ARIB TR-T12-26.904 V11.0.0

Improved Video Coding Support

(Release 11)

Refer to “Notice” in the preface of ARIB TR-T12 for Copyrights.

3GPP TR 26.904 V11.0.0 (2012-09)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Improved Video Coding Support (Release 11)

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.
The present document has not been subject to any approval process by the 3GPP Organizational Partners and shall not be implemented.
This Specification is provided for future development work within 3GPP only. The Organizational Partners accept no liability for any use of this Specification.
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices.

Keywords

GSM, UMTS, codec, LTE, video

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Copyright Notification

No part may be reproduced except as authorized by written permission.
The copyright and the foregoing restriction extend to reproduction in all media.

© 2012, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TTA, TTC).
All rights reserved.
UMTS™ is a Trade Mark of ETSI registered for the benefit of its members
3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners
LTE™ is a Trade Mark of ETSI currently being registered for the benefit of its Members and of the 3GPP Organizational Partners
GSM® and the GSM logo are registered and owned by the GSM Association

Contents

Foreword 5
1 Scope 6
2 References 6
3 Definitions and abbreviations 7
3.1 Definitions 7
3.2 Abbreviations 8
4 General 8
4.1 Introduction 8
5 Use Cases 8
5.1 2D Video Use Cases 8
5.1.1 Adaptive HTTP Streaming and Caches 8
5.1.2 UE Power Saving and Fast Stream Switching in MBMS 9
5.1.3 Graceful Degradation 9
5.1.3.1 Rate Adaptation in PSS When Entering Bad Reception Conditions 9
5.1.3.2 Graceful Degradation in MBMS Services When Entering Bad Reception Conditions 9
5.1.3.3 Graceful Degradation in Traffic Congestion 9
5.1.3.4 Combined Support of Heterogeneous Devices and Graceful Degradation 9
5.2 Stereoscopic 3D Video Use Cases 10
5.2.1 Stereoscopic 3D Video Delivery 10
5.2.2 External Viewing 3D Experience 10
5.2.2.1 Introduction 10
5.2.2.2 Video Eyewear 3D Experience 10
5.2.2.3 Mobile Terminal Connected to a 3DTV Set 11
6 Evaluation of Solutions 12
6.1 2D Use Cases 12
6.1.1 Enabling Codecs and Formats 12
6.1.1.1 Scalable Video Coding 12
6.1.1.1.1 Introduction 12
6.1.1.1.2 Solution Configuration 14
6.1.2 Solution Integration Approaches 14
6.1.2.1 Rate Adaptation for PSS using SVC with Priority-Based Transmission Scheduling 14
6.1.2.2 Rate Adaptation using SVC MGS Scalability 16
6.1.2.2.1 Rate Adaptation Approach 16
6.1.2.2.2 Example 17
6.1.2.3 Unequal Error Protection with SVC in eMBMS 19
6.1.2.4 SVC Layer Aware Transmission in eMBMS 19
6.1.2.5 Fast Stream Switching in eMBMS 20
6.1.3 Performance Evaluation 21
6.1.3.1 Unequal Error Protection in eMBMS 21
6.1.3.2 SVC Layer Aware Transmission for Coverage Improvement in eMBMS 25
6.1.3.3 SVC Layer Aware Transmission for Capacity Improvement in eMBMS 28
6.1.3.3.1 Introduction 28
6.1.3.3.2 Evaluation Setup 28
6.1.3.3.3 Capacity Improvement 29
6.1.3.3.4 Performance Evaluation 32
6.1.3.4 Graceful Degradation for MBMS Using SVC 37
6.1.3.4.1 Introduction 37
6.1.3.4.2 Test system 38
6.1.3.4.2 Test sequences 40
6.1.3.4.4 Transmission Schemes 41
6.1.3.4.5 Quality metric 42
6.1.3.4.6 Simulation Results 43
6.1.3.5 Coding Results using KTA 45

6.1.3.5.1 Experimental Setup ... 45
6.1.3.5.2 Results ... 46
6.1.3.6 Coding Results using JSVM ... 51
6.1.3.6.1 JSVM ... 51
6.1.3.6.2 Experimental Setup ... 51
6.1.3.6.3 Sequences ... 51
6.1.3.6.4 Coding Tools ... 51
6.1.3.6.5 Results ... 52
6.1.3.6.5.1 320x240 (QVGA) SNR Scalability (Configuration 1) ... 52
6.1.3.6.5.2 320x240 (QVGA) SNR Scalability (Configuration 2) ... 55
6.1.3.7 Caching Efficiency Improvement with SVC for Adaptive HTTP VoD ... 58
6.1.3.7.1 Overview ... 58
6.1.3.7.2 Effect of Multiple Representations on the Caching Efficiency ... 59
6.1.3.7.3 Scalable Video Coding and Impact on the Caching Efficiency ... 60
6.1.3.7.4 Caching Algorithm ... 61
6.1.3.7.5 Congestion Control ... 61
6.1.3.7.6 Performance Targets ... 62
6.1.3.7.7 Simulation Results ... 62
6.2 Stereoscopic 3D Video ... 66
6.2.1 Enabling Codecs and Formats ... 66
6.2.1.1 Introduction ... 66
6.2.1.2 Packing Formats ... 66
6.2.1.2.1 Frame Compatible Video ... 66
6.2.1.2.2 Temporal Interleaving ... 67
6.2.1.3 Multi-view Video ... 68
6.2.2 Performance Evaluation ... 68
6.2.2.1 Performance Evaluation of the Compression Efficiency ... 68
6.2.2.1.1 Simulation Setup ... 68
6.2.2.1.2 Performance Evaluation ... 69

7 Conclusions ... 71

Annex A: Assumptions for Simulation Method for Solutions on MBMS Services ... 73

Annex B: Impact of Screen Size on Stereoscopic Video ... 80

B.1 Geometry of Stereoscopic Video ... 80

B.2 Depth Range ... 80

B.3 Effect of Display Size Changes ... 81

B.4 Discussion ... 82

Annex C: Real World Statistics of VoD User Request ... 83

Annex D: Change history ... 85

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document provides an analysis of the future video capability requirements of streaming and multicast/broadcast services. The purpose of the present document is two-fold. On the one hand, it studies the options to upgrade the minimal requirements for video reception and decoding. On the other hand, it studies use cases for support of more advanced UEs. The ultimate target of this study item is to recommend solutions for efficiently providing video support commensurate with UE and user capabilities and needs in PSS and MBMS services.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TS 26.346: "Multimedia Broadcast/Multicast Services (MBMS); Protocols and Codecs".

[2] 3GPP TS 26.234: "Transparent End-to-End Packet Switched Streaming Service (PSS); Protocols and Codecs".

[3] ITU-T Recommendation H.264 (03/09), "Advanced video coding for generic audiovisual services" | ISO/IEC 14496- 10:2009 Information technology—Coding of audiovisual objects— part 10: Advanced Video Coding".

[4] T. Schierl, Y. Sanchez de la Fuente, C. Hellge, and T. Wiegand: "Priority-based Transmission Scheduling for Delivery of Scalable Video Coding over Mobile Channels," 3rd European Symposium on Mobile Media Delivery (EUMOB), London, 2009.

[5] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[6] 3GPP TR 25.814 (V7.1.0): "Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA) (Release 7)".

[7] H.264/AVC Reference Software, http://iphome.hhi.de/suehring/tml/download/jm17.2.zip.

[8] KTA Software, http://iphome.hhi.de/suehring/tml/.

[9] Nokia MVC Software, http://research.nokia.com/page/4988.

[10] M. Luby, T. Gasiba, T. Stockhammer, M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," Broadcasting, IEEE Transactions on, Vol. 53, Issue 1, Part 2, pp235-246, March 2007.

[11] O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 42083, 21 pages, 2006.

[12] A. P. Couto da Silva, P. Rodrıguez-Bocca, and G. Rubino, "Optimal quality-of-experience design for a P2P multi-source video streaming," in *Proceedings of the IEEE International Conference on Communications (ICC '08)*, pp. 22–26, Beijing, China, May 2008.

[13] Cornelius Hellge, Thomas Schierl, Jörg Huschke, Thomas Rusert, Markus Kampmann, Thomas Wiegand: Graceful degradation in 3GPP MBMS Mobile TV services using H.264/AVC temporal scalability; Eurasip Journal on Wireless Communications and Networking; August 2009.

[14] JSVM reference software, version 9.17, available via CVS from "garcon.ient.rwth-aachen.de:/cvs/jvt".

[15] VCEG-AJ10r1: "Recommended Simulation Common Conditions for Coding Efficiency Experiments".

[16] R. Skupin, C. Hellge, T. Schierl and T. Wiegand, "Fast Application-level Video Quality Evaluation for Extensive Error-Prone Channel Simulations", *15th International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD)*, Miami, 2010.

[17] G. Liebl et al., "Simulation platform for multimedia broadcast over DVB-SH", *3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools)*, Malaga, 2010

[18] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no.9, pp 1103–1120, 2007.

[19] H. Hoffmann (EBU), T. Itagaki (Brunel University, UK), D. Wood (EBU), "Quest for Finding the Right HD Format: A New Psychophysical Method for Subjective HDTV Assessment," *SMPTE Motion Imaging Journal*, Issue: 04 April, 2008.

[20] 3GPP TR 26.902: "Video Codec Performance (Release 7)", June 2007.

[21] D. Hong, D. De Vleeschauwer, F. Baccelli, "A Chunk-based Caching Algorithm for Streaming Video", *Proceedings of the 4th Workshop on Network Control and Optimization*, Ghent (Belgium), November 29 – December 1, 2010.

[22] Y. Sanchez, T. Schierl, C. Hellge, T. Wiegand, D. Hong, D. De Vleesschauwer, W. Van Leekwijck, Y. Lelouedec, "Improved caching for HTTP-based Video on Demand using Scalable Video Coding, " *Consumer Communication & Networking Conference 2011 (CCNC 2011)*, Special Session on IPTV and Multimedia CDN, Las Vegas, Nevada, USA, 9-12 January 2011.

[23] Laurent Chauvier, Kevin Murray, Simon Parnall, Ray Taylor, James Walker, "Does size matter: the challenges when scaling stereoscopic 3D content", *Proc. IBC 2010 Conference*, http://www.nds.com/pdfs/3DTV-DoesSizeMatter_IBC2010Award.pdf

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [5] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [5].

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [5] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [5].

AVC	Advanced Video Coding
BLER	BLock Error Rate
CDN	Content Delivery Network
CGS	Coarse Grain Scalability
ESR	Erroneous Seconds Ratio
GOP	Group Of Pictures
JSVM	Joint Scalable Video Model
KTA	Key Technical AreasMBSFN Multicast Broadcast Single Frequency Network
MCS	Modulation and Coding Scheme
MGS	Medium Grain Scalability
MVC	Multi-view Video Coding
NAL	Network Abstraction Layer
PBTS	Priority Based Transmission Scheduling
PLR	Packet Loss Rate
PSNR	Peak Signal to Noise Ratio
SVC	Scalable Video Coding
TTI	Transmission Time Interval
UCC	Used Cell Capacity
UEP	Unequal Error Protection

4 General

4.1 Introduction

This Technical Report studies use cases and solutions for both video scalability and 3D stereoscopic video and investigates their performance in a variety of setups using 3GPP's streaming and multicast/broadcast services. Subclause 5.1 introduces use cases on 2D service, and the codec solutions enabling the use cases are described in subclause 6.1.1. Subclause 6.1.2 introduces some applications integrating the 2D solutions and codecs, and the performance of the solutions is evaluated in subclause 6.1.3. Stereoscopic 3D use cases are introduced in subclause 5.2. Enabling codecs for the 3D use cases are described in subclause 6.2.1, and the performance is evaluated in subclause 6.2.2. This document includes two attachment files which are the config files used in the evaluations of codecs. Annexes for helping understanding the simulation conditions are also included. Finally conclusion based on the study of this TR is presented in subclause 7.

5 Use Cases

5.1 2D Video Use Cases

5.1.1 Adaptive HTTP Streaming and Caches

This use case considers HTTP-based streaming delivery of video content. Caching of popular content can significantly decrease the average and peak load within a 3GPP backbone. Using HTTP streaming, caching can be performed by standard HTTP caches.

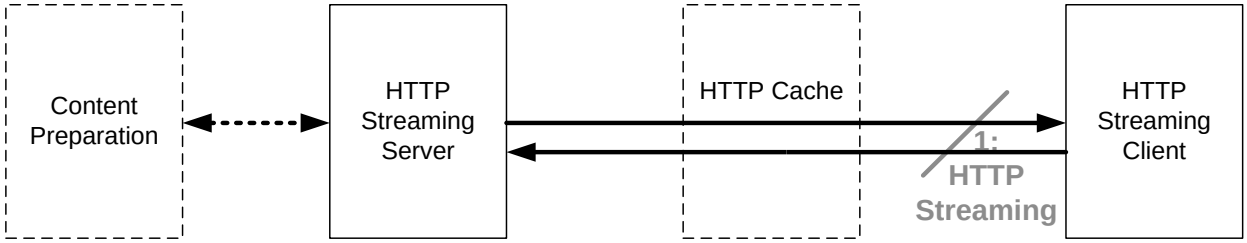

Figure 1: System architecture for adaptive HTTP streaming [2].

In this use case the media coding, especially the video coding (e.g. multi-layered SVC compared to multi bitrate versions of H.264/AVC single layer coding), and its integration into HTTP Streaming framework will be evaluated with respect to improvement in usage of originating server, backbone and caches, and in general its impact on the system, including the impact on encoders and clients. Furthermore the effect of rate adaptation will be evaluated in such scenarios.

5.1.2 UE Power Saving and Fast Stream Switching in MBMS

Efficient power usage is an important criterion in providing MBMS TV service. When the TV stream is transmitted continuously, UE should receive data continuously in active mode, as a result, battery power is consumed. Typical method used for UE power saving is scheduling the transmission and sleep period that UE may turn-off radio component during the sleep interval. This requires discontinuous transmission of MBMS streams. However, a trade-off is that user may experience long delay when switching between streams, if the sleep interval is increased Therefore, it should be able to support efficient power usage of UE as well as fast content switching.

5.1.3 Graceful Degradation

5.1.3.1 Rate Adaptation in PSS When Entering Bad Reception Conditions

A mobile TV service may have to cope with varying reception conditions at the UE to avoid service interruptions. A desired behaviour would be to apply by rate adaptation of the video stream to the achievable service bit rate. Since a reduced media rate results in a reduced video play out quality, such a video stream adaptation should be performed in a graceful way. Therefore, the service should allow a fine granular rate adaptation to avoid abrupt quality changes in an efficient way

5.1.3.2 Graceful Degradation in MBMS Services When Entering Bad Reception Conditions

In contrary to a PSS service, an MBMS service cannot adapt to individual receivers need. That is, users entering difficult reception conditions may experience sudden service interruption instead of soft degradation of e.g. video quality. To keep users satisfied when switching from PSS services to MBMS, a Graceful Degradation of the broadcast service is a desired feature. Such a feature can be applied to a broadcast service by allowing differentiation transmission robustness for different parts of the video stream. The service should allow minimum acceptable quality to the user perception at the service coverage configured by operator.

5.1.3.3 Graceful Degradation in Traffic Congestion

In a situation where multiple service users converge in a cell, available bandwidth of the cell depletes quickly. In such case, service to lately incoming UEs may be refused, or all UEs in the cell may suffer severe quality degradation. The situation can be improved when bandwidth of the streams can be reduced with graceful quality degradation using IVS. The service quality is recovered as congestion state of the cell is relieved.

5.1.3.4 Combined Support of Heterogeneous Devices and Graceful Degradation

It is expected, that there will be a coexistence of a variety of device capabilities within 3GPP system and each of these devices may be in different reception conditions. Therefore to cope with both of these challenges in an efficient way, a service should be able to support the heterogeneous devices and to provide Graceful Degradation behaviour at the same time.

5.2 Stereoscopic 3D Video Use Cases

5.2.1 Stereoscopic 3D Video Delivery

Stereoscopic 3D video content is becoming increasingly available. A steadily growing share of professionally produced content is captured in stereoscopic 3D format. On the other hand, mobile devices with 3D rendering capabilities will gradually enter the market. Since capturing clean stereoscopic 3D video is extremely challenging, it is expected that the main short-term usage of these device capabilities will be for the consumption of professionally produced stereoscopic 3D content. Figure 2 depicts an example setup for the distribution of stereoscopic 3D content. While 3D capable devices will enjoy the stereo video, it should be possible to author so that legacy devices can consume the same content in 2D.

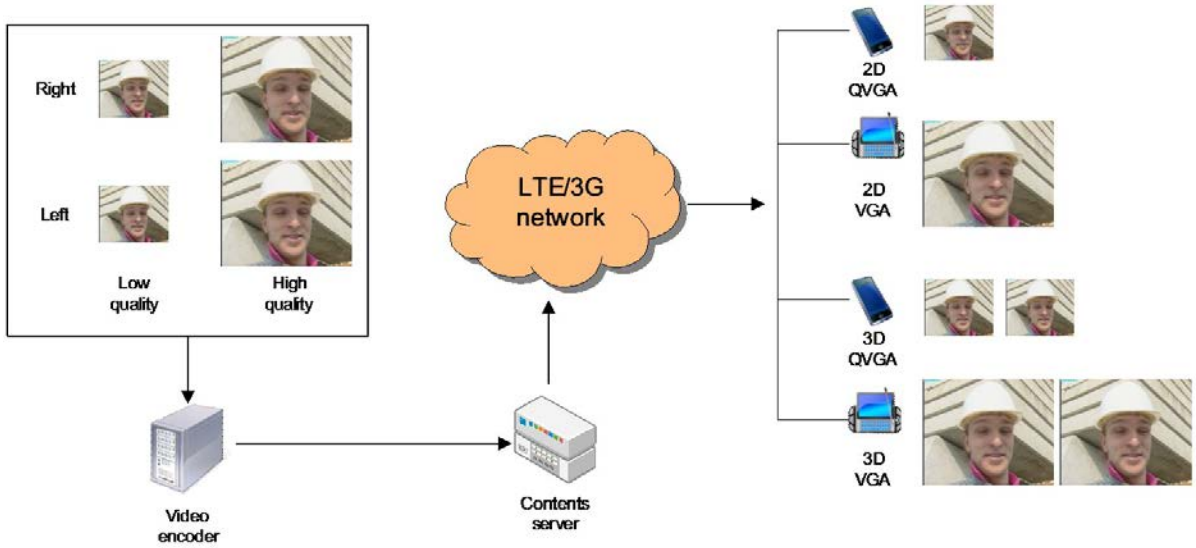

Figure 2: Example scenario of distribution of 2D and stereoscopic 3D video

Services such as PSS and MBMS provide the right channels for distributing the content to 3D capable mobile devices. The specified delivery options include multicast, RTP streaming, adaptive HTTP streaming and progressive download.

This use case may be enabled through different video coding solutions such as H.264/MVC [3] and frame-compatible H.264/AVC (with SEI signaling). These solutions will be studied and their performances will be evaluated.

It is in the scope of the study to consider not only coding and backwards-compatibility, but also the suitability of mobile devices in general for viewing 3D content (considering issues such as screen size, viewing distance, and resolution, for example). It is also in scope to consider whether 3D content from other domains could be re-targeted or whether the mobile environment might need custom 3D content preparation. Finally, consideration of whether different mobile devices might need different content (not just, for example, different encodings or resolutions), is in scope.

5.2.2 External Viewing 3D Experience

5.2.2.1 Introduction

The following use cases are based on the same access conditions as presented in 5.2.1. They propose the ability to decode a 3D video content directly on the UE with using an external display to provide the 3D experience.

5.2.2.2 Video Eyewear 3D Experience

This use case describes a 3D experience provided thanks to video glasses (also called video eyewear headsets) compatible with stereoscopic video. During the recent years, progress has been achieved on the ability to use such video glasses in order to simulate large screen viewing experience. When connected to a mobile terminal receiving a stereoscopic video the video glasses display the left view on the left eye and the right view on the right eye. Each eye receiving a different view, the depth is provided to the user.

Figure 3 illustrates the current use case.

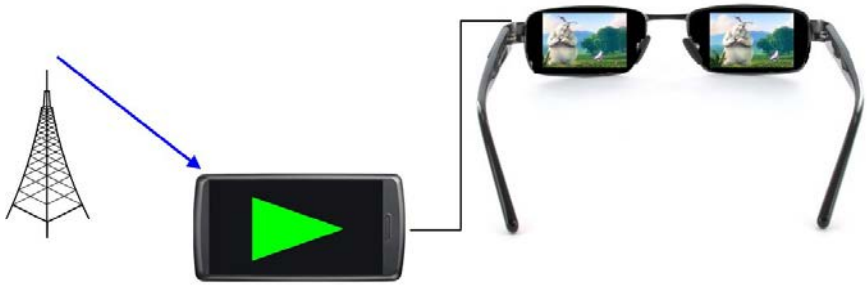

Figure 3: Use case of 3D content viewed on video glasses

5.2.2.3 Mobile Terminal Connected to a 3DTV Set

This use case can be associated to the mobile 3D concept in the way it enables 3D experience when receiving a video content over the 3GPP access network. A user wants to watch a 3D movie on its 3D compatible TV set at home. He may take advantage of its LTE coverage to get the streamed video which is decoded in its mobile terminal. The terminal has a digital connectivity which enables the connection with a 3DTV display (e.g. via a micro-HDMI/HDMI cable). In this use case the mobile terminal acts as a mobile Set top box.

Figure 4 hereafter illustrates the current use case.

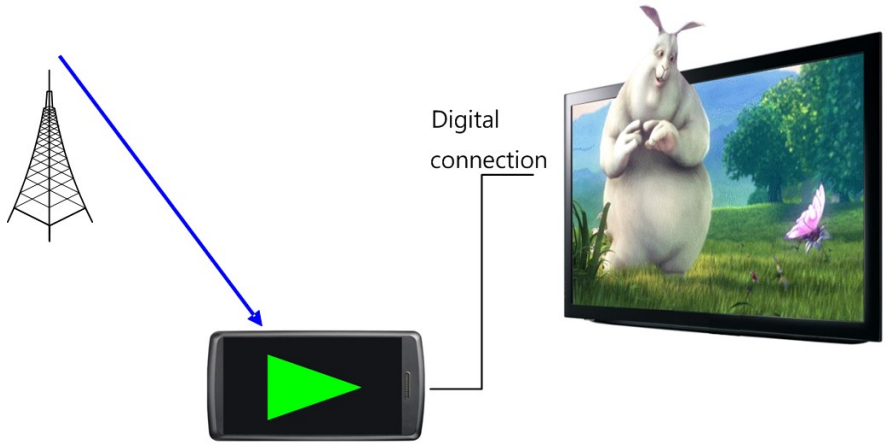

Figure 4: Use case of 3D content viewed on a 3D TV set

6 Evaluation of Solutions

6.1 2D Use Cases

6.1.1 Enabling Codecs and Formats

6.1.1.1 Scalable Video Coding

6.1.1.1.1 Introduction

Scalable Video Coding (SVC) [3] has been defined as an extension to the H.264/AVC [3] video coding standard. SVC enhances H.264/AVC with a set of new profiles and encoding tools that may be used to produce scalable bitstreams. SVC supports three different types of scalability: spatial scalability, temporal scalability, and quality scalability. Temporal scalability is realized using the already existing reference picture selection flexibility in H.264/AVC [3] as well as bi-directionally predicted B-pictures. The prediction dependencies of B-pictures are arranged in a hierarchical structure. Furthermore, appropriate rate control is used to adjust the bit budget of each picture to be proportional to its temporal importance in a procedure called quantization parameter cascading. The slightly and gradually reduced picture quality of the hierarchical B-pictures has been shown not to significantly impact the subjective quality and the watching experience, while showing high compression efficiency. Figure 5 shows an example of the realization of temporal scalability using hierarchical B-pictures. The example shows 4 different temporal levels, resulting in one base layer and 3 temporal enhancement layers. This allows the frame rate to be scaled by a factor up to 8 (e.g. from 60Hz to 7.5Hz). This approach has the drawback that it incurs a relatively high decoding delay that is exponentially proportional to the number of temporal layers, since the pictures have to be decoded in a different order than their display order. As the coding gain also diminishes with the increasing number of hierarchy levels, it is not appropriate to generate a high number of temporal layers. An alternative to the above mentioned approach for temporal scalability is the use of low-delay uni-directional prediction structures, hence avoiding the out-of-display-order decoding at the cost of reduced coding efficiency.

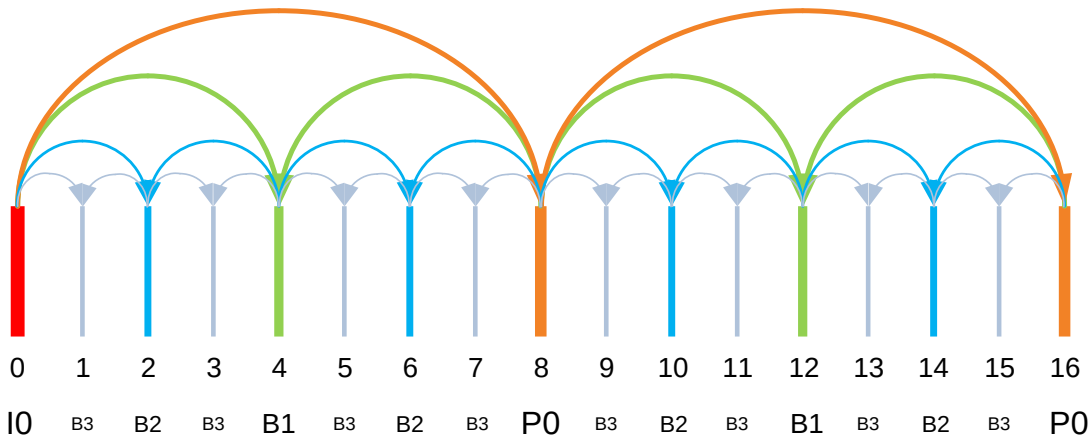

Figure 5: Temporal scalability with hierarchical B-picture structure in SVC

Spatial scalability is the most important scalability type in SVC. It enables encoding a video sequence into a video bit stream that contains one or more subset bit streams and where each of these subsets provides a video at a different spatial resolution. The spatially scalable video caters for the needs of different consumer devices with different display capabilities and processing power. Figure 6 depicts an example for a prediction structure for spatial scalability (QCIF to CIF resolution). The spatial scalability layer is enhanced with an additional temporal scalability layer that doubles the frame rate at the CIF resolution.

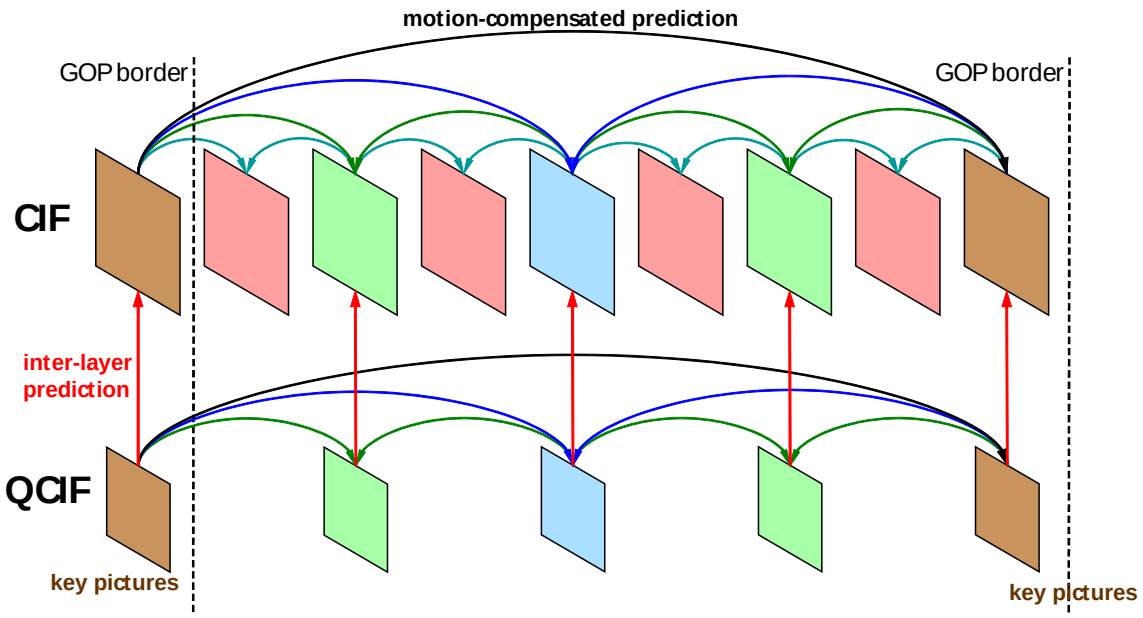

Figure 6: Example prediction structure for spatial scalability

SVC defines three different inter-layer prediction modes that are designed to enable the single-loop low complexity decoding at the decoder. In other words, motion compensation is performed only once at the target layer at the decoder. The inter-layer prediction tools are inter-layer INTRA (texture) prediction, inter-layer motion prediction, and inter-layer residual prediction.

Inter-layer INTRA prediction enables texture prediction from the base layer at co-located macro-blocks (after upsampling). It is restricted to INTRA coded macroblocks at the lower layer. The up-sampling of the macroblock texture is performed using well-specified up-sampling filters (a 4-tap filter for Luma samples and bi-linear filter from chroma samples). Inter-layer motion prediction implies prediction of the base layer motion vector from the co-located INTER-coded macro-block (after upsampling) of the lower layer. The prediction involves all components of the motion vector: the macro-block partitioning structures, the reference picture indices, and the x- and y- components representing the motion direction. Finally, the inter-layer residual prediction allows inter-layer prediction from the residual after INTER-prediction at the lower layer. At the decoder side, the residual information of the target layer is built up by summing all correctly up-scaled residuals of the lower dependent layers.

The third prediction type in SVC is quality scalability. Quality scalability enables the achievement of different operation points, each yielding a different video quality. Coarse Grain Scalability (CGS) is a form of quality scalability that uses the same tools as the spatial scalability, hence operating in the spatial domain. Alternatively, Medium Grain Scalability (MGS) may be used to achieve quality scalability performing the inter-layer prediction at the transform domain. Two techniques are advocated for MGS scalability: splitting number of transform coefficients and encoding difference of transform coefficients quantized using different quantization parameters. MGS significantly reduces the complexity at encoder and decoder. CGS may be seen as a variant of spatial scalability where the spatial scaling factor is set to one. Quality scalability may be used to address different use cases such as rate adaptation or for offering a high quality pay service.

SVC MGS scalability offers an increased flexibility for bit stream adaptation and error robustness. Scalable streams providing a variety of bit rates can be effectively encoded. For MGS coding in SVC, two new features have been introduced: *motion-compensated prediction for the base layer from the enhancement* layer and the supported of so-called *key pictures*.

These concepts are illustrated in Figure 7.

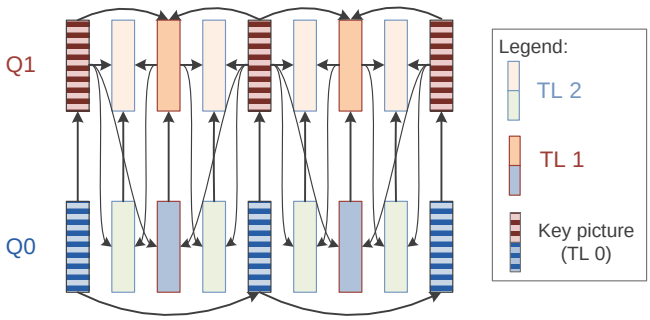

Figure 7: Key picture concept of SVC for hierarchical prediction structures

The first feature enables a simple but effective drift control for hierarchical prediction structures. For each picture it is signalled whether the base layer representation (when available) or the enhancement layer representation of the reference pictures is employed for motion-compensated prediction. Pictures that use the base layer representation for motion-compensated prediction are called *key pictures* (see Figure 7). This has the advantage that if any enhancement layer data are lost, no drift occurs between encoder and decoder reconstruction.

