ETSI TR 103 343 V1.1.1 (2015-12)

PowerLine Telecommunications (PLT);
Powerline HDMI® analysis for very short range link
HD and UHD applications

Reference
DTR/PLT-00044

Keywords
powerline, video

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Powerline Telecommunications (PLT).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

This Technical Report investigates how to transmit over the powerline medium HD and UHD contents that typically are exchanged through the HDMI® cable between transmitter video sources like Blu-ray™ players or set-top boxes and receiver video sinks like video displays. The report also presents earlier findings. The scope of the Technical Report is providing the technical elements needed to establish a PHDMI specification. The report is structured as follows: in clause 4, the requirements in terms of target bit rate to be fulfilled by a PHDMI technology are presented for different HD and UHD formats together with a set of PLT links used for testing purposes. It has to be underlined that the initial target for phase 1 was only HD. It was however estimated that the market is rapidly moving towards UHD and it was hence decided to enlarge the scope to also cover UHD. In clause 5, some target use cases are described: these scenarios highlight the potential fields of application of a PHDMI technology. Clause 6 presents the schemes that have been scrutinized as potential PHDMI technologies. Two of them are tandem schemes, i.e. systems that separate the channel encoding part from the source encoding part, they are based upon the serial concatenation of a compression encoder (based upon JPEG 2000 or Dirac) specifications and a OFDM power line modem operating in the 2 MHz to 100 MHz band that is able to provide SISO-based or MIMO-based communication. Besides tandem scheme, a joint scheme relying on the SoftCast paradigm has been also considered: it accommodates joint source and channel encoding. These communication schemes have been tested on HD and UHD videos both on flat channel with AWGN and on realistic PLT links (both SISO, MIMO 2×2 and MIMO 2×3): results are reported in terms of video quality metrics in clause 7 and clause 8. In particular, it is worth noticing that a realistic long video sequence was furnished by France Télévision for this analysis. Clause 9 shows how to optimize the source encoder parameters for the tandem schemes: an interesting point that it is also evaluated in this clause is the resiliency to errors at the PLT level, i.e. the investigation to see if it is possible to tolerate some errors at the PLT level without requiring retransmission of wrong packets. Clause 10 presents a scheme of a rate controller: it is the PHDMI component that manages the compression encoder rate as a function of eventual changes of the PLT rate during the video transmission. Transmit and receive buffer requirements are also put in evidence. Conclusion and final recommendations are reported in clause 11 of the present document.

NOTE: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

1 Scope

The present document addresses Short Range Powerline modems for Very High Bit Rate links for both HDMI® 1.x and HDMI® 2.0 interfaces.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

Not applicable.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] "High-Definition Multimedia Interface Specification Version 1.4b", October 2011.

[i.2] Consumer Electronic Association®: "A DTV profile for uncompressed high speed digital interfaces", CEA-861-D, July 2006.

[i.3] ETSI TR 101 562-3: "PowerLine Telecommunications (PLT); MIMO PLT; Part 3: setup and statistical results of MIMO PLT channel and noise measurements".

[i.4] L. Yonge, J. Abad, K. Afkhamie, et al.: "An overview of the HomePlug AV2 technology", Journal of Electrical and Computer Engineering, volume 2013, Article ID 892628, 20 pages, 2013. Doi:10.1155/2013/892628.

[i.5] H. Chaouche, F. Gauthier, A. Zeddam, M. Tlich and M. Machmoum: "Time domain modeling of powerline impulsive noise at its source", Journal of Electromagnetic Analysis and Applications, 3(9):9, 2011.

[i.6] J.A. Corteés, L. Diéz, F.J. Cañete, and J.J. Sanchez-Martinez: "Analysis of the indoor broadband power-line noise scenario", IEEE Transactions on Electromagnetic Compatibility, 52(4):849-858, November 2010.

[i.7] V. Degardin, M. Lienard, A. Zeddam, F. Gauthier and P. Degauque: "Classification and characterization of impulsive noise on indoor powerline used for data communications", IEEE Transactions on Consumer Electronics, 48(4):913-918, November 2002.

[i.8] D. Umehara, S. Hirata, S. Denno and Y. Morihiro: "Modeling of impulse noise for indoor broadband power line communications", in Proc. International Symposium on Information Theory and its Applications, ISITA2006, 2006.

[i.9] D. Veronesi, R. Riva, P. Bisaglia, F. Osnato, K. Afkhamie, A. Nayagam, D. Rende and L. Yonge: "Characterization of in-home MIMO power line channels", IEEE International Symposium on Power Line Communications and Its Applications (ISPLC), pp.42-47, April 2011.

[i.10] A. Tomasoni, R.Riva and S.Bellini: "Spatial correlation analysis and model for in-home MIMO power line channels", IEEE International Symposium on Power Line Communications and Its Applications (ISPLC), pp.286-291, 2012.

[i.11] D. Rende, A. Nayagam, K. Afkhamie, L. Yonge, R. Riva, D. Veronesi, F. Osnato and P. Bisaglia: "Noise correlation and its effect on capacity of inhome MIMO power line channels", IEEE International Symposium on Power Line Communications and Its Applications (ISPLC), pp.60-65, 2011.

[i.12] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies: "Image coding using wavelet transform", IEEE Transactions on Image Processing, volume 1, pp.205-220, April 1992.

[i.13] J. Shapiro: "Embedded image coding using zerotrees of wavelet coefficients", IEEE Transactions on Signal Processing, volume 41, pp. 3445-3462, December 1993.

[i.14] A. Said and W. Pearlman: "A new, fast, and efficient image codec based on set partitioning in hierarchical trees", IEEE Transactions on Circuit and Systems for Video Technology, volume 6, pp.243-250, June 1996.

[i.15] N. Adami, A. Signoroni and R. Leonardi: "State-of-the-art and trends in scalable video compression with wavelet-based approaches", IEEE Transactions on Circuit and Systems for Video Technology, volume 17, number 9, pp.1238-1255, 2007.

[i.16] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps" Journal of Fourier Analysis and Applications, vol.4, no.3, pp.247-269, 1998.

[i.17] D. Taubman and A. Zakhor: "Multirate 3-d subband coding of video", IEEE Transactions on Image Processing, volume 3, pp.572-588, September 1994.

[i.18] J.-R. Ohm: "Three-dimensional subband coding with motion compensation", IEEE Transactions on Image Processing, volume 3, pp.559-571, September 1994.

[i.19] S.-J. Choi and J. Woods: "Motion-compensated 3-D subband coding of video", IEEE Transactions on Image Processing, volume 8, pp.155-167, February 1999.

[i.20] B. Pesquet-Popescu and V. Bottreau: "Three-dimensional lifting schemes for motion compensated video compression", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), volume 3, pp.1793-1796, 2001.

[i.21] S.-T. Hsiang and J. Woods: "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling", IEEE International Symposium on Circuits and Systems, volume 3, pp.662-665, 2000.

[i.22] D. Taubman: "High performance scalable image compression with EBCOT", IEEE Transactions on Image Processing, volume 9, pp.1158-1170, 2000.

[i.23] M. Kieffer and P. Duhamel: "Joint source-channel coding and decoding", 2014.

3 Symbols and abbreviations

3.2 Symbols

For the purposes of the present document, the following symbols apply:

b	bit
dB	decibel
G	Giga
Hz	Hertz
M	Mega
s	second
2D	Bi-dimensional
3D	Three-dimensional
64K	65536

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AVC	Advanced Video Coding
AWGN	Additive White Gaussian Noise
BBC	British Broadcasting Corporation
BCJR	Bahl Cocke Jelinek and Raviv
BER	Bit Error Rate
CE	Consumer Electronics
CEA	Consumer Electronics Association®
CR	Compression Ratio
DC	Direct Current
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding with Optimized Truncation
EMI	Electro Magnetic Interference
EZBC	Embedded Zero-Block Coding
EZW	Embedded Zero-Tree Wavelet coding
FEC	Forward Error Correction
GoP	Group of Pictures
HD	High Definition
HDMI®	HD Multimedia Interface
HDTV	High Definition TeleVision
HEVC	High Efficiency Video Coding
ICT	Irreversible Component Transformation
ITU	International Telecommunication Union
L1	Level 1
L2	Level 2
LCD	Liquid Crystal Display
LLSE	Linear Least Square Error
LS	Lifting Schemes
MAC	Medium Access Control
MC	Motion Compensation
MCTF	Motion Compensated Temporal Filtering
MIMO	Multiple Input Multiple Output
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
MV	Motion Vector
NAL	Network Abstraction Layer
OFDM	Orthogonal Frequency Division Multiplexing
PE	Protective Earth
PHDMI	Powerline HDMI
PHY	PHYsical

PLT	PowerLine Telecommunications
PSD	Power Spectral Density
PSNR	Peak SNR
QAM	Quadrature Amplitude Modulation
QF	Quality Factor
QoS	Quality of Service
RCT	Reversible Component Transformation
RDO	Rate Distortion Optimization
SD	Standard Definition
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SPHIT	Set Partitioning in HIerarchical Trees
SSIM	Structural SIMilarity
SVC	Scalable Video Coding
TDMA	Time Division Multiple Access
UHD	Ultra HD
UHDTV	Ultra High Definition TeleVision
VLC	Variable Length Code
VS	Video Source
WSVC	Wavelet-based Scalable Video Coding
WT	Wavelet Transform

4 PLT HDMI® bit rate targets

4.1 Introduction

The recent increase in HD video contents has brought the need to develop communication standards capable of multi-gigabit per second throughput, like HDMI® and Display Port. Consumer electronics (CE) users also want the flexibility provided by wireless connections to set up and reconfigure multimedia systems, and to eliminate wired connections required by HD multimedia systems, like home theatres.

Driven by these needs, the CE industry is developing formats capable of delivering uncompressed video, at the necessary data rates, via wireless and wireline connections.

Simultaneously, PLT devices have become widespread and, in the latest specifications, are capable of data rates at the physical (PHY) layer up to 1 Gbit/s for single input single output (SISO) implementations and up to 2 Gbit/s for multiple input multiple output (MIMO) implementations.

These data rates are obtained for optimum channel conditions : taking into account PHY and medium access (MAC) layer PLT overheads, it is clear that they are insufficient for streaming uncompressed HD video or for transferring HD contents, like a HD film, as fast as would be desirable.

Uncompressed HD video transmission requires very high bit rates, up to 2 to 4 Gbit/s for Full HD video.

Uncompressed HD video transmission avoids compression at the transmitter and decompression at the receiver, therefore providing:

a) lower latency which permits timing sensitive applications like multimedia applications and gaming;

b) higher interoperability between devices because, unlike compressed video transmission, the receiver device just displays the video content and does not need to be able to decode the video codec;

c) no degradation in picture quality due to compression losses in the transmission.

The technologies specifically designed for video transmission organize the source devices (transmitters) and the sink devices (receivers) into a short range powerline video network, that allows for example:

- Point to point uncompressed video transmission ;
- Point to multi point uncompressed video transmission.

The video network has three type of devices: video sink (HD display), video sources (set top box, Blu-ray™ player) and devices that can perform both tasks (desktop).

The characterization of digital video and audio signals is important to assess the network data rate requirements. A digital video signal data rate is defined by: resolution, i.e. the total number of pixels of each image, normally referred as number of horizontal pixels by vertical pixels on the screen; colour depth, i.e. the number of bits used to represent each of the three colours of a pixel; refresh rate, i.e. number of times per second the image is completely reconstructed on the screen; progressive or interlaced formats, i.e. the way lines of an image are displayed in the refreshing cycles, progressive formats display all the lines on all the refresh cycles, while interlaced ones display even and the odd lines in alternated refresh cycles.

Currently available HD video formats are referred as "720p", "1080i", and "1080p". These terms indicate the number of lines and the display method used. Images used in HD video formats have a 16:9 aspect ratio, resulting in wider images than the conventional 4:3 aspect ratio used in Standard Definition (SD) video. The number of pixels, np, in each HD image can be calculated by:

$$np = (16/9) \times nl \quad (1)$$

where nl is the number of lines, indicated by the video format designation. From this equation, it is possible to calculate that 720p images are formed by 1280 × 720 pixels; 1080i and 1080p images are formed by 1920 × 1080 pixels.

The bit rate required to transmit "Full HD" video, vbr, with progressive display can be calculated from:

$$vbr = np \times ncchannels \times cdepth \times rfreq \quad (2)$$

where np is the number of pixels, ncchannels is the number of colour channels, cdepth is the number of bits used to represent each colour and rfreq is the display refresh frequency. Video signals also contain audio information and digital audio data is defined by: number of audio channels (nac), sampling rate(srate) and number of bits used to quantify each audio sample (sdepth). The audio bit rate, abr, can be calculated from these three quantities :

$$abr = nac \times srate \times sdepth \quad (3)$$

Using these relations, it is possible to calculate the required bit rate to transmit an uncompressed video and audio signal. The net bit rate for Full HD video and audio is shown.

$$vbrFullHD = 1920 \times 1080 \times 3 \times 8 \times 60 = 2{,}99 \text{ Gbit/s} \quad (4)$$

$$abrFullHD = 8 \times 192 \text{ k} \times 24 = 36{,}8 \text{ Mbit/s} \quad (5)$$

NOTE: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

4.2 Targets for HD support

In table 1 essential Powerline High-Definition Multimedia Interface (PHDMI) video formats for HD video delivery are reported (information video bit rates 1,1 Gbit/s to 3 Gbit/s). They have been chosen after investigating both the HDMI® specification [i.1] and available datasheets of HD sources like Blu-ray™ players or set top boxes. In table 2 other formats that are typically also supported by the aforementioned applications or similar are reported (information video bit rates 0,4 Gbit/s to 0,5 Gbit/s). Clearly, there are also many other video formats that an HD capable PHDMI technology could support, the scope of table 1 and table 2 only being indicating the most important ones.

