "The pareto envelope-based selection algorithm for multiobjective optimization," in *International Conference on Parallel Problem Solving from Nature*, pp. 839–848, Springer, 2000.

[162] M. Erickson, A. Mayer, and J. Horn, "The niched pareto genetic algorithm 2 applied to the design of groundwater remediation systems," in *Evolutionary Multi-Criterion Optimization*, pp. 681–695, Springer, 2001.

[163] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[164] Y. Jiang and M. S. Pattichis, "A dynamically reconfigurable architecture system for time-varying image constraints (drastic) for motion jpeg," *Journal of Real-Time Image Processing*, pp. 1–17, 2014.

[165] C. I. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *IEEE Transactions on medical imaging*, vol. 22, no. 7, pp. 902–912, 2003.

[166] "Epiphan video, dvi2usb 3.0 software documentation, drivers and software." https://www.epiphan.com/products/. (Accessed on 12/13/2016).

[167] Z. Antoniou, A. Panayides, M. Pattichis, S. Stavrou, E. Kyriacou, A. Spanias, A. Constantinides, and C. Pattichis, "Adaptive emergency scenery video communications using hevc for responsive decision support in disaster incidents," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 173–176, IEEE, 2015.

[168] "R: The R project for statistical computing." https://www.r-project.org/. (Accessed on 12/14/2016).

[169] W. R. Rice, "Analyzing tables of statistical tests," *Evolution*, vol. 43, no. 1, pp. 223–225, 1989.

[170] T. J. Ypma, "Historical development of the newton–raphson method," *SIAM review*, vol. 37, no. 4, pp. 531–551, 1995.

[171] M. Crisfield, "Accelerating and damping the modified newton-raphson method," *Computers & structures*, vol. 18, no. 3, pp. 395–407, 1984.

[172] M. Inc, "x265 documentation." http://x265.readthedocs.io/en/1.7/, 2015. (Accessed on 12/14/2016).

[173] R. R. Americas, *Mobile Broadband Explosion*. 2013. (Accessed on 08/21/2017).

[174] J. Chakareski, "Adaptive multiview video streaming: challenges and opportunities," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 94–100, 2013.

[175] M. M. Nasralla, C. T. Hewage, and M. G. Martini, "Subjective and objective evaluation and packet loss modeling for 3D video transmission over LTE networks," in *Telecommunications and Multimedia (TEMU), 2014 International Conference on*, pp. 254–259, IEEE, 2014.

[176] C. T. Hewage, M. G. Martini, and N. Khan, "3D medical video transmission over 4g networks," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, p. 180, ACM, 2011.

[177] "Laboratory for Image and Video Engineering - The University of Texas at Austin." `http://live.ece.utexas.edu/index.php`. (Accessed on 11/14/2017).

[178] "TruNET wireless." `www.fractalnetworx.com`, 2017.