6.1.1.1.2 Solution Configuration

For the purposes of improved video support in 3GPP services, a profile of SVC [3] is selected that allows backwards compatibility to basic terminals. This is inherently provided by SVC by requesting the base layer to be H.264/AVC [3] compatible. Furthermore, it has to be ensured that the base layer also conforms to the minimal requirements for basic services. In order to ensure the conformance with the constrained baseline profile of H.264/AVC [3]. SVC has to be used according to the Scalable Baseline profile with the same constraints.

Additionally, the level selection for a base layer has to be aligned with the minimal level requirements for 3GPP services. For enhancement layers, the level selection is proposed to be set to level 3, which has the following characteristics:

Table 1: Limitations of the proposed SVC level 3

Maximum macroblocks/second	**Maximum Frame Size in MBs**	**Maximum Bitrate**	
40500	1620	10 *Mbps*	
Format	**Luma Width**	**Luma Height**	**Frame Rate**
QCIF	176	144	172
QVGA	320	240	135
WQVGA	400	240	108
CIF	352	288	102.3
HVGA	480	320	67.5
nHD	640	360	45
VGA	640	480	33.8
525 SD	720	480	30
625 SD	720	576	25

The proposed solution should be optional for service provider and for UE. Appropriate mechanisms to properly announce and setup the session (either including or excluding enhancement layers) are available or should be extended. If UE supports SVC and it detects that the service also provides SVC enhancement layer(s), then the UE is able to consume the service at an improved quality/resolution.

6.1.2 Solution Integration Approaches

6.1.2.1 Rate Adaptation for PSS using SVC with Priority-Based Transmission Scheduling

This solution integration is related to the use case "Rate adaptation in PSS when entering bad reception conditions" (subclause 5.1.3.1).

In order to overcome outages and phases with reduced bit rate, a priority-based transmission scheduling (PBTS) algorithm is proposed to be used to pre-buffer larger amounts of more important data for longer playouts than data with less importance for the resulting video playout quality. The adaptation of the transmission scheduling and the media rate is only based on buffer status reports from client to PSS server as depicted in Figure 8.

Figure 8: Transmission scheduling and media rate adaptation based on priority based buffer status reports

Typically, the size of a UEs buffer is fixed which is assumed in this scenario. The maximum buffering time is depicted in Figure 8 for a standard buffer with one media quality and a priority based buffer with exemplary two quality levels, either temporal, spatial or quality levels or combination of those.

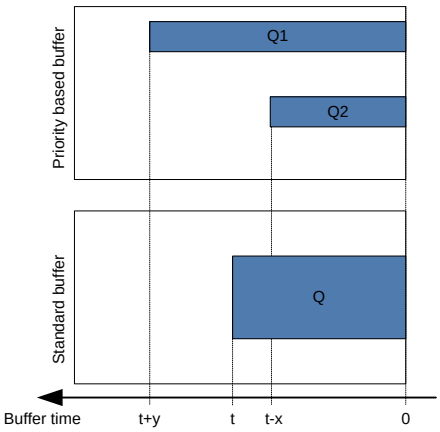

Figure 9: Priority (PBTS) buffer using different qualities (*Q1* and *Q2*) vs. standard buffer with one quality (Q), with *t+y* respectively *t* being the maximum sustainable outage time

In this example, the maximum buffer time for the standard buffer is *t*, which is dependent on the bit rate of the video stream (Q). The priority buffer allows to prebuffer a longer time of the lowest quality level (Q1) *t+y* by reducing the prebuffer time of the higher quality level (Q2) to *t-x*, where *t+y* and *t-x* depend on the bit rate of the quality levels.

To fill up a standard buffer, the PSS server uses a transmission scheduling in decoding order of the video stream. Whereas to fill up a priority based buffer, the PSS server uses a priority based transmission scheduling, where it first fills up the lowest quality level to *t+y* and after that the higher quality layer to *t-x*. After that it switches to the standard transmission scheduling in decoding order.

When the UE enters difficult reception conditions, the available bit rate may no longer be sufficient for the transmission of the highest quality. Having a standard buffer, in such a case users would experience a video outage. In case of having a buffer filled with a priority scheduling algorithm, the high quality data in the buffer runs out earlier than lower qualities. Using SVC, the PSS server would adapt the media stream bit rate to the available service bit rate by dropping quality layers, which still allows to keep the buffer state of the lowest quality level fully filled. Compared to the use of a standard buffer, the highest quality runs out even faster with the priority based approach. Nevertheless, the priority based scheduling allows for keeping the playout alive during longer outages than in the standard case.

Dependent on the buffer reports, the PSS streaming server adapts the media stream bit rate to the quality of the available service bit rate. If the clients' reception condition allows a higher quality, the transmission scheduling is adapted to allow rebuffering of the priority buffer to the maximum quality of the available service bit rate.

Although PBTS can be based on H.264/AVC temporal scalability (AVC-PBTS), SVC has the handy advantage to allow a bit rate reduction using quality or spatial scalability instead of relying on pure temporal scalability as described in [4].

6.1.2.2 Rate Adaptation using SVC MGS Scalability

6.1.2.2.1 Rate Adaptation Approach

The SVC bit rate adaptation approach is based on multiple Operation Points within the same MGS-encoded scalable stream. An Operation Point (OP) is defined as a unique combination of temporal and quality levels, where reasonable combinations of different pairs of Temporal ID and Quality ID are selected in a way so that OPs can be dropped from the bit stream one after the other. An SVC MGS bit-stream consists of different NAL (Network Abstraction Layer) units with different importance for the decoding process. A Quality ID indicating the quality level and a Temporal ID indicating the temporal level are included in the NAL unit SVC header extension of each SVC NAL unit.

Tables 3 and 4 show for example a number of reasonable combinations of different pairs of Temporal ID and Quality ID regarding SVC encoding of a video sequence with 2 MGS quality layers and respectively 4 (GOP 8) and 5 (GOP 16) temporal levels. Operation points are chosen in such a way that at least the base layer stream remains and the frame rate of a stream isn't reduced. If necessary, streams with reduced frame rates can be also built.

From one row to the next row below, exactly one additional pair of {T, Q} is dropped from the SVC bit stream. The bit rate of a video stream is decreased by choosing of the next highest Operation Point (OP) for this stream and dropping all packets which are not needed for decoding the content at this OP. In this way an efficient bit rate adaptation is achieved.

Table 2: Reasonable operation points with SVC MGS scalability (T, Q layer combination) (GOP 8, 2 MGS quality layers)

	T:	0		1		2		3	
OP	Q:	0	1	0	1	0	1	0	1
0		●	●	●	●	●	●	●	●
1		●	●	●	●	●	●	●	x
2		●	●	●	●	●	x	●	x
3		●	●	●	x	●	x	●	x
4		●	x	●	x	●	x	●	x

Table 3: Reasonable operation points with SVC MGS scalability (T, Q layer combination)**(GOP 16, 2 MGS quality layers)**

	T:	0		1		2		3		4	
OP	Q:	0	1	0	1	0	1	0	1	0	1
0		●	●	●	●	●	●	●	●	●	●
1		●	●	●	●	●	●	●	●	●	x
2		●	●	●	●	●	●	●	x	●	x
3		●	●	●	●	●	x	●	x	●	x
4		●	●	●	x	●	x	●	x	●	x
5		●	x	●	x	●	x	●	x	●	x

Figure 10 shows an example for the composition of each frame within a GOP for the OPs highlighted in the Table 2. It becomes apparent that dropping the quality layers Q1 of the temporal levels T2 and T3 reduces the video rate from x to y (OP2). Further, dropping the quality layer Q1 of all temporal levels results to the video bit rate of *z* (OP4) where $z < y < x$. In general, several OPs can be selected for bandwidth optimization. It should be noted that the base layer should be selected at an acceptable quality for the viewer, in order to limit the quality degradation of the video.

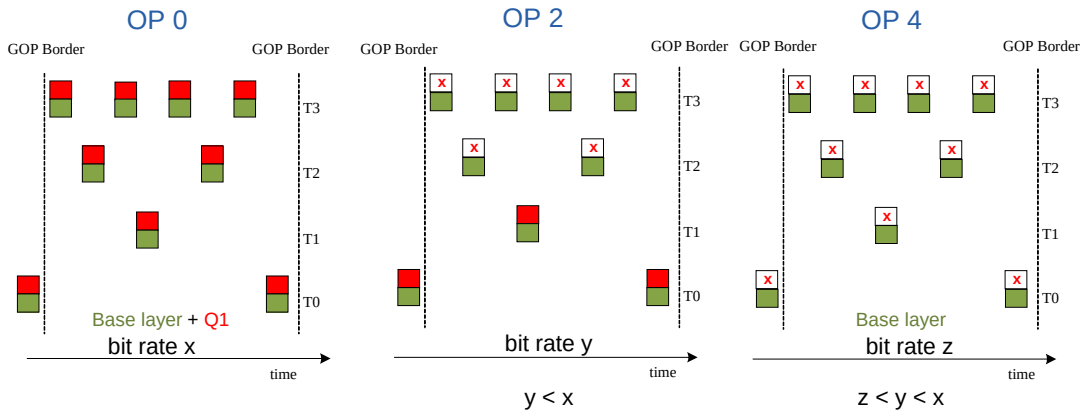

Figure 10: Temporal and Quality layer combination for OP 0, 2 and 4

6.1.2.2.2 Example

Table 4 illustrates an example of the rate adaptation approach using the SVC MGS scalability with three quality layers and multiple operation points for the given video test sequence *IceDance* [19] at 720p resolution and 50 fps (GOP 8).

For each combination of remaining NAL units, corresponding to a certain temporal and quality layer, the average bit-rate and the corresponding average PSNR value are given in the rightmost columns. The rows are sorted according to the resulting PSNR value in descendent order. Note that the frame rate is kept constant until OP8. As soon as only the base layer bit-stream remains (OP 8, shaded in green), dropping of the temporal layers starts, and hence streams with reduced frame rates can be built.

Table 4: Operation points with SVC MGS scalability (T, Q layer combination)(IceDance 720p@50fps [3], GOP 8)

OP	T:	0			1			2			3			Rate [*kbps*]	Average PSNR [*dB*]
	Q:	0	1	2	0	1	2	0	1	2	0	1	2		
0		●	●	●	●	●	●	●	●	●	●	●	●	6977,50	39,87
1		●	●	●	●	●	●	●	●	●	●	●	x	6335,76	39,11
2		●	●	●	●	●	●	●	●	x	●	●	x	5766,06	38,47
3		●	●	●	●	●	x	●	●	x	●	●	x	5263,88	37,94
4		●	●	●	●	●	x	●	●	x	●	x	x	4976,12	37,26
5		●	●	x	●	●	x	●	●	x	●	x	x	2820,75	36,08
6		●	●	x	●	●	x	●	x	x	●	x	x	2586,06	35,52
7		●	●	x	●	x	x	●	x	x	●	x	x	2353,05	34,94
8		●	x	x	●	x	x	●	x	x	●	x	x	1305,15	33,29
9		●	x	x	●	x	x	●	x	x	x	x	x	1104,53	28,43
10		●	x	x	●	x	x	x	x	x	x	x	x	925,40	24,28
11		●	x	x	x	x	x	x	x	x	x	x	x	761,55	21,16

Figures 11~14 represent a distribution of PSNR values of each frame of the *IceDance* sequence for several operation points. Table 5 illustrates results for Operation Points 0, 2, 6, and 8 of the *IceDance* video sequence regarding the PSNR of Average Normalized Square Difference (PANSD). Further information on PANSD can be found in [20].

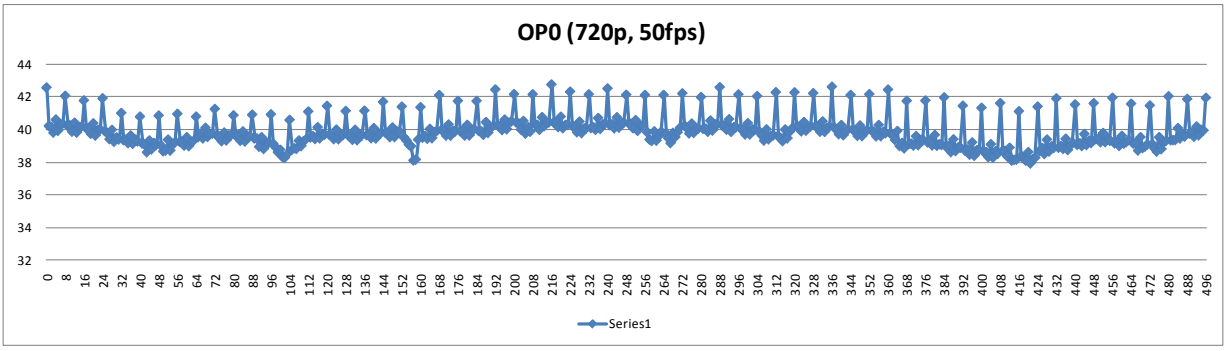

Figure 11: PSNR of each picture for OP0 (IceDance 720p@50fps, GOP 8)

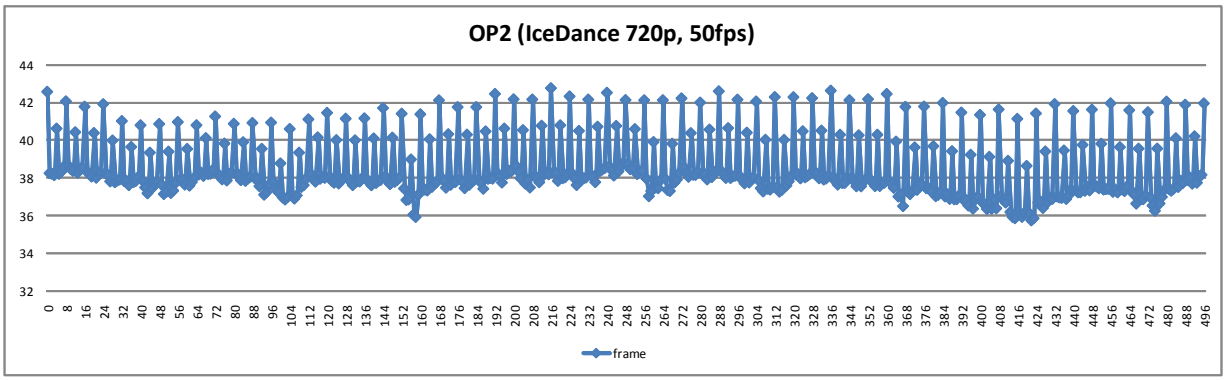

Figure 12: PSNR of each picture for OP2 (IceDance 720p@50fps, GOP 8)

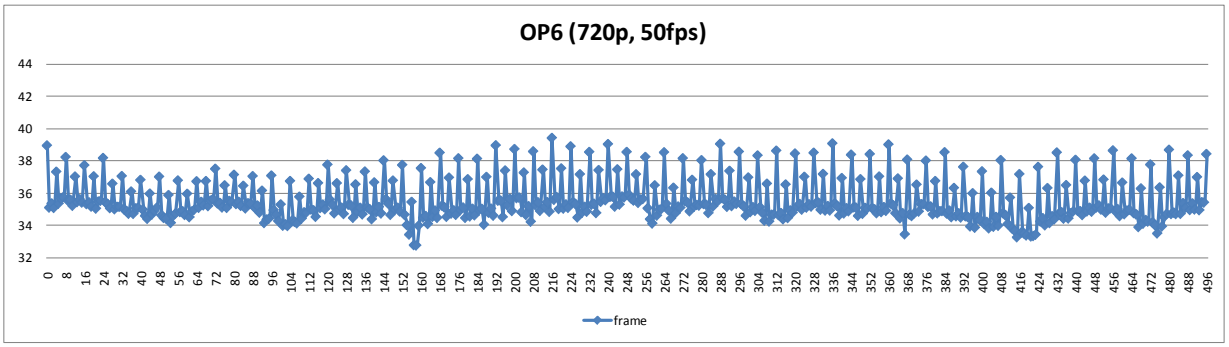

Figure 13: PSNR of each picture for OP6 (IceDance 720p@50fps, GOP 8)

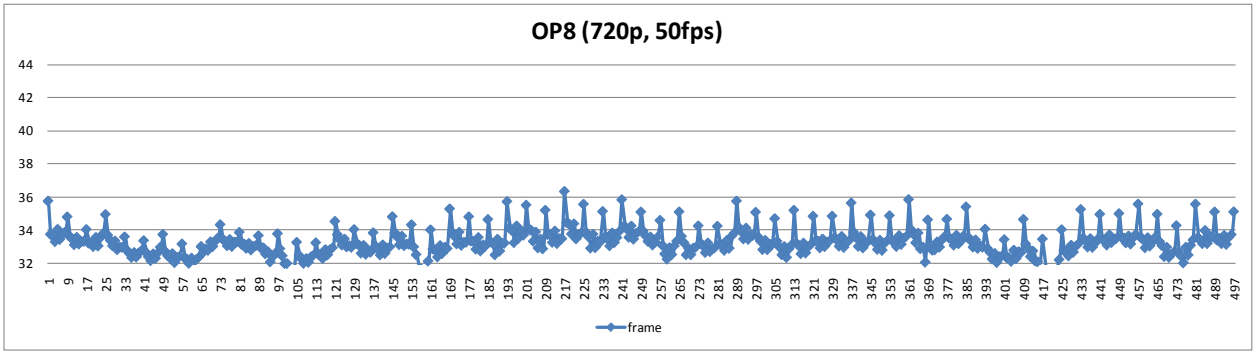

Figure 14: PSNR of each picture for OP8 (IceDance 720p@50fps, GOP 8)

Table 5: OP 0, 2, 6, 8 with SVC MGS scalability (Average MSE and PANSD) (IceDance 720p@50fps, GOP 8)

	T:	0			1			2			3			Rate	Average	PANSD
OP	Q:	0	1	2	0	1	2	0	1	2	0	1	2	[kbps]	MSE	[dB]
0		●	●	●	●	●	●	●	●	●	●	●	●	6977,5	6,833	39,784
2		●	●	●	●	●	●	●	●	x	●	●	x	5766,1	9,781	38,227
6		●	●	x	●	●	x	●	x	x	●	x	x	2586,1	18,925	35,360
8		●	x	x	●	x	x	●	x	x	●	x	x	1305,2	30,979	33,220

6.1.2.3 Unequal Error Protection with SVC in eMBMS

The presented solution is related to the use cases "Graceful Degradation in MBMS services when entering bad reception conditions" (subclause 5.1.3.2) and "Combined support of heterogeneous devices and Graceful Degradation" (subclause 5.1.3.4).

The layered structure of SVC allows for transmission of the video in separate network streams. Thereby, SVC allows services providing different quality steps either by temporal, spatial, quality scalability or combination of those. Using unequal error protection (UEP), such a service can provide different quality levels of different robustness, which allows for Graceful Degradation behaviour in MBMS scenarios. An exemplary UEP scheme is depicted in Figure 9, where the more important layer (Base) has a higher protection than the enhancement layers.

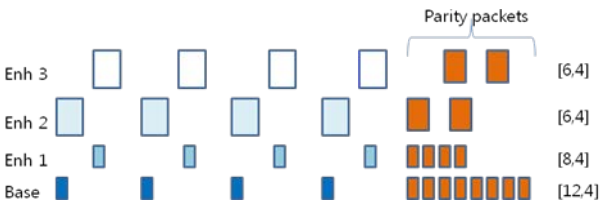

Figure 15: UEP (Unequal Error Protection): Important packets are protected with higher code rate

In the exemplary scenario in Figure 16, there are two layers, using quality, spatial or temporal scalability or combinations of those, with different robustness. UEs in good reception conditions will receive the highest quality and UEs entering worse reception conditions can still receive the base layer, which results in a drop in quality when entering bad reception conditions.

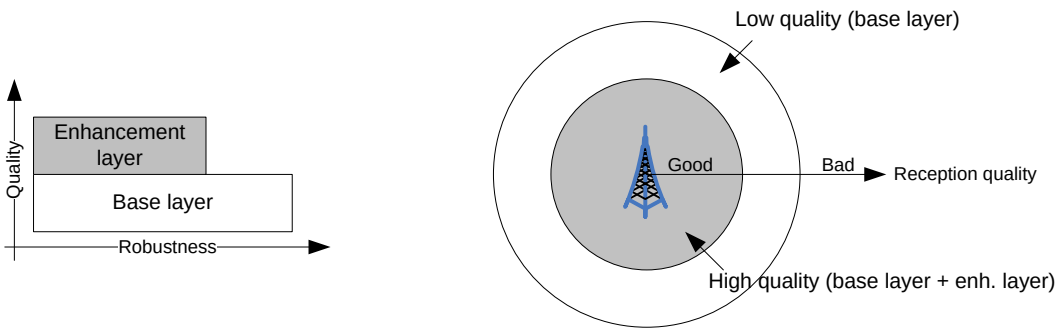

Figure 16: MBMS service with graceful degradation behaviour using unequal error protection with SVC either with temporal, spatial or fidelity scalability or combinations of those

Such a differentiation in robustness of the scalable layers can be applied by a MBMS service at the application layer using different code rates at the application layer forward error correction (AL-FEC).

6.1.2.4 SVC Layer Aware Transmission in eMBMS

In this subclause, we assume that multiple MBMS radio bearers of different MCS (Modulation and Coding Scheme) levels [6] can be allocated to each SVC layer. The high-priority base layer can be transmitted using robust, but low rate MCS channel, while the enhancement layers can be transmitted using high rate MCS channels. The combined effect of

allocating multi-level MCS channels for SVC is that UEs in an area of good signal strength may receive all base and enhancement layers, however the UEs in an area of poor signal strength may only receive base layer data. Compared to the case where uniform MCS level is assigned to MBMS bearers, the multi-level MCS allocation for SVC is adaptive to channel condition and provides graceful quality degradation.

For example, Figure 17 shows typical MBMS bearer allocation, that H.264/AVC [3] single layer stream is allocated to a radio bearer of 16 QAM modulation. Assuming that eNodeB signal power is set to cover 90% of the MBMS service area, UEs may loss data or experience service outage in the rest of 10% area with this MCS allocation.

Figure 18 shows the case of SVC channel allocation where the radio resource is divided to carry SVC layers in different MCS channels. The base layer is transmitted using robust QPSK modulation, hence the signal can reach almost entire area of MBMS cells. The remaining radio resource is given to enhancement layers, therefore the enhancement layer needs much higher rate channel. In this example, 64 QAM modulation channel is allocated for transmitting enhancement data. Since the coverage of 64QAM signal is smaller (e.g. less than 80%) than QPSK, only the UEs in 80% area may receive high quality video. The quality may degrade in the rest of 20% area, however it will be no worse than the minimum level (i.e. base quality).

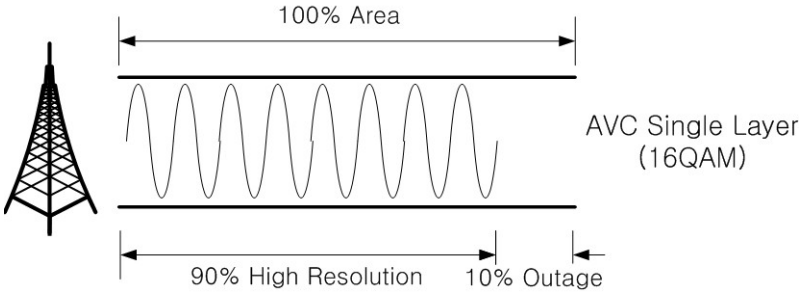

Figure 17: Single level MCS allocation for H.264/AVC

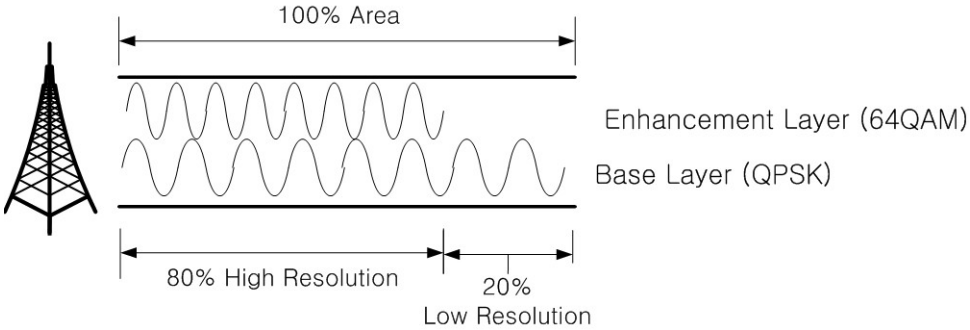

Figure 18: Multi-level MCS allocation for SVC layers

6.1.2.5 Fast Stream Switching in eMBMS

In this subclause, a solution integration approach for fast MBMS stream switching is presented.

In TS 22.246 subclause 5.1, it is stated that "*The MBMS service shall add no more than 1 second when switching between different TV streams to any delay introduced with regards to the coding of the TV stream. It shall be possible for an operator to configure the MBMS Television service so that the typical switching time, from the end user's perspective, does not exceed 2 seconds.*"

In order to comply with the stringent latency requirement, a solution, as depicted in Figure 19, is proposed that a bundle of base streams is used for instant decoding of low resolution video while high resolution video is being switched.

In Figure 19, it is assumed the content providers of MBMS TV service support scalable video. The MBMS server (e.g. BMSC) separates base layer streams from the received video streams and collects them into a bundle of base streams (i.e. preview stream). The preview stream and enhancement streams are transmitted in distinct MBMS bearers. Baseline UEs are able to decode low quality video using the preview stream, while advanced type UEs receive both the preview stream and the enhancement stream, and decode the high quality video as following description.

In the initial service start-up phase, the advanced type UEs receive the preview data and buffer them for sufficient period of time that can exceed the latency for performing stream switching. This stream switching latency usually includes the period for buffering, de-interleaving, FEC recovery and decoding the high quality video, etc. When the sufficient amount of preview data is stored, the advanced type UE starts to receive enhancement data, and extracts base layer data from the preview buffer, and decodes the scalable video. Old preview data is disposed as new preview data is buffered.

When the user requests MBMS stream switching, the UE retrieves the base data of the requested TV stream from the buffered preview data and decodes low quality video instantly. While the low quality video is being played, the UE performs stream switching to receive the enhancement data stream. High quality video is recovered soon after the enhancement stream switching is completed. As a result, the user does not experience latency for stream switching except the initial decoding delay of low quality video.

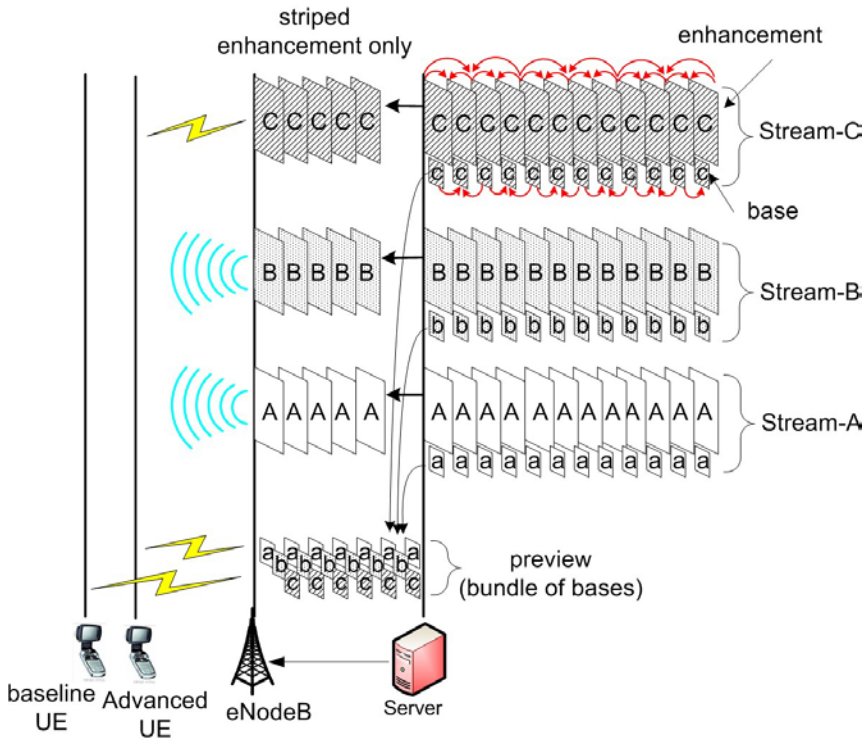

Figure 19: Fast MBMS stream switching using scalable video

It is note that the similar feature of fast stream switching may also be achieved using simulcast of H.264/AVC.

6.1.3 Performance Evaluation

6.1.3.1 Unequal Error Protection in eMBMS

In this subclause, evaluation result of SVC UEP (Unequal Error Protection) method against single layer H.264/AVC is presented. The related use case is presented in subclause 5.1.3, and a solution of UEP is described in subclause 6.1.2.3.

In this evaluation, MBSFN channel of 9*Mbps* throughput in 7 sectors layout is applied commonly to the video streams. Only the ratio of application layer FEC packets is manipulated to test the UEP performance. In this experiment, Raptor code as in TS26.346 Multimedia Broadcast/Multicast Service (MBMS) is used as an FEC method.

In order for fair comparison, the PSNR of H.264/AVC encoded source file and SVC two layer files are produced to be identical (=35.4*dB*). Due to slightly high coding overhead of SVC, the file size of SVC becomes 7% ~ 15% larger than H.264/AVC file. Foreman QCIF and CIF sequences are encoded with the JSVM 13.1. The bitrate of H.264/AVC stream is 398kbps, while those of SVC base layer and enhancement layer are 48*kbps* and 383*kbps*, respectively if PSNR is set to be identical. Bitrate of SVC in total is 431*kbps*, which is 8% more than that of H.264/AVC.

Since equal amount of radio resources should be allocated for transmitting the AVC and SVC streams, the numbers of FEC redundancy packets are adjusted to make the total amount of physical blocks of the two streams identical.

Therefore, FEC stream of 126*kbps* is added to H.264/AVC stream, and FEC streams of 90*kbps* is added to SVC stream, as a result, total bitrate of both codecs become 524 *kbps* (= *video+parity*).

The code rate of H.264/AVC single layer including the FEC overhead is 0.72. The protection period of FEC is 4 GoP length (=2 seconds), of which the size of GoP is 16 in 30*Hz* frame rate.

SVC two layer file is generated in 1:8 ratio of base: enhancement. The code rate of base layer including the FEC overhead is 0.41, and 0.87 in enhancement layer. Hence the base layer protection is enforced while sacrificing the enhancement protection.

Detail of the sample file specification is further described below.

Table 6: Sample files specification used in the evaluation

Codec	AVC	SVC		
		Enhance layer	Base layer	(Altogether)
Sequence	Foreman	Foreman		
Resolution / Frame rate	CIF / 30*Hz*	CIF / 30*Hz*	QCIF / 15*Hz*	
PSNR [*dB*]	35.4	35.4	27.5	35.4
Bit-rate [*kbit/s*]	397.7	383.0	47.9	430.9
File Size (*bits*)	848,112	816,768	101,816	918,584
# of Packets ***(=k)*** (512 byte/packet ± α)	208	200	25	225
Parity packets ***(=n-k)*** (Raptor FEC)	82	29	36	47
Sum of Packets ***(=n)***	290 (=208+82)	229 (=200+29)	61 (=25+36)	290 (=229+61)
FEC Code Rate ***(=k/n)***	0.72	0.87	0.41	***n/a***

% Common Factors:

GOP size: 16

FEC Protection Period: 4 GoP (=2 seconds)

MBSFN Layout: 7 sector layout (ISD=500 m)

Physical Channel: MCS-3, 64 QAM, 1/2 rate (=9 *Mbps* throughput)

The MBMS channel loss model described in Annex A is applied in this experiment. The MBSFN signal transmission area and the service reception area are identical in this layout, therefore the video quality at the border cells of the MBSFN area are also considered in the evaluation. 9 *Mbps* throughput channel (i.e. 64QAM modulation and 1/2 coding rate) is selected to apply the block loss rate equally to the AVC and SVC streams. In consequence, PSNR performance, as described in following equation, is measured at each coverage point.

In the example, the number of H.264/AVC video packets is 208 ***(=k),*** and the number of parity packets for it is 82. Therefore the coding ratio ***(n, k) = (290, 208),*** where ***n*** is total sum of the packets.