NOTE: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

Table 1: Essential video formats for HD support

Video format @60 Hz	720p	1080i	1080p
Video format @50Hz	720p	1080i	
Video format @24 Hz			1080p

Table 2: Important video formats for HD support

Video format @60 Hz	720x480p	640x480p
Video format @50Hz		720x576p

Table 3 reports the target information video bit rate for the video formats reported in table 1 and table 2 considering chrominance subsampling of 4 :4 :4 and 24 bits per pixel. Table 3 also reports, as a reference, the bit rate that HDMI® needs to transport video, audio and control information. The difference among the two bit rates is mainly due to two features :

- HDMI® introduces an 8/10 channel encoding module; and
- HDMI® uses the video blanking zone to transport audio and control data.

For instance, in accordance with CEA-861-D standard [i.2], in the case of 1080p (1920×1080) at 60 fps, HDMI® transmits a 2200×1125 format using the extended 280 45 matrix for audio and control data. Considering that the audio information bit rate that is needed amounts to about 37 Mbit/s, hence it is negligible compared to the video, it results that the total informative content has a required bit rate which is quite similar to the information video bit rate reported in table 3.

Table 3: Target bit rates for HD support

Video format	Frame rate (fps)	≈ Information video bit rate (Gbps)	≈ HDMI® bit rate (Gbps)
1080p	60	2,99	4,46
1080i	60	1,49	2,23
720p	60	1,33	2,23
1080i	50	1,24	2,16
1080p	24	1,19	2,23
720p	50	1,11	2,23
720×480p	60	0,50	0,81
720×576p	50	0,50	0,81
640×480p	60	0,44	0,76

Accordingly, a PHDMI technology has at least two possibilities:

1) The PHDMI video source module passes the audio, video and control informative content to the source power line PHDMI module which includes its signal processing before transmitting it to the power line. At the receiver, the power line PHDMI sink module processes the received information and furnishes it to the PHDMI video sink module (information content approach).

2) The PHDMI video source module passes the HDMI® content (including all the HDMI® overhead) to the power line PHDMI source module which includes its signal processing before transmitting it to the power line. At the receiver, the power line PHDMI module processes the content and furnishes it to the PHDMI video sink which applies HDMI® signal processing to recover video and audio information (*HDMI® content approach*);

Both approaches have advantages and drawbacks.

The information content approach has the following advantages :

1) It allows targeting lower bit rates on the powerline.

2) PHDMI source and sink have only PLT signal processing on the information content. The information content approach has the following drawback:

   - The control part at the sink PHDMI PLT module should be able to present to the PHDMI video sink module the information content in a proper way in order to enable synchronizing the received audio and video content with the sink characteristics.

The HDMI® content approach has the following advantages :

1) Control information is already furnished by HDMI® processing. The HDMI content approach has the following drawbacks : 1) The bit rate to target on the powerline is higher (up to 100 % higher).

2) PHDMI source and sink have both HDMI and PLT processing.

The analysis presented in clause 7 will follow the information content approach, also because the available testing video sequences are in a video format that does not include blanking. Note however that it is evident from table 3 that a PHDMI technology HD capable and able to sustain information video bit rates ≤ 3 Gbit/s will automatically be able to sustain most (all but one) of the HDMI® bit rates. Note also that if the PHDMI technology is UHD capable (see clause 4.3), it can deal with both approaches for HD.

4.3 Targets for UHD support

On 12th of November 2014, CEA® announced that it sets licensing agreement for UHD logos to be used by manufacturers, a sign that the UHD era has started. For the scopes of the present document, the definition reported in table 4 is used:

Table 4: UHD video format

UHD video format	3840×2160p

With UHD it is expected that a variety of combinations of parameter will be adopted in the future, which are more difficult to predict than the more consolidated ones for HD. In table 5 target bit rates for some selected combination parameters are presented for UHD. As in clause 4.2, both the information video bit rate and the HDMI® bit rate needed to transmit video, audio and control data are reported. The HDMI® bit rate is computed taking into account HDMI blanking transmission (4400×2250 extended matrix) and HDMI® 8/10 channel encoding and it is about 50 % higher than the information video bit rate for UHD. As in clause 4.2, a PHDMI technology UHD capable could follow the information content approach or the HDMI® content approach. The advantages and the drawbacks are the same as highlighted in clause 4.2 and they will be not repeated here. Results presented in clause 8 will follow the information content approach. This theoretically will allow also complying with the target of the HDMI® content approach for several combinations of the parameters.

Table 5: UHD target bit rate for different parameter combinations

Video format	Frame rate (fps)	Chrominance subsampling	Bits per pixel	≈Information video bit rate (Gbps)	≈Bit rate HDMI® (Gbps)
UHD	60	4:4:4	24	11,94	17,82
UHD	60	4:2:2	36	11,94	17,82
UHD	60	4:2:2	30	9,95	14,85
UHD	50	4:4:4	24	9,95	14,85
UHD	50	4:2:2	36	9,95	14,85
UHD	60	4:2:0	36	8,96	13,37
UHD	50	4:2:2	30	8,29	12,38
UHD	60	4:2:2	24	7,96	11,88
UHD	60	4:2:0	30	7,46	11,14
UHD	50	4:2:0	36	7,46	11,14
UHD	24	4:4:4	36	7,17	10,69
UHD	50	4:2:0	30	6,22	9,28
UHD	24	4:4:4	30	5,97	8,91
UHD	50	4:2:0	24	4,98	7,43
UHD	24	4:4:4	24	4,78	7,13
UHD	24	4:2:2	24	3,18	4,75

4.4 500 test links and bit rate

Earlier projects have made extensive characterization of in home MIMO PLT links [i.3]. MIMO has been adopted by different standardization organization as a basis for new technologies (HomePlug® AV2 [i.4] by HomePlug® and G.hn MIMO by ITU-T). As demonstrated in clause 4.3, the lowest target PHDMI throughput for UHD is greater than 3 Gbps: current MIMO technologies furnish a maximum throughput lower than 2 Gbit/s. Moreover, it has to be noticed that this maximum throughput is achieved at the PHY layer and in ideal SNR conditions.

According to the use cases (see clause 5), the present work has involved the creation of an internal SNR database on 500 links to be used for testing the different PHDMI techniques through this document. Each link relates to SISO, MIMO 2×2 and MIMO 2×3 on the 2 MHz to 100 MHz band. It should be noted that eigenbeamforming has been used for MIMO. Eigenbeamforming has been selected by the HomePlug® consortium as a basis for the HomePlug® AV2 technology. SNRs for an exemplary link are reported in figure 1.

Result that will be presented in clause 7 and clause 8 have been obtained with very long simulations: for instance the UHD video "FTV" (see clause 8) is more than 8 minutes long and targets an information video bit rate of 5 Gbit/s. It could take more than one week for a test on a single PLT link. Hence, the test links have been accurately chosen in order to provide some variety of performance. One can perceive this variety by looking at figure 2 that shows the throughput that can be obtained at the PLT application level in the case of a tandem scheme (see clause 6.2): it goes from 50 Mbit/s (the worst SISO case) to about 400 Mbit/s (the best MIMO 2×3 case). Note that it is expected that links associated to higher throughputs exist: a choice has been made in order to analyse power line conditions that can allow being confident in the definition of the PHDMI technology.

For tandem schemes (see clause 6.3), one has to properly use the compression capability of the source encoder to pass from the PLT rates shown in figure 1 (or higher) to the bit rate requirements shown in clauses 4.2 and 4.3. For the joint schemes (see clause 6.3), the bit rate requirements of clauses 4.2 and 4.3 are jointly achieved by the compression and channel encoding part. As it will be shown in clause 7 and clause 8, in the present context of video delivery, a comparison between tandem schemes and joint schemes can be done by analysing the displayed video quality.

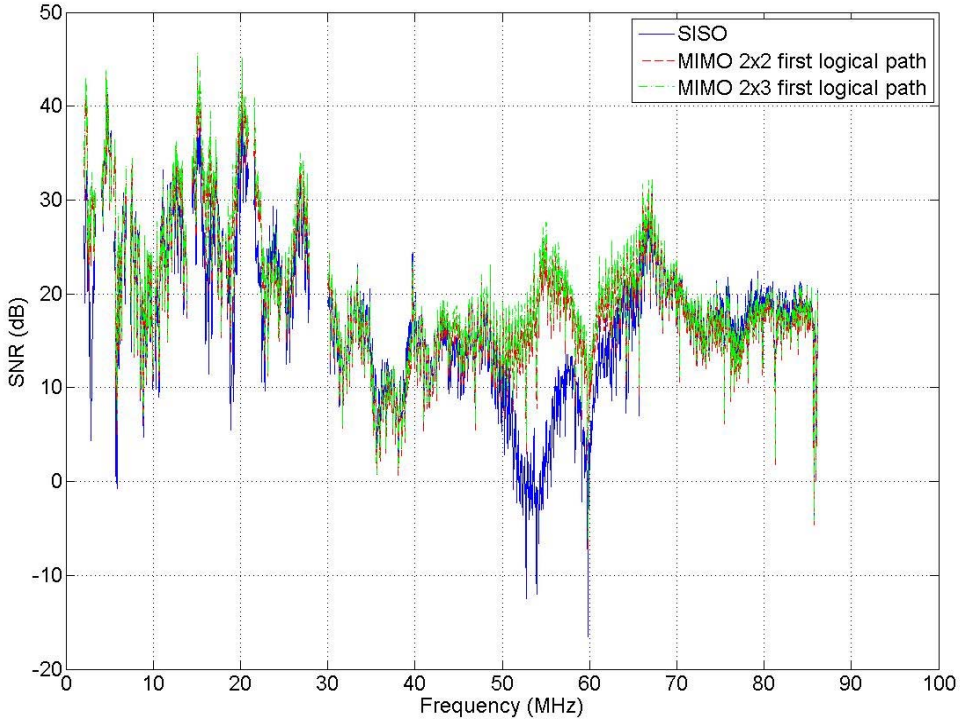


Figure 1: Example of a PLT link SNR extracted SNR database (MIMO reported only on the first logical path)

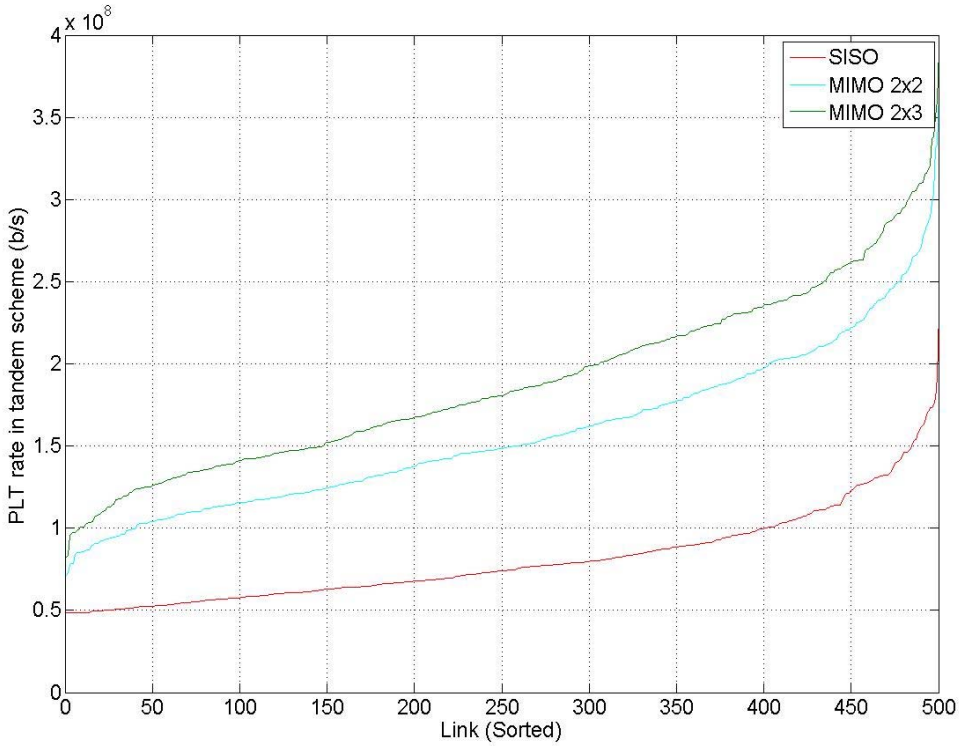


Figure 2: PLT rate in tandem schemes at the application level for the links of the present project

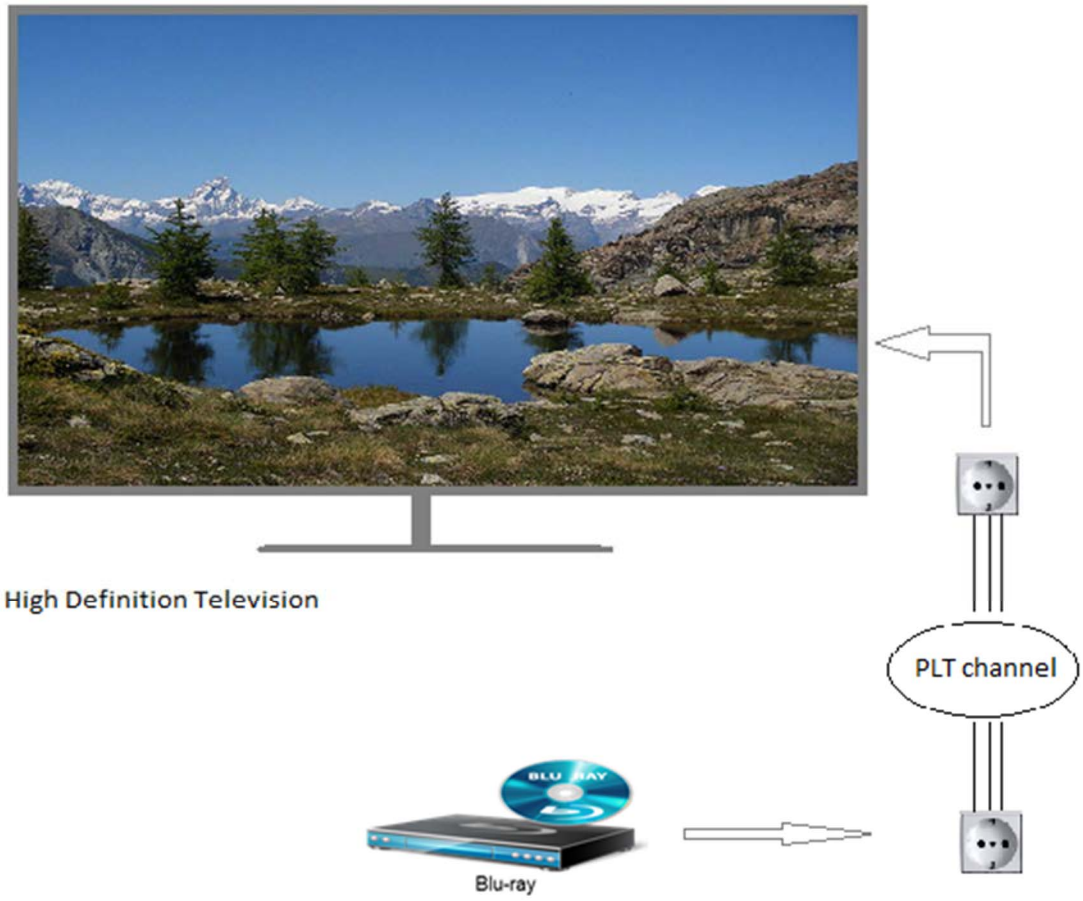
5 Use cases

5.1 Introduction

Clauses 5.2 to 5.7 describe some of the possible use cases that can be addressed with a PHDMI technology. As HD and UHD applications are growing in number, it is obvious that the listed use cases are not exhaustive. However they will be representative of the potential multimedia equipments and scenarios that could benefit from the introduction of a PHDMI technology.