According to [10], the failure probability of Raptor is calculated as following Equation (1).

$$P_f(m,k) = \begin{cases} 1 & m < k \\ 0.85 \times 0.567^{m-k} & m \geq k \end{cases} \quad (1)$$

In the Equation (1), ***m*** is the number of packets including video and parity received correctly through the radio channel, and ***k*** is the number of original video packets before transmission. Note that ***k*** doesn't include the number of parity packets. It is an important characteristic of Raptor code that failure probability is subject to ***m-k*** regardless of ***k***.

Assuming that we are measuring PSNR, GOP by GOP, the PSNR of AVC single layer (i.e. $PSNR_{single}$(dB)) is calculated as following Equation (2).

$$PSNR_{single}(\mathrm{dB}) = \sum_{m=0}^{n} P_{sm} \cdot \left(PSNR_0 \cdot P_f(m,k) + PSNR_s \cdot \left(1 - P_f(m,k)\right)\right) \quad (2)$$

In the above Equation (2), it is assumed that a damaged GOP is replaced by the last decoded frame of previous GOP, hence the $PSNR_o$ denotes the PSNR of the freezed GOP. P_{sm} is the probability that m packets are received successfully among n transmitted packets. This probability is typically calculated using Poison function. $PSNR_s$ is the original undamaged PSNR of the GOP.

The PSNR of SVC (i.e. $PSNR_{scalable}$(dB)) is calculated as Equation (3) when it consists of only 2 layers.

$$\begin{aligned} PSNR_{scalable}(dB) \\ &= \sum_{i=0}^{n_b}\sum_{j=0}^{n_e} P_{bi}\cdot P_{ej}\cdot (PSNR_0\cdot P_f(i,k_b) + PSNR_b\cdot\left(1-P_f(i,k_b)\right)\cdot P_f(j,k_e) \\ &+ PSNR_e\cdot(1-P_f(i,k_b))(1-P(j,k_e))) \end{aligned} \quad (3)$$

where, n_b: number of original base layer packets + parity packets for base layer:

k_b: number of base layer packets

n_e: number of original enhancement layer packets + parity packets for enhancement layer

k_e: number of enhancement layer packets

P_{bi}: probability that i packets are received successfully among n_b transmitted packets

P_{ej}: probability that j packets are received successfully among n_e transmitted packets

$PSNR_0$: PSNR of freezed GOP, when the whole GOP is damaged

$PSNR_b$: PSNR of original undamaged base layer GOP

$PSNR_e$: PSNR of original undamaged enhancement layer GOP

In Equation (3), note that $P_f(i, k_b)$ is the failure probabilities of Raptor decoding given that i packets are received out of n_b transmitted packets. Similarly, $P_f(j, k_e)$ is the failure probabilities of Raptor decoding given that j packets are received out of n_e transmitted packets.

Figure 20 shows the evaluation result of PSNR performance at each coverage point. Note that the coverage in this context is the ratio of area that can guarantee the level of PSNR in the 7 sector MBSFN area.

In the Figure 20, it is observed that the source file PSNR (=35.4*dB*) of both the H.264/AVC stream and the SVC stream are maintained up to 45% coverage. The PSNR of SVC (solid red curve) degrades to 27.5 *dB* which is the PSNR of original base layer.

H.264/AVC (dotted line) results in the same PSNR until 55% coverage. PSNR degrades thereafter. PSNR for SVC shows three PSNR levels, one below 45% of 35.4*dB*, one between 45% and 60% coverage of 27.5*dB* and one for greater 60% coverage.

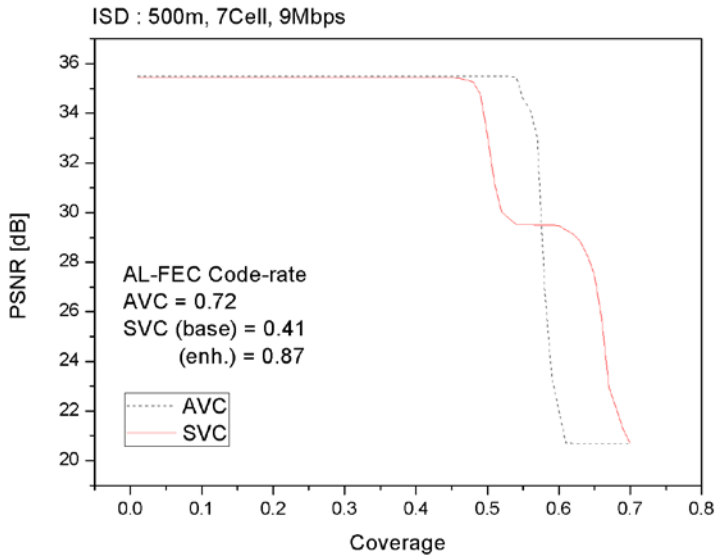

Figure 20: Comparison of PSNR curves of H.264/AVC and SVC

Figure 20 ~ Figure 22 shows coverage vs. PSNR curves in 19 sector layout and 37 sector layout. Although the range of performance variation may slightly be reduced, the effect of graceful quality degradation of SVC is observed identical in different sector layouts.

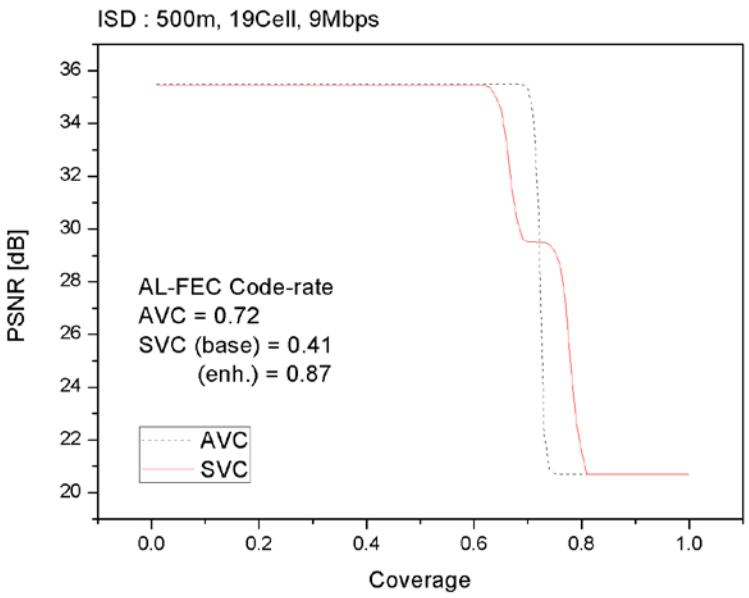

Figure 21: PSNR curves in 19 MBSFN sector layout

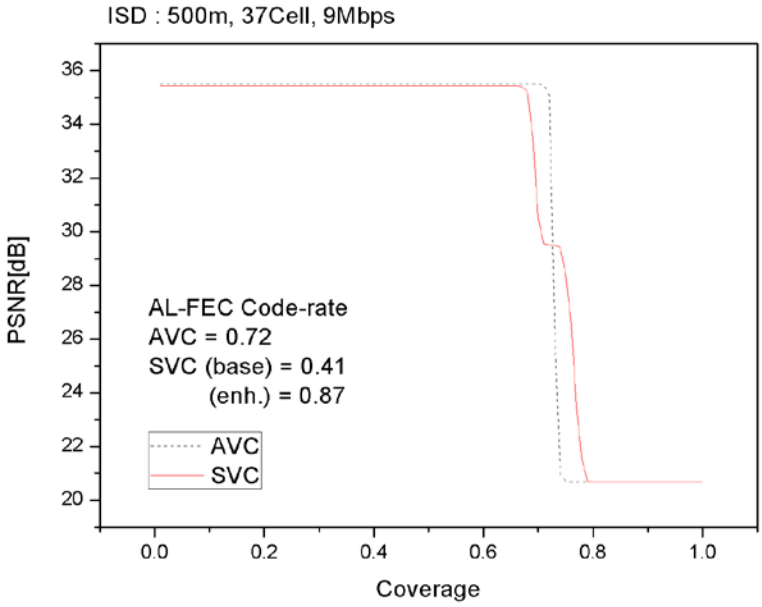

Figure 22: PSNR curves in 37 MBSFN sector layout

6.1.3.2 SVC Layer Aware Transmission for Coverage Improvement in eMBMS

In subclause 6.1.2.4, a solution for SVC layer aware bearer allocation is described. In this subclause, the effect of Differentiated Modulation (i.e. layer aware bearer allocation) when combined with application layer UEP (Unequal Error Protection) method is presented. Similar to the case of UEP, the PSNR performance of SVC, as described in equations (2) and (3) in subclause 6.1.3.1, is evaluated to the case of single layer H.264/AVC when the MBSFN channel loss model is applied.

In this evaluation, two MBSFN channels of 9*Mbps* throughput and 3*Mbps* throughput are used for carrying SVC enhancement layer stream and base layer stream respectively. The MBSFN signal transmission area and the service reception area are identical in this layout, therefore the video quality at the border cells of the MBSFN area are also considered in the evaluation. In addition, FECs using Raptor code is applied to the two streams in several different ratios to test the combined effect of UEP. H.264/AVC stream is transmitted using 6*Mbps* throughput channel. Due to different channel efficiency, the number of physical blocks used for carrying the streams may be different. In order for fair comparison, the same number of physical blocks are used for transmitting SVC streams and AVC stream. Table 3 describes the MCS levels used in the three physical channels and physical block size.

Table 7: MCS levels, data rates and physical block size

MCS	Modulation	Code Rate	Data rate (*Mbps* in 5 MHz)	Block Size (Bytes/BLK)
1	QPSK	1/2	3.0	375
2	16QAM	1/2	6.0	750
3	64QAM	1/2	9.0	1125

Soccer CIF and 4CIF sequences are encoded with the JSVM 13.1. The size of H.264/AVC encoded file is 4,845,608 bits (PSNR = 35.2*dB*), and the size of SVC encoded file is 5,082,762 bits (enhancement=4,565,728, base=517,064). The number of FEC packets added to AVC file is 119, hence the code rate of H.264/AVC single layer is 0.91. The protection period of FEC is 4 GoP length (=2 seconds), of which the size of GoP is 16 in 30 *Hz* frame rate. Since 6*Mbps* throughput channel is used for H.264/AVC, total 889 physical blocks are transmitted through the radio channel. FEC redundancy packets for SVC layers are produced to match the physical block usage of 889.

Three different FEC ratios are tested to evaluate the effect of UEP. In the Case-1 test, 113 FEC packets are given to enhancement layer (code rate = 0.91) and 115 FEC packets are assigned for base layer (code rate=0.52). As a result, base layer pretention is enhanced while sacrificing enhancement layer protection. In the Case-2 test, enhancement layer

data transmitted via 9*Mbps* channel is protected more with FEC packets, and in the Case-3 test, the FEC coding ratio of the enhancement layer and base layer are relatively even.

Detail of the sample file generation and FEC rates are described in Table 8.

Table 8: Sample files specification used in the evaluation

Codec	AVC	SVC			UEP scenario
		Enhance layer	Base layer	(Altogether)	
Sequence	Soccer	Soccer			
Resolution / Frame rate	4CIF / 30 *Hz*	4CIF / 30 *Hz*	CIF / 30 *Hz*		
PSNR [*dB*]	35.2 *dB*	35.2 *dB*	30.2 *dB*	35.2 *dB*	
Bit-rate [*Mbps*]	2.27*Mbps*	2.14*Mbps*	0.24 *Mbps*	2.38 *Mbps*	
Physical channels (Throughput)	MCS-2 16QAM (6*Mbps*)	MCS-3 64QAM (9*Mbps*)	MCS-1 QPSK (3*Mbps*)		
physical blocks/sec (without parity)	404	254	86	340	
physical blocks/sec (with parity)	443	279	164	443	Case 1
		349	94		Case 2
		314	129		Case 3
FEC Code Rate ***(=k/n)***	0.91	0.92	0.52		Case 1
		0.73	0.91		Case 2
		0.81	0.66		Case 3

% Common Factors

GOP size: 16

FEC Protection Period: 4 GoP (=2 seconds)

MBSFN Layout: 7 sector layout (ISD=500 m)

Figure 23 shows the evaluation result of PSNR performance at each coverage point in 7 MBSFN sector layout.

In the Figure 23, it is observed that the source file PSNR (=35.2*dB*) of both the H.264/AVC stream and the SVC stream are maintained up to 45% area. The PSNR curves of the three UEP cases degrade in different pattern respectively to the coding ratios of enhancement layers. The Case-1 curve falls first because enhancement layer protection is weaker than base layer protection. The Case-2 curve falls next and followed by Case-2 curve in the order of FEC coding ratio of enhancement layer.

The video quality of H.264/AVC (dotted line) drops quickly to the minimum level after the 67% coverage area, however the PSNR of SVC streams maintain 30.2 *dB* up to 95% ~ 98% coverage. It is also observed that the effect of base layer protection by FEC is relatively minimal in the three cases, although the Case-2 curve drops slightly earlier than others.

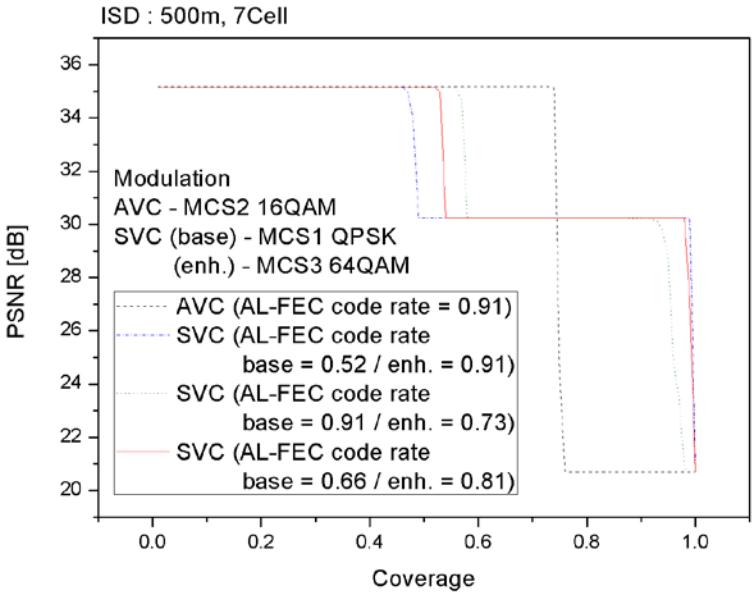

Figure 23: Comparison of PSNR curves of H.264/AVC and SVC

Figure 24 ~ Figure 25 show coverage vs. PSNR curves in 19 sector layout and 37 sector layout. In the following figures, only the Case-3 coding ratio of even distribution of FEC packets is tested. Although the performance disparity between the AVC and SVC is slightly reduced, the effect of graceful quality degradation of SVC is identified and the PSNR is higher than AVC in 75% ~ 90% area.

Table 9: MCS level, coding rates and number of parity packets

Codec	AVC	SVC	
		Enhancement layer	Base layer
MCS-level (channel throughput)	MCS-2 16QAM (*6Mbps*)	MCS-3 64QAM (*9Mbps*)	MCS-1 QPSK (*3Mbps*)
Case-3 Code rate (Parity packets)	Code rate = 0.91 (119)	Code rate = 0.81 (266)	Code rate = 0.66 (64)

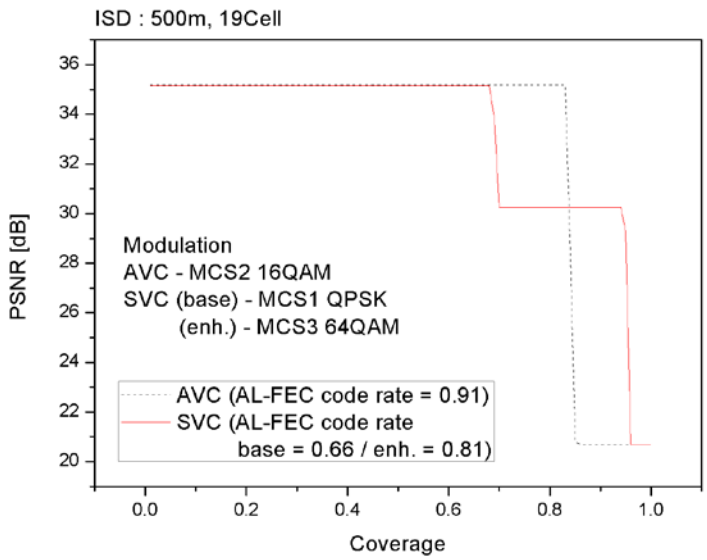

Figure 24: PSNR curves in 19 MBSFN sector layout

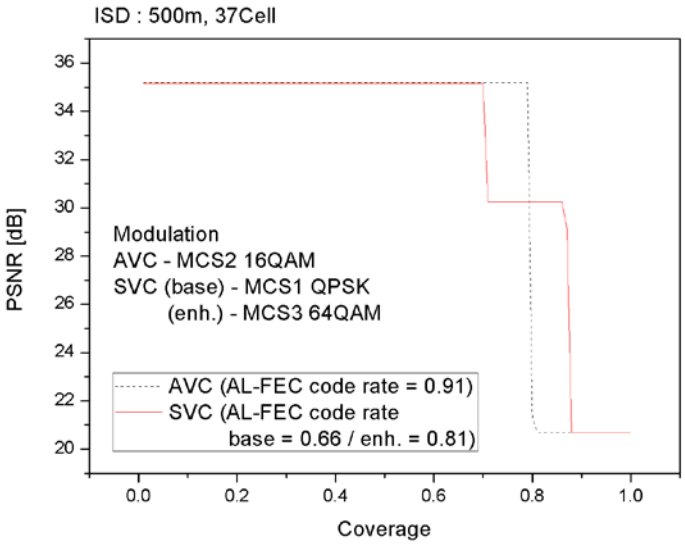

Figure 25: PSNR curves in 37 MBSFN sector layout

6.1.3.3 SVC Layer Aware Transmission for Capacity Improvement in eMBMS

6.1.3.3.1 Introduction

MBMS service delivery over MBSFN cannot adapt to the reception of individual receivers. Delivery of Scalable Video Coding (SVC) coded layered video data with different modulation and coding schemes (MCS) for the individual layers could be used to cope with varying reception conditions by providing physical layer unequal error protection (UEP).

Hierarchically layered video, such as SVC coded video, allows separate transmission of video layers that can be decoded with graceful degradation on the UE. Multi-level MCS allocation schemes can be used to realise physical layer unequal error protection (PL-UEP) for the individual SVC layers.

Based on the MCS schemes and BLER performance data in Figures A.2 to A.6 in Annex A, the presented results analyze the theoretical gain in terms of additional services or capacity when using multi-level MCS SVC transmission with physical layer UEP compared to single MCS AVC transmission. It furthermore compares the theoretical results using SVC with a similar setup using simulcast.

6.1.3.3.2 Evaluation Setup

The presented setup targets a reduction of the overall required transmission cost of a service by the use of SVC in combination with PL-UEP. The general idea is to provide a basic quality using more robust MCS and the quality enhancement layer using less robust MCS. In comparison with a single layer service in the more robust MCS, such a service could give the same robustness in terms of continuous playout while allocating less resources in the more robust and more expensive channel. The cost reduction is gained by providing lower quality video to the users within bad reception conditions. UE with good reception receive the SVC base and enhancement layer stream with highest quality while UE with bad reception may only receive the lower quality SVC base layer. The percentage of users with bad reception depends on the difference in coverage of the chosen MCS schemes for base and enhancement layer. Figure 26 shows an exemplary setup, where the AVC single layer and the SVC base layer is allocated to MCS 1 and the enhancement layer in MCS 2. The figure shows the resulting difference in terms of coverage.

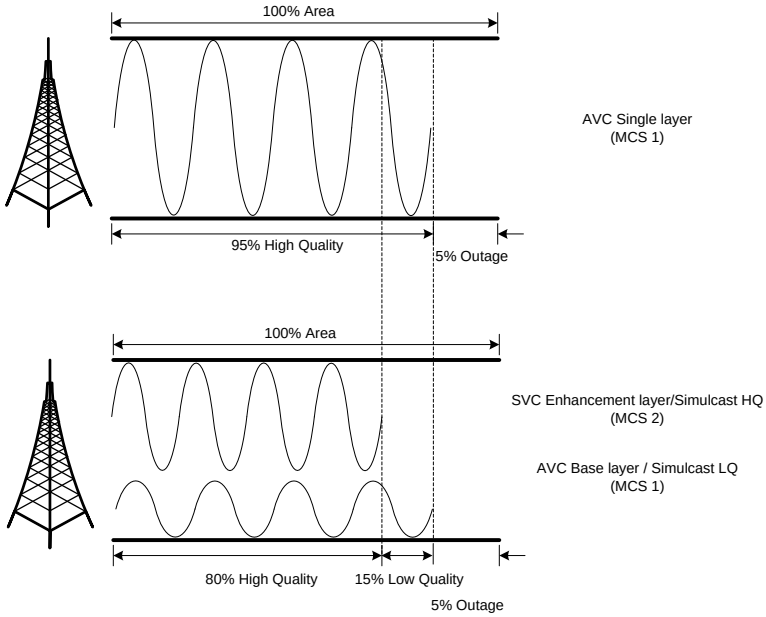

Figure 26: Exemplary setup considered in the evaluation

Three different transmission scenarios are under consideration. For scenario A, AVC transmission with MCS 1 serves as reference for SVC transmission using MCS 1 for the SVC base layer and MCS 2 for the SVC enhancement layer. AVC Simulcast transmission of low quality (LQ) streams with MCS 1 and high quality (HQ) streams with MCS 2 is evaluated accordingly. Scenario B continues in this manner with higher MCS schemes, as can be seen from Table 10.

Table 10: MCS levels for AVC and SVC layers for scenario A, B and C

Scenario	*AVC*	*SVC base layer Simulcast LQ*	*SVC enhancement layer Simulcast HQ*
A	MCS 1	MCS 1	MCS 2
B	MCS 2	MCS 2	MCS 3

6.1.3.3.3 Capacity Improvement

Transmission with multi-level MCS setup directly affects the achievable data rate for a given bandwidth. For SVC and Simulcast transmission, MCS are (time- or frequency-) multiplexed according to the SVC base layer ratio or the ratio of Simulcast LW to Simulcast HQ bitrates, which will be referred to as multiplex rate in the following. Thus, the channel capacity for a given constant bitrate changes according to the multiplex ratio. For instance, with 50% average base layer ratio of all SVC services in scenario A, 50% of MCS 1 data rate for base layer (= 0.5 *Mbps*) plus 50% of MCS 2 data rate for enhancement layer (= 1.5 *Mbps*) are available. AVC Simulcast transmission with a high quality AVC stream of twice the bitrate in the low quality AVC stream behaves accordingly. This leads to 2 *Mbps* total channel capacity for multi-level MCS SVC transmission while single MCS AVC transmission with MCS 1 allows 1 *Mbps* at the same coverage.

A wide range of multiplex ratios has been considered in order to provide multiple operation points with varying quality for the SVC base layer and the Simulcast LQ stream. Note that the selection of optimal operation point is considered to be up to the needs of the service providers. For the selected scenarios, gains in terms of additional channel capacity can be observed for SVC and AVC Simulcast transmission compared to AVC transmission due to the multi-level MCS allocation, as depicted in Figure 27.

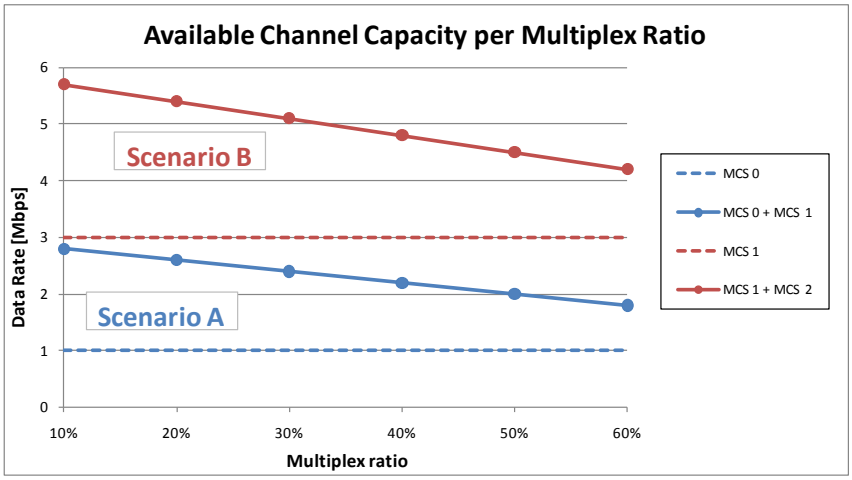

Figure 27: Available data rate per scenario and SVC base layer ratio

The reference AVC stream for all following calculations has a bitrate of 500 *kbps*. In order to provide a satisfying quality of the SVC base layer and of the Simulcast LQ stream, a bitrate range from 100 to 300 *kbps* has been considered, preserving equal quality of both. The SVC enhancement layer and the Simulcast HQ stream are assumed to have the same quality as the AVC reference stream. For overall evaluation, it is further necessary to consider the coding overhead introduced by SVC. Table 11 gives an exemplary calculation of gains in terms of additional services with multi-level MCS SVC transmission compared to the 500 *kbps* reference AVC stream in scenario A. Available channel capacity for AVC is 1 *Mbps* for MCS 1 in the selected scenario and a SVC overhead of 10% is assumed. In Table 11 UCC refers to used channel capacity which is the required channel allocation ratio per service. UCC is calculated by the overall media bitrate divided by the AVC or SVC channel capacity. E.g. for the AVC reference stream with 500 *kbps* and the AVC channel capacity of 1 *Mbps* the AVC UCC is 50%. For SVC case at a multiplex ratio of 18% the SVC UCC is calculated by the overall SVC media bitrate of 550 *kbps* divided by the SVC Channel Capacity of 2636 *kbps* which leads to a SVC UCC of 21%.The number of services per channel can be calculated by 1/UCC.

Table 11: Exemplary calculation of SVC gains for scenario A

AVC bitrate [kbps]	*Multiplex ratio*	*SVC BL bitrate [kbps]*	*SVC EL bitrate [kbps]*	*AVC chan. capacity [kpbs]*	*AVC UCC*	*SVC Chan. Capacity [kbps]*	*SVC UCC*	*AVC services per chan.*	*SVC services per chan.*	*Difference [services]*	*SVC gain [services]*
500	*18%*	*100*	*450*	*1000*	*50%*	*2636*	*21%*	*2*	*4.79*	*2.79*	*139.67%*
500	*27%*	*150*	*400*	*1000*	*50%*	*2455*	*22%*	*2*	*4.46*	*2.46*	*123.14%*
500	*36%*	*200*	*350*	*1000*	*50%*	*2273*	*24%*	*2*	*4.13*	*2.13*	*106.61%*
500	*45%*	*250*	*300*	*1000*	*50%*	*2091*	*26%*	*2*	*3.80*	*1.80*	*90.08%*
500	*55%*	*300*	*250*	*1000*	*50%*	*1909*	*29%*	*2*	*3.47*	*1.47*	*73.55%*

Figures 28 to 31 show the gain of multi-level MCS SVC (solid lines) and AVC Simulcast (dashed lines) transmission compared to single MCS AVC transmission in terms of additional services for all defined scenarios and varying SVC overheads from 0% to 30%. SVC base layer and Simulcast LQ stream bitrate from 100 to 300 kbps have been selected to represent all reasonable operation points.

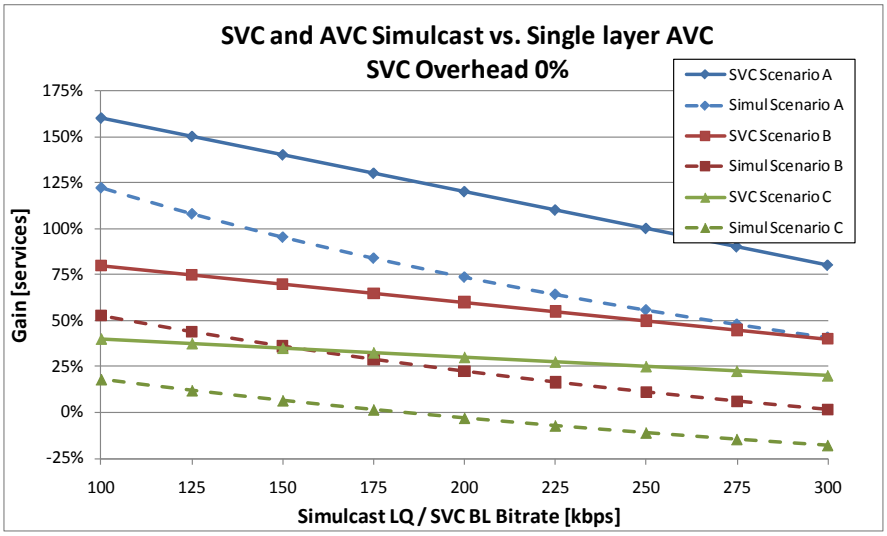

Figure 28: Gain in terms of additional services with SVC overhead of 0%

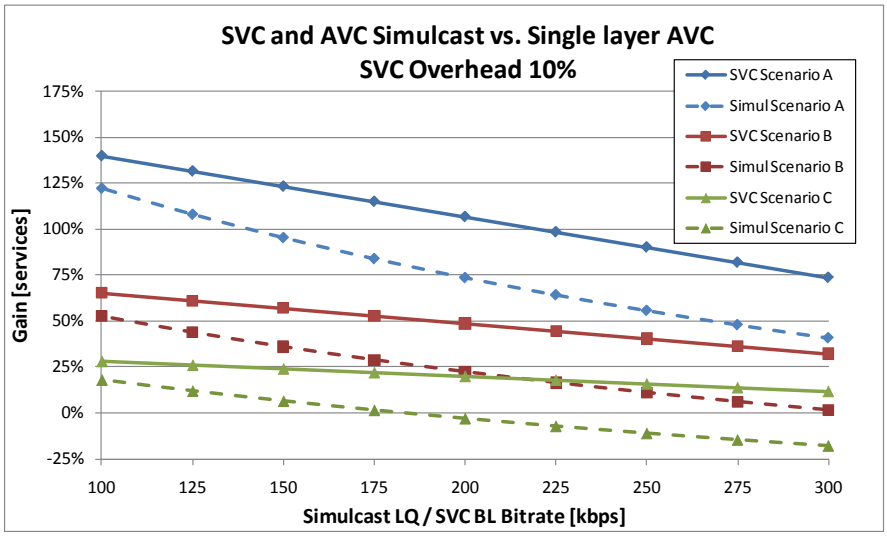

Figure 29: Gain in terms of additional services with SVC overhead of 10%

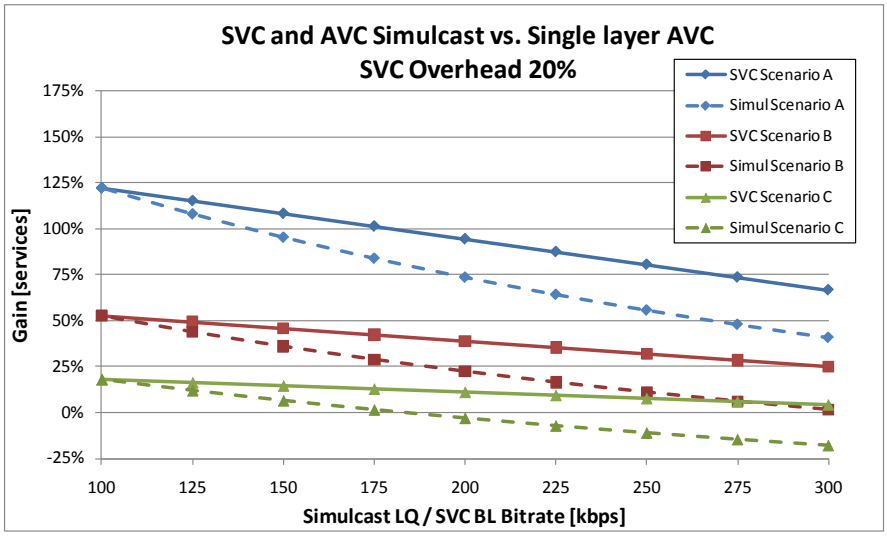

Figure 30: Gain in terms of additional services with SVC overhead of 20%

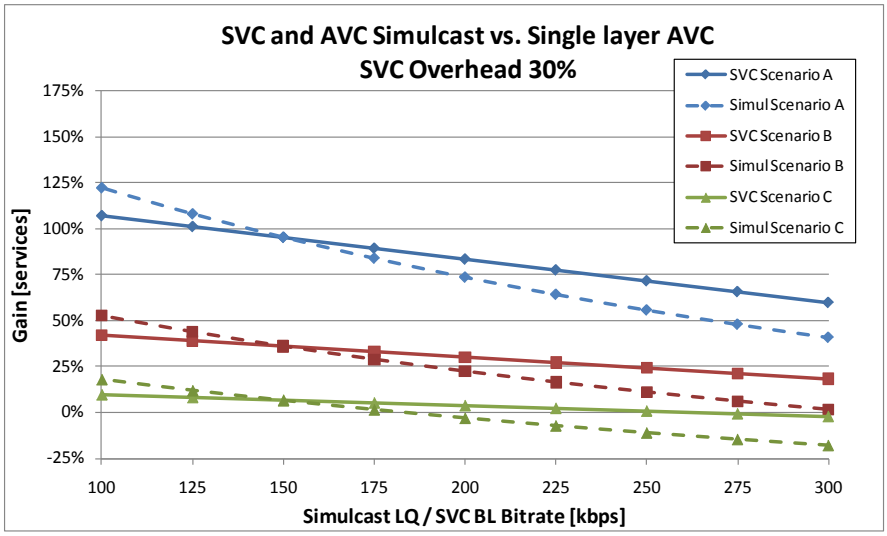

Figure 31: Gain in terms of additional services with SVC overhead of 30%

Figures A.2 to A.5 in Annex A give an estimate on the coverage of a specific MCS setting given a minimum BLER. In order to establish a satisfying quality in terms of image fidelity and continuous playout of video without additional application-layer FEC, it is assumed that a BLER of less than 0.001 has to be maintained. According to these constrains, costs for multi-level MCS allocation with SVC or Simulcast can be estimated for the specified scenarios. The difference in coverage is calculated as the percentage of measured area that is provided only low quality video. Table 8 give the coverage of all MCS schemes at a BLER of 0.001 for different amounts of cooperating MBSFN cells and the cost in terms of coverage.

Table 12: Coverage of MCS schemes and costs of scenarios

Cooperating Cells	*MCS 1 Coverage*	*MCS 2 Coverage*	*MCS 3 Coverage*	*MCS 4 Coverage*
7 Cell	98%	85%	60%	32%
19 Cell	95%	87%	75%	48%
37 Cell	86%	83%	74%	57%

6.1.3.3.4 Performance Evaluation

This evaluation considers the most promising scenarios A and B which are listed again in Table 10. Each scenario is investigated within an MBSFN area that consists of 19 cooperating sectors of an 57 sectors cell layout.

A set of target coverage areas was defined by a circle around the middle sector, thus all user trajectories proceed within this target area. It covers ~96% of the area of cooperating sectors in Scenario A and ~83% in Scenario B, thereby omitting areas with bad reception in the AVC / SVC base layer MCS, which are predominately located nearby the edge of the cooperating sectors in the channel model. This is done to ensure a well performing AVC transmission as reference for evaluation of multi-level MCS SVC transmission.

The error trace files are generated in a simulation environment where 100 UEs randomly traverse the MBSFN area with the given restrictions and cover a distance of 600 m, therefore total traversing distance is 60 km (= 100 UE x 600 m). The simulation settings are described in detail in Annex A. Figure 32 shows the trajectories of all UEs movement in a 19 sector layout (thick blue lines) with scenario A (left) and scenario B (right) within the target coverage area (green circle), where the colours of the trajectory represent the values of BLER measurements. The top plots show the MCS used for AVC reference and SVC base layer, whereas the bottom plots depict BLER values achieved by the MCS used for the SVC Enhancement layer.

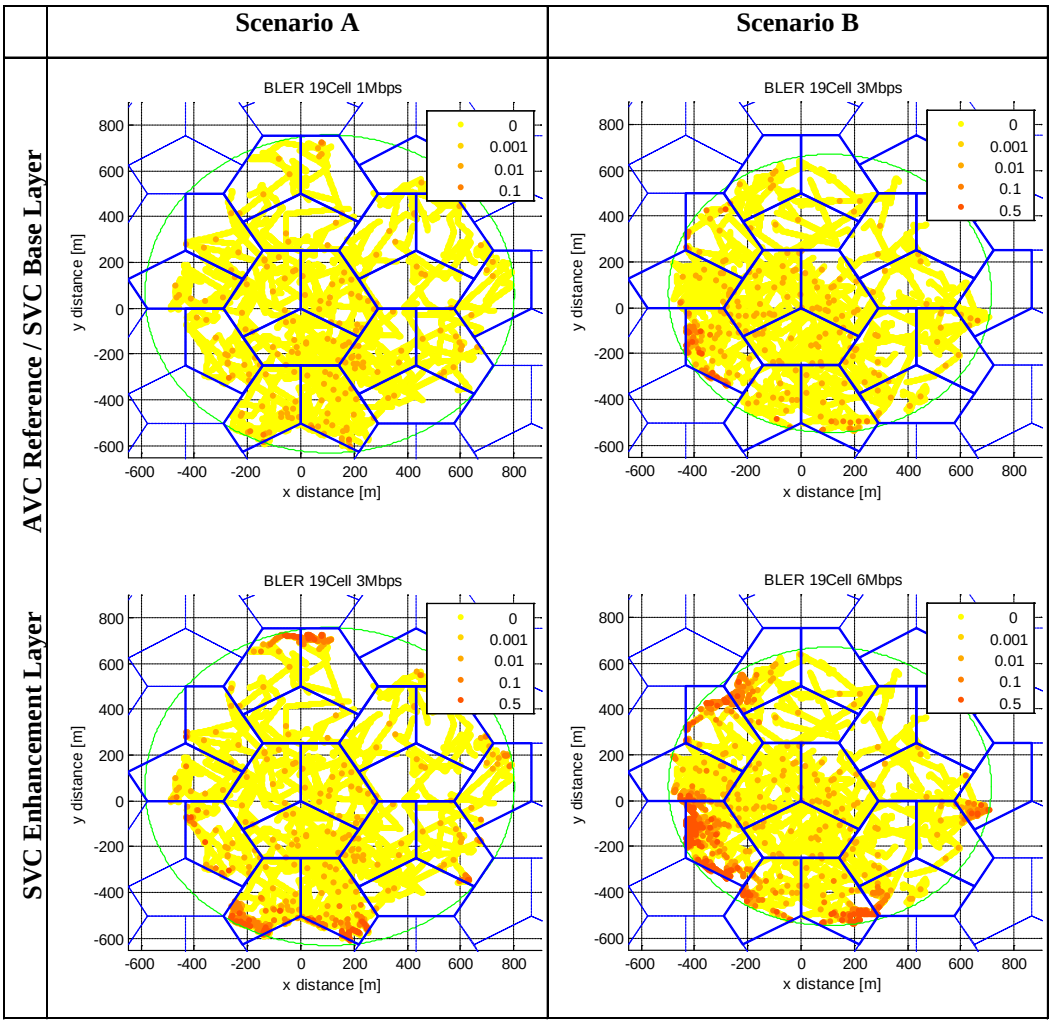

Figure 32: Trajectories of 100 UEs in 19 sector cell area with MCS1 (top) and MCS2 (bottom)

The distance between the measurement points is 1m. A pedestrian walking speed is assumed, where the UE is moving at approximately 3.6 km/h (= 1 m / sec). Since the distance between the measurement points is 1m, the period between measurements is 1 sec, and the total measurement time is 16,7 hours (= 100UEs x 600 measurement points x 1 sec / 3600). It is assumed that the BLER of a measurement point generally represents the average BLER experienced by the UE until moving 1 m to next measurement point.

We compare a single layer AVC stream with an SVC stream using CGS with 2 layers. The source video consists of a concatenation of 4 sequences, where each sequences has been encoded with a constant bitrate for AVC and slightly higher bitrate for SVC in order to achieve a total average bitrate of approximately 270kbit for AVC and 300kbit for SVC, leading to a SVC overhead of about 10%. The GOP size is 8 frames at a frame rate of 15fps. The SVC Streams have been encoded with two unique base layer ratios of roughly 30% and 50%. All coding parameters can be found in Table 13.

Table 13: Bitrate and video quality of simulated video sequences for AVC Reference and SVC Streams with 30% and 50% base layer ratio

AVC Reference			SVC Base Layer Ratio ~30%					SVC Base Layer Ratio ~50%				
	AVC		SVC EL		SVC BL			SVC EL		SVC BL		
Seq. Name	Bitrate [kbit]	PSNR [*dB*]	Bitrate [kbit]	PSNR [*dB*]	Bitrate [kbit]	PSNR [*dB*]	BL ratio	Bitrate [kbit]	PSNR [*dB*]	Bitrate [kbit]	PSNR [*dB*]	BL ratio
city	226.94	36.10	235.34	36.43	67.71	29.73	0.29	258.68	36.22	135.35	33.17	0.52
crew	282.82	35.24	316.17	34.94	90.20	30.50	0.29	312.41	34.79	154.75	32.66	0.50
harbour	278.09	30.75	303.30	30.77	83.48	26.32	0.28	308.65	30.58	153.87	28.43	0.50
soccer	278.43	35.77	317.51	35.74	98.22	31.02	0.31	322.05	35.67	160.02	33.06	0.50
average	**266.57**	**34.47**	**293.08**	**34.47**	**84.90**	**29.39**	**0.29**	**300.45**	**34.32**	**151.00**	**31.83**	**0.50**

All encodings provide approximately similar video quality in terms of PSNR. The enhancement layer PSNR of the SVC stream with 30% base layer ratio matches the AVC PSNR precisely, whereas the SVC stream with 50% base layer ratio has 0.15*dB* lower PSNR quality on encoding side. This is due to the limited rate control of the used JSVM reference encoder and has to be considered when interpreting simulation results.

The selected video streams are encapsulated in RTP packets according to their specific RTP payload format and subsequently into IP streams. Real-time transmission of transport-blocks was simulated according to the BLER measurements over time of the recorded UE traces as described in Annex A. This process is repeated 100 times for each UE with a time-seeded random generator in order to obtain statistically relevant results.

Averaged frame-wise peak-SNR (PSNR) of the transmitted, decoded video is used for the evaluation of video quality on the UE. Measurement of playout robustness utilizes Erroneous Seconds Ratio (ESR). ESR is the ratio of video seconds that contain at least an erroneous frame to the length of the video in seconds. We use (1-ESR), where a measurement of 100% corresponds to an error free video and 0% translates into erroneous frames within every second of the video.

Subsequently, quality evaluation of the transmission results was done as described in [16]. This approach features precalculation of a PSNR database based on real decoding with error concealment techniques. Further evaluation of video quality in terms of PSNR is conducted on packet-level, thus omitting redundant video decoding operations. ESR measurements take coding dependencies of frames within the video into account.

Due to the lower robustness for the SVC enhancement layer, users will experience a graceful degradation behaviour when entering areas with bad reception conditions of the enhancement layer. The simulation results allow analyzing the video quality degradation experienced by users for each scenario during traversing the user trajectories with transmission scheduling scenario A and B using AVC, or SVC encodings. Figure 27 shows the cumulative distribution of average user PSNR for both scenarios, whereas Figure 33 shows the cumulative distribution of video quality (PSNR) degradation of SVC compared to AVC. Note that SVC stream with 50% base layer ratio has 0.15*dB* lower PSNR quality on encoding side, which has to be considered when interpreting Figures 27 and 33. Furthermore, Figure 34 shows the cumulative distribution of play out robustness (1 - ESR) for AVC reference and SVC, where both SVC base layer ratios are represented with a single line due to equal performance and for the sake of legibility.

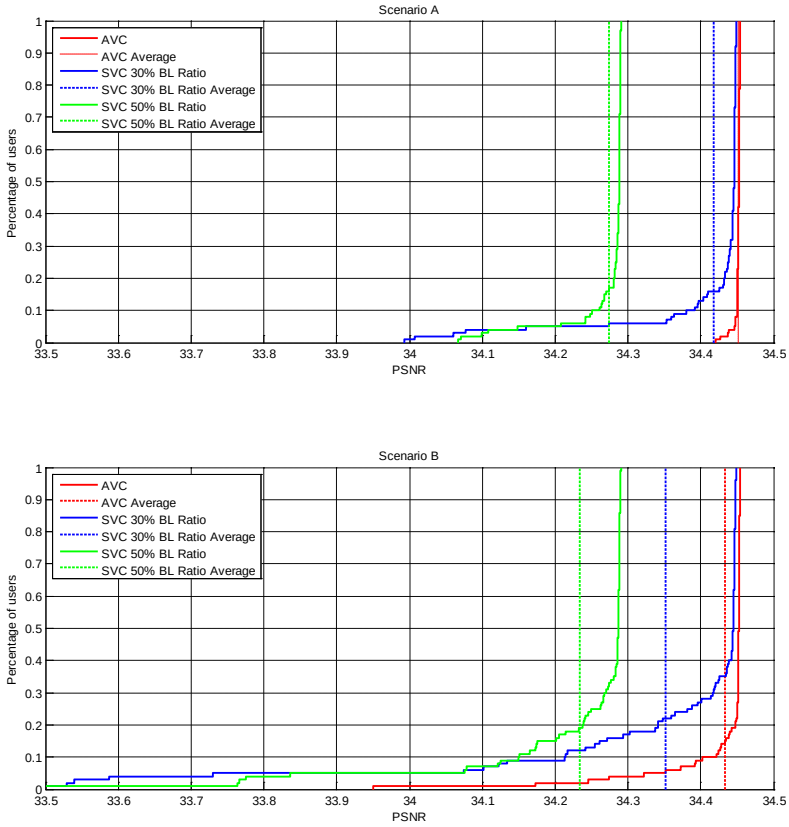

Figure 33: Cumulative distribution of average PSNR using AVC or SVC (30% BL ratio and 50% BL ratio) for transmission scenario A and scenario B within a 19 Cells sector cell area

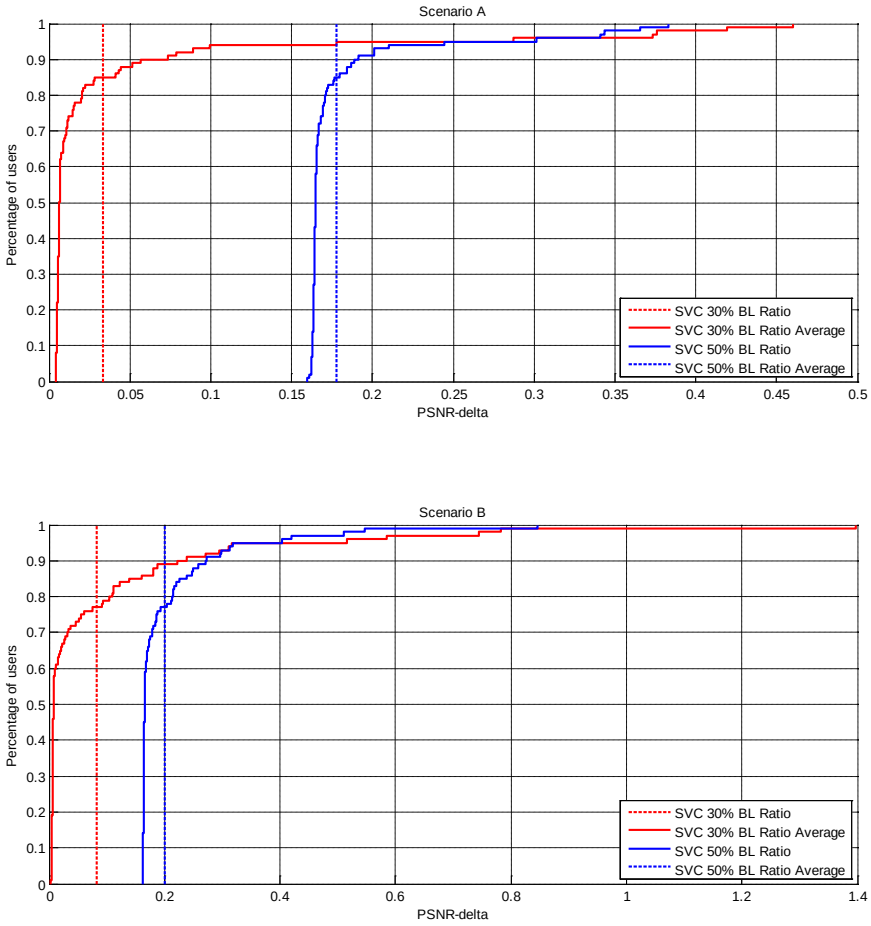

Figure 34: Cumulative distribution of PSNR degradation using AVC or SVC (30% BL ratio and 50% BL ratio) for transmission scenario A or B within a 19 Cells sector cell area

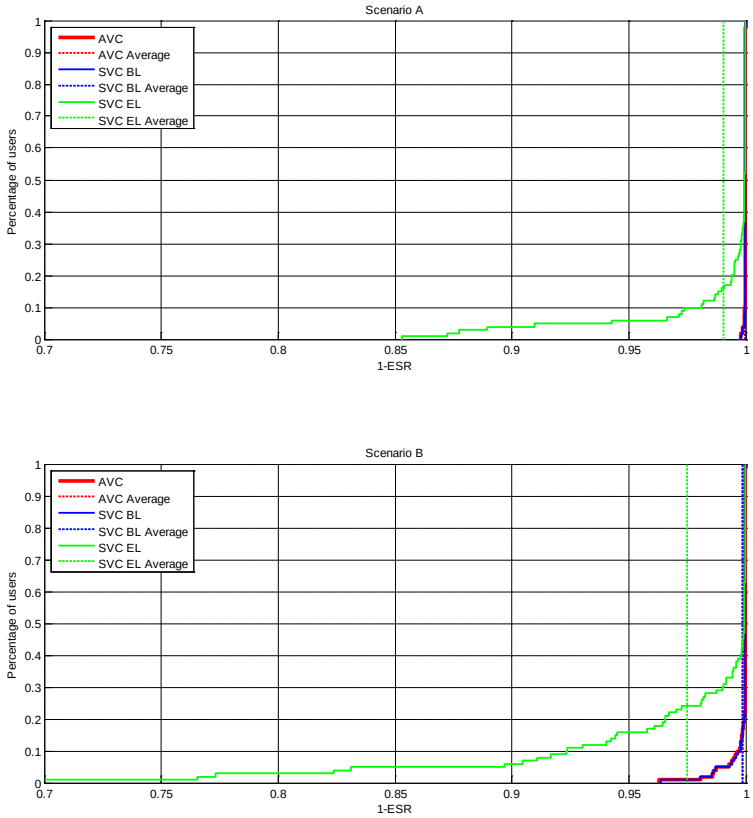

Figure 35: Cumulative distribution of play out robustness (1 - ESR) using AVC or SVC for transmission scenario A or B within a 19 Cells sector cell area

The results in Figure 27 and Figure 33 show that the average quality degradation of both scenarios is below 0.2 dB in terms of PSNR. Less than 5% of users experience a quality degradation of more than 0.3dB PSNR and the maximum PSNR degradation observed in the simulation setup is 0.46 dB for scenario A and 1.4dB for scenario B, where the latter is experienced by an outlying single user with 99% of users experiencing a quality degradation below 0.8 dB PSNR.

The (1-ESR) measurements in Figure 34 show the difference of base layer (SVC ESR_BL) and enhancement layer play out robustness (SVC ESR_EL). It can be seen that the play out robustness of AVC reference is preserved for the SVC base layer when using multi-level MCS SVC transmission.

6.1.3.4 Graceful Degradation for MBMS Using SVC

6.1.3.4.1 Introduction

In this subclause, a test system for graceful degradation for MBMS Rel-6 is presented. Quality metrics for degraded video are introduced. Furthermore, test results for graceful degradation in MBMS are given.

6.1.3.4.2 Test system

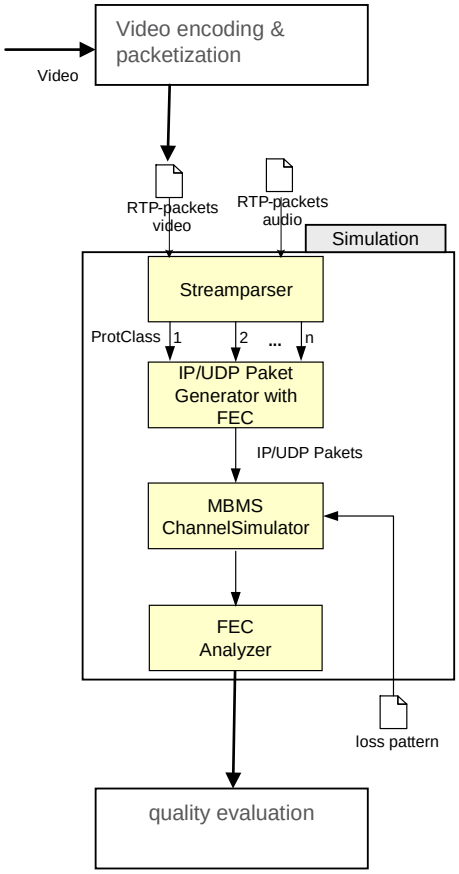

Figure 36: Test system for graceful degradation in MBMS

A MBMS simulation chain has been built (Figure 36), which simulates transmission of video and audio data over multiple streams over one MBMS channel. Each of these streams can be protected with an MBMS standard compliant Raptor FEC by different code rates. The MBMS channel is simulated by using traces of MBMS loss patterns. The loss patterns contain transport block (TB) loss probabilities for different transmission powers and different bearer rates. Figure 36 shows the test system. The Raptor FEC needs a small amount of additional received symbol overhead λ for successful decoding. During the following simulations this overhead is set to $\lambda = 3\ \%$ of the number of source symbols. One FEC code block extends over *2s* considering the resulting bitrate after RTP encapsulation.

The size of a transport block (TB) is fixed to 82 Byte. The number of TBs in each TTI and the length of each TTI depend on the selected bearer rate. Due to common channel coding and interleaving of the TBs in a TTI, in our simulations either all TBs in a TTI are lost or all are not lost.

Table 14 depicts the settings for each bearer rate.

Table 14: MBMS parameters for simulating different bearer rates

Bearer rate	TTI duration	TB Size	TBs/TTI
64 kBit/s	80 ms	82 Byte	8
128 kBit/s	80 ms	82 Byte	16
256 kBit/s	40 ms	82 Byte	16

The MBMS simulation is based on loss patterns measured in a simulated MBMS Rel-6 system for different transmission power and bearer rates. The main radio network simulation assumptions are listed in Table 15. Details of the simulation assumptions can be found in [13]. 500 users are dropped randomly and then traces are recorded for 40s while users are moving. As users move with only 3km/h, users move only a few *10s* of meters, so considering the inter-site distance of 1500m users can be macroscopically regarded as stationary. Mainly the fast fading changes during a trace.

Only the 128 *kBit/s* and 256 *kBit/s* bearers are used for the simulations. For the 128 *kBit/s* case we used loss patterns with transmission power from -13*dB* to -5*dB* (relative to $P_{max}=17.4W$) and for the 256 *kBit/s* we used loss patterns with transmission power from -10*dB* to -2*dB*. Note that a double bearer rate requires approx. 3*dB* higher transmission power to provide similar loss behaviour.

Table 15: Radio network simulation parameters

Property	Value	Remarks
Cell layout	Hexagonal grid, 3-sector	
Intersite distance	1500 m	
Antenna model	Max gain 18 *dBi*, electrical + mechanical tilt: 6 + 2 degrees	Horizontal and vertical patterns
Propagation model	pathloss $L=15.3 + 37.6*\log_{10}(D)$	D in [m] L in [*dB*]
Channel model	Vehicular A, 3 *km/h*	
BS maximum output power, P_{max}	17.4 *W*	non MBMS channels transmit are allocated as much power that the total output power reaches the maximum.
Common Pilot Channel power	10% of P_{max}	
Soft combining	enabled, maximum 3 cells	

Depending on the selected bearer rate the data of the media stream is mapped on the MBMS transport blocks. The losses for each TTI are simulated by comparing a random value with the probability of the utilized loss pattern. If TTI is lost, all TBs of this TTI are lost too.

Figure 37 depicts the mapping of the RTP packets into the MBMS transport blocks. First RTP packets are fragmented to fit the MTU size. The resulting RTP fragment units are packed together with the parity packets into IP packets. Then the IP packets are mapped into the transport stream of the MBMS service.

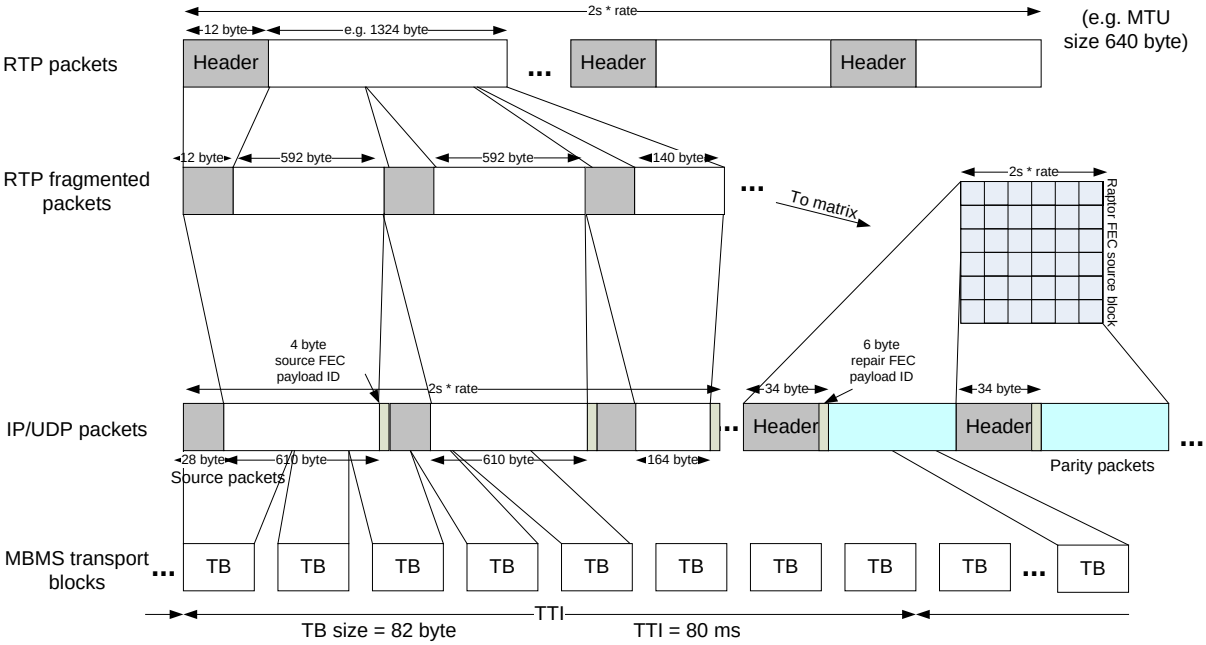

Figure 37: Mapping of RTP packets

6.1.3.4.2 Test sequences

Three different test sequences are considered for simulation where each of them containing an audio track. The associated properties are depicted in Table 16.

Table 16: Properties of simulated video sequences

	Resolution	Frame rate	Number of frames
wineyard	qCIF (176x144)	25 Hz	1617
stronger	qCIF (176x144)	25 Hz	1617
reuter	qCIF (176x144)	12.5 Hz	317

For both H.264/AVC and SVC encoding the JSVM 8.8 software was used in the simulations. We used a hierarchical coding structure with a GOP size of 16 and an I-frame period of around 2 seconds. In the SVC encodings, we used SNR scalability with one MGS enhancement layer. The quantization parameters were selected such that H.264/AVC and SVC bit streams (including both base and enhancement layer) yielded similar PSNR values.

Audio encoding parameters are fixed for all sequences and test runs. The bit rate is set to 32 *kBit/s* and the sample frequency to 48 *kHz*.

Figure 38 and Figure 39 depict the resulting bit rates for H.264/AVC and SVC with one MGS layer encoding.

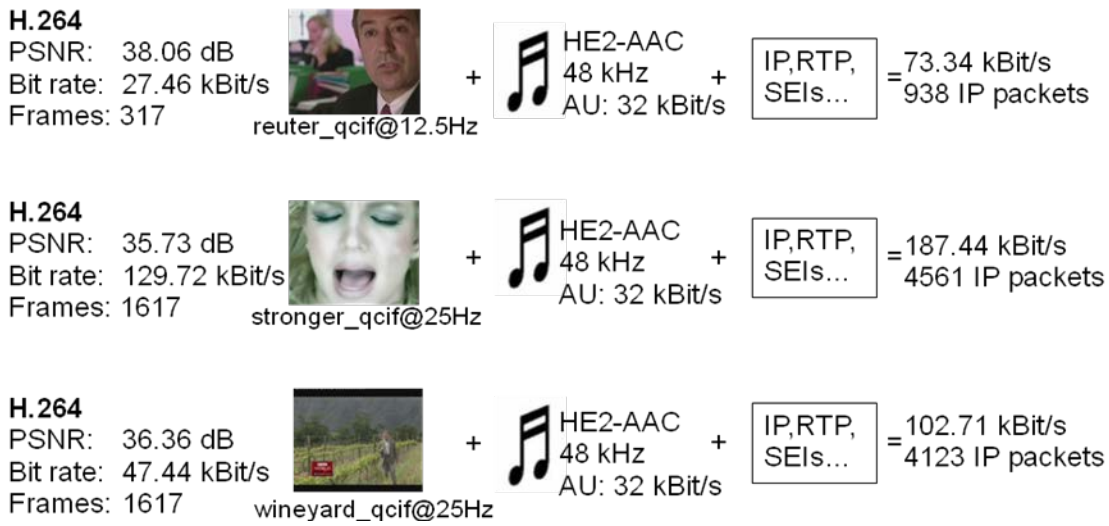

Figure 38: Video sequence parameters for H.264/AVC compliant base layer with GOP 16

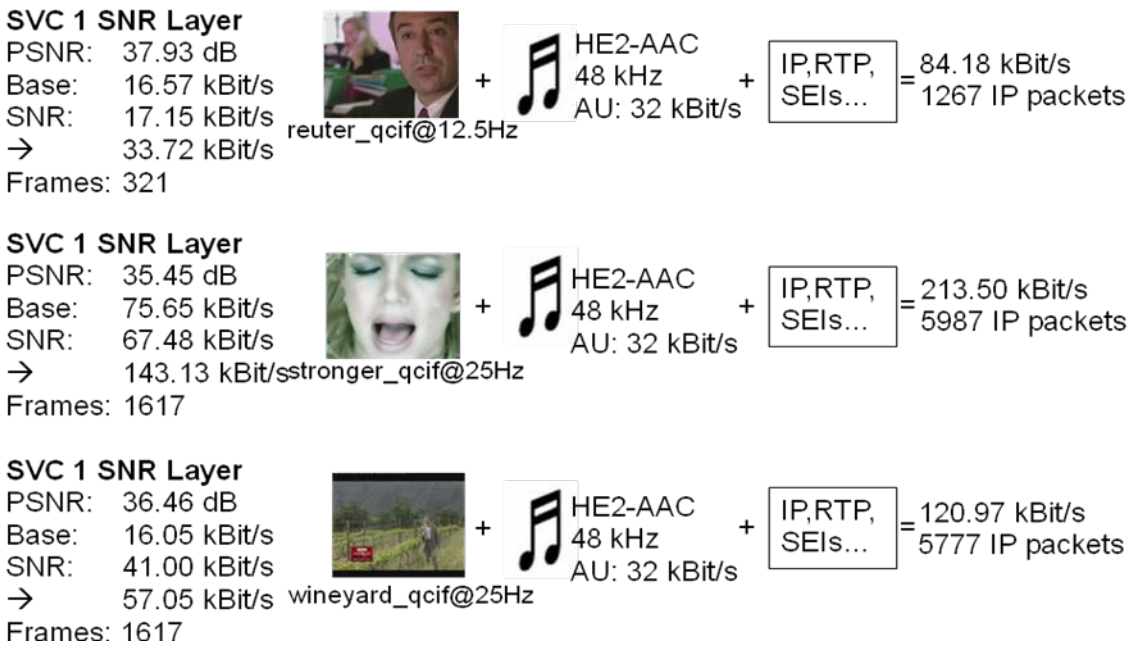

Figure 39: Video sequence parameters for SVC with 1 SNR layer with GOP 16

Figure 40 shows the signalling header overhead of H.264/AVC and SVC. Header compression is not applied.

reuter_qcif@12.5Hz

	Frames	Total bytes	IP packets	Header Overhead
H.264	317	237715 Byte	938	≈ 5.0 %
SVC 1 Enh. Layer	317	272891 Byte	1267	

stronger_qcif@25Hz

	Frames	Total bytes	IP packets	Header Overhead
H.264	1617	1522024	4561	≈ 3.5 %
SVC 1 Enh. Layer	1617	1733653	5987	

wineyard_qcif@25Hz

	Frames	Total bytes	IP packets	Header Overhead
H.264	1617	834098 Byte	4123	≈ 7.0 %
SVC 1 Enh. Layer	1617	982259 Byte	5777	

Figure 40: Signalling header overhead H.264/AVC and SVC

6.1.3.4.4 Transmission Schemes

Six different transmission schemes are simulated. Two of them provide single layer and four multilayer transmissions. The issue is to compare different settings of multi layer transmission providing graceful degradation behaviour by the use of application layer Raptor FEC or power spreading between the different video layers (2–6) with single layer transmission with and without additional application layer FEC (1–2) . The abbreviations for each transmission scheme later used in the results subclause are:

1) Single layer transmission
("**SingleLayer**")

2) Single layer transmission with additional raptor FEC
("**SingleLayerFEC**")

3) Layered transmission with unequal error protection
("Unequal Error Protection" (**UEP**))

4) Layered transmission over different transmission power
("Unequal Transmit Power" (**UTP**))

5) Layered transmission over different transmission power and additional FEC (**UTP_FEC**)

6) Layered transmission over different transmission power and additional unequal error protection (**UTP_UEP**)

Each transmission scheme has a certain transmission cost which is affected by the transmission power and the total content bit rate. As metric for the necessary transmission cost we define the "Used cell capacity" (Ucc) metric. A Ucc value of 1 means, that the full transmission capacity of a cell is necessary for transmitting one "Content channel" with the selected transmission scheme. A Content channel defines the transmission of one audio/video stream with the selected transmission scheme.