5.2 Use case 1: Blu-ray™ digital television

In this scenario, HD or UHD TV contents are transferred from a Blu-ray™ player to a video screen of a digital television. With PHDMI, a true Plug&Play solution can be enabled.

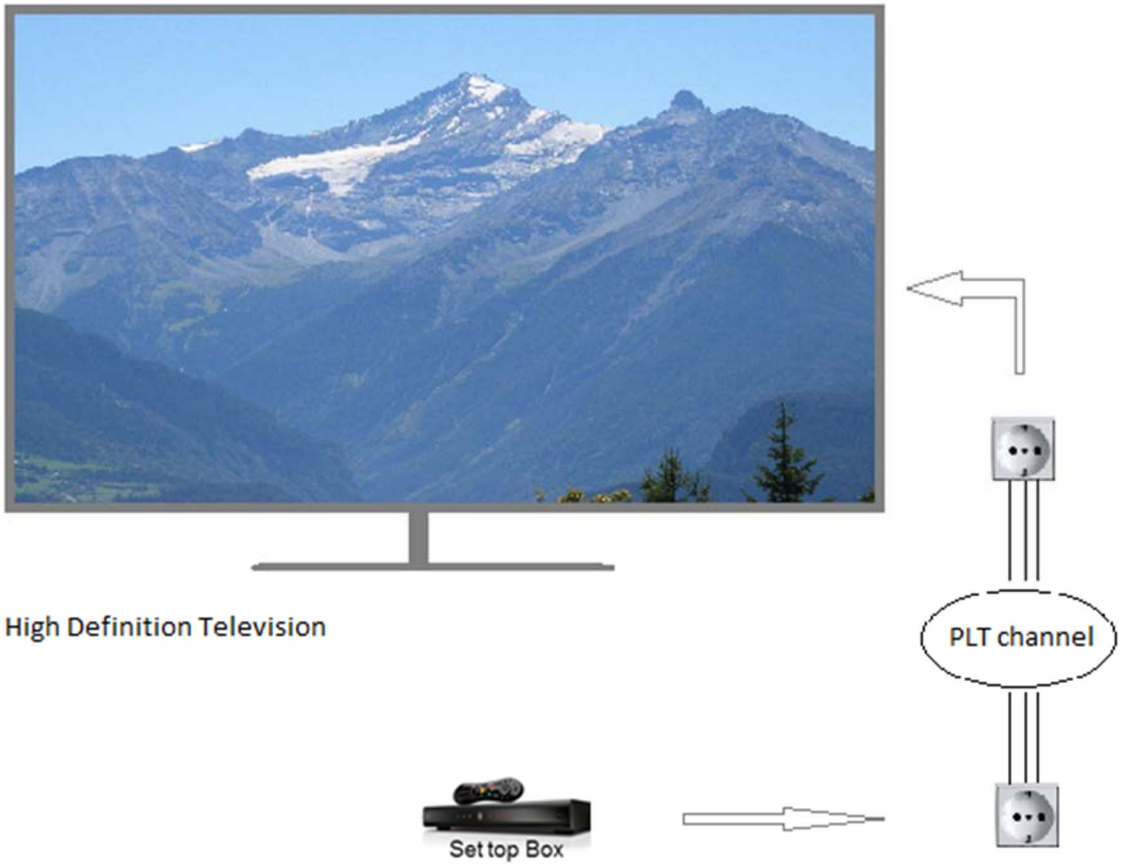


NOTE 1: Source: Blu-ray™ player. Sink: digital television.
NOTE 2: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

Figure 3: PHDMI use case 1

5.3 Use case 2: set-top box-digital television

In this scenario, HDTV or UHDTV contents are transferred from a Set-Top box to a video screen of a digital television.

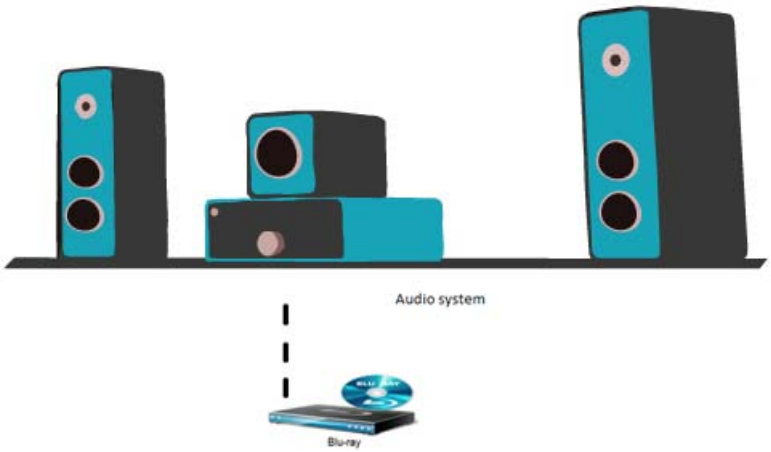


NOTE: Source: set-top box. Sink: digital television.

Figure 4: PHDMI use case 2

5.4 Use case 3: high resolution audio equipment links

In this scenario, high quality audio is transferred from a Blu-ray™ device to an audio amplifier system.

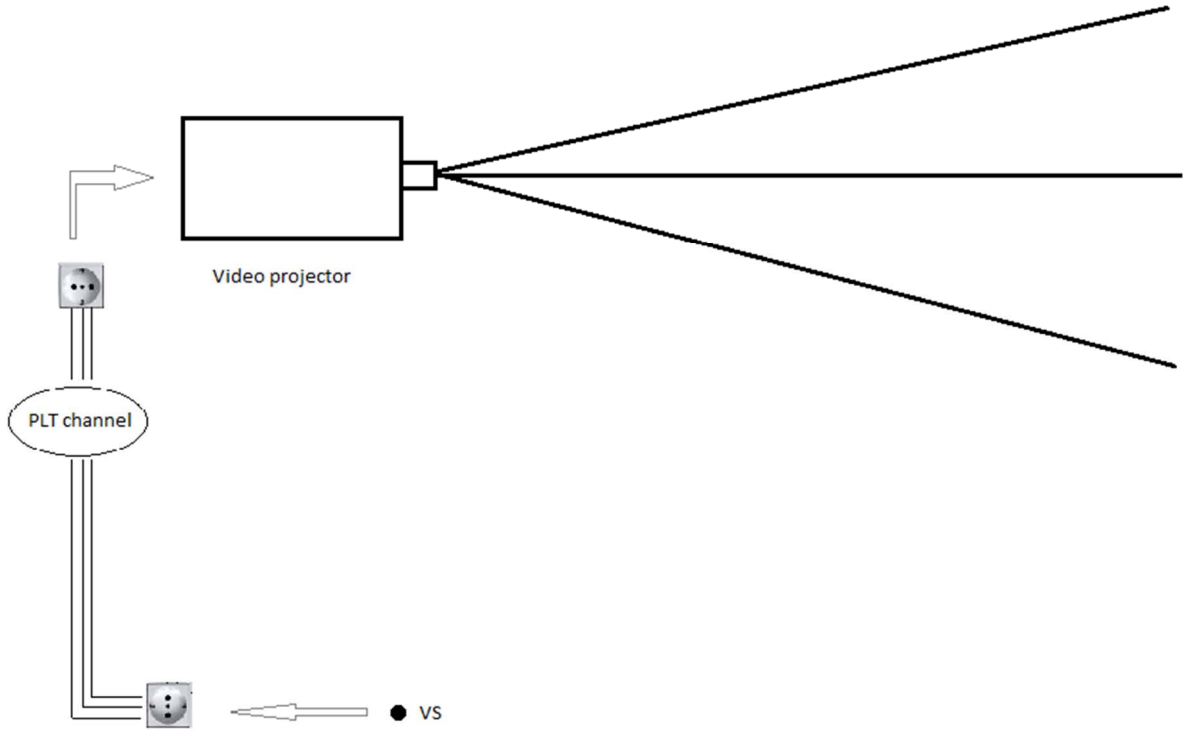


NOTE 1: Source: Blu-ray™ player. Sink: audio system.
NOTE 2: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

Figure 5: PHDMI use case 3

5.5 Use case 4: video source-video projector

In this scenario, a video source transfers audio/video contents to a video projector.

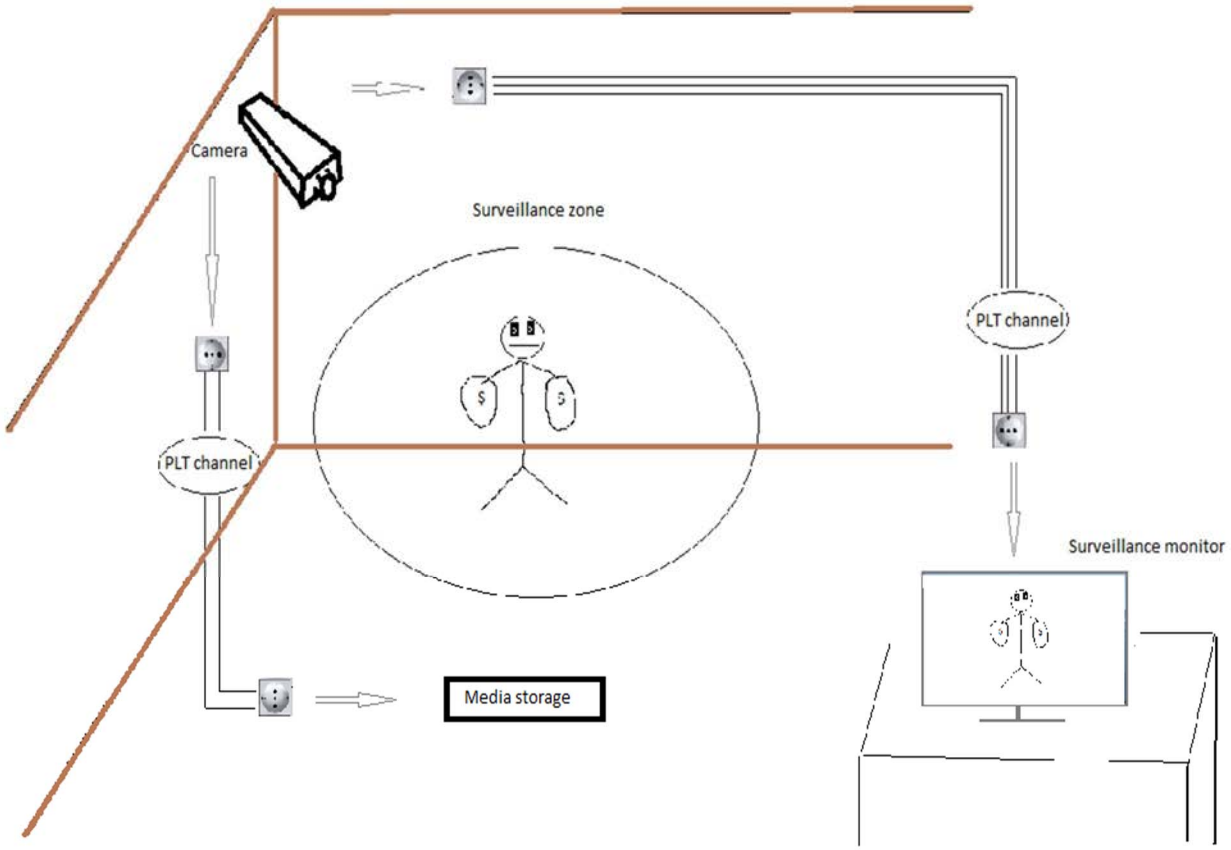


NOTE: Source: a generic video source (VS). Sink: a video projector.