For instance, if Ucc value is about 0.3 there can be three "Content channels" with the same characteristic (bit rate) provided in a certain cell.

Used cell capacity (Ucc):

Percentage of total cell capacity used for transmission of one content channel.

Example calculation with transmission scheme 6:

$power_x$ *(x = layer):* $power_1 = -5dB$; $power_2 = -7dB$

bit rate (including FEC): $b_1 = 194\ kBit/s$; $b_2 = 48\ kBit/s$;

Power fraction:

$pf_1 = 10^{power1/10} = 10^{-5/10} = 0.32$;

$pf_2 = 10^{power2/10} = 10^{-7/10} = 0.20$

Percentage of cell capacity used for payload transmission = 80 % (20% pilots/control channels)

$Bearer\ rate_{12} = 256\ kBit/s$

→ $channels_1 = 80\% / pf_1 = 0.8 / 0.32 = 2.5$
→ $channels_2 = 80\% / pf_2 = 0.8 / 0.20 = 4.0$
→ $Total\ cell\ capacity_1 = channels_1 * bearer\ rate_1 = 2.5 * 256\ kBit/s = 640\ kBit/s$
→ $Total\ cell\ capacity_2 = channels_2 * bearer\ rate_2 = 4.0 * 256\ kBit/s = 1024\ kBit/s$

$Percentage\ of\ used\ cell\ capacity_1 = b_1/Total\ cell\ capacity_1 = 0.30$

$Percentage\ of\ used\ cell\ capacity_2 = b_2/Total\ cell\ capacity_2 = 0.05$

Ucc $= percentage\ of\ used\ cell\ capacity_1 + percentage\ of\ used\ cell\ capacity_2 = 0.35$

Content channels $= round(1 / Ucc) = 3$

6.1.3.4.5 Quality metric

One major challenge is the quality evaluation of the received media stream. The Peak Signal to Noise Ratio (PSNR) measure is commonly used in the area of video coding. According to [21], PSNR is not suited for evaluating the effect of packet losses or freezing frames of a video. Taking into account the work in [19] and [20] we defined appropriate objective quality categories from maximum to inacceptable quality based on three different measured values described below:

Measured values:

1) **Lost video play out**
percentage of freeze frame which reflects the amount of losses in the SVC base layer.
→ a value of 0.3 means 30 % of whole stream is affected by errors

2) **Lost audio play out**
percentage of time where there is no audio
→ a value of 0.3 means 30 % of audio is lost

3) **Playout frames with reduced quality**
percentage of non decodable SNR layer without freeze frames
→ a value of 0.3 means 30 % of all non referenced frames are lost

The four introduced quality categories try to reflect the scalability behaviour using SVC with one SNR layer. The user experiences the appropriate quality if the already described metrics lie in the following defined ranges.

Four quality categories:

1. Maximum:

Lost video play out	*< 0.02 &&*
Lost audio play out	*< 0.02 &&*
Playoutframes with reduced quality	*< 0.02*

2. Medium:

Lost video play out	*< 0.02 &&*
Lost audio play out	*< 0.02 &&*
Playoutframes with reduced quality	*< 0.7*

3. Minimum:

Lost video play out	*< 0.1 &&*
Lost audio play out	*< 0.1 &&*
Playoutframes with reduced quality	*<= 1*

4. Inacceptable:

Lost video play out	*>= 0.1 ||*
Lost audio play out	*>= 0.1*

To get an overview of the received quality in a transmission cell, we define the "Coverage" metric which shows the percentage of users receiving at least a certain quality.

EXAMPLE**:** 250 out of the total of 500 users achieve constraints of medium quality → Coverage of simulated transmission scheme at medium quality is 50 % coverage.

6.1.3.4.6 Simulation Results

For SVC, two layer transmission schemes are applied whereas audio and video base layer belong to one transmission layer (with higher FEC protection and/or higher transmission power) and the video SNR layer to another transmission layer.

For the different transmission schemes, FEC code rates (in the range between 0.4 and 1.0) and transmission power levels are varied. The plots in this subclause show coverage at the y-axis and Ucc at the x-axis.

Results are given for the sequence "Reuter". Figures 41 to 43 show the results for all transmission settings for each quality category.

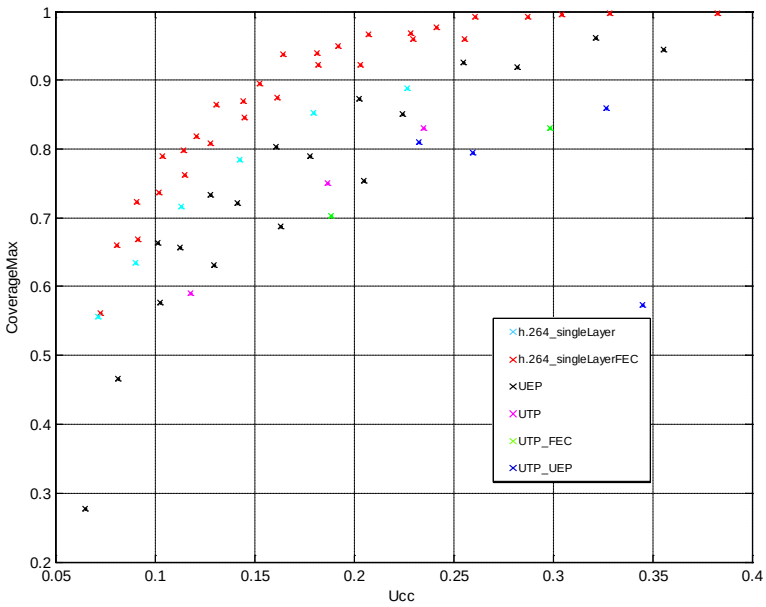

Figure 41: Coverage of maximum quality sorted by transmission schemes

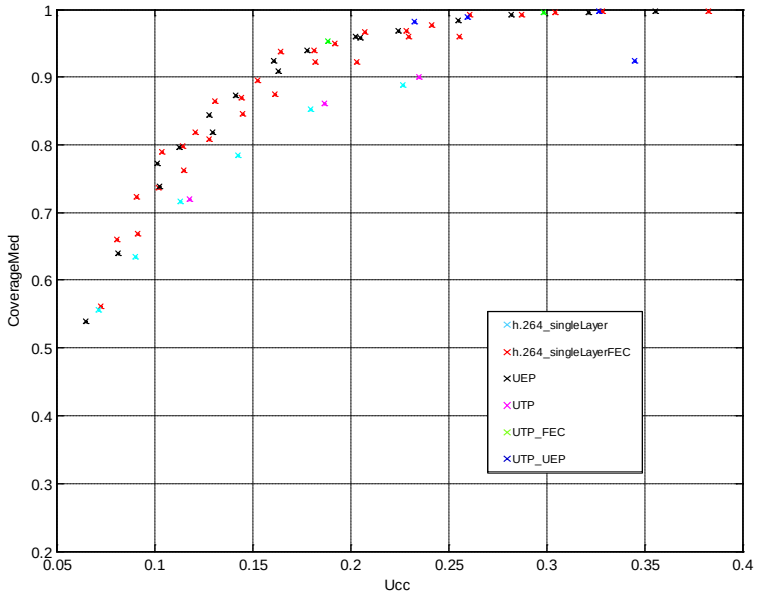

Figure 42: Coverage of medium quality sorted by transmission schemes

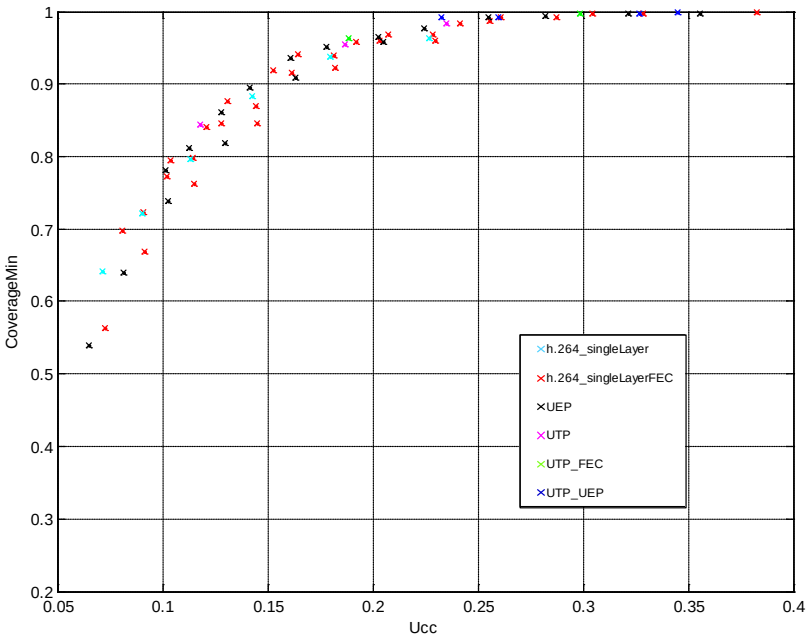

Figure 43: Coverage of minimum quality sorted by transmission schemes

6.1.3.5 Coding Results using KTA

6.1.3.5.1 Experimental Setup

In this subclause, QVGA/VGA encoding using SVC and an optimized H.264/AVC encoder is compared.

For the experiments, the publicly available JSVM 9.17 and KTA 2.3 software packages [14] and [8] are used. The simulation settings were aligned with the VCEG recommended simulation conditions [15].

Four publicly available test sequences were used (Table 17). For each sequence a basis resolution (QVGA) and an enhanced resolution (VGA) is used, so as to compare SVC spatial scalability against H.264/AVC coding.

Table 17: Test sequences

Sequence	Basis resolution	Enhanced resolution
CrowdRun	QVGA, 12.5 *Hz*	VGA, 25*Hz*
Seeking		
Crew	QVGA, 15 *Hz*	VGA, 30*Hz*
Soccer		

For SVC encoding, the scalable baseline profile is used. For H.264/AVC encoding, baseline profile for encoding of the basis resolution and high profile for encoding of the enhanced resolution is used. The common coding tools that were used in the SVC and H.264/AVC simulations are shown in Table 18. (A complete description of the coding settings for JSVM and KTA are provided in attachment A). Note that KTA was operated in H.264/AVC mode, which means that only H.264/AVC compliant coding tools were used.

Table 18: Common coding settings

Coding options	Basis resolution	Enhanced resolution
B pictures	No	Yes
8x8 transform and intra prediction	No	Yes
Entropy coding	CAVLC	CABAC
Number of active reference pictures for list 0	2	2
Number of active reference pictures for list 1	Na	2
Deblocking filter	Yes	Yes
Weighted prediction	No	No
Prediction structure	IPP	IbBbP
Intra period	12 frames for 12.5 fps sequences 14 frames for 15 fps sequences	24 frames for 25 fps sequences 28 frames for 30 fps sequences
Search range	64	64

For KTA simulations, a QP range of 22-37 was used. For JSVM simulations, {37, 33, 29, 25} was used as base layer QPs, and a QP offset of 2, i.e. the corresponding enhancement layer QPs were {39, 35, 31, 27}.

6.1.3.5.2 Results

Figures 44 to 51 and Table 19 ~ Table 22 show the results for the four test sequences.

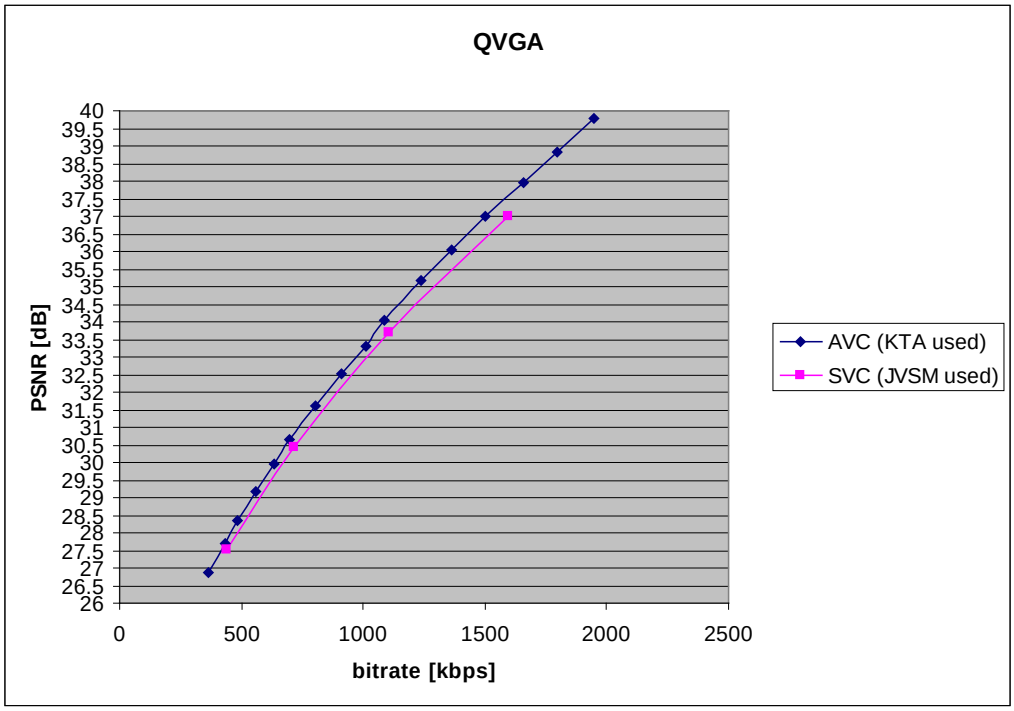

Figure 44: Sequence CrowdRun: PSNR results for QVGA SVC and QVGA AVC. JSVM was used for SVC encoding and KTA for AVC encoding

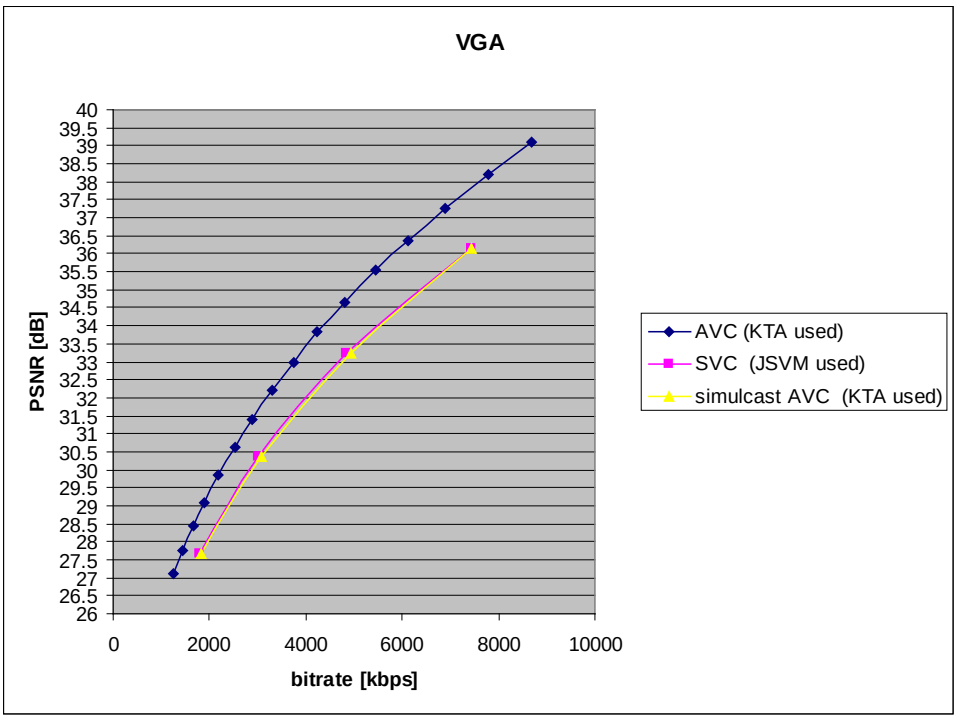

Figure 45: Sequence CrowdRUN: PSNR results for VGA SVC, VGA AVC and AVC simulcast (QVGA+VGA). JSVM was used for SVC encoding and KTA for AVC encoding

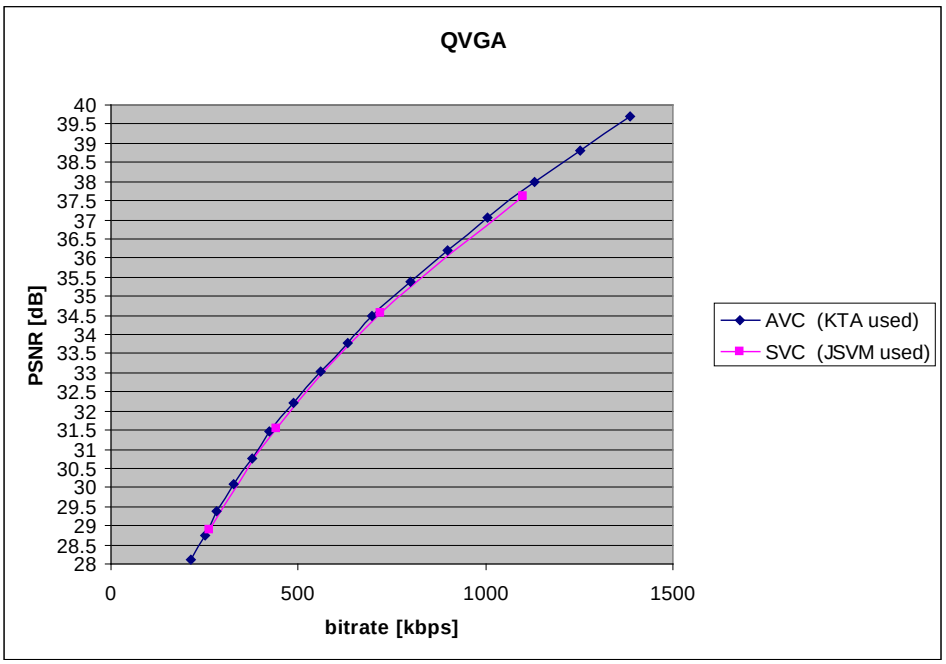

Figure 46: Sequence Seeking: PSNR results for QVGA SVC and QVGA AVC. JSVM was used for SVC encoding and KTA for AVC encoding

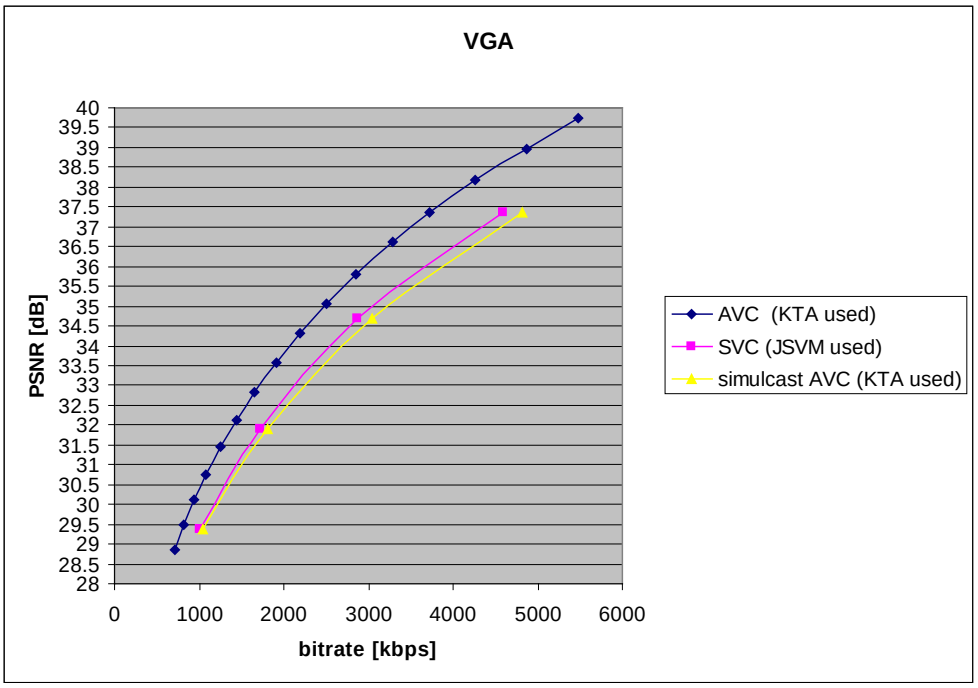

Figure 47: Sequence Seeking: PSNR results for VGA SVC, VGA AVC and AVC simulcast (QVGA+VGA). JSVM was used for SVC encoding and KTA for AVC encoding

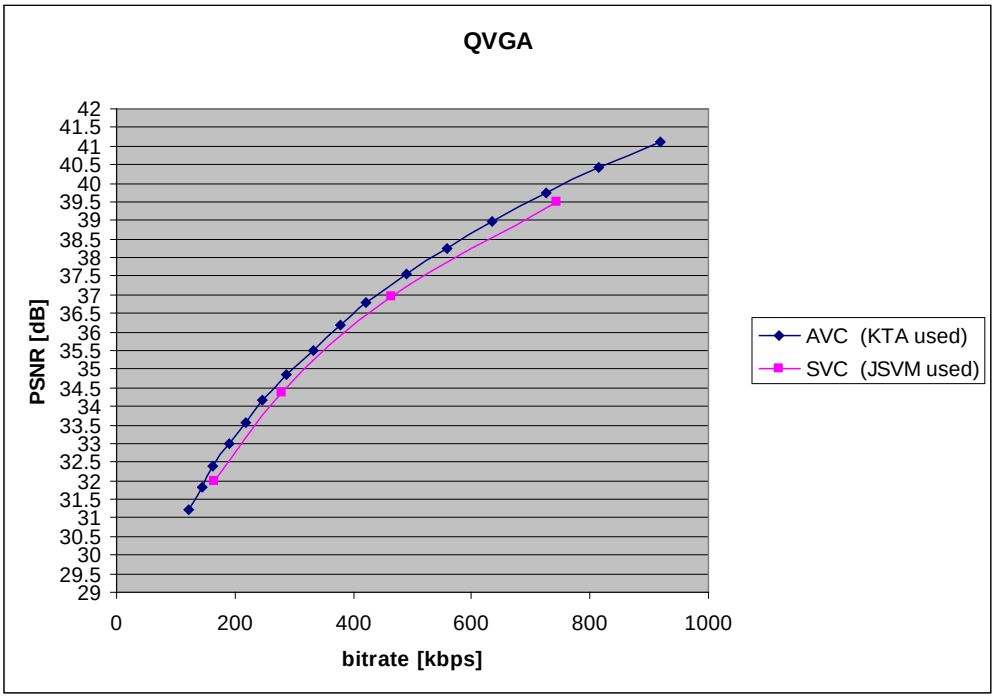

Figure 48: Sequence Crew: PSNR results for QVGA SVC and QVGA AVC. JSVM was used for SVC encoding and KTA for AVC encoding

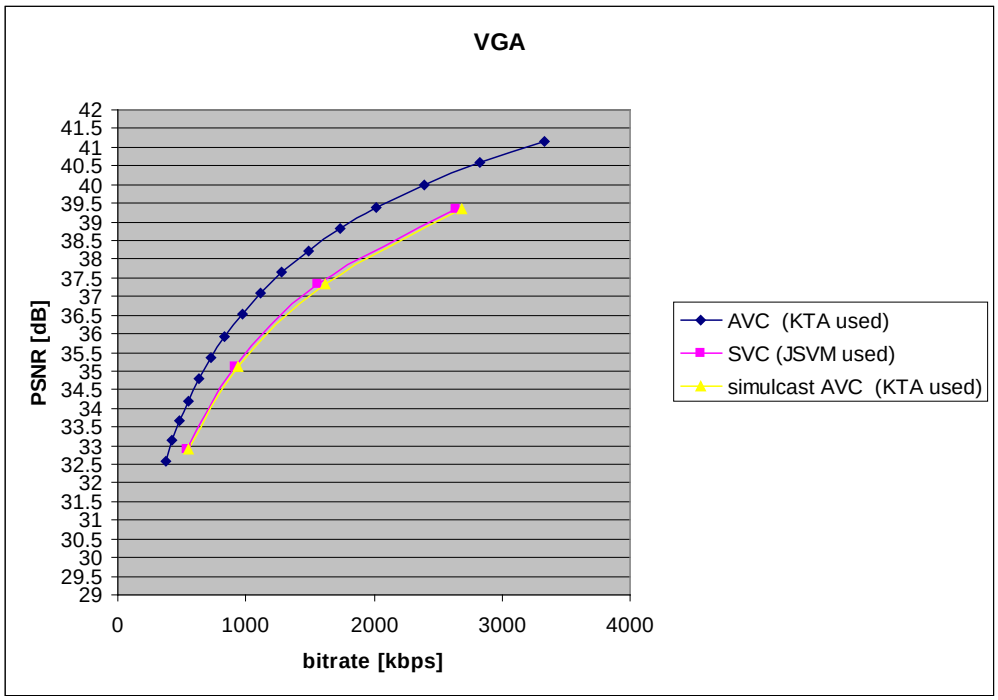

Figure 49: Sequence Crew: PSNR results for VGA SVC, VGA AVC and AVC simulcast (QVGA+VGA). JSVM was used for SVC encoding and KTA for AVC encoding

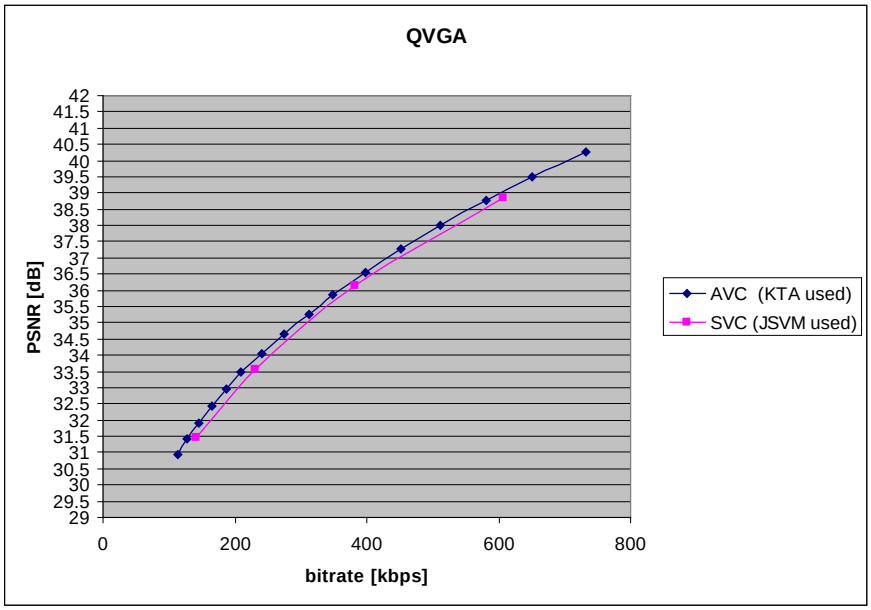

Figure 50: Sequence Soccer: PSNR results for QVGA SVC and QVGA AVC. JSVM was used for SVC encoding and KTA for AVC encoding

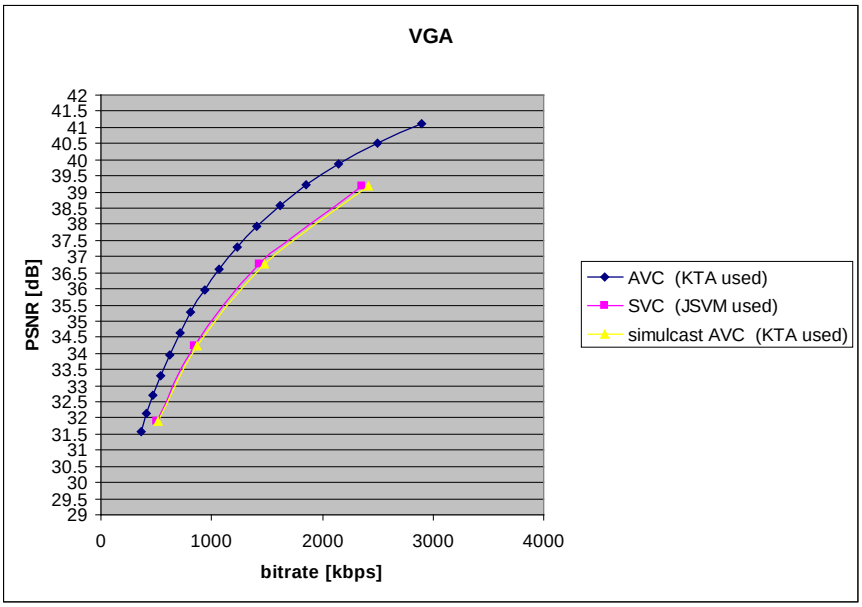

Figure 51: Sequence Soccer:PSNR results for VGA SVC, VGA AVC and AVC simulcast (QVGA+VGA). JSVM was used for SVC encoding and KTA for AVC encoding

Table 19: Results for sequence CrowdRun. JSVM was used for SVC encoding and KTA for AVC encoding

QVGA SVC				VGA SVC				Simulcast AVC	
Bitrate [*kbps*]	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**SVC gain over AVC simulcast [%]**
440.8	27.5	420.4	4.85	1797.8	27.7	1408.9	27.60	1829.3	1.72
718.5	30.4	677.0	6.13	3018.0	30.3	2396.0	25.96	3073.1	1.79
1106.6	33.7	1047.1	5.68	4847.9	33.2	3878.9	24.98	4926.0	1.59
1597.9	37.0	1501.7	6.40	7425.9	36.1	5930.6	25.21	7432.3	0.09
Average			5.76				25.94		1.30

Table 20: Results for sequence Seeking. JSVM was used for SVC encoding and KTA for AVC encoding

QVGA SVC				VGA SVC				Simulcast AVC	
Bitrate [*kbps*]	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**SVC gain over AVC simulcast [%]**
262.2	28.9	256.1	2.41	1003.6	29.4	795.7	26.14	1051.7	4.57
442.3	31.5	429.9	2.88	1718.5	31.9	1376.5	24.84	1806.4	4.87
718.2	34.6	704.3	1.97	2871.1	34.7	2333.8	23.02	3038.1	5.50
1101.6	37.6	1082.5	1.77	4599.6	37.4	3732.6	23.23	4815.1	4.47
Average			2.26				24.31		4.85

Table 21: Results for sequence Crew. JSVM was used for SVC encoding and KTA for AVC encoding

QVGA SVC				VGA SVC				Simulcast AVC	
Bitrate [*kbps*]	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**SVC gain over AVC simulcast [%]**
165.3	32.0	148.7	11.21	543.8	32.9	400.2	35.90	548.8	0.91
278.7	34.4	256.5	8.65	917.7	35.1	684.1	34.15	940.6	2.43
463.3	37.0	432.6	7.09	1554.8	37.3	1181.8	31.56	1614.4	3.69
744.8	39.5	697.2	6.82	2633.7	39.3	1994.6	32.05	2691,7	2.15
Average			8.44				33.41		2.30

Table 22: Results for sequence Soccer. JSVM was used for SVC encoding and KTA for AVC encoding

QVGA SVC				VGA SVC				Simulcast AVC	
Bitrate [*kbps*]	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**PSNR Y**	**AVC bitrate [*kbps*]**	**cost over AVC [%]**	**Bitrate [*kbps*]**	**SVC gain over AVC simulcast [%]**
140.5	31.4	128.9	9.02	500.3	31.9	390.2	28.20	519.1	3.62
230.9	33.6	215.2	7.26	845.1	34.2	656.5	28.72	871.7	3.06
382.6	36.1	365.9	4.58	1425.6	36.8	1106.5	28.84	1472.4	3.17
606.9	38.9	588.7	3.09	2361.6	39.2	1832.9	28.85	2421.6	2.48
Average			5.99				28.65		3.08

On average, SVC induces average bit rate costs of 5.6% and 28.1% over non-scalable H.264/AVC at QVGA and VGA resolution, respectively. The average SVC bit rate reduction over H.264/AVC simulcast is 2.9%.