Figure 6: PHDMI use case 4

5.6 Use case 5: video surveillance in Smart Cities

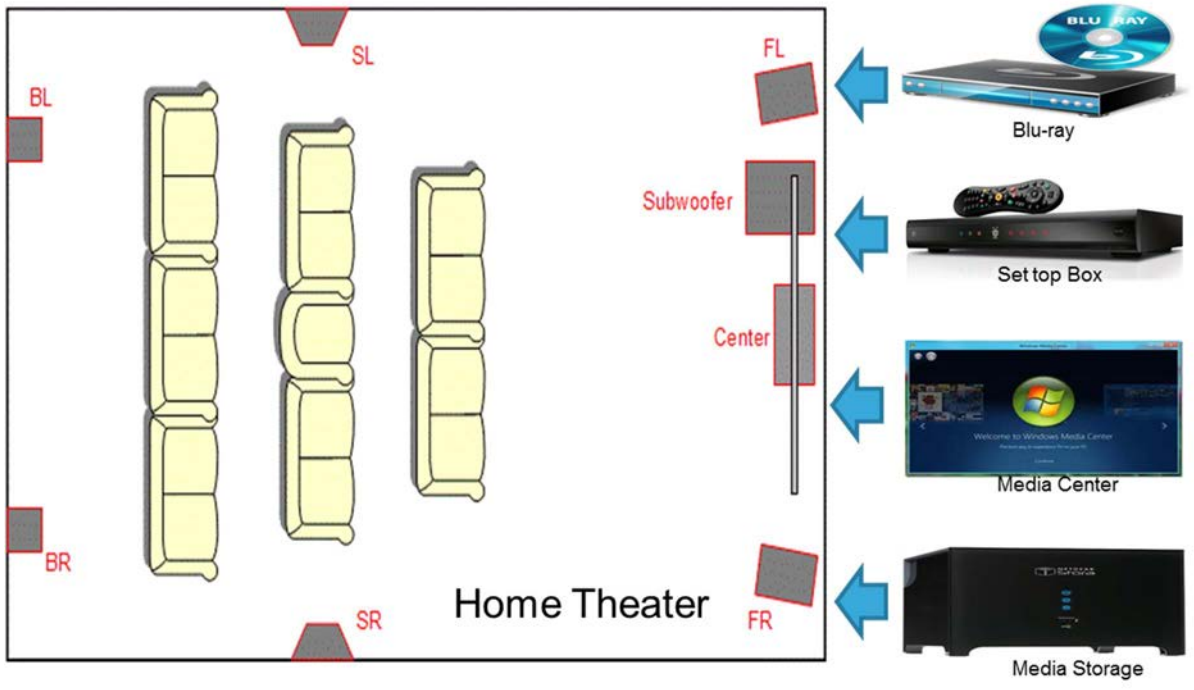
Smart Cities needs connected smart devices to provide services such as granting security. In this scenario, a surveillance camera sends HD or UHD contents to a media storage device. The media storage device is typically close to the camera. The surveillance camera can be equipped with a motion sensor to properly activate and consequently reduce the contents that are recorded in the media storage system. The video stored by the media storage is a high quality HD or UHD video allowing accurate investigation in case of need. For real time monitoring, the video camera can send its content to one or more surveillance monitors (only one terminal shown in figure 7). As the distance among the camera and the monitors increases, PHDMI lower video bit rates (see clause 4) are more easily supported.



NOTE: Source: a surveillance camera. Sinks: a media storage system and/or a surveillance monitor(s).

Figure 7: PHDMI use case 5

5.7 Use case 6: home theater system



NOTE: Data flow diagram for home theater.

Figure 8: PHDMI use case 6

Figures 8 and 9 describe the use case where users enjoy the high definition video/audio in a home theater. A Blu-ray™ player or a set top box or a Media Center or Media storage provides the video/audio stream to a LCD display or a project set and surround sound speakers system.

NOTE: Blu-ray™ is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

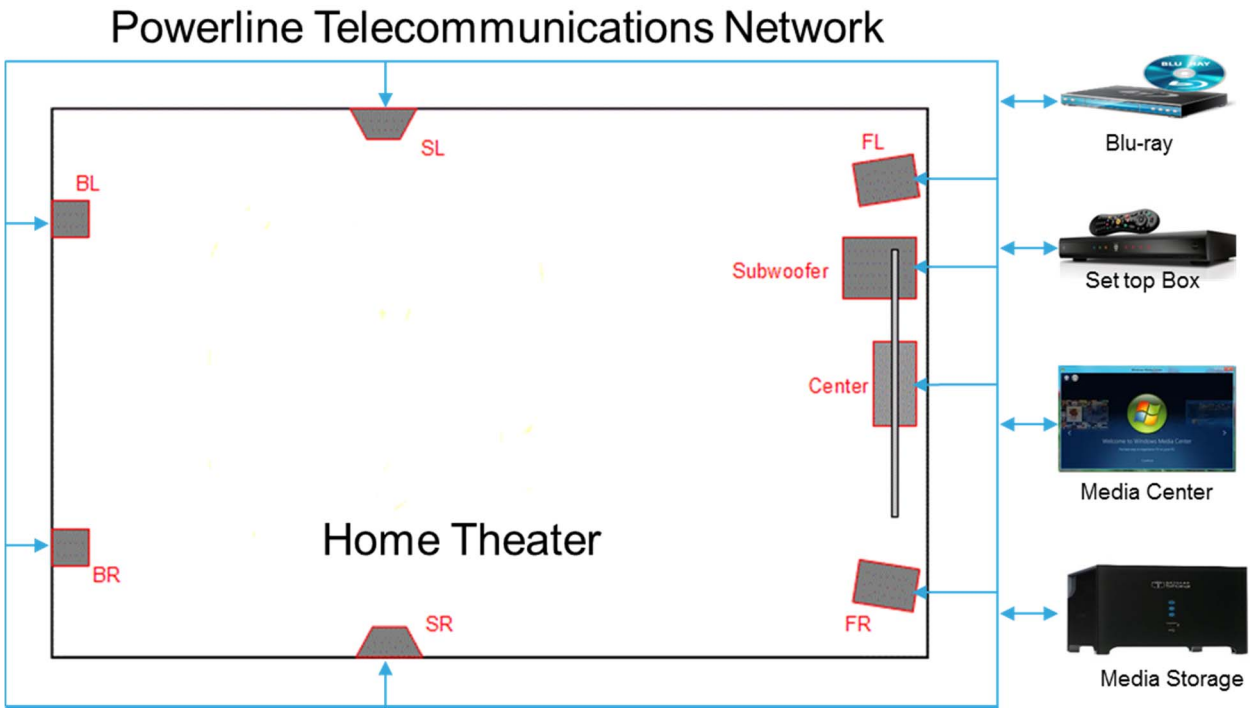


Figure 9: Data flow on PLT network for home theatre

6 Analysed schemes

6.1 Introduction

Clauses 6.2 and 6.3 describe possible PHDMI schemes. They can be subdivided into two categories: tandem schemes (presented in clause 6.2) and joint schemes (presented in clause 6.3).

6.2 Tandem schemes

6.2.1 General

For the purposes of the present document, a tandem scheme is defined as a communication scheme that keeps separated the source encoding part and the channel encoding communication part. The typical realization of a tandem scheme consists in the concatenation of a source encoder with a PLT modem. In clause 6.2.2, the source encoder is based upon the JPEG 2000 standard while the PLT modem uses the 2 MHz to 100 MHz frequency band for communication in TDMA mode. In clause 6.2.3, the source encoder is based upon the Dirac standard while the PLT modem uses the 2 MHz to 100 MHz frequency band for communication in TDMA mode.

6.2.2 JPEG 2000+PLT on 2 MHz to 100 MHz

A possible choice for the PLT modem is, according to the current state of the art, a modem operating in the 2 MHz to 100 MHz frequency band. Examples are modems following the HomePlug® AV2 and ITU-T G.hn specification. As per the source encoder, a possibility, described in this clause, is an encoder compliant to the JPEG 2000 standard. The JPEG 2000 encoder compresses the individual frames of the video before sending them to the PLT modem. The JPEG 2000 encoder includes several blocks: preprocessing, multicomponent transformations, discrete wavelet transforms (DWT), quantization, and entropy coding.

Preprocessing: The image components (YUV) are decomposed into rectangular non-overlapping blocks, or tiles, which are compressed independently. This allows for easy access to specific regions in the image. The next step of the preprocessing is the DC level shifting, where the same quantity 2^P-1, where P is the precision of the component, is subtracted from all samples in the component tile.

Wavelet transforms: The DWT is used to process the tile components into different decomposition levels. The JPEG 2000 standard supports two component transformations, the irreversible component transformation (ICT), used for lossy coding, and the reversible component transformation (RCT), used for both lossless and lossy coding. The default irreversible transform is implemented by means of the Daubechies (9,7) floating point wavelet, while the default reversible transformation is implemented by means of the (5,3) integer wavelet.

Quantization: All coefficients are then quantized using the uniform scalar quantization with dead-zone, with one quantization step size allowed per subband. The operation is lossy, unless the quantization step is 1 and the coefficients are integers, as produced by the reversible integer (5,3) wavelet. The step size can be chosen in such a way that a given quality is achieved. A fixed bitrate is achieved if the coefficients are finely quantized, and the embedded bitstream is truncated to match the desired rate.

Entropy coding: After quantization, each subband is divided into rectangular blocks. Three rectangles corresponding to the same region (one from each subband at each resolution level) form a packet partition location or precinct. Each precinct is further divided into non-overlapping rectangles, called code blocks, which form the input to the entropy coder. Entropy coding is performed independently on each code-block. The coding is a context-dependent arithmetic coding of bitplanes.

Within each subband the code blocks are traversed in raster order. The code blocks are coded a bit plane at a time starting from the most significant bit plane with a non-zero element to the least significant bit plane. The individual bit planes are coded within three coding passes, each collecting contextual information about the bit plane data. This contextual information and its internal state are used by the arithmetic coder to generate a compressed bit stream.

Several parameters affect the performance of the JPEG 2000 encoder.

- Choice of wavelet: in the lossy compression case, the wavelet (9,7) gives a better space-frequency representation, although it is more complex than the (5,3) wavelet
- Size of code blocks: since code blocks are the encoding unit, their size affects how well the context modelling represents the neighbourhoods. At the same time, using more code blocks allows for parallel processing.

6.2.3 Dirac+PLT on 2 MHz to 100 MHz

As in clause 6.2.2, the PLT modem is, according to the current state of the art, a modem operating in the 2 MHz to 100 MHz frequency band. As per the source encoder, in this clause an encoder compliant to the Dirac standard is considered. Dirac is a wavelet-based Open Source video codec developed by the BBC. It uses motion estimation, motion compensation and discrete wavelet transform. After quantization and binarization, variable length coding or arithmetic coding is used for entropy coding.

Similar to other codecs such as H.264 and HEVC, in Dirac there are three types of frames that form the coding unit, or a GoP (Group of Pictures). The I frames are intra coded, Level 1 (L1) frames are predicted upon earlier I or L1 frames, and Level 2 (L2) frames are predicted from the I and L1 frames closest in time, both earlier and later.

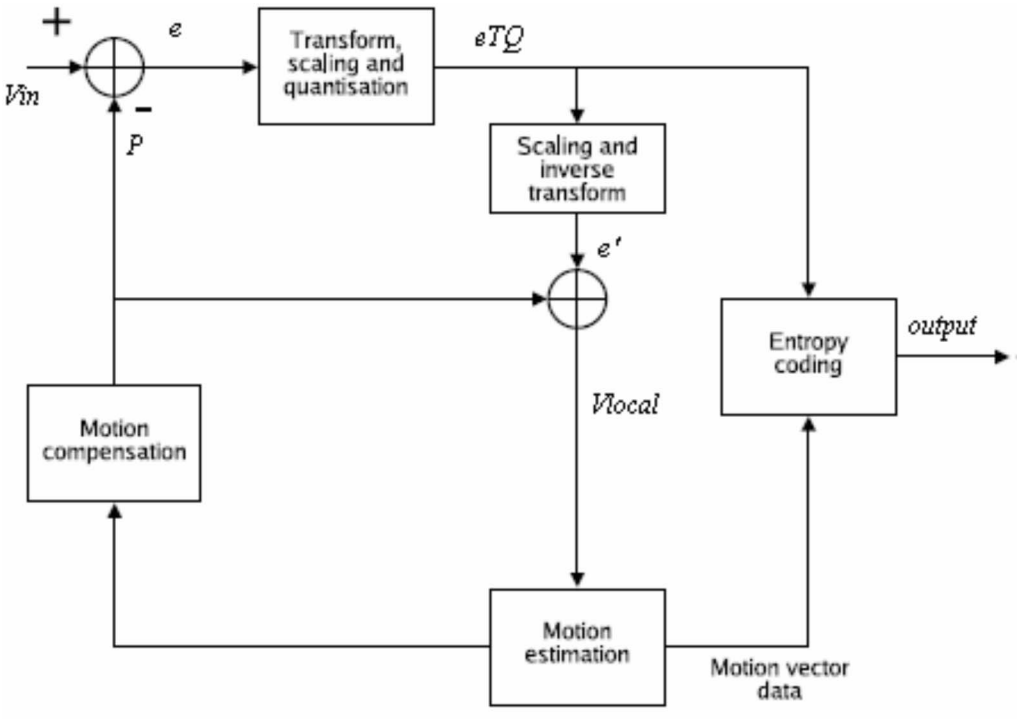


Figure 10: Dirac encoder architecture

Motion estimation and compensation: Dirac implements hierarchical motion estimation. The current and reference frames are repeatedly down converted by a factor of two in both dimensions, and motions vector are found at different levels. The number of down conversion levels depends on the size of the frame. Motion estimation is performed first in the lowest resolution level and the vectors are used as a guide for searching at the higher level, until reaching the size of the original image. Each block at the lower resolution level corresponds to four blocks at immediate higher resolution level. Thus it provides a guide motion vector to the 4 blocks at immediate higher resolution level. The search pattern used in lowest level is diamond shape with the search range 5 and all other levels use a square shape search pattern with search range 1. Pixel accurate motion vectors can be refined to sub-pixel accuracy up to 1/8 of a pixel.

Rate-Distortion optimization: In RDO optimization, one tries to minimize $Distortion(p)+\lambda Rate(p)$ over some parameter *p*. In video coding bit rate constraints are global, and the bit stream consists of many elements coded independently which have to hit the bit rate in total. By choosing a point of constant rate-distortion slope independently for each element of the bitstream, one obtains that the distribution of bits between elements is optimal.

RDO in motion estimation is also necessary. Motion vector coding is associated to coding the residues, since the distortion that a particular motion field introduces can be corrected by the residue coding. So there is a trade-off between bits spent on motion vectors and bits spent on correcting the residue. In general, at high bit rate there are diminishing returns from improving the motion field, and at low bit rate there are much better returns from having a good motion field.

In Dirac, the overall trade-off factor λ is derived from a value named "QF" given at the encoder, or derived from a starting QF by the rate-control algorithm in the case of constant bit-rate. The initial assignation of values λ for the different type of pictures is:

$$\begin{aligned} I\_lambda &= \frac{1}{16}10^{\frac{10-QF}{2.5}} \\ L1\_lambda &= 32 * I\_lambda \\ L2\_lambda &= 256 * I\_lambda \end{aligned} \quad (6)$$

These lambda variables are used for quantizer selection in I, L1 and L2 pictures. From these, motion estimation lambdas have been derived by experiment:

$$L1\_me\_lambda = 2.0 * \sqrt{L1\_lambda}$$
$$L2\_me\_lambda = L1\_me\_lambda \quad (7)$$

Several parameters affect the performance of the Dirac encoder:

- Choice of wavelet filter: several available filters offer a trade-off between space-frequency representation and implementation complexity
- Motion vector estimation: sub-pixel accuracy can offer better estimation at the expense of increased computation time
- GoP length and structure: these parameters can be tuned to obtain a trade-off between better compression in terms of rate-distortion and decoding delay.

6.3 Joint schemes

6.3.1 General

For the purposes of the present document, a joint scheme is defined as a communication scheme that performs together (at least at the same abstraction layer) both source and channel encoding. In clause 6.3.2 the SoftCast PLT scheme on the 2 to 100 MHz frequency band will be presented.

6.3.2 SoftCast PLT on 2 MHz to 100 MHz

SoftCast (see figure 11) divides a video into several Groups of Pictures (GoP). At the encoder side, a full-GoP 3D-DCT is performed on each GoP to group the energy in the GoP. Then the matrix of 3D-DCT coefficients is divided into many chunks, where chunks are small 2D blocks on the DCT plane. The chunks with the most energy are kept to be transmitted. Their number is determined by the available bandwidth constraints.

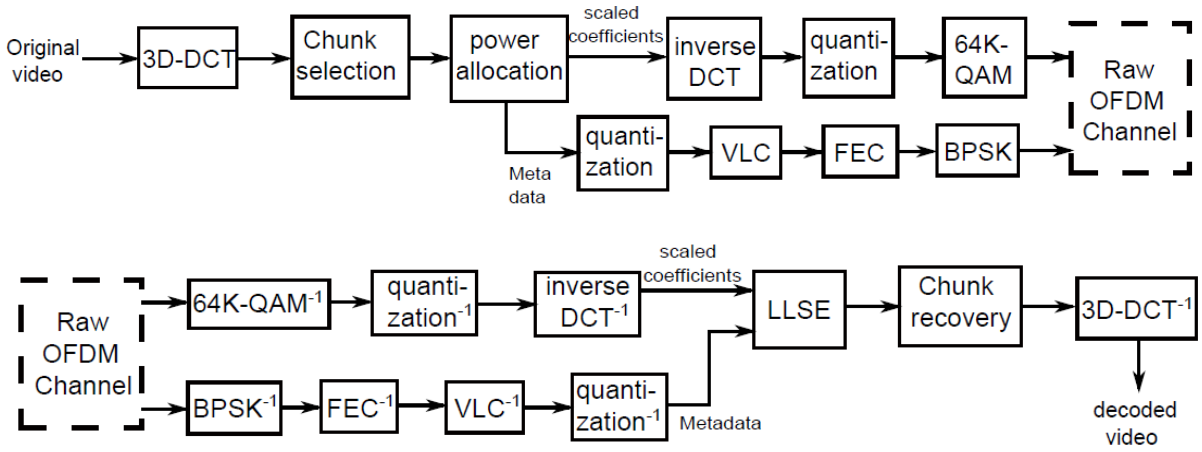


Figure 11: SoftCast scheme, transmitter and receiver structure

Linear scaling is applied on the selected chunks to optimally allocate transmission power among chunks. The scaling coefficients are calculated by minimizing the mean square reconstruction error at the decoder side given a certain transmission power constraint. Then a one-dimensional inverse DCT is used on the selected chunks to get equal-energy slices so that one chunk loss will not cause severe decoding problems. No channel coding is used. A high-density modulation (64K-QAM without gray mapping) is used to linearly modulate the samples of each slice. The I and Q components of the digital signal are then transmitted using OFDM-based physical layer. On the decoder side a Linear Least Square Error (LLSE) estimator is used to reconstruct the original frames. This estimator requires an estimate of the noise introduced by the OFDM subchannels.

Some side information generated by the encoder needs to be transmitted as metadata, to be used at the decoder side to reconstruct an estimate of the original video sequence. These metadata are:

- Format of video sequence (length, width, color subsampling, frame rate, etc.)
- Size of GoP (8 bits / GoP)
- Position of transmitted chuncks (1 bit / chunk)
- Mean and variance of transmitted chunks (2 × 8 bits / chunk)

The last two types of metadata may be variable-length coded and channel coded to protect them from transmission errors. Then a low-density modulation can be used to transmit the coded metadata of some OFDM subchannels. The type of modulation and coding scheme used for the metadata may depend on the characteristics of the channel and on the level of reliability required.

6.4 Short summary of clause 6

This clause has described three most prospective schemes for baseline PLT HDMI® technical evaluations. Two kinds of which are tandem schemes and consist on the combinations of forward error correction based OFDM powerline communication techniques (SISO or MIMO) on 2 MHz to 100 MHz frequencies together with video compression techniques (JPEG 2000 and Dirac). The third is a joint scheme, i.e. it merges the channel and source encoding: it is a combination of high density analog OFDM powerline communication technique with 3D DCT video (Softcast). The three schemes are compared in clauses 7 and 8.

7 Results for HD videos

7.1 Introduction

Clause 7.2 presents results obtained by the schemes considered in clause 6 with HD videos. The test sequences used for evaluating the schemes are illustrated in clause 7.2. Results with homogenous PLT per subcarrier SNRs are presented in clause 7.2.3 while results with more realistic PLT per subcarrier SNRs are shown in clause 7.2.4.

7.2 Considered test sequences

7.2.1 General

Experiments have been performed on three HD sequences: "DucksTakeOff", "ParkJoy", and "YachtRide" (see figure 12) that are part of the MPEG database and a sequence furnished by France Télévision labelled in what follows as "FTV". These video sequences are considered difficult to encode and have the following characteristics:

- DucksTakeOff: ducks on a lake, still camera, slow movement of ducks followed by sudden take off, water rippling.

  Resolution: 1920×1080. Duration: 500 frames. Frame rate: 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from https://media.xiph.org/video/derf/

- ParkJoy: people in a park, running towards the right, camera following the movement, obstructed by trees, another person appears on the right, running towards left.

  Resolution: 1920×1080. Duration: 500 frames. Frame rate: 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from https://media.xiph.org/video/derf/

- YachtRide: yacht moving from left to right, camera following movement, blurred houses in the background, detailed foam of the water and reflection of water on the boat.

  Resolution: 1920×1080. Duration: 300 frames. Frame rate: 60 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from http://ultravideo.cs.tut.fi/#testsequences

Figure 12: Video sequences used: DucksTakeOff (on the left), ParkJoy (on the center) and YacktRide (on the right)

- FTV: sequence containing more than 8 minutes worth of real-life scenes: scenery, news, concert, sports, interview, etc. (for some picture of this sequence, see the clause 8.2. In fact, this is "transposed" from an original UHD sequence).

  Resolution: 1920×1080. Duration: 26123 frames. Frame rate: assumed 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel. The sequence has been HEVC encoded at 8 Mb/s and decoded.

  Video furnished by France Télévision.

7.2.2 Used quality metrics

The used quality criteria are the Peak Signal-to-Noise Ratio (PSNR), based on the Mean Square Error (MSE), and the Structural Similarity (SSIM), which has been designed to better characterize human visual perception.

These two criteria are computed between the original video (x) and the decoded one (y):

- The PSNR (expressed in dB) is given by (for image pixels coded in 8 bits) :

$$PSNR_{dB}(x,y) = 10log_{10}\frac{255^2}{MSE(x,y)} \tag{8}$$

where MSE corresponds to the mean square error between the two images.

- The SSIM (comprised between 0 and 1) is given by:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{9}$$

where μ_x is the average of x, μ_y the average of y, σ_x the standard deviation of x, σ_y the standard deviation of y and σ_{xy} the covariance of x and y. C_1 and C_2 are two variables to stabilize the division with weak denominator.

These criteria are evaluated using the toolbox qPSNR for libav that calculates both PSNR and SSIM on all the video sequences.

7.2.3 Results with homogeneous PLT per carrier SNRs

This clause describes results obtained in a framework where all PLT carriers experience the same SNR.

The test sequences described in clause 7.2 have been used to evaluate the tandem schemes described in clause 6.2. In particular, as per clause 6.2.2, they have been compressed with JPEG 2000 (see clause 6.2.2) to achieve several target bit rates, namely 100 Mb/s, 150 Mb/s, 200 Mb/s, and 250 Mb/s at the application layer, for transmission over the PLT channel. These bit rates may be obtained assuming a flat Gaussian channel with a bandwidth ranging from 1,8 MHz to 86 MHz and with an SNR given in table 6:

Table 6: PLT rate at the application layer as a function of channel SNR

Channel SNR (dB)	5 dB to 5,8 dB	8,5 dB to 9,5 dB	11,2 dB to 12,3 dB	14 dB to 15,4 dB
PLT Rate	100	150	200	250

Parameters used for JPEG 2000 are: one tile, lossless compression, 6 levels of resolution, code-block size of 64 × 64, one layer of quality. To achieve the target bit rates, sequence "ParkJoy" (1920×1080 at 50fps) has been compressed with compression ratios 11,86:1, 7,92:1, 5,93:1 and 4,75:1. The PSNR of the luminance component (Y-PSNR) for these bit rates is reported in figure 13. Results for the other sequences are included in annex A.

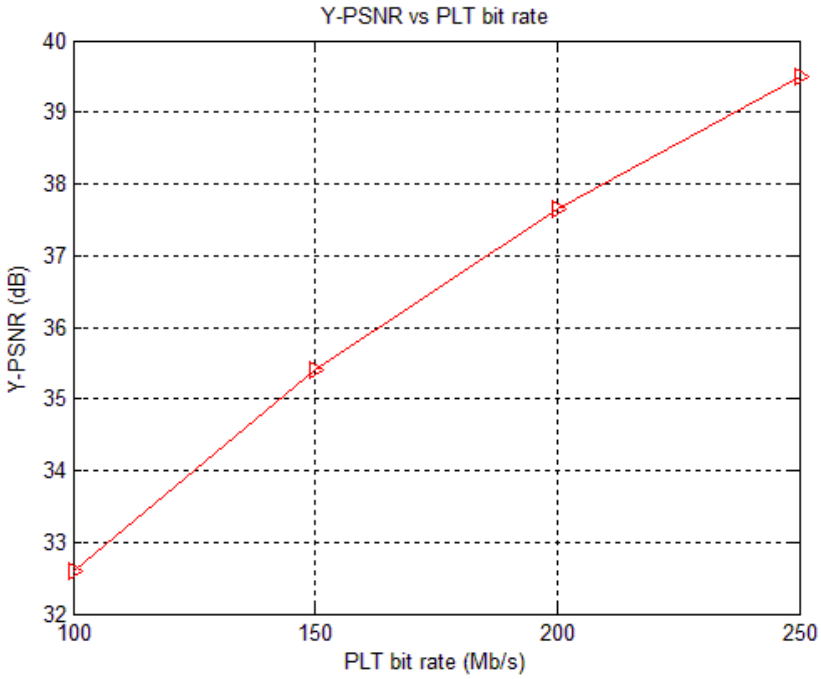


Figure 13: Experimental results for "ParkJoy" using the tandem scheme described in clause 6.2.2

It can be noted that, considering the tandem scheme involving JPEG 2000 (see clause 6.2.2), the Y-PSNR increases regularly with the available bandwidth. In the simulations, the channel has been assumed error-free. The fact that each frame is independently encoded prevents propagation of the transmission error from one frame to the other. Moreover, this reduces the coding and decoding latency.

The adaptation of the bit rate is made easy by the embedded nature of the bit stream produced by JPEG 2000.

Similar scenario has been considered for the Dirac tandem scheme of clause 6.2.3. In the experiments, a GoP of 6 frames with the structure I L2 L2 L1 L2 L2 has been used. The rest of the parameters used are the default ones: Deslauriers-Dubuc (9, 7) wavelet filters for the wavelet transforms for the intra frames and for the motion-compensated residual frames, ¼ pixel accuracy for motion vectors, and context adaptive arithmetic coding for the entropy coding.

The test sequences described in clause 7.2 have been also used to evaluate the performance of the joint schemes described in clause 6.3.

As per the SoftCast joint scheme described in clause 6.3.2, the metadata are here assumed to be perfectly received. Moreover, the bandwidth required for their transmission has been neglected.

In figure 14, the average PSNR of the luminance (Y-PSNR) of the considered video sequence is reported for various channel SNRs. For comparison, the performance of the tandem scheme of clause 6.2.2 is also reported in red, while the performance of the tandem scheme of clause 6.2.3 is reported in green. For the three considered video sequences, for the SoftCast scheme, the Y-PSNR increases almost linearly with the channel SNR. This increase in quality does not require any modification of the transmitted content, which is generated independently of the channel conditions. The quality of the received video increases thus gracefully with the quality of the channel.

At values of the channel SNR corresponding to a target bitrate of 150 Mb/s (8,5 dB to 9,5 dB), one sees that the Dirac tandem scheme is about 1,7 dB better in PSNR than the JPEG 2000 tandem scheme and about 3,5 dB better than the SoftCast joint scheme.

Note that to get this performance, the tandem schemes requires a good knowledge of the available channel rate at the encoder. This is not required by SoftCast.