6.1.3.6 Coding Results using JSVM

6.1.3.6.1 JSVM

The JSVM (Joint Scalable Video Model) software is the reference software for the Scalable Video Coding (SVC) project of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). JSVM can be used to encode standard compliant H.264/AVC streams as well as scalable video streams.

6.1.3.6.2 Experimental Setup

In the conducted experiments, we used the JSVM software package version 9.17 [14].

The presented encodings are comprised of a 2-layer SVC based stream with SNR or resolution scalability. The respective resolutions used in our tests were 320x240 to 320x240 (QVGA/QVGA) for SNR scalability (Configuration 1 and configuration 2).

6.1.3.6.3 Sequences

For the QVGA test the Sequences and their corresponding configurations are depicted in Table 23.

Table 23: Sequences and their configurations for QVGA SNR scalability

Configuration 1	Configuration 2	Sequence
- AVC: QVGA @ 12.5 fps High Profile - AVC: QVGA @ 25 fps High Profile - SVC BaseLayer: QVGA @12.5 fps High Profile (Identical to AVC QVGA) - SVC Enh. Layer: QVGA @ 25 fps Scalable High Profile	- AVC: QVGA @ 25 fps High Profile - AVC: QVGA @ 25 fps High Profile - SVC BaseLayer: QVGA @25 fps High Profile (Identical to AVC QVGA) - SVC Enh. Layer: QVGA @ 25 fps Scalable High Profile	AlohaWave
		CrowdRun
		ParkJoy
		Umbrella

6.1.3.6.4 Coding Tools

Tables 24 and 25 display the coding tools used for Single Layer, and SVC encoding.

Table 24: Coding tools for single layer coding

Coding tools for single layer coding (& simulcast)	High profile
B pictures	Yes
8x8 transform & intra pred.	Yes
entropy coding	CABAC
GOP size	16
Random Access Point distance	1,28 s

Table 25: Coding tools for SVC coding

Coding tools for SVC Scalable High Profile	Enhancement Layer
B pictures	Yes
8x8 transform & intra pred.	Yes
entropy coding	CABAC
GOP size	16
Random Access Point distance	1,28 s

6.1.3.6.5 Results

The Rate-Distortion (RD) curves for the selected sequences are shown in subclauses 6.1.3.6.5.1 and 6.1.3.6.5.2

The plots show the performance of SVC and AVC encodings. Each plot shows the Rate-Distortion (RD) curve for the base layer (AVC 0) and the enhancement layer (SVC 1) as well as for the corresponding single layer (AVC 1) providing the same quality level as the SVC enhancement layer and the accumulated simulcast curve.

Sample configuration files used to generate these results can be found in the Attachment B.

The video sequences used in this subclause are publicly available.

6.1.3.6.5.1 320x240 (QVGA) SNR Scalability (Configuration 1)

320x240 Scalable High Profile (SVC) vs. AVC 320x240 High Profile with CGS SNR – scalability.

Base-layer at half frame-rate and a quantization point difference of 4:

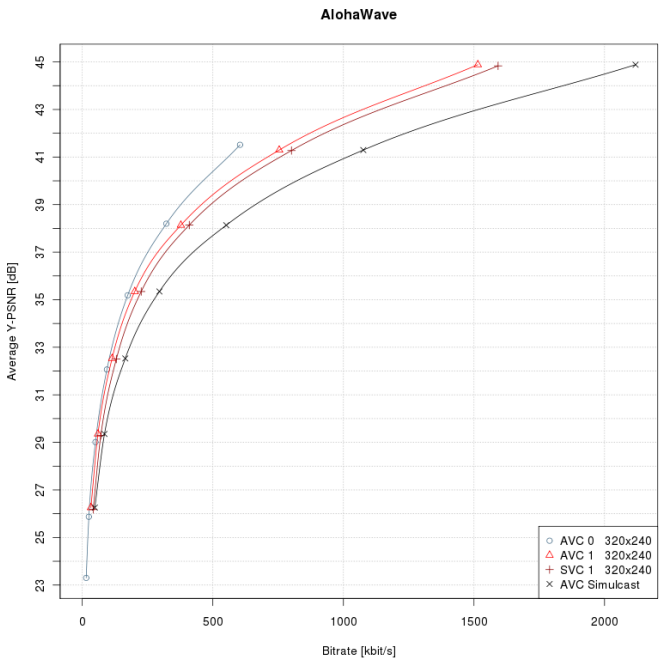

Figure 52: AlohaWave

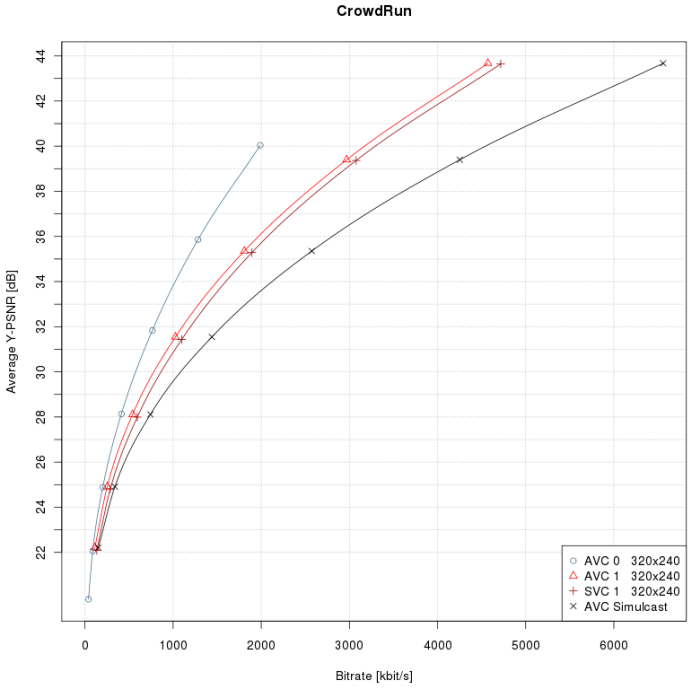

Figure 53 : CrowdRun

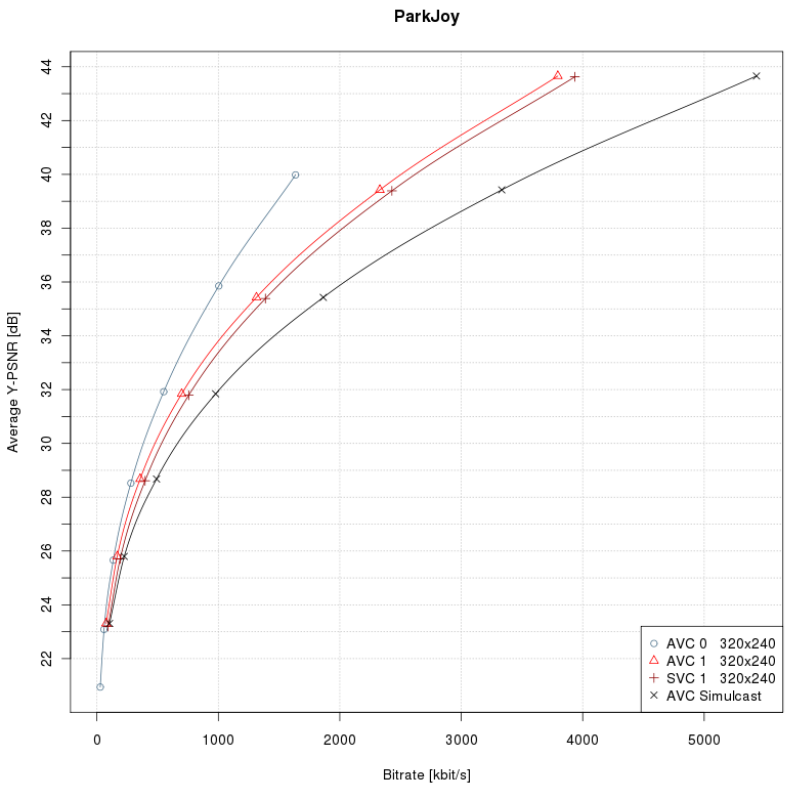

Figure 54 : ParkJoy

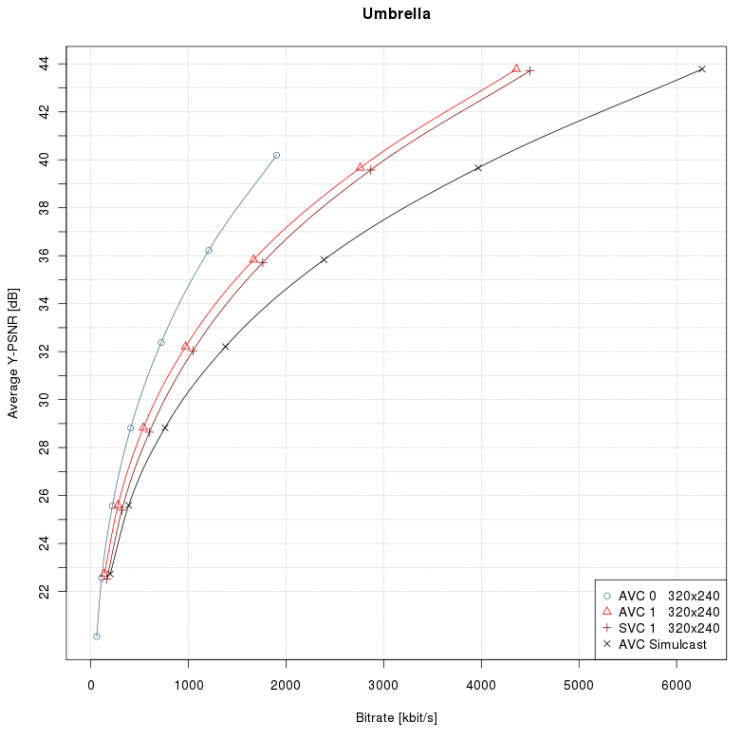

Figure 55 : Umbrella

Table 26: Results of AlohaWave

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
129.92	32,48	113,08	14.90	163.01	20.30
227.81	35,39	203,08	12.18	298.48	23.68
404.69	38,08	373,08	8.47	544.94	25.74
803.18	41,29	753,08	6.65	1074.54	25.25

Table 27: Results of CrowdRun

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
610.09	28,13	542,40	12.48	745.45	18.16
1112.98	31,52	1022,40	8.86	1434.41	22.41
1914.16	35,37	1812,40	5.61	2579.12	25.78
3090.84	39,41	2972,40	3.98	4257.29	27.40

Table 28: Results of ParkJoy

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
403.02	28,68	356,55	13.03	491.69	18.03
761.35	31,83	696,55	9.30	975.43	21.95
1399.41	35,44	1316,55	6.29	1868.36	25.10
2436.78	39,41	2326,55	4.74	3328.81	26.80

Table 29: Results of Umbrella

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
619.64	28,82	538,79	15.01	760.13	18.48
1071.74	32,20	968,79	10.63	1377.07	22.17
1790.63	35,84	1668,79	7.30	2391.58	25.13
2895.69	39,66	2758,79	4.96	3967.15	27.01

6.1.3.6.5.2 320x240 (QVGA) SNR Scalability (Configuration 2)

320x240 Scalable High Profile (SVC) vs. AVC 320x240 High Profile with CGS SNR – scalability.

Base-layer at full frame-rate and a quantization point difference of 4.

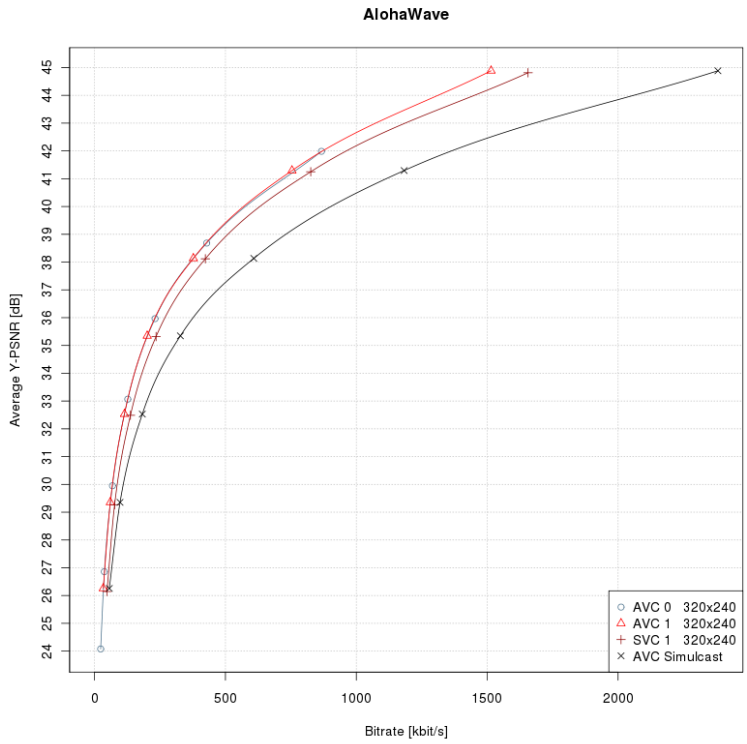

Figure 56: AlohaWave

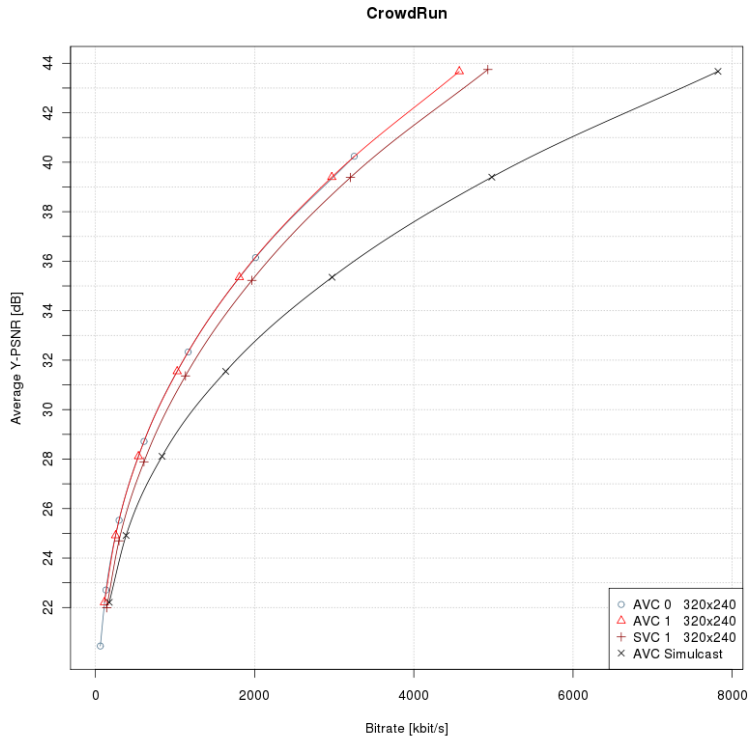

Figure 57: CrowdRun

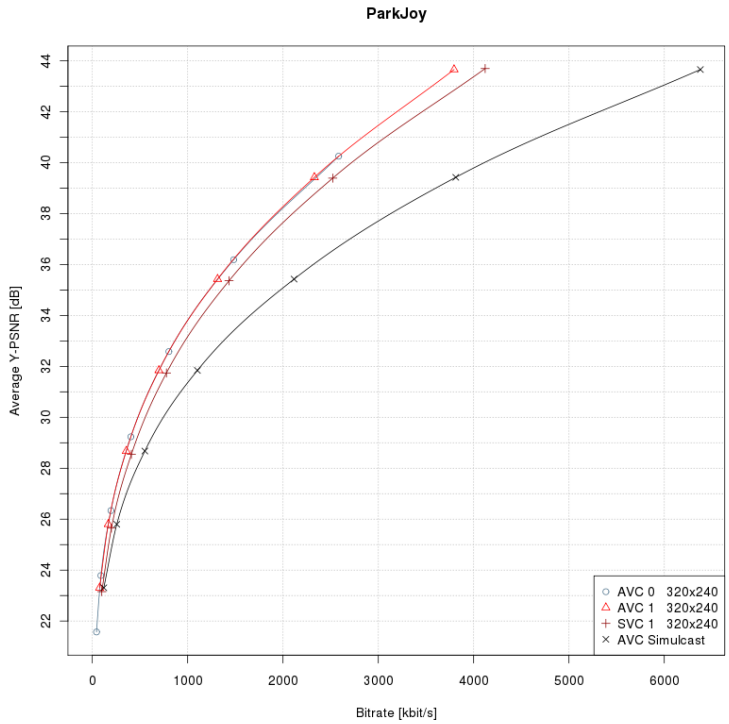

Figure 58: ParkJoy

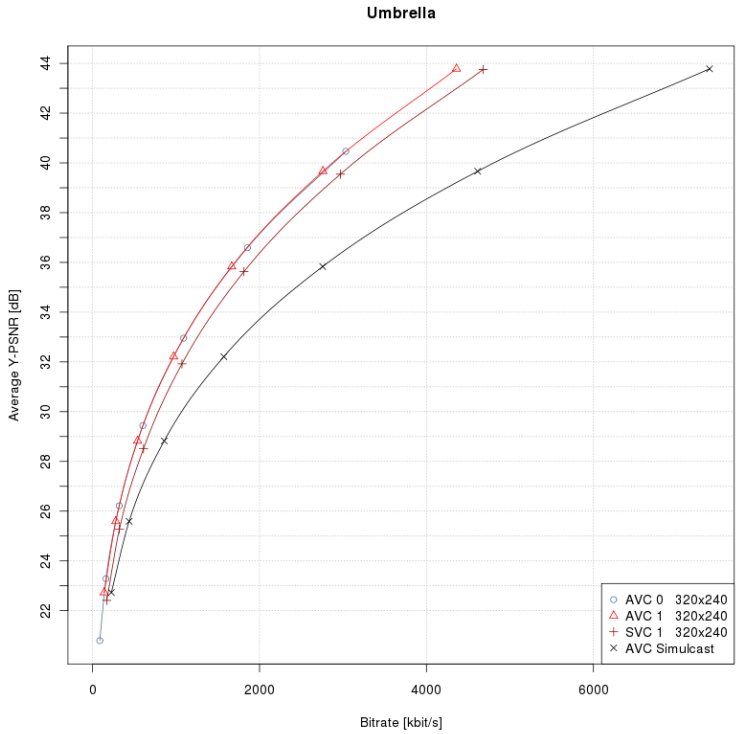

Figure 59: Umbrella

Table 30: Results of AlohaWave

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
136.88	32,48	113,08	21.05	180.66	24.23
238.33	35,39	203,08	17.36	331.21	28.04
419.94	38,08	373,08	12.56	601.27	30.16
831.84	41,29	753,08	10.46	1180.72	29.55

Table 31: Results of CrowdRun

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
638.78	28,13	542,40	17.77	840.31	23.98
1159.17	31,52	1022,40	13.38	1629.82	28.88
2001.61	35,37	1812,40	10.44	2980.79	32.85
3211.44	39,41	2972,40	8.04	4986.23	35.59

Table 32: Results of ParkJoy

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
421.77	28,68	356,55	18.29	551.61	23.54
789.90	31,83	696,55	13.40	1099.49	28.16
1451.64	35,44	1316,55	10.26	2121.56	31.58
2528.89	39,41	2326,55	8.70	3807.23	33.58

Table 33: Results of Umbrella

SVC				Simulcast AVC	
Bitrate [*kbps*]	Y-PSNR [*dB*]	AVC bitrate [*kbps*]	cost over AVC [%]	Bitrate [*kbps*]	SVC gain over AVC simulcast [%]
643.72	28,82	538,79	19.47	857.53	24.93
1115.75	32,20	968,79	15.17	1568.79	28.88
1864.75	35,84	1668,79	11.74	2760.56	32.45
3009.03	39,66	2758,79	9.07	4612.60	34.76

6.1.3.7 Caching Efficiency Improvement with SVC for Adaptive HTTP VoD

6.1.3.7.1 Overview

Adaptive HTTP streaming in a VoD system takes advantage of the widely deployed network caches to relieve video servers from sending the same content to a high number of users in the same access network. Since the connection characteristics may vary over the time, with adaptive Streaming over HTTP, a technique that has been recently proposed, video clients may dynamically adapt the requested video quality for ongoing video flows, to match their current download rate as good as possible. One possibility to provide adaptive streaming over HTTP is to encode multiple representations of each of the videos with H.264/AVC at the server and offer them side-by-side. Another is offering all these representations embedded in one file via Scalable Video Coding (SVC). The presented simulations compare the impact of multiple chunk based content representations on the caching efficiency either using H.264/AVC or SVC. Similar simulation results within an HTTP based progressive download scenario have already been presented in [22].

Figure 60 schematically shows a network over which a video library is offered by a Video on Demand (VoD) service. The operator of the access network (i.e., the cloud in the figure), offers connectivity to its customers via access links and connects to the Internet (where the content library is offered on an origin server by a third party) over a "transit" link, in the following referred to as the cache feeder link. In that way the customers of the access network operator can access video content, in particular the movies on the origin server. The network operator deploys a proxy and a cache in its network to minimize the amount of transmitted data through the "transit" link relieving the server of having to send an extremely high amount of video data. Since the cache is usually too small to host the complete video library and the content library on the origin video server often changes, the video files that are stored in the cache at every moment need to be carefully selected. This is accomplished by an appropriate caching algorithm.

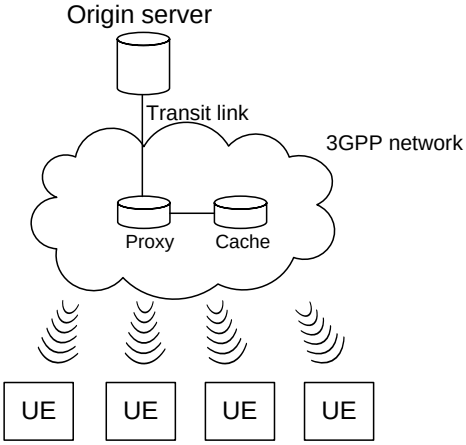

Figure 60: A typical network, hosting a cache, over which content is offered.

There are many different cache replacement algorithms that have been proposed over the last years that optimize the caching performance based on some special criteria. Most algorithms make decisions based either on how recently an object has been requested or on how frequently an object has been requested over a time period or a combination thereof. In [21] the chunk-based delivery (video files downloaded in smaller parts thereof, i.e. chunks/segments) is exploited in a caching context. In this work the chunks that will be consumed in a near future with a high probability are predicted, assuming that it is very likely that a user playing chunk n of a given video file at the current moment will play chunk ***n+k*** of the same video file ***k*** time instants later.

6.1.3.7.2 Effect of Multiple Representations on the Caching Efficiency

In this contribution, we consider the scenario where users may request a certain video clip in one of a possible set of resolutions or quality versions. Hence, each video offered by the origin server must be encoded in a given number (N) of bit rates. These N versions can be encoded separately with AVC and offered side by side, a scenario we refer to as "Multi-Representation VoD (MR-VoD)", or can be embedded in a multi-layer representation which allows for further separation into file subsets (layers) using SVC, a scenario we refer to as "SVC-VoD". We discuss the impact of the former first and comment on the latter.

Compared to the scenario in which only one version is offered (which we refer to as the "Single-Representation VoD (SR-VoD)" scenario), in the MR-VoD scenario, the requests for a particular video clip are distributed over its *N* versions.

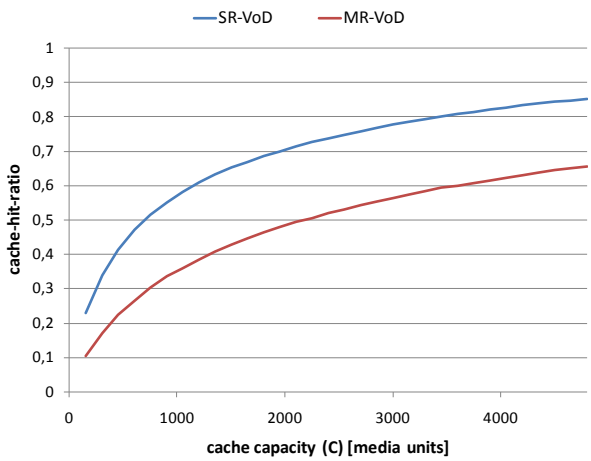

Figure 61: Caching efficiency reduction result of offering a higher variety of representations (e.g. 4) for each file

If each of the versions associated with a video clip is requested with more or less equal probability, the ranking in the MR-VoD scenario is almost the same as in the scenario with only one version: instead of occurring only once, each video clip occurs N times in that ranking, but with high probability in a block of N consecutive ranks. A consequence of this is that if a certain version of a video is cached it is highly likely that all other versions need to be cached as well. Consequently, in order to attain the same cache-hit-ratio in the MR-VOD scenario as in the SR-VOD scenario, the cache should be able to store all N versions of the video instead of just one. Since storing N versions side by side requires more storage, a larger cache size/capacity is needed to attain the same hit ratio. Conversely, if the same cache capacity is used, a lower cache-hit-ratio results, as illustrated in Figure 61. Note that based on a similar reasoning (and as described in more detail in clause 6.1.3.7.3) the SVC-VoD scenario could attain the same hit ratio with practically the same cache size.

6.1.3.7.3 Scalable Video Coding and Impact on the Caching Efficiency

The main difference between MR-VoD and SVC-VoD is illustrated in Figure 62. It can be seen that by using SVC much more video clips at different representations can be stored in the cache, while with MR-VoD many files have to be removed from the cache to obtain additional space for the new incoming files or versions of them.

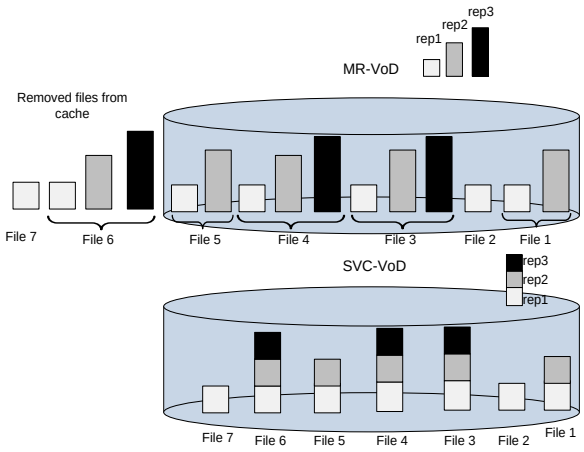

Figure 62: Caching performance comparison for MR-VoD and SVC-VoD

When considering a VoD service with multiple available representations based on layers of SVC, first the amount of data that has to be transmitted to and stored in the cache is reduced compared to the MR-VoD case, and second, more clients request the same data (layers) since clients requesting different representations of a same video clip are expecting to receive a set of layers, where some layers are common for all of those requests, e.g. the base layer. Thus, the HTTP request for a certain quality results in a multiple HTTP request for each of the mentioned layers and all requests for a single content incorporate at least the base layer representation. Consequently, the probability of a cache-hit for files containing the lowest layers of SVC streams, which most of the users are interested in, is increased.

Note that requesting multiple layers for each segment could be done in twofold manner. One is within one single TCP connection requesting each layer after the other. The first approach (single TCP connection) introduces additional buffering requirement (for lower layer segments) at the clients, in addition to the playout delay, since client has to wait until highest layer segment is received before playout can start. Another possibility could be to setup parallel TCP connections. The second approach (parallel TCP connections) would reduce (but not eliminate) this buffering and playout delay but introduces additional HTTP overhead (new connection per layer). The simulation results analyze the effect of different video codings on the caching efficiency. Therefore, the impact on the buffering requirements at the clients and the additional HTTP overhead is not considered here.

6.1.3.7.4 Caching Algorithm

The performance of the cache is here analyzed for two different caching algorithms (operating on chunks):

- LRU: where the most recently requested chunks are kept in the cache.
- CC: An algorithm described in [1] that takes into account the number of guaranteed hits of chunks (if the HTTP streaming client keeps on selecting the same version as it currently does), which uses an improved movie content scoring algorithm that combines the LRU and LFU basics.

In case of considering SVC there are n chunks per time interval, where n corresponds to the number of layers. In other words, the layers are transmitted and stored in the cache separately and therefore count as different objects for the cache-hit-ratio evaluation. In case of offering the n version side by side via AVC, each time interval has n independent versions, in the sense that if one version is cached and another is requested no cache hit can be counted.

6.1.3.7.5 Congestion Control

Clients (on the same access network) of a multimedia service typically share (transport and caching) resources with other multimedia clients and/or users downloading any type of data from the Internet, which produces some cross-traffic in the network causing congestion. This results in a temporarily reduced available download rate for the clients of the service.

These clients (HTTP streaming-clients) detect these variations in the connection rate available to them and adapt the bit rate at which they download their ongoing video stream, by requesting the following chunks/segments in an appropriate version. Therefore, every time a user requests a new chunk of a video an additional decision has to be made with respect to which version it will be download. This choice depends on:

- The capability of the terminal of the user.
- The congestion state between the cache and the end user (i.e., the access). If requesting the version that a user wants to download would congest the link, this request is downgraded as many times as needed to alleviate congestion.

On the access link other services run (i.e., a user may be downloading a large file, may be browsing the web, etc.) besides the HTTP streaming video client streaming a video. This type of congestion can occur any time of the day and is not necessarily restricted to peak hours. The model for this type of congestion that we have simulated in this paper is shown in Figure 63. This figure illustrates a Markov-chain with four states corresponding to four possible download rates and selected OPs.

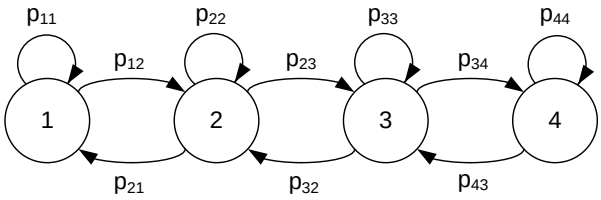

Figure 63: Model for congestion due to cross-traffic

In fact we assume that the cross traffic on the access link which is the result of sharing this link with one or more HTTP streaming clients or any other client requesting data from the Internet is such that the HTTP streaming client requesting the version in the next slot, can be described by a Markov chain.

As seen in Figure 63 this Markov chain consists of four states where the transition probabilities p_{ij} of the transition matrix $P=[p_{ij}]$ with $|j-i|>1$ are set to zero, i.e. it is only possible to go from a state to its neighbour states. The rest of the parameters (represented in the figure) were set to values that lead to realistic situations.