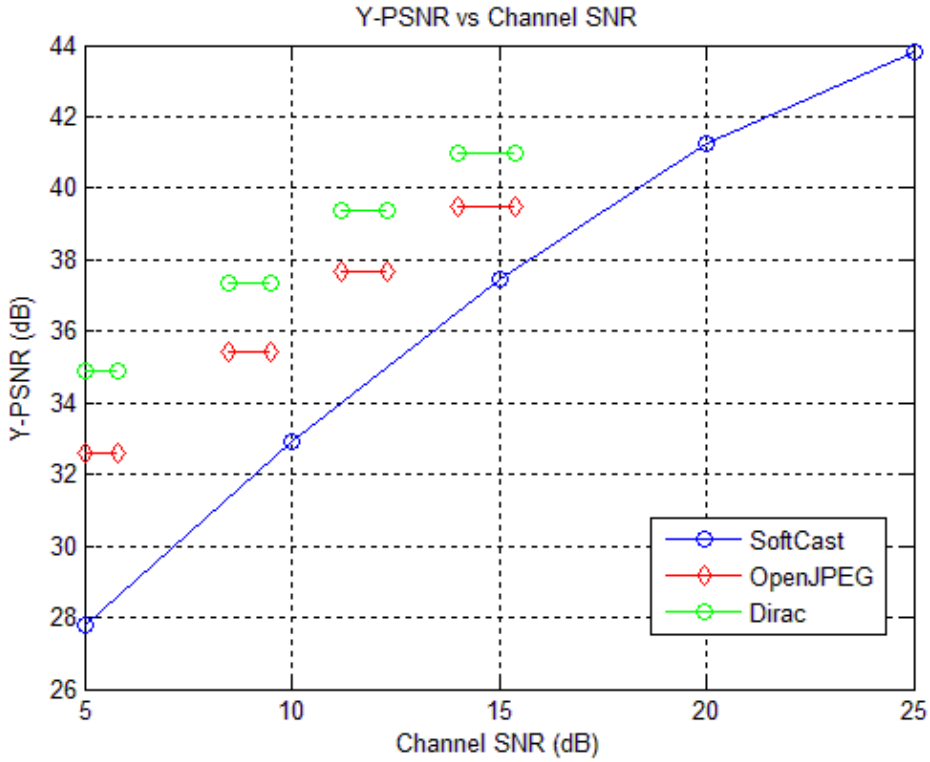


Figure 14: Experimental results for "ParkJoy" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Results in terms of mean Y-SSIM are reported in table 7 and in annex A for all the test sequences of clause 7.2 and all the schemes of clause 6. It can be observed that the general trend highlighted using the PSNR metric is confirmed by the Y-SSIM analysis.

Table 7: Y-SSIM experimental results for "ParkJoy" using the schemes described in clause 6.

SNR	5	10	12	15	20	25
Mean Y-SSIM-OpenJPEG	0,9003	0,9339	0,9561	0,9678		
Mean Y-SSIM-Dirac	0,9434	0,9606	0,9709	0,9774		
Mean Y-SSIM-SoftCast	0,8058	0,9044		0,9602	0,9837	0,9920

7.2.4 Results with more realistic PLT per carrier SNRs

This clause describes results obtained on 500 PLT links. SISO, MIMO 2×2 and MIMO 2×3 communication schemes have been considered, where MIMO schemes adopt eigenbeamforming. As far as tandem schemes of clause 6.2 are concerned, the PLT rate obtained on these links range from 50 Mb/s to 400 Mb/s (see clause 4.4).

Figure 15 compares the JPEG 2000 tandem scheme and the Dirac tandem scheme in terms of Y-PSNR versus the used compression ratio for the "FTV" sequence, which has been HEVC encoded at 8Mb/s and decoded. The reference sequence for the quality measures is thus the HEVC decoded sequence. As the SoftCast joint scheme furnished worse performance than tandem scheme as shown in clause 7.2.3 and in annex A, it was decided to limit the tests on the very long FTV sequence on realistic PLT SNRs to the tandem schemes. Results with SoftCast on the "FTV" sequence will be found on clause 8.2.3.

For JPEG 2000 the parameters used were the (9,7) wavelet, 6 resolutions, and code blocks of size 64×64. For Dirac, the following settings have been used: filter DD(13,7), GoP size of 24, and motion vector precision of ½.

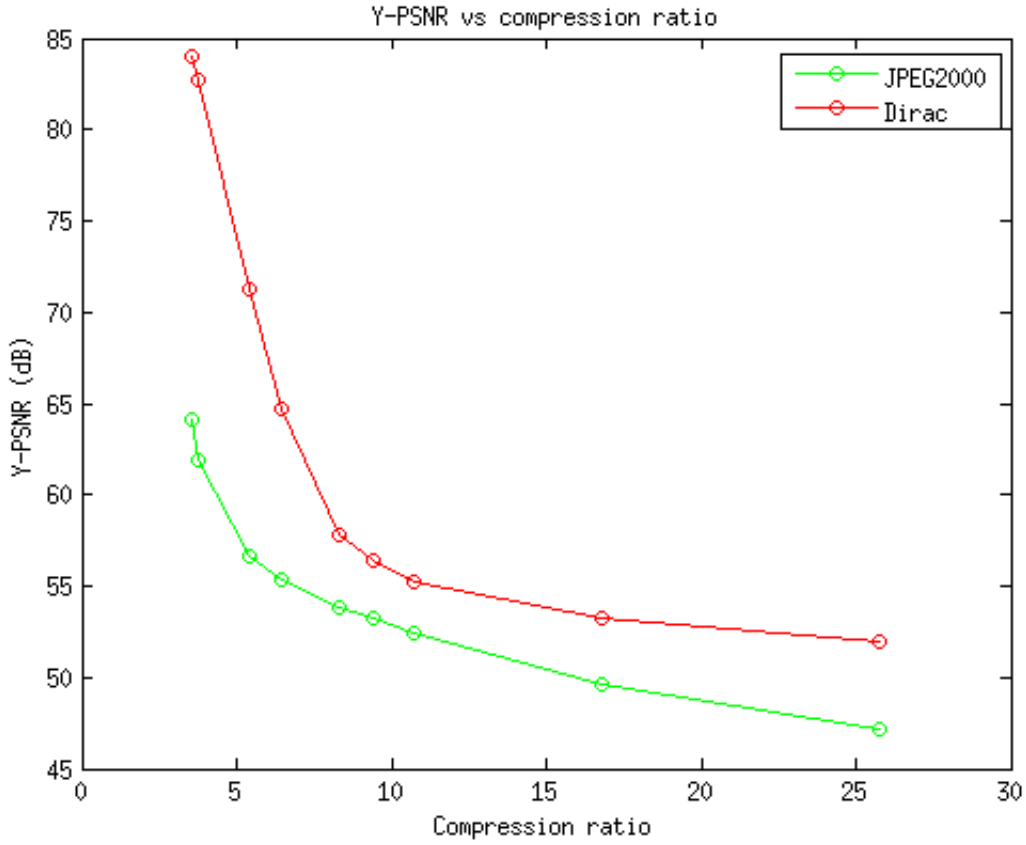


Figure 15: PSNR experimental results for "FTV" using the tandem schemes described in clause 6.1

Table 8 reports the results in terms of mean Y-SSIM for three specific PLT links. Table 8 also allows performing a comparison among PLT modems using SISO, MIMO 2×2 and MIMO 2×3, though it is not the main scope of this clause.

The used PLT links achieve the following bitrates at the application layer:

- SISO: [48,28 74 231,2] Mbit/s
- MIMO 2×2: [116,43 132,41 331,37] Mbit/s
- MIMO 2×3: [149,67 192,2 350,45] Mbit/s

Table 8: Y-SSIM experimental results for HD "FTV" (HEVC encoded and decoded) using the tandem schemes described in clause 6.2

Compression Ratio	3,55	3,76	5,38	6,47	8,31	9,40	10,69	16,82	25,77
Mean Y-SSIM JPEG 2000	0,9963	0,9958	0,9928	0,9907	0,9873	0,9854	0,9834	0,9726	0,9588
Mean Y-SSIM Dirac	0,9982	0,9980	0,9967	0,9957	0,9941	0,9934	0,9929	0,9912	0,9893
Link	3	3	3	2	1	2	1	2	1

Figure 15 shows that while both schemes achieve good PSNR, Dirac is notably better than Jpeg2000 at very small and very high compression ratios. Table 8 gives more input about the Quality of Experience (QoE).

In general, a Y-SSIM in the range [0,97 - 1] indicates a QoE with virtually no losses to a non-expert eye, a value in the range 0,90 to 0,97 indicates that artefacts are visible, but the quality is still good to the viewer. If the Y-SSIM < 0,90 usually the visual experience cannot be accepted.

7.2.5 Short summary of clause 7 and related results in annex A

This clause (and the related results reported in annex A) has compared the performance results by evaluating the three PHDMI techniques presented in clause 6 on HD videos. The evaluations are initially conducted with homogeneous SNR characteristics on three HD video sequences and then extended to the SNR database related to 500 PLT links. For three video sequences, "DucksTakeOff", "ParkJoy" and "YachtRide", the tandem schemes have more advantages over the joint scheme by the quality of Y-PSNR and Y-SSIM, especially considering that the joint scheme has been assumed without any MAC overhead. However, visual examination revealed that the joint scheme is also able to provide sufficient quality. The evaluation of the two tandem schemes on the very long "FTV" sequence revealed a preference of the Dirac-based scheme. As it will be seen in clause 8, this preference will be more marked.

8 Results for UHD videos

8.1 Introduction

Clauses 8.2.2 and 8.2.3 present results obtained by the schemes considered in clause 6 with UHD videos. The test sequences used for evaluating the schemes are illustrated in clause 8.2.1. Results with homogenous PLT per subcarrier SNRs are presented in clause 8.2.2 while results with more realistic PLT per subcarrier SNRs are shown in clause 8.2.3.

8.2 Considered test sequences

8.2.1 General

Experiments have been performed on four UHD sequences: "DucksTakeOff", "ParkJoy", "Crowd_run" and a sequence by France Television labelled in what follows as "FTV". These video sequences have the following characteristics:

- DucksTakeOff: ducks on a lake, still camera, slow movement of ducks followed by sudden take off, water rippling.

  Resolution: 3840×2160. Duration: 500 frames. Frame rate: 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from https://media.xiph.org/video/derf/

- ParkJoy: people in a park, running towards the right, camera following the movement, obstructed by trees, another person appears on the right, running towards left.

  Resolution: 3840×2160. Duration: 500 frames. Frame rate: 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from https://media.xiph.org/video/derf/.

- CrowdRun: people on a running race on a field, slight movement of the camera towards the left, then to the right

  Resolution: 3840×2160. Duration: 500 frames. Frame rate: 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel

  Retrieved from https://media.xiph.org/video/derf/.

- FTV: sequence containing more than 8 minutes worth of real-life scenes: scenery, news, concert, sports, interview, etc. (see figure 16)

  Resolution: 3840×2160. Duration: 26123 frames. Frame rate: assumed 50 fps (progressive).

  Color subsampling: 4:2:0, format YUV, 24 bits/pixel. The sequence has been HEVC encoded at 18Mb/s and decoded.

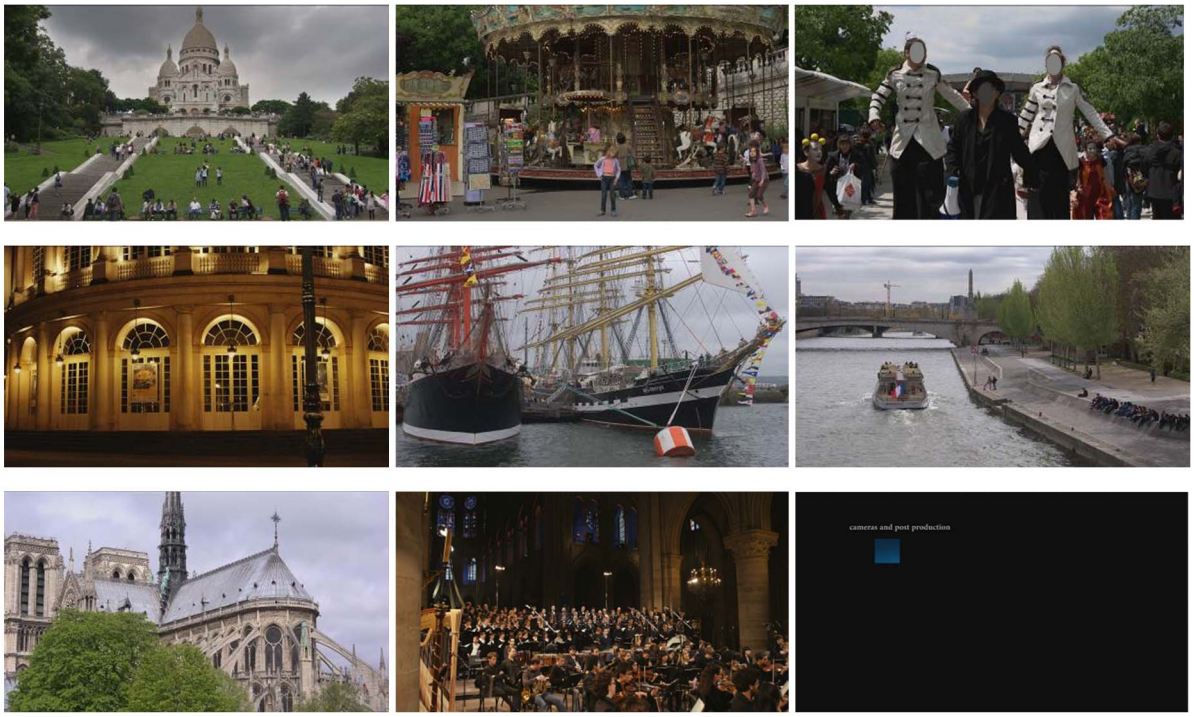


Figure 16: Example of scenes as part of the "FTV" real-life sequence

8.2.2 Results with homogeneous PLT per carrier SNRs

This clause describes results obtained in a framework where all PLT carriers experiment the same SNR.

The test sequences "ParkJoy" and the first 600 frames of the "FTV" sequence (in this clause with 4:4:4 colours subsampling, i.e. with information video bit rate of 10 Gbit/s) described in clause 8.2 have been used to evaluate the tandem schemes described in clause 6.1 and the joint scheme described in clause 6.3. Figures 17 and 18 compare the JPEG 2000 tandem scheme, the Dirac tandem scheme and the SoftCast joint scheme. Similarly to clause 7.2.2, the SoftCast scheme assumes a PLT PHY rate of 1 Gbit/s while for the JPEG 2000 and Dirac schemes PLT rates at the application level are: 100 Mb/s, 150 Mb/s, 200 Mb/s and 250 Mb/s. In figure 17, the PSNRs obtained by the tandem schemes are clearly better than the PSNRs obtained with the SoftCast joint scheme. This confirms the trend already observed for HD videos (see clause 7.2.2 and in particular the results obtained with the homologous "ParkJoy" HD video). Comparing the two tandem schemes, in figure 17 JPEG 2000 appears to be better than Dirac as the SNR increases (note that on the homologous "ParkJoy" HD video in clause 7.2.2, Dirac was always better than JPEG 2000). In figure 18, related to the "FTV" video, the Dirac tandem scheme offers best performance. The JPEG 2000 tandem scheme is also worse than the SoftCast joint scheme. This last fact has been attributed to the fact that the first 600 frames of the "FTV" video are quite static in character.

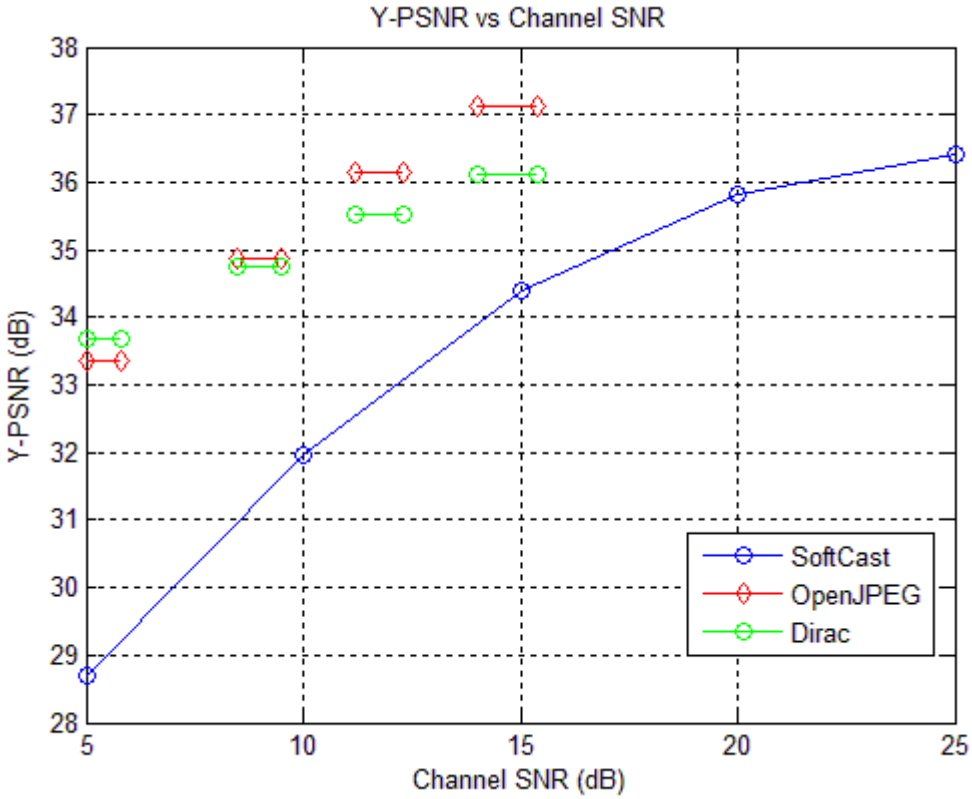


Figure 17: PSNR experimental results for "ParkJoy" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

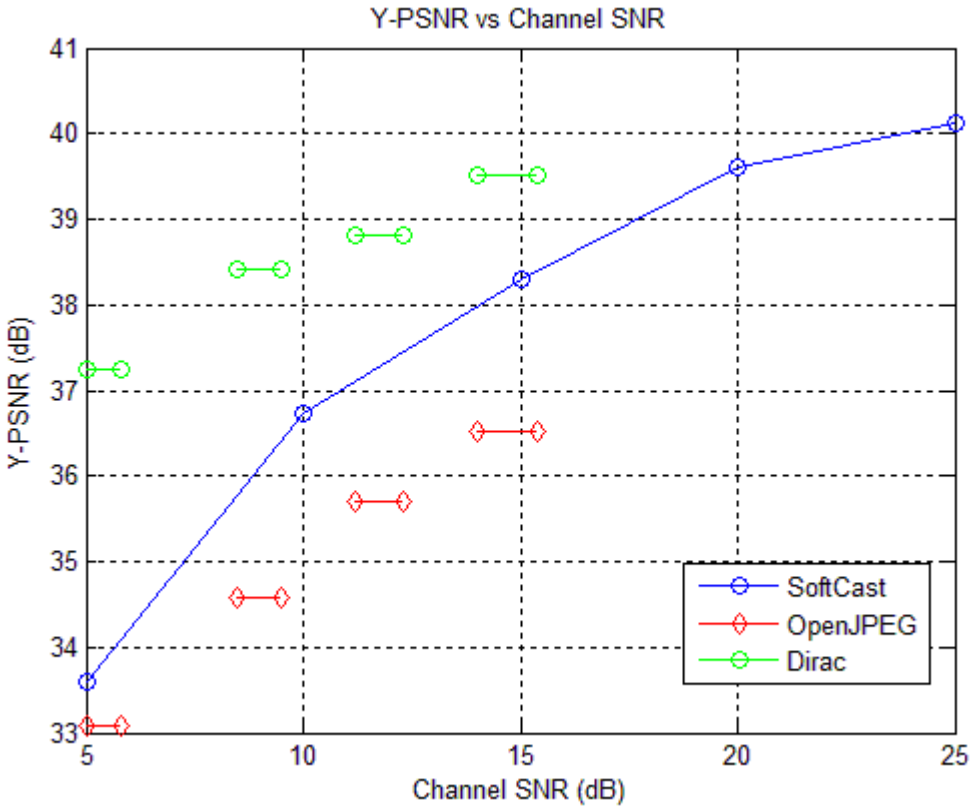


Figure 18: Experimental results for the first 600 frames of "FTV" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Results in terms of mean Y-SSIM are reported in table 9 for the "ParkJoy" video. These results are in line with the ones depicted in figure 21, with the tandem schemes outperforming SoftCast. It has to be noted, however, that visual examination of the decoded "ParkJoy" revealed that even SoftCast at 5 dB yields acceptable results (only little variations of the colour components have been observed). In table 10, mean Y-SSIM results are reported for the "FTV" video and confirm the findings of figure 22 with SoftCast being better than OpenJPEG.

Table 9: Y-SSIM experimental results for "ParkJoy" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

SNR	5	10	12	15	20	25
Mean Y-SSIM-OpenJPEG	0,8817	0,9047	0,9202	0,9328		
Mean Y-SSIM-Dirac	0,8946	0,9103	0,9197	0,9262		
Mean Y-SSIM-SoftCast	0,7751	0,8597		0,9074	0,9312	0,9392

Table 10: Y-SSIM experimental results for "FTV" using the joint scheme described in clause 6.3.2 and the tandem scheme described in clause 6.2.2

SNR	5	10	12	15	20	25
Mean Y-SSIM-OpenJPEG	0,8539	0,8885	0,9076	0,92		
Mean Y-SSIM-Dirac	0,9342	0,9466	0,9482	0,9547		
Mean Y-SSIM-SoftCast	0,8636	0,9286		0,948	0,9625	0,9673

8.2.3 Results with more realistic PLT per carrier SNRs

This clause describes results obtained on 500 PLT links. SISO, MIMO 2×2 and MIMO 2×3 communication schemes have been considered, where MIMO schemes adopt eigenbeamforming. As far as tandem schemes of clause 6.2 are concerned, the PLT rate obtained on these links range from 50 Mb/s to 400 Mb/s (see clause 4.4). These PLT rate assumes the possibility of packet retransmission at the PLT level, which is an assumption supported by state of the art modems.

The test sequences described in clause 8.2 have been used to evaluate the tandem schemes described in clause 6.2 and the joint scheme described in clause 6.3. The "FTV" video considered in this clause has color subsampling equal to 4:2:0. All the test sequences have been HEVC encoded and decoded at 18 Mb/s. Reference for quality metric (PSNR, SSIM) computation is the HEVC decoded video.

Figures 19 and 20 compare the JPEG 2000 tandem scheme, the Dirac tandem scheme and the SoftCast joint scheme in terms of Y-PSNR versus the used compression ratio for the "ParkJoy" and the "FTV" sequence. Results for other sequences are included in annex A. In figure 19, it can be seen that for low compression ratios the JPEG 2000 scheme (green curves) is generally best. Instead, for medium to high compression ratio, the use of the Dirac scheme (red curves) gives best results. For the "FTV" video (see figure 20), this behaviour is not confirmed with the Dirac scheme outperforming JPEG 2000 on the whole compression ratio range. Moreover, figures 19 and 20 highlight that the SoftCast joint scheme performs worst (note also that SoftCast results are at the PHY level). Similar considerations can be drawn looking at tables 11 and 12 where results in terms of mean Y-SSIM are reported. For these tables, it should be observed that Dirac reported results are related to a compression ratio which is not exactly the one reported in the first row (it was not possible to match it exactly. Dirac results are generally obtained with a little higher compression ratio). For three specific PLT links, here numbered 1, 2 and 3, tables 11 and 12 also allow performing a comparison among PLT modems using SISO, MIMO 2×2 and MIMO 2×3, though it is not the main scope of this clause.

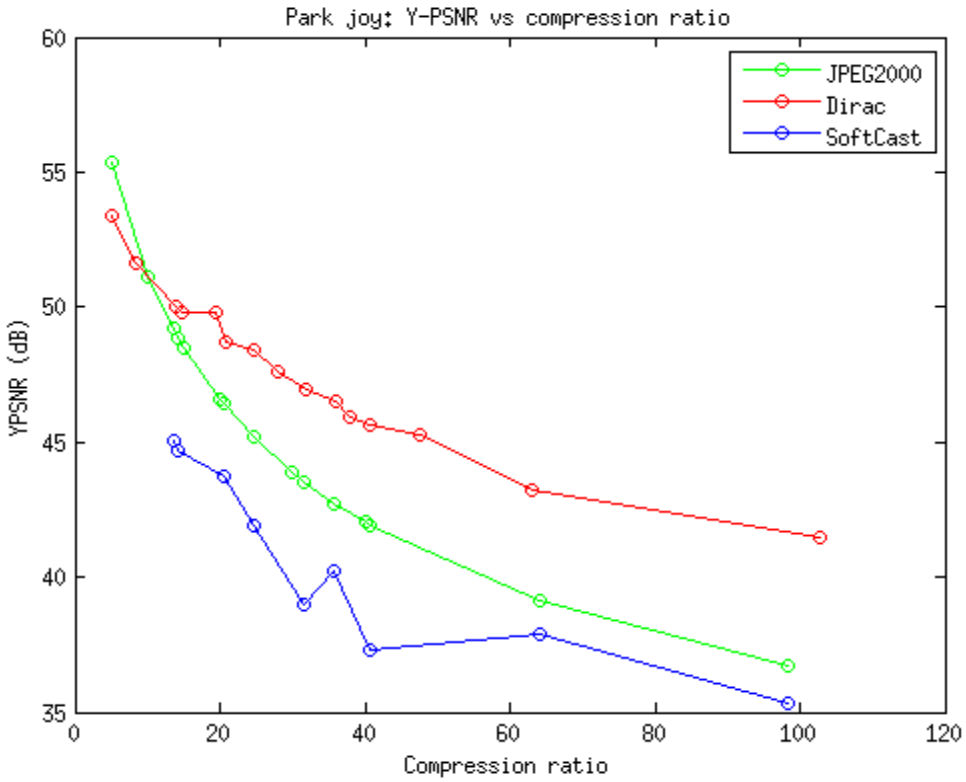


Figure 19: PSNR experimental results for UHD "ParkJoy" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

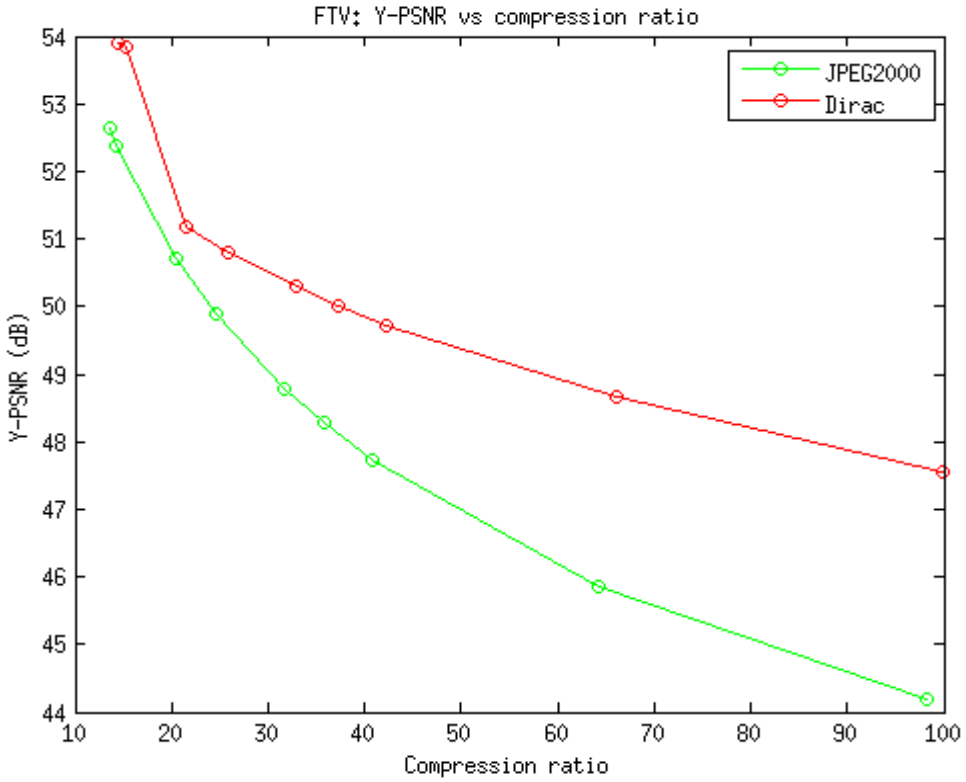


Figure 20: PSNR experimental results for "FTV" using the tandem schemes described in clause 6.2