The most important parameters to take into account to consider whether the selected values correspond to a realistic situation or not are the mean state sojourn time (mean duration of being in a state: $E[t_i]$) and average percentage of time in each of the states (p_i), which can be derived easily from the transition probabilities, as shown in ***Eq.(4)*** and ***Eq.(5)***.

$$E[t_i] = \sum_{t_i=0}^{\infty} (t_i + 1) * p_{ii}^{\,t_i} * (1 - p_{ii}) = \frac{1}{1 - p_{ii}} \tag{4}$$

$$p_i : \pi * P = \pi \tag{5}$$

where, $\pi=\{p_1,p_2,p_3,p_4\}$ is the left eigenvector of P (associated with eigenvalue 1), a.k.a. steady state vector, which fulfils

$$\sum_{i=1}^{4} p_i = 1 \tag{6}$$

The simulation time step in the presented Markov-chain model corresponds to the selected chunk size, since the adaptation is performed by the HTTP streaming clients on a chunk basis.

6.1.3.7.6 Performance Targets

In order to compare the system where the different version of a video are offered encoded in AVC side by side with the system in which the versions are embedded in one SVC stream, we consider cache-hit-ratio and cache capacity:

- The cache-hit-ratio: calculated on a chunk basis, or when SVC is considered on smaller objects, corresponding to each of the layers of each of the chunks. It represents the percentage of these objects that can be served from the cache and do not need not to be transported over the cache feeder link.
- The cache capacity is measured in media units, which are equivalent to the size of a video clip of 90 minutes at 500 *kbps* (1 media unit=337.5 MB).

6.1.3.7.7 Simulation Results

The results presented in the following show the performance of the system comparing both multiple representations encoded with AVC (MR-VoD) offered side-by-side and multiple representations encoded with SVC (SVC-VoD). The rate distribution for the different video representations is summarized in the table below with an SVC overhead of 10% using bit rate adaptation with quality scalability and one quality layer as similarly shown in [22].

Each of the video clips is offered at four different encoding bitrates. The bitrate assumptions for AVC and SVC encodings are summarized in Table 34.

Table 34: Rate distribution for the video representations

	Rep. 1	Rep.2	Rep.3	Rep. 4
AVC	500 *kbps*	1000 *kbps*	1500 *kbps*	2000 *kbps*
SVC	500 *kbps*	1066 *kbps*	1633 *kbps*	2200 *kbps*

The chunk length is 10s.

The results shown in Table 3 correspond to the case where the bottleneck is the access link, as a consequence of some cross-traffic produced by other users. The transition probabilities can be found in the following transition matrix.

$$P = \begin{bmatrix} 0.996 & 0.004 & 0 & 0 \\ 0.004 & 0.992 & 0.004 & 0 \\ 0 & 0.004 & 0.992 & 0.004 \\ 0 & 0 & 0.004 & 0.996 \end{bmatrix}$$

The shown transition probabilities correspond to an adaptive HTTP client that spend on average an equal percentage of time in each state of 25%, in the following referred to as heavy cross traffic.

Table 35: Cache-hit-ratio for congestion in access links

Cache capacity (media units)	LRU		CC	
	AVC	SVC	AVC	SVC
500	30.9 %	45.6 % (+14.7%)	42.9 %	56.6 % (+13.7%)
1000	42.1 %	58.2 % (+16.1%)	52.0 %	64.5 % (+12.5%)
2000	54.6%	69.0% (+14.4%)	61.5%	72.0% (+11.5%)

The results in Table 35 show the difference between the use of AVC and SVC for both caching algorithms LRU and CC. Different versions of the requested videos are stored in the cache which leads to a spoilage of the available storing capacity of the cache when a single layer codec is considered, whereas when SVC is used the available resources are much more efficiently used. Furthermore, the hit-ratio increases due to the fact that many users make requests for the same data since, even though they may be interested in different version of the same video, their requests are split into multiple request, one associated with each layer that they are requesting. Since the layers built on top of each other, a user requesting layer ***k***, needs to request layer 1 to ***k-1*** too. In particular the base layer is requested by everyone.

Since the difference between both AVC and SVC are more disparate for this case we have conducted the simulations for a higher range of values for cache capacity (***C***) only focusing on the LRU caching algorithm, leading to the results shown in Figure 64.

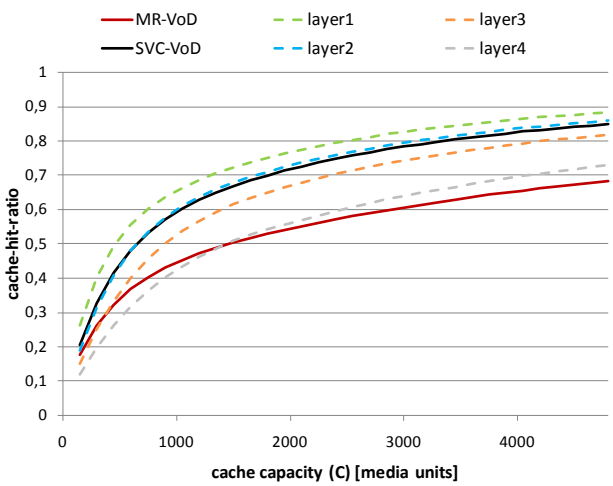

Figure 64: Congestion due to heavy cross traffic

In this figure, the cache-hit-ratio over cache capacity is depicted. It can clearly be seen how the use of SVC improves the performance of the system in terms of cache-hit-ratio compared to the use of MR-VoD. It is also noticeable that the

cache-hit-ratio for the AVC case is even lower than for the highest layer (layer 4) when SVC is used almost for all cache capacity values, since the storage capacity at the cache runs out faster with the higher diversity in requested files due to using the MR-VoD approach. Furthermore, the caching performance for the base layer is significantly higher compared to the other files and layers as the number of request for this is higher than for the other layers or different representations when AVC is considered.

The increased cache hit ratio leads to an reduced traffic through the "transit" link, which is shown in Figure 65 for SVC-VoD and MR-VoD for heavy cross traffic and LRU algorithm.

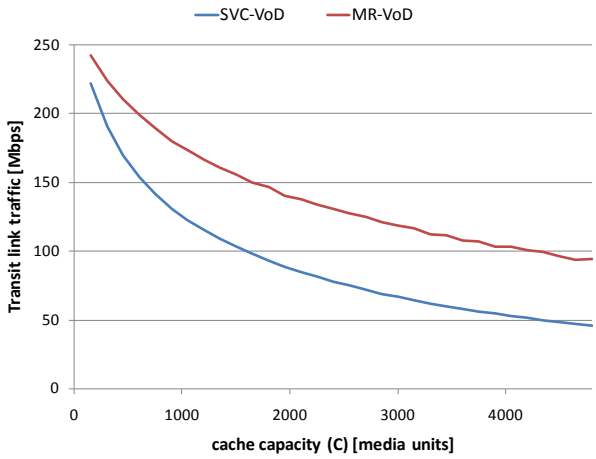

Figure 65: Average traffic through the "transit" link

The number of representations influences the saved traffic on the transit link as shown in Figure 66.

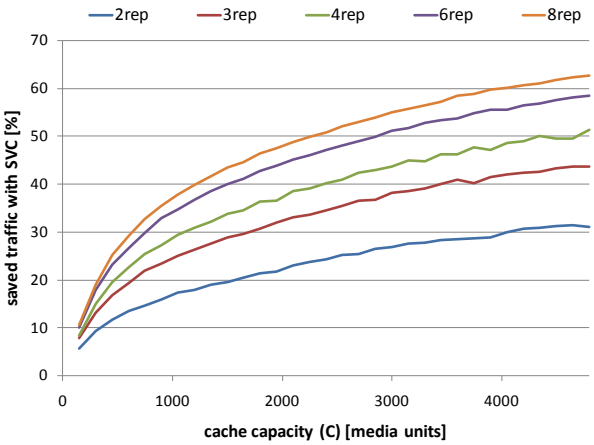

Figure 66: Saved traffic in the transit link with the use of SVC for different number of representations

In Figure 67 it is shown how the SVC penalty influences the performance of the cache hit-ratio.

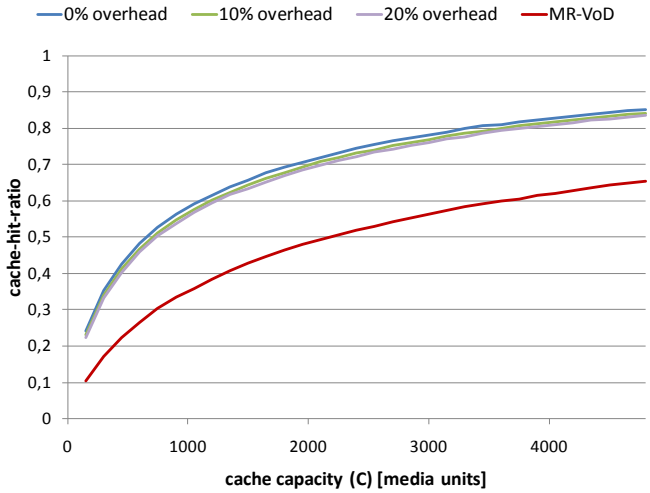

Figure 67: Cache hit-ratio for different SVC encoding overhead (0% to 20 %)

Figure 68 shows the cash-hit-ratio for a different set up of the simulation, simulating a situation with less heavy cross traffic, resulting in the HTTP streaming client residing in the highest state (4) more often. In this case, the percentage of time in each state is unequal with $\boldsymbol{p}=\{9.1\%, 9.5\%, 19.1\%, 62.3\%\}$, as well as the mean state sojourn time $\boldsymbol{E}[t_i]=(approx.)\{2s, 2s, 10s, 40s\}$, which may be closer to that which may happen in the reality. The correspondent transition matrix is shown below:

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.096 & 0.9 & 0.004 & 0 \\ 0 & 0.002 & 0.985 & 0.013 \\ 0 & 0 & 0.004 & 0.996 \end{bmatrix}$$

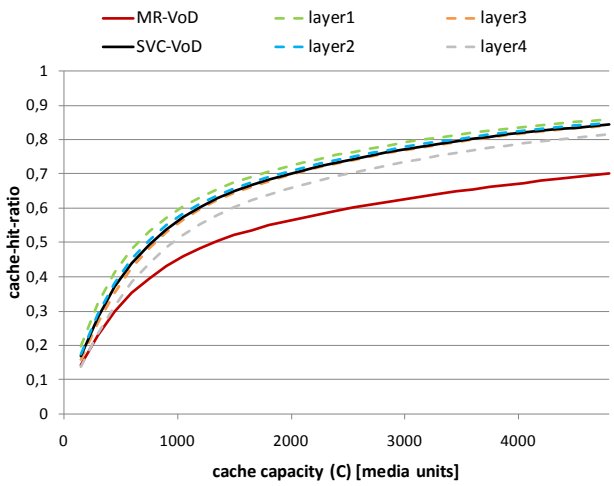

Figure 68: Cache hit-ratio for congestion due to light cross traffic

Although the variety of versions requested for this set up is supposed to be lower than in the case before, the gains of SVC-VoD compared to MR-VoD are still noticeable. Due to this reduced variability the MR-VoD performs slightly better than before but still quite poorly when compared to SVC-VoD. It can be also clearly seen how the cache-hit-ratio for the base layer is reduced (layer 1) and the cache-hit-ratio for the highest layer is increased (layer 4). If we keep on reducing the congestion all lines would converge.

The SVC penalty influences the traffic on the last mile. It is influenced by the SVC coding penalty itself but also by the congestion behaviour, since the base layer does not involve any overhead. Table 36 shows the average overhead for the two scenarios for two different assumed SVC overheads. With the assumed SVC overhead of 10% for heavy congestion it is 6.4% since the base layer is requested more often and 8.6% with light congestion, since the highest quality is

requested more often. With an assumed SVC overhead of 20% the last mile overhead for the heavy congestion case is 12.7% and the light congestion of 17.2%.

Table 36: SVC last mile overhead

	Heavy congestion	Light congestion
Last mile overhead (10% SVC overhead)	6.4%	8.6%
Last mile overhead (20% SVC overhead)	12.7%	17.2%

6.2 Stereoscopic 3D Video

6.2.1 Enabling Codecs and Formats

6.2.1.1 Introduction

There are 2 major ways of formatting the views of a stereoscopic video: spatial compression and temporal interleaving. Other formats such color shifting and 2D+Depth are possible but are either outdated or still subject to research and development.

Finally, the left and right views may also be encoded as separate views, possibly exploiting redundancies between the two views to enhance the compression efficiency. This technique is standardized by MPEG as part of the H.264/AVC standard.

6.2.1.2 Packing Formats

6.2.1.2.1 Frame Compatible Video

This technique uses spatial compression to pack the two views of the stereoscopic video into a single frame (thus the name frame compatible). This allows the usage of deployed encoding and transport infrastructure and keeping similar bandwidth requirements at the cost of information loss. The two views are first down-sampled and then packed. The down-sampling may be performed horizontally, vertically, or diagonally. The packing may use a side-by-side, top-bottom, interleaved, or checkerboard format. The different alternatives are illustrated in the following figures.

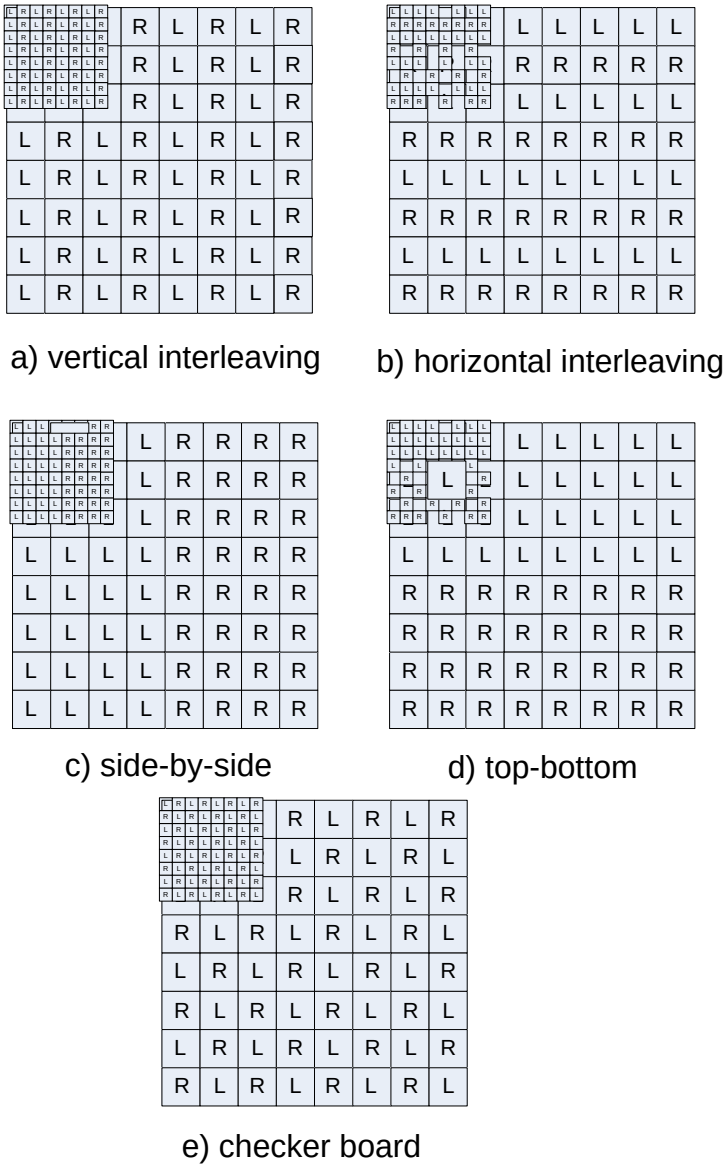

Figure 69: Spatial packing formats

6.2.1.2.2 Temporal Interleaving

In temporal interleaving, the video is encoded at double the frame rate of the original video. Each pair of subsequent pictures constitutes a stereo pair (left and right view). The rendering of the time interleaved stereoscopic video is typically performed at the high frame rate, where active (shutter) glasses are used to blend the incorrect view at each eye. This requires accurate synchronization between the glasses and the screen.

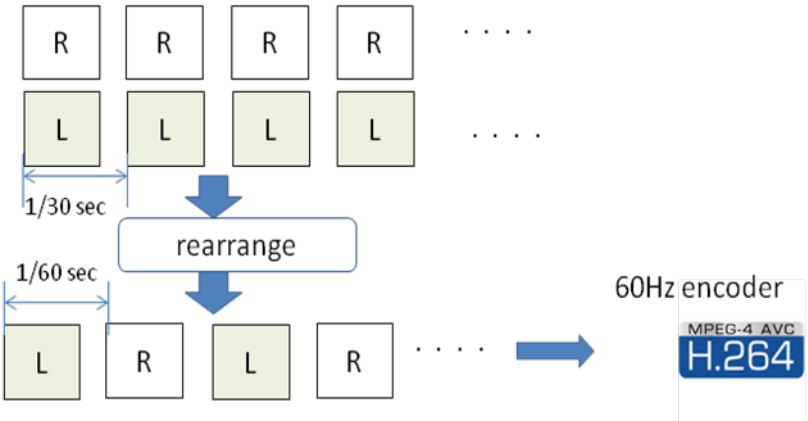

Figure 70: Temporal interleaving

6.2.1.3 Multi-view Video

MVC [3] has recently been standardized for the compression of multiple view video as an addition to the H.264/AVC standard family. In MVC, the views from different cameras are encoded into a single bit-stream that is backwards compatible with single view H.264/AVC. MVC introduces new coding tools to exhibit the spatial redundancy among the different views.

MVC is able to efficiently compress stereoscopic video in a backwards compatible manner and without compromising the view resolutions. The NAL units from the secondary view are ignored by legacy decoders as the NAL unit type will not be recognized. If the server is aware of the UE capabilities, it can omit sending NAL units from the secondary view to a device that does not support 3D or does not have enough bitrate to deliver both views.

The following figure depicts a possible prediction chain for a stereoscopic video.

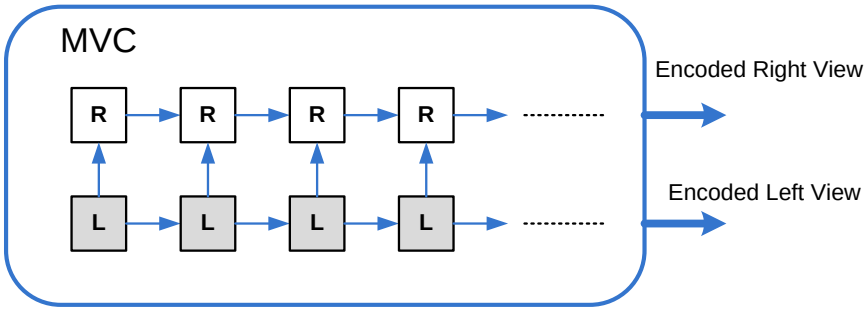

Figure 71: MVC encoding with inter-view prediction

6.2.2 Performance Evaluation

6.2.2.1 Performance Evaluation of the Compression Efficiency

6.2.2.1.1 Simulation Setup

The following formats for stereoscopic 3D video are compared:

- Side-by-Side frame packing
- Top-Bottom frame packing
- Vertical Interleaving frame packing
- Horizontal Interleaving frame packing
- Separate Left and Right view encoding

- Multi-view Video Coding (MVC)

For the different frame packing formats, the left and right views are sub-sampled to yield a packed frame that has the same resolution as the original view resolution.

For the AVC encoding of the packed formats, JM [7], KTA [8], as well as the Nokia AVC encoder have been used. The open source Nokia MVC [9] encoder has been used to encode the MVC sequences.

The following encoding parameters have been used:

- No B pictures to maintain compatibility with Baseline profile
- Fixed QP for I and P pictures: 20-34
- Reference Frames: 2
- GOP period: 30 pictures
- Baseline profile conformance
- Motion estimation search range: 16

The test sequences that have been used are:

- Alt Moabit: 432x240, 100 pictures
- Book Arrival: 432x240, 100 pictures
- Door Flowers: 432x240, 100 pictures
- Leaving Laptop: 432x240, 100 pictures

All sequences may be downloaded from the MPEG FTP server.

For the down-sampling, the tool from the JSVM and JMVM reference software has been used. The tool implements a dyadic down-sampling filter.

For evaluating the performance, PSNR has been calculated over the different sequences. For the case of frame packing, the PSNR is calculated compared to the original frame-packed video sequence. For MVC and the separate view encoding, the PSNR is calculated for the left and right views separately and then averaged.

6.2.2.1.2 Performance Evaluation

The following figures depict the Rate-Distortion curves for the different frame packing and compression configurations and for the different video sequences.

Figure 72 depicts the results for the Alt Moabit video sequence.

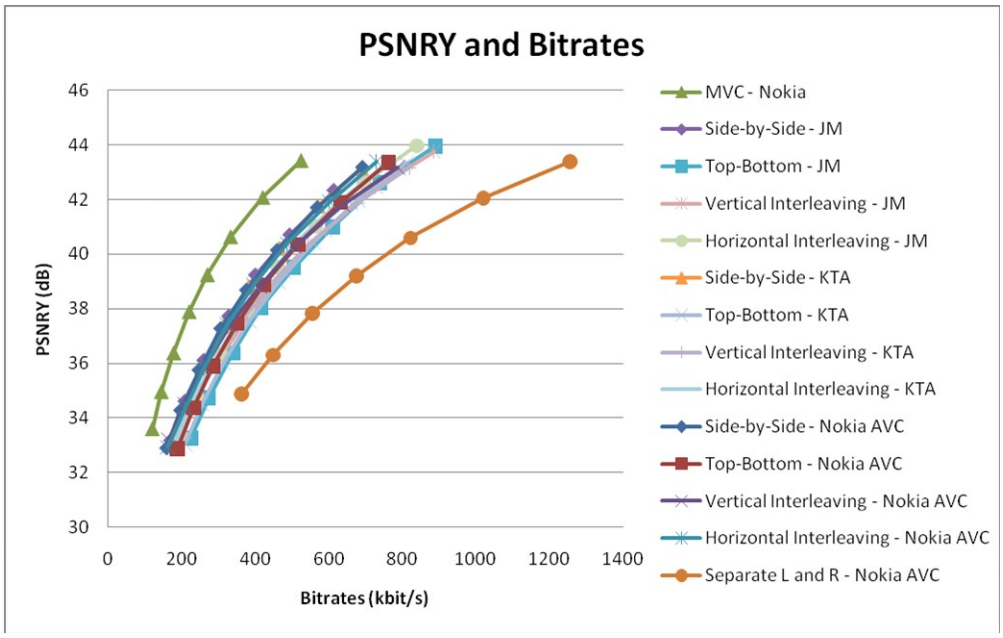

Figure 72: Alt Moabit video sequence

Figure 73 depicts the results for the Book Arrival video sequence.

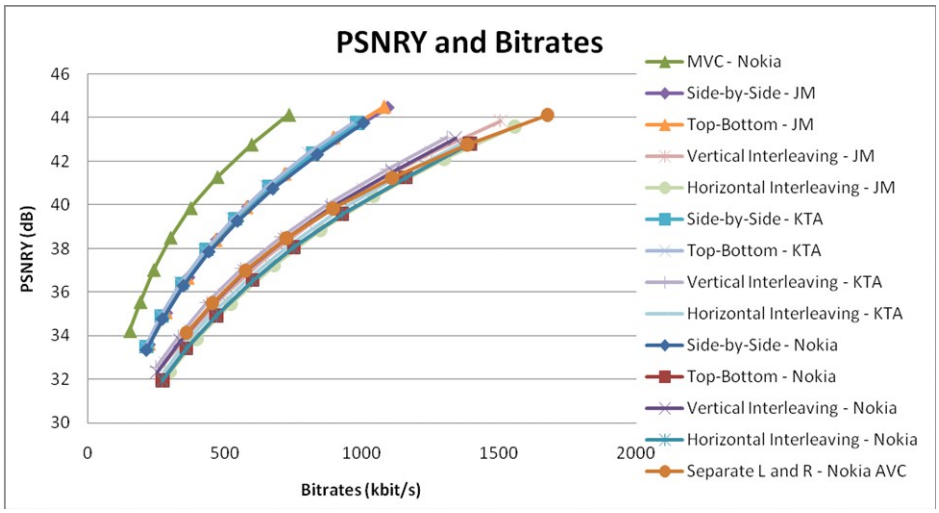

Figure 73: Book arrival

Figure 74 depicts the results for the Door and Flowers video sequence

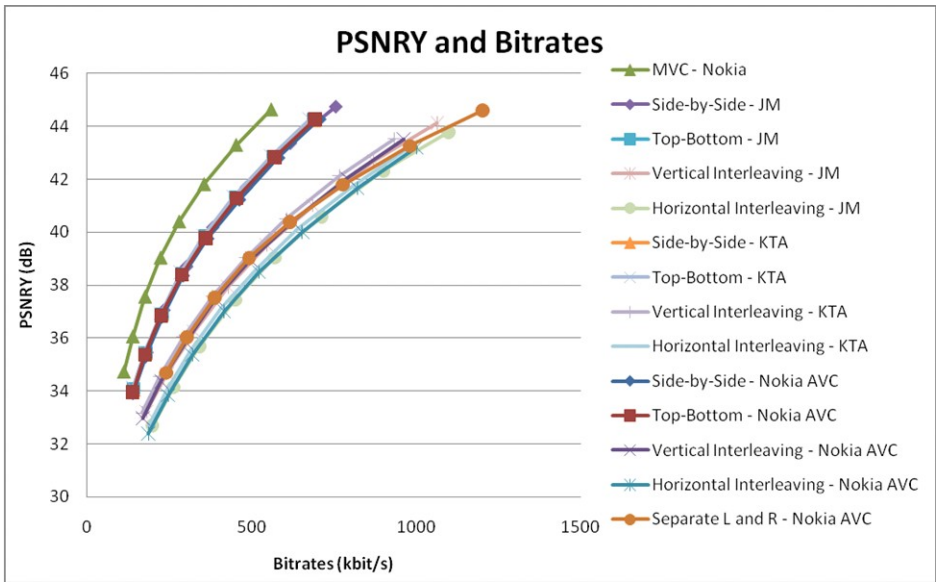

Figure 74: Door and flowers

Figure 75 depicts the results for the Leaving Laptop video sequence

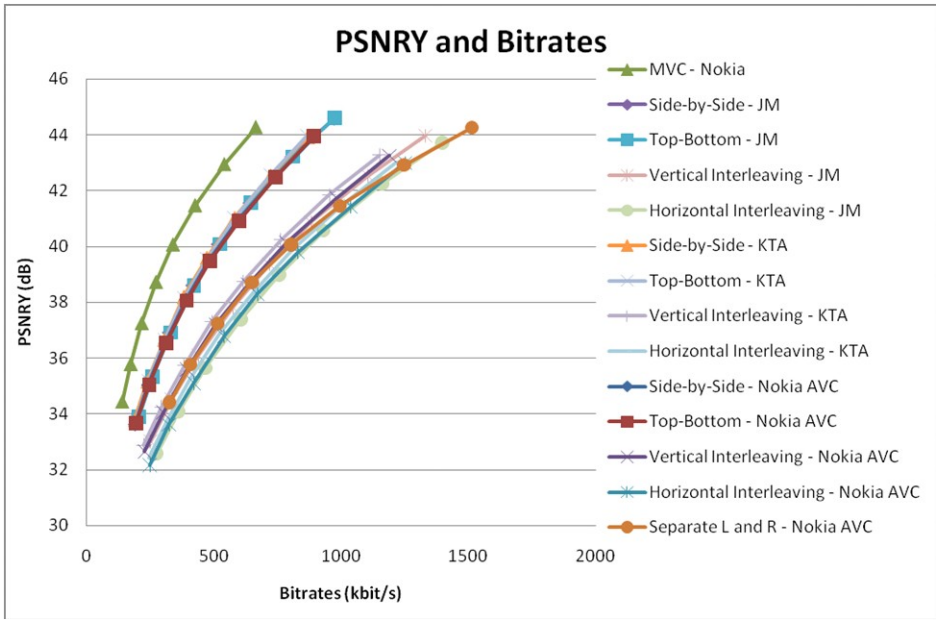

Figure 75: Leaving laptop

7 Conclusions

This Technical Report provides a set of use cases for video services in PSS and MBMS environments. The 2D use cases have been selected in order to highlight the potentially added value of scalable video coding. The 3D use cases provide initial considerations of video formats and codecs to be used in order to convey stereoscopic 3D video to a 3D capable device.

In the MBMS context, scalable video coding is compared to single layer H.264 AVC as recommended in TS26.346. Specifically, the potential benefits of SVC have been evaluated when combined with transport layer features, e.g. unequal error protection modes, layer aware transmission, in order to provide graceful degradation behaviour at client side.

In the PSS context, the ability to optimize caching and CDN traffic load has been discussed when SVC is used for adaptive HTTP streaming services.

Despite potential gains having been shown for selected use cases, some drawbacks have also been reported regarding the coding efficiency of SVC. Some concerns were raised on the implementation complexity of SVC, however this is out of the scope of the TR. More detailed analysis would be needed in order to evaluate this complexity issue on both the codec and transport level.

All the SVC performance results are either based on theoretical simulation models or make use of the SVC reference software. At the time this report was generated SVC was facing a lack of available commercial encoding/decoding solutions.

The evaluation tools collected in the TR for SVC may serve as a basis for more detailed investigations in the future.

There are no changes expected on the 3GPP specifications as a result of this TR.

Regarding the mobile 3D use cases, different representation formats have been considered in this study, such as full resolution per view, and several frame-packing arrangements, such as side-by-side or top-and-bottom (also called "frame compatible") formats. The video codecs under consideration included H.264 AVC and its 3D extension called MVC. The performance evaluation highlighted the coding efficiency of MVC. However, beyond coding efficiency other aspects, such as the rendering technology need to be taken into account (e.g. parallax barrier, lenticular network, external display device, etc.).

The first 3D capable mobile devices are expected to be available in the market in 2011. They will allow further study of the video formats (codec, bitrate, source representation format) and (most important) evaluate their associated quality of experience. Backward compatibility with 2D services and devices may have to be considered. Additional work in 3GPP is encouraged in this area to identify the potential areas and interfaces where 3GPP can provide relevant specifications to support 3D services in the context of 3GPP services. This technical report should be considered as the basis for such future work on 3D.

Annex A: Assumptions for Simulation Method for Solutions on MBMS Services

This Annex A presents assumptions for simulation study for solutions within MBMS services (i.e. MBSFN). The information in the present document has been collected with the best knowledge that was available at the time when the present document was produced and may not necessarily represent a realistic MBMS deployment. It is up to the reader of the TR to identify if the parameters in this Annex are relevant for their use. Note that the packet loss pattern proposed in this Annex A is time uncorrelated model.

The cell layouts frequently found in performance studies in RAN working groups are similar as Figure A.1. These layouts are composed of 19 cells of which each cell consists of 3 sectors. Therefore, total number of sectors is 57.

Figure A.1 shows 4 cases of MBSFN sector deployments over 57 sectors. The sectors of MBSFN transmission mode are synchronized in transmission time, frequency band, modulation and channel coding rate. The effect of synchronized MBSFN transmission is increased spectral efficiency. Therefore UEs surrounded by MBSFN cells achieve good signal quality as the size of MBSFN area becomes large. Other surrounding sectors are all interference sectors.

In Figure A.1, MBSFN participating sectors are increased from single sector (1/57 case), 7 sectors which is a formation of a centre sector surrounded by a ring of MBSFN cooperating sectors (7/57 case), 19 sectors (19/57 case) and 37 sectors (37/57 case).

The performance metric measured in this layout is coverage versus BLER. The "coverage" denotes normalized ratio of measured area to the size of entire MBSFN area (i.e. total size of MBSFN sectors). Therefore, 50% coverage in single sector deployment usually means only half area of a sector size. However, 50% coverage in an area consists of 37 MBSFN sectors may encompass the area of 7 sectors. The signal strength degrades gradually from centre of the MBSFN area to the edge because the interference from surrounding cells is increased. Therefore BLER (Block Error Rate) is generally increased as the coverage is increased. Figures in A.1 show scatter graphs of BLER level in different MBSFN layouts and channels. In the figures, it is illustrated that 64 QAM signal of 10% loss rate (purple dots) may only cover less than 20% area in single sector layout, however in 7 sector layout, the coverage of 10% loss rate increases to 45%, and it becomes 65% in 19 sectors layout, 75% in 37 sectors layout. The red dots' area is high-loss rate area due to strong interference. BLER figures of 16QAM and QPSK channel in the case of 19 sector layout are also described.

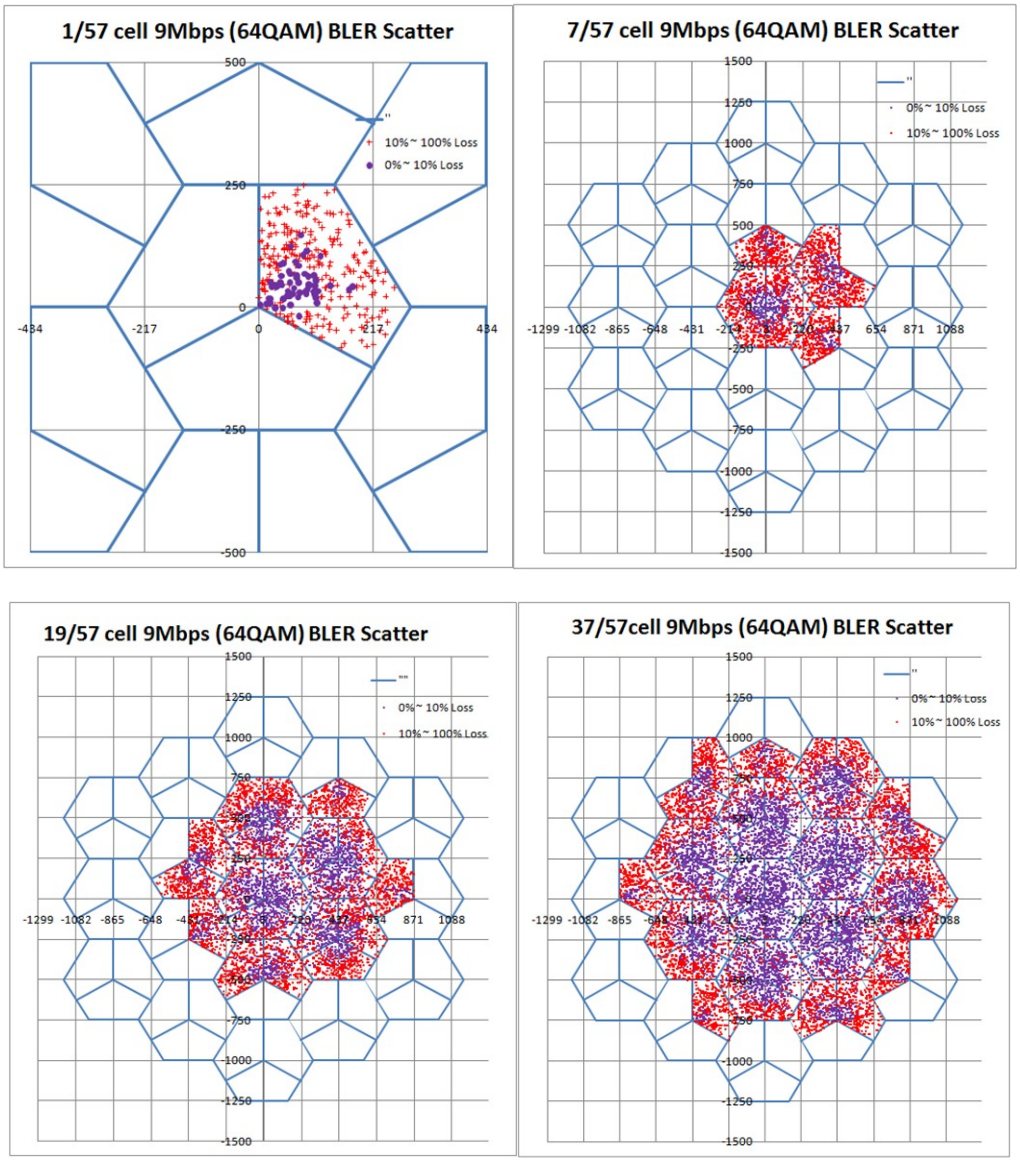

Figure A.1: MBSFN layouts composed of 1, 7, 19, 37 sectors in 57 sector area

Table A.1 is the configuration for channel level simulation. These are also generally accepted assumptions in RAN WG1 documents.

Table A.1: Simulation Configuration

Parameter	Value
Number of Cells	19 cell wraparound layout (3 sectors each)
The number of MBSFN cooperation cells	1, 7 19, 37
Interference	2 tier interfering cells except MBSFN cells
Number of users per cell	10
Bandwidth	5 *MHz*
Number of Rx Antennas	2
Number of Tx Antennas	1
TTI	1 *ms*
FFT Size	512
Number of guard carriers	212
Number of pilot sub-carriers per symbol	50
Number of data sub-carriers per symbol	250
Number of OFDM symbols per TTI	12
Cyclic prefix	128 (16.6 μs)
BS power	43 *dBm*
MCS	QPSK 1/6, 1/2 16QAM 1/2 64QAM 1/2, 4/5
Channel estimation loss	1 *dB*
Channel Model	SCM – urban macro 8 degree
ISD	500m, 1732m
Link-to-System Mapping	Constrained Capacity Effective SNR

Two types of cell density models are considered. The urban macro dense deployment model uses inter-site distance (ISD) 500m, and the sparse model uses *ISD = 1732m*. The pedestrian mobility speed of UE is limited to *3km/hr*.

There are 4 combinations of channel modulation and coding schemes (MCS) tested to generate the BLER trace. Table A.2 summarizes the MCS settings, information data rates (i.e. channel throughput) available to application layer and physical block size. Note that a physical block in LTE channel corresponds to subframe of *1 msec*. Therefore the size of block may range from *125 bytes/block* to *1125 bytes/block* respectively to each MCS level. If a block contains corrupted bit, the block is counted as error.

Only the downlink performance is measured and uplink feedback channel is not defined in this broadcast channel model.

Table A.2: MCS levels, data rates and physical block size

MCS	Modulation	Code Rate	Data rate (*Mbps* in 5 MHz)	Block Size (Bytes/BLK)
1	QPSK	1/6	1.0	125
2	QPSK	1/2	3.0	375
3	16QAM	1/2	6.0	750
4	64QAM	1/2	9.0	1125

Figures A.2 to A.6 show the BLER curves of the 4 MCS channels in various cell layouts. Figure A.2 is the BLER curves in single MBSFN sector (*ISD=500m*). The graph shows that almost *90%* of the single sector area can be guaranteed less than *0.1%* of BLER, if MCS-1 channel of 1*Mbps* throughput (i.e. QPSK and 1/6 rate coding) is used for application. If one wants to increase the channel throughput to 3*Mbps* (i.e. QPSK and 1/2 rate coding), the coverage

drops to 65%. The highest throughput channel of 9*Mbps* (i.e. 64QAM and 1/2 rate coding) may only cover 10% area if BLER is less than 0.10%.

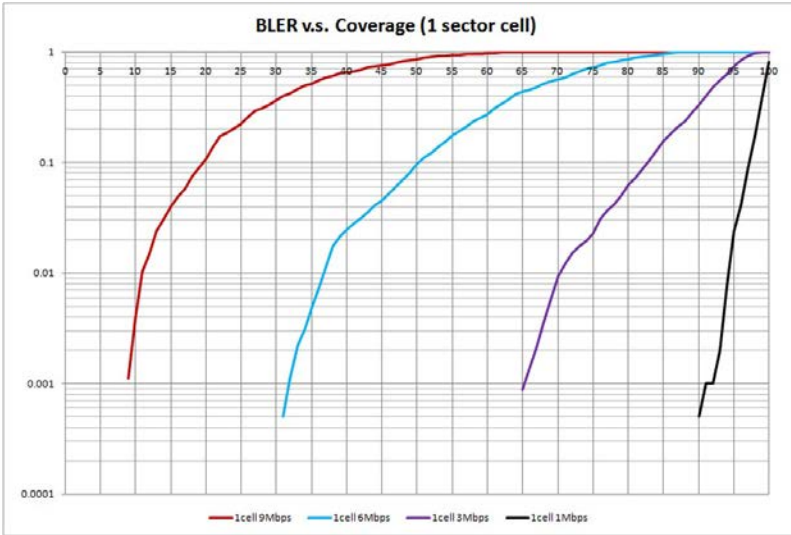

Figure A.2: BLER in Single Sector (ISD=500m)

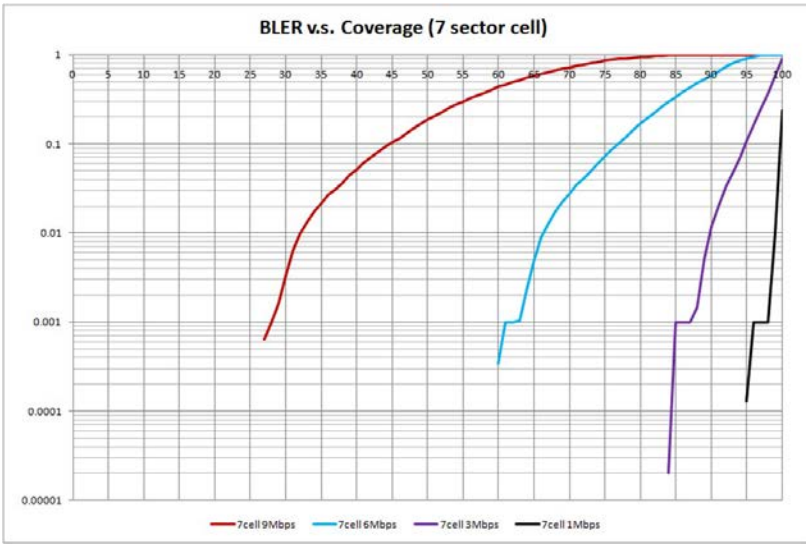

Figure A.3: BLER in 7 Sector (ISD=500m)

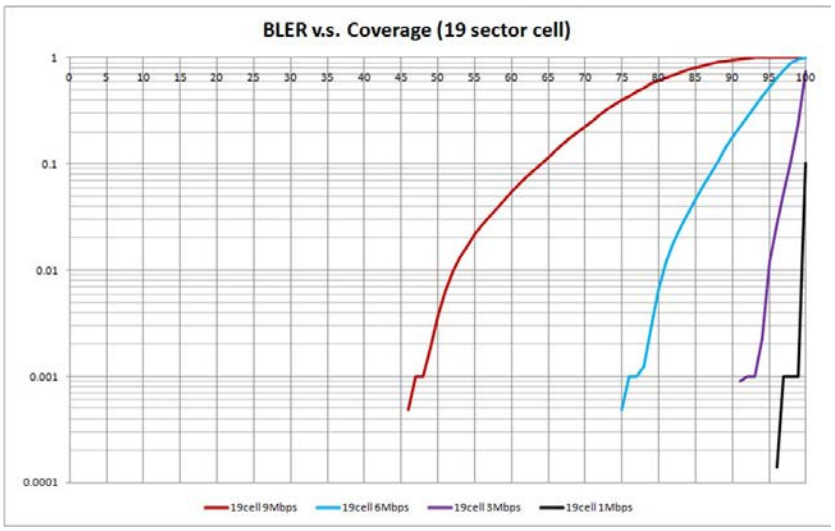

Figure A.4: BLER in 19 Sector (ISD=500m)

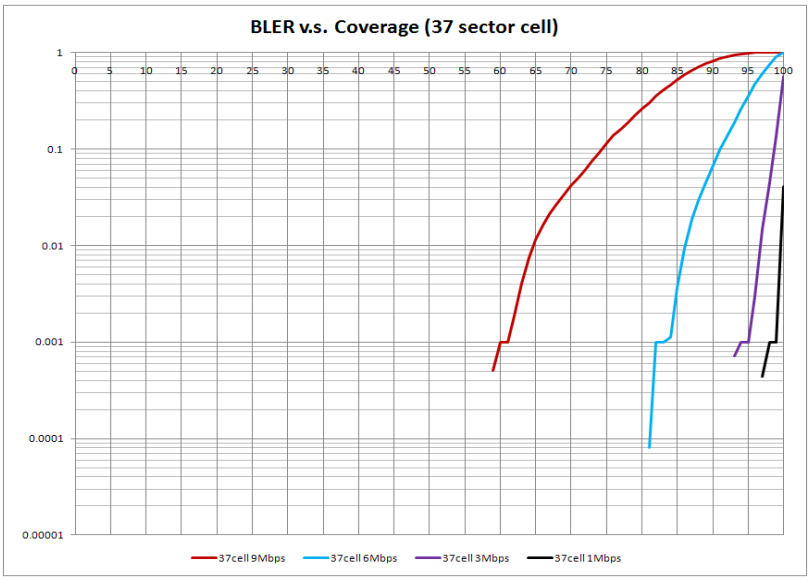

Figure A.5: BLER in 37 Sector (ISD=500m)

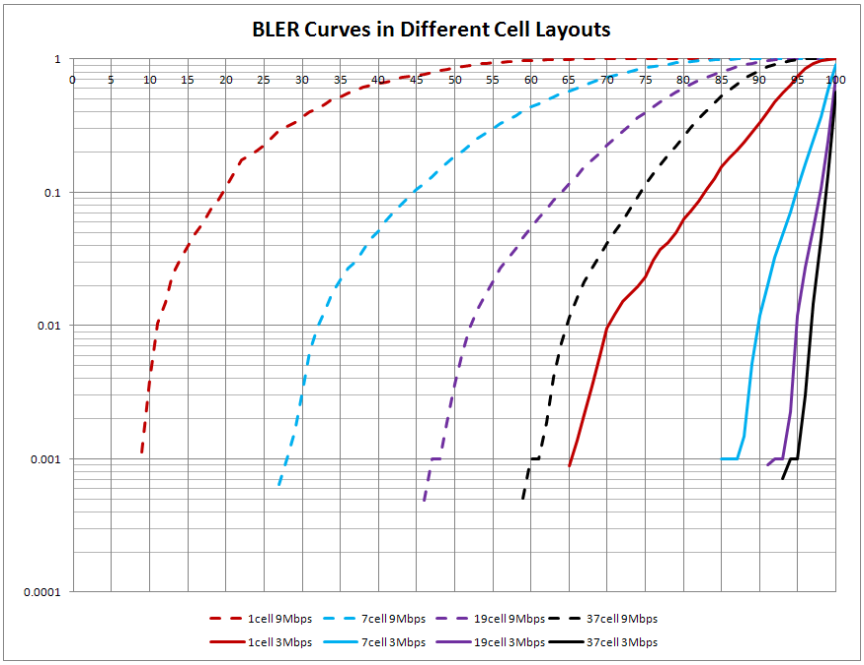

Figure A.6: BLER Curves in Different Layouts (9*Mbps* and 3*Mbps*)

The presented results so far within Annex A assumed that the MBSFN area is equal to the MBMS service reception area. An alternative for the simulation setup include an increase of the MBSFN area beyond the intended MBMS service reception area. The MBSFN area can be made equal to the size of the intended reception area plus one or more ring of cells.

Figure A.7 shows the simulation setup with an MBSFN area size of 19 cells surrounded by interfering cells. For the moment we assume that the surrounding cells transmit unicast data. 3 reference circles close to the border of the MBSFN area are also shown in Figure A.7. Figure A.8 shows the scatter plot of SINR (*dB*-averaged over frequency domain) from the simulation scenario in Figure A.7. The radius of the 3 reference circles is shown as vertical lines. It can be seen that below 500m distance the mean of the SINR distribution versus the distance is quite constant. At a distance larger than 500m a strong drop of the SINR is noted. This strong drop can be avoided if the MBSFN area is extended beyond the service reception area resulting in a more uniform SINR within the reception area. In order to get similar simulation results, locations in the border cells of the MBSFN area may be excluded from the evaluation of reception locations.

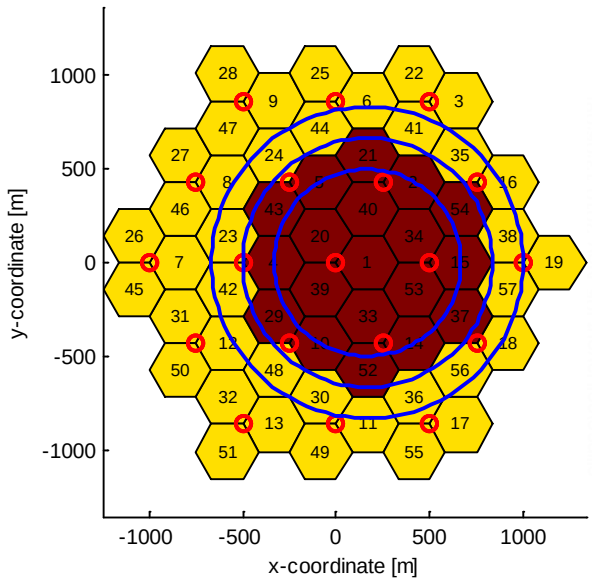

Figure A.7: Simulation scenario: 19 cells in MBSFN area

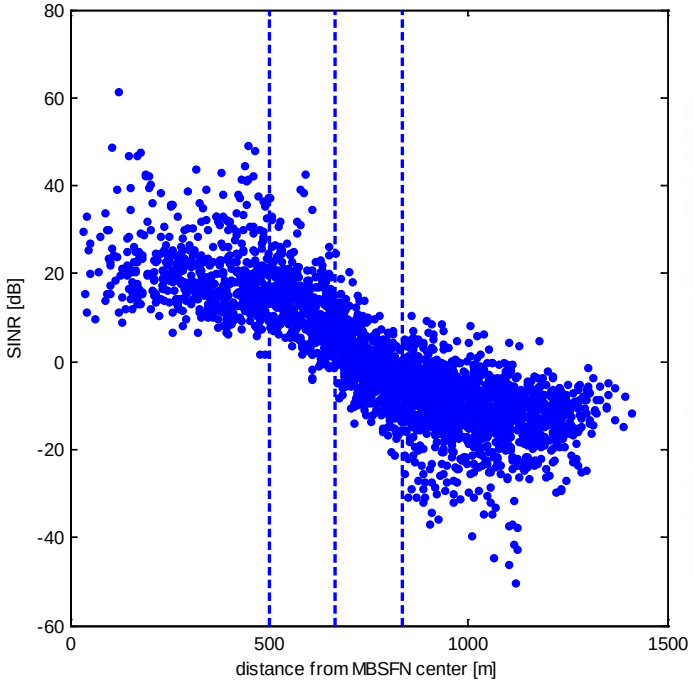

Figure A.8: SINR versus distance from MBSFN center; interfering unicast cells

In another simulation setting the cells surrounding the considered MBSFN reception area are assumed to belong to another MBSFN. In this case, techniques such as interference rejection combining (IRC) in the UE are more efficient, because the signals from all the cells of the adjacent MBSFN area coherently aggregate (as long as they arrive within the cyclic prefix) and thereby the adjacent MBSFN area is seen as one single large interfering cell. IRC is most efficient in this case of a dominant single interferer. Figure A.9 shows the SINR results for this scenario. Compared with Figure A.8 the SINR is significant higher. Therefore, the assumptions of simulations should distinguish whether the cells outside of the considered MBSFN area transmit unicast data or belong to another MBSFN area.

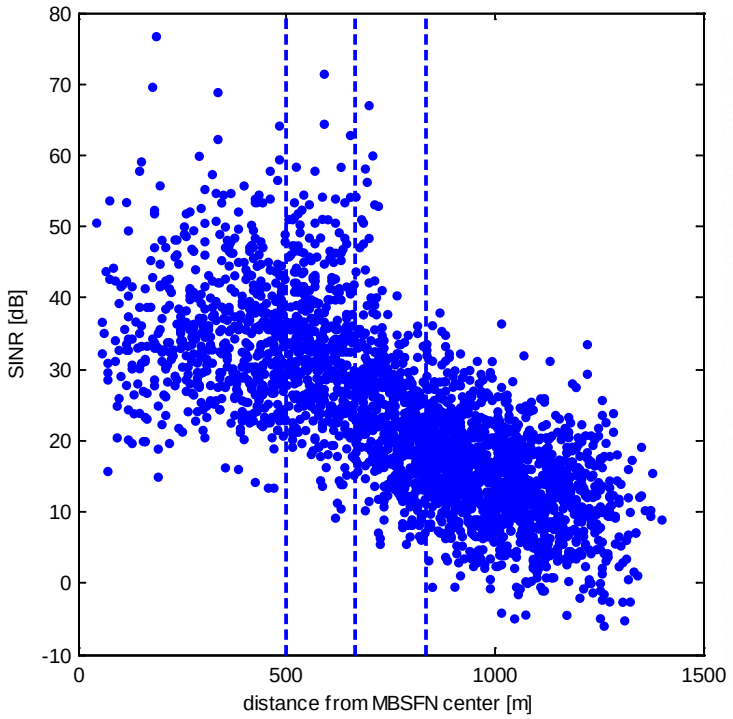

Figure A.9: SINR versus distance from MBSFN center; interfering cells from a second MBSFN

Annex B:
Impact of Screen Size on Stereoscopic Video

B.1 Geometry of Stereoscopic Video

In case of stereoscopic video, two separate images (one for the left eye and one for the right eye) are provided. Figure B.1 shows the basic geometry in case of stereoscopic video and how the depth of an object is perceived.

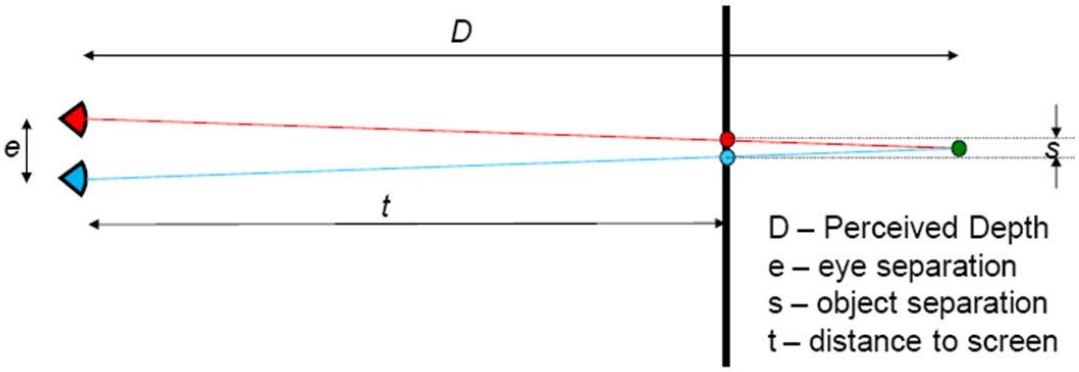

Figure B.1: Geometry of stereoscopic video

Using simple geometry, the following formula using the symbols from Figure B.1 is derived

$$D = \frac{et}{e - s} \qquad (.B.1)$$

The formula shows that the perceived depth is inversely proportional to the separation of the renditions. Scaling the separation (as in the case of a display size reduction) does not have a linear impact on the depth at which the object is perceived.

B.2 Depth Range

Figure B.2 shows the perceived depth placement for several objects based on the separation of the left and right rendition of the objects. Each object is also labelled with the separation of its left and right renditions on the screen as a multiple of eye separation. From (***Eq.*** B.1), it can be noted that an object with rendition separation ***s=e*** would appear at infinite depth.

Typically, objects in a scene appear in a certain depth range. Placing objects at the border or outside the depth range will limit the comfort of the viewing experience.

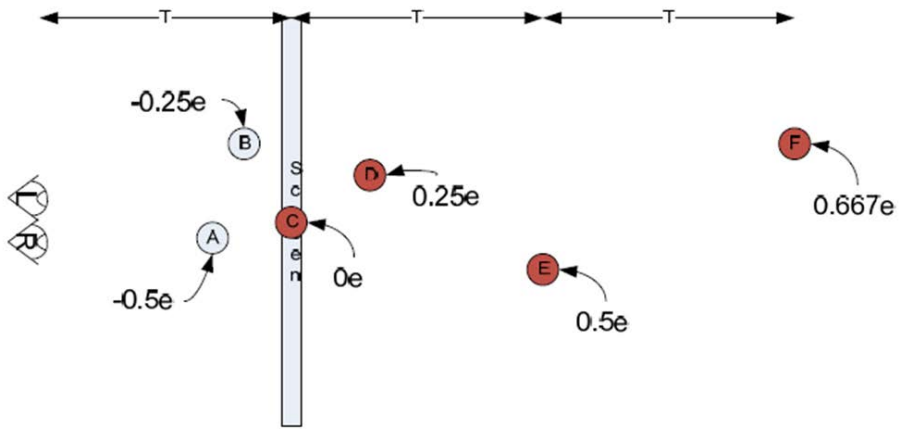

Figure B.2: Depth placements of objects and the related left-right separations (from [23])

B.3 Effect of Display Size Changes

Now, the stereoscopic image pair is displayed on a screen with a reduced size. Since the screen size scales, the separations of the object renditions scale, and so the perceived depth of the objects.

Figure B.3 shows the effect on the depth range when the objects are displayed on a screen half of the size, showing the objects at their newly labelled positions. One result of this scaling is that objects that were previously at infinity are moved to a depth of $2t_2$ (this is shown as ∞ in Figure B.3), where t_2 is the viewing distance from the scaled screen.

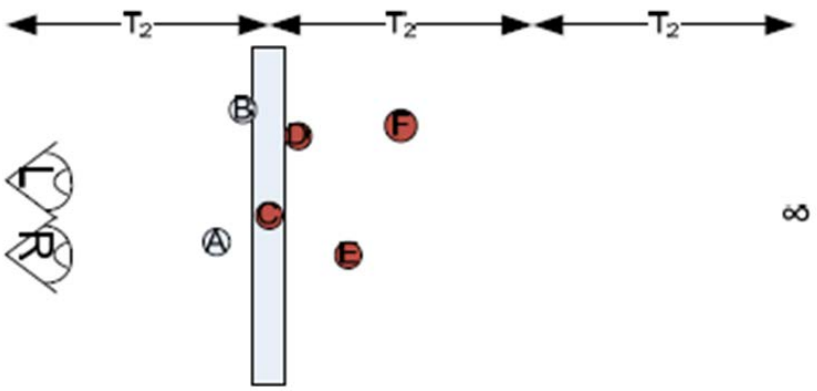

Figure B.3: Scene from Figure B.2, as perceived on a screen of half size (from [23])

As seen from Figure 3 reducing the screen size compresses the depth range. The perceived depths for the original and half screen sizes are listed in Table B.1.

It should be noted that due to the reduced screen size the viewing distance t_2 is significantly smaller than the viewing distance t_1.

Table B.1: Variation in perceived depth as screen size scales

Object	Separation	Depth	
		Original size	Half size
A	-e/2	$2t_1/3$	$4t_2/5$
B	-e/4	$4t_1/5$	$8t_2/9$
C	0	t_1	t_2
D	+e/4	$4t_1/3$	$8t_2/7$
E	+e/2	$2t_1$	$4t_2/3$
F	+2e/3	$3t_1$	$3t_2/2$
∞	e	∞	$2t_2$

B.4 Discussion

Looking at the results from Figure B.3 and Table B.1 the following effects are observed when the screen size is reduced:

- All objects seem to move towards the screen.
- Distant objects move closer (e.g. from ∞ to 2t2)
- Depth range is non-linear compressed
- An object travelling with constant speed towards the viewer seems to change speed (accelerating when approaching screen, slowing down when moving away from screen)
- Impossible to place objects at infinity distance
- "Close" objects move even closer to the viewer (and also to the screen)
- Objects may move out of stereoscopic comfort zone.

Taken these effects into account, corrections of object separations may be necessary to improve the perceived depth of stereoscopic video content, if content that is produced for a larger screen is displayed on the 3D display of a mobile device. Proposals for such depth correction can be found in [23].

Annex C:
Real World Statistics of VoD User Request

This annex describes real world statics of VoD service used in the evaluation of this TR. The statistics have been measured within the time period of one month. The provided VoD service offers a wide variety of movies of more than 5000 files among which the users can make their selection from. In these statistics an average of about 3400 requests per day is reported. Further statistics on the data are given in Figure C.1, Figure C.2, Figure C.3, and Table C.1.

Figure C.1 shows the number of requests issued by all users for all files grouped per hour. A clear diurnal and weekly pattern can be observed. Day 1, 8, 15, 22 and 29 seem to have the largest peaks and these days were identified as Saturdays.

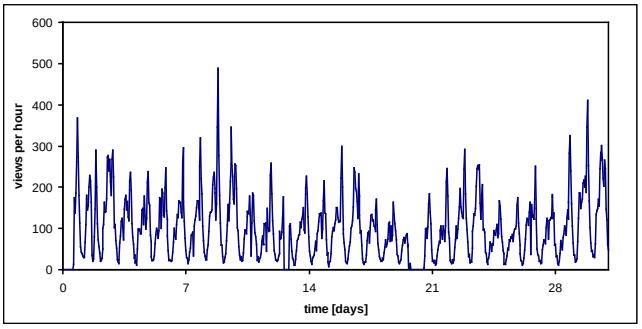

Figure C.1: Requests statistics

Figure C.2 shows the average over 30 days of the diurnal pattern. That is, the evolution over each individual day was cut out of the evolution shown in Figure C.1 and these 30 curves were averaged. It can be seen that the peak demand occurs at 8pm.

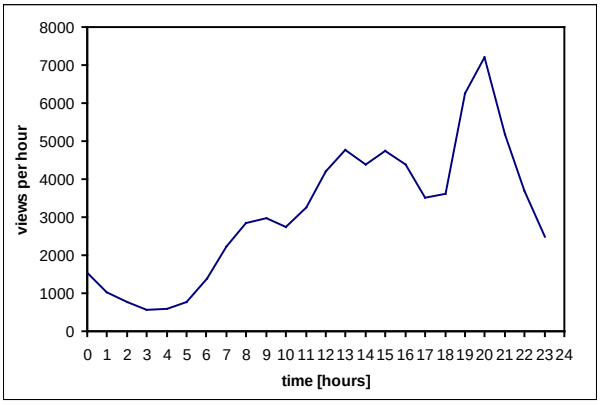

Figure C.2: Number of requests per hour averaged over 30 days

Figure C.3 shows the popularity evolution for the 10 most popular multimedia objects. The number of requests for a particular media object is accumulated over a day (so that diurnal effects cannot be seen). The weekly patterns can be observed with peaks on the Saturdays and although not very prominent some multimedia objects expose an aging effect, i.e., as time goes by the interest in them decreases.

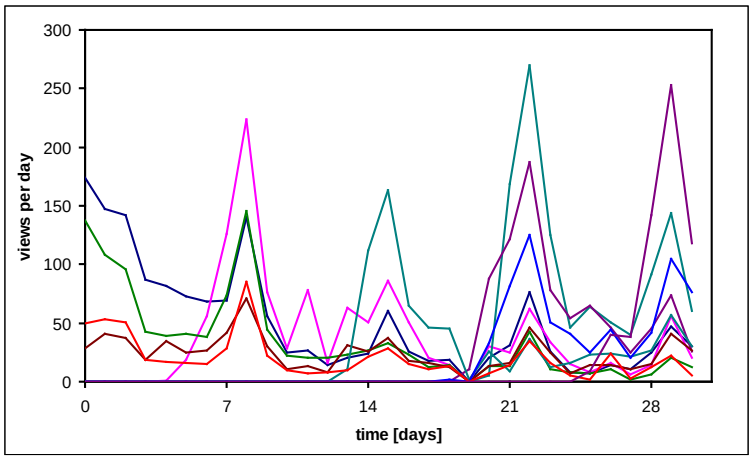

Figure C.3: Number of requests accumulated over one day for the 10 most popular files

Table C.1 shows the distribution of the requests for the films.

Table C.1: *Nb* of requests for the films within the time period of 31 days

Nb of requests	% of nb of films	Cumulative percentage
300 - 691	0,312%	0,312%
200 - 299	0,502%	0,814%
100- 199	1,679%	2,493%
30-99	7,201%	9,694%
1-29	76,355%	86,049%
0	13,952%	100%

Annex D:
Change history

Change history							
Date	**TSG SA#**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2010-12	50	SP-100786			Presented for information at TSG SA#50		1.0.0
2011-03	51	SP-110048			Presented for approval at TSG SA#51	1.0.0	2.0.0
					Approved at TSG SA#51	2.0.0	10.0.0
2012-09	57				Version for Release 11	10.0.0	11.0.0