Table 11: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM JPEG 2000	0,9889	0,9880	0,9815	0,9763	0,9670	0,9612	0,9544	0,9232	0,8843
Mean Y-SSIM Dirac	0,9904	0,9901	0,9876	0,9859	0,983	0,9812	0,979	0,9696	0,9583
Mean Y-SSIM SoftCast	0,9739	0,9716	0,9664	0,9588	0,9225	0,9412	0,8975	0,9182	0,8594
Link	3	3	3	2	1	2	1	2	1

Table 12: Y-SSIM experimental results for UHD "FTV" (HEVC encoded and decoded) using the tandem schemes described in clause 6.2

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM JPEG 2000	0,9871	0,9863	0,9801	0,9761	0,9702	0,9668	0,9631	0,9482	0,9310
Mean Y-SSIM Dirac	0,9898	0,9896	0,9878	0,9870	0,9856	0,9849	0,9840	0,9805	0,9761
Link	3	3	3	2	1	2	1	2	1

Figures 21 and 22 report frame by frame results in terms of Y-PSNR and Y-SSIM for the two tandem schemes of clause 6.2 using SISO on link 3 of table 12. Dirac appears offering more stable quality, but it has to be noticed that there are some zones of the video where JPEG 2000 performs better. Figures 23 and 24 reports a similar frame by frame comparison for the two tandem schemes using MIMO 2×2 on link 1 of table 12. Again Dirac appears more stable in the average with some zones where JPEG 2000 is better.

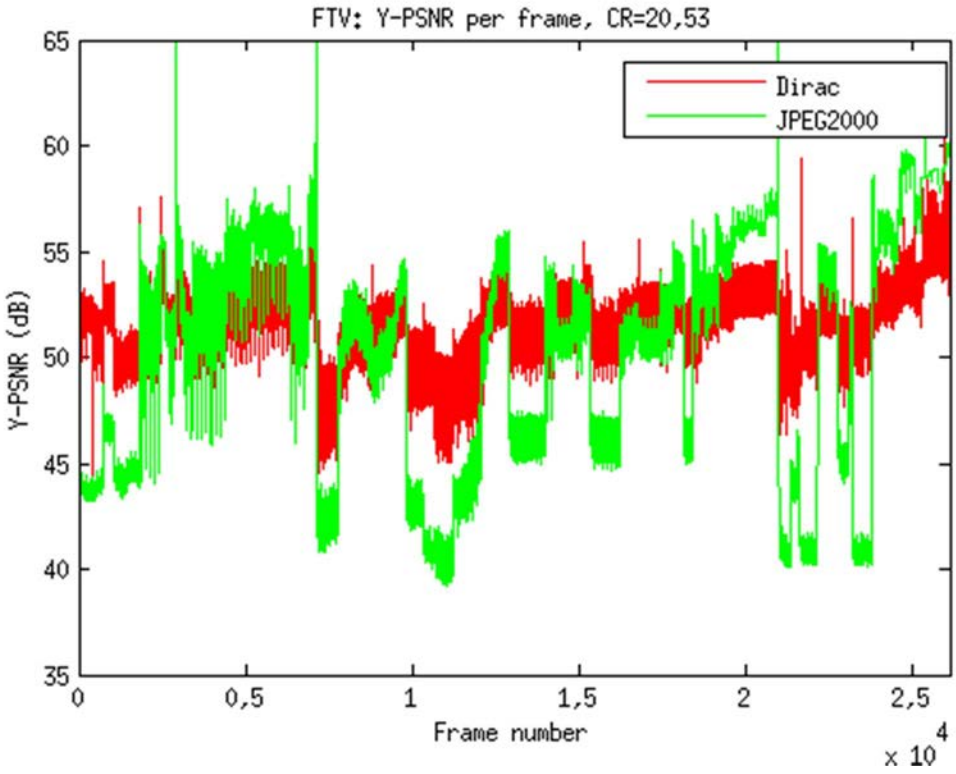


Figure 21: PSNR frame by frame experimental results for "FTV" using the tandem schemes described in clause 6.2 on link 3 (SISO) of table 12

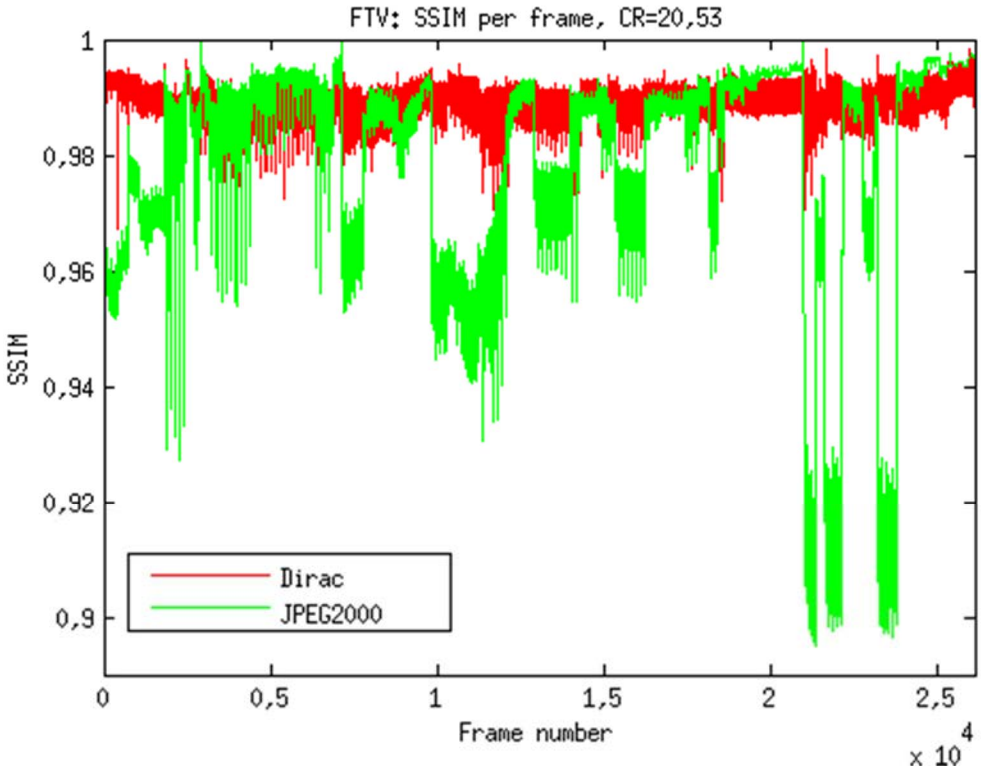


Figure 22: Y-SSIM frame by frame experimental results for "FTV" using the tandem schemes described in clause 6.2 on link 3 (SISO) of table 12

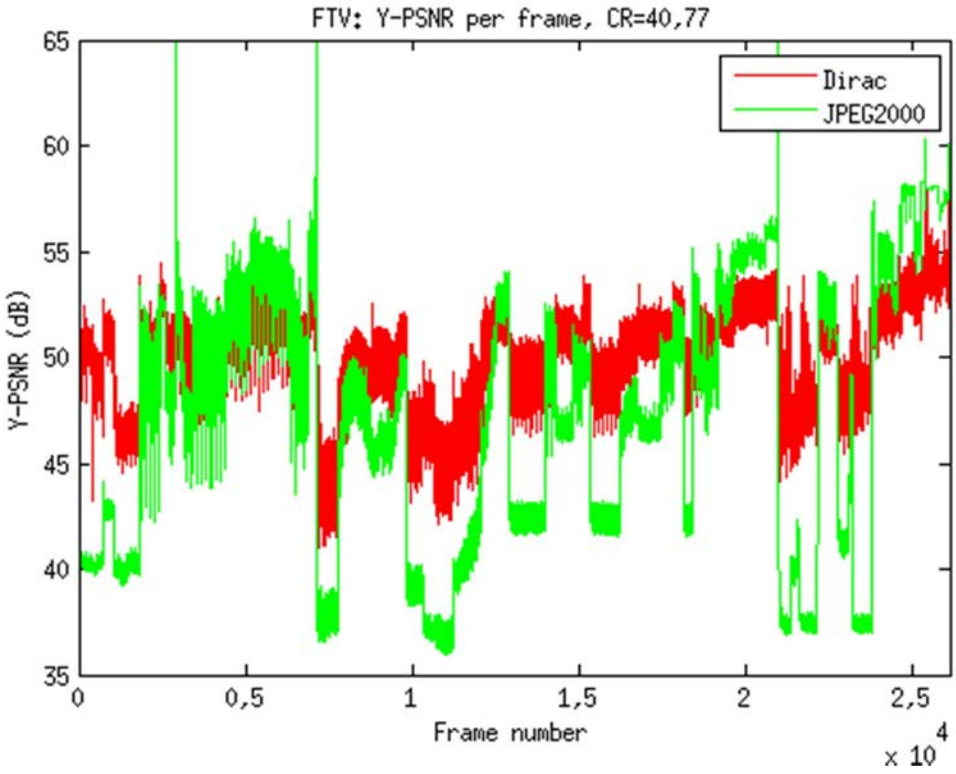


Figure 23: Y-PSNR frame by frame experimental results for "FTV" using the tandem schemes described in clause 6.2 on link 1 (MIMO 2×2) of table 12

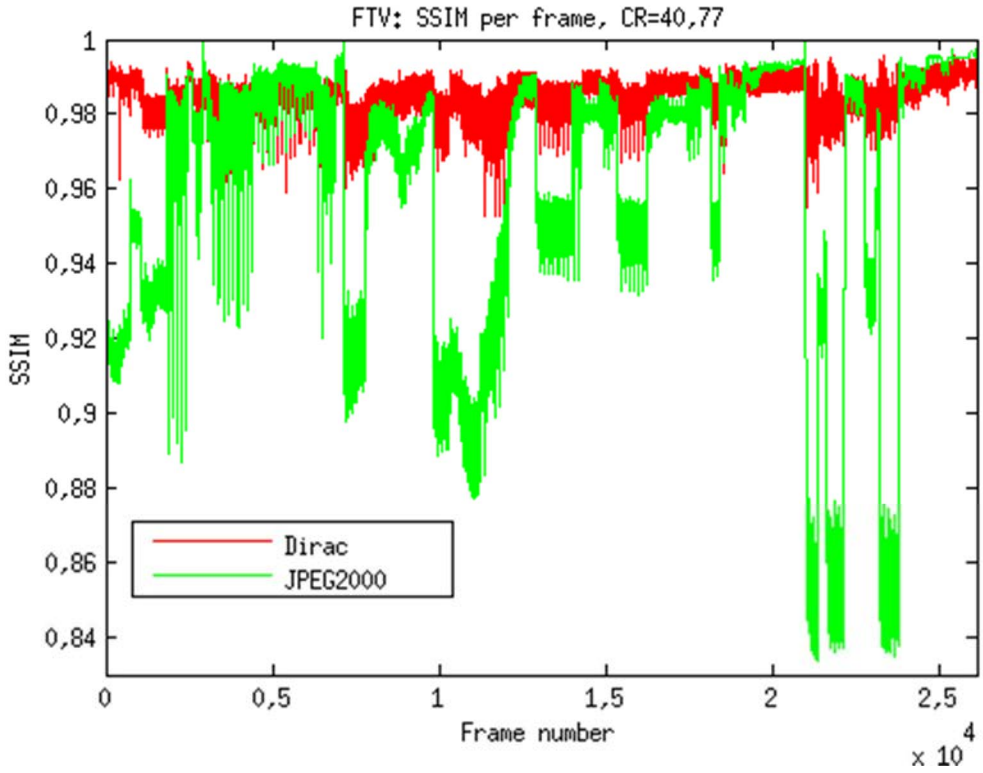


Figure 24: Y-SSIM frame by frame experimental results for "FTV" using the tandem schemes described in clause 6.2 on link 1 (MIMO 2×2) of table 12

8.2.4 Short summary of clause 8 and related results in annex A

This clause (and related results in annex A) compares the performance results by evaluating the three PHDMI techniques presented in clause 6 on UHD videos. The evaluations are initially conducted with homogeneous SNR characteristics on three UHD video sequences and then extended to the SNR database related to 500 PLT links. Besides known UHD video sequences "DucksTakeOff", "ParkJoy", "CrowdRun", the clause presented results on the real video "FTV" offered by France Télévision which is a more than 8 minutes sequence. The considered information video bit rate is mostly 5 Gbit/s. Tandem schemes at the PLT level have performance that varies between 50 Mbit/s to 400 Mbit/s, hence the compression encoder part for these scheme has been analysed up to a compression factor of 100. For the known video sequences, the tandem schemes are much better than the joint scheme. The "FTV" video performance shows that the Dirac tandem scheme is better than the JPEG 2000 scheme. It is important to observe that in this clause, it has been assumed that a feedback is present at the PLT level between the PHDMI sink and the PHDMI source.

9 Tandem schemes optimization

9.1 Introduction

Scope of clauses 9.2 and 9.3 is to analyse tandem schemes from two perspectives. In clause 9.2, some possible choices for the parameters of the source encoder are investigated. In clause 9.3, the resiliency of the source encoder to PLT errors is taken into consideration.

9.2 Source encoder parameter optimization

In figure 25, a comparison in terms of PSNR among wavelet (9,7) and wavelet (5,3) has been performed for JPEG 2000. The considered video is the UHD video ParkJoy described in clause 8.2. The video has been HEVC encoded and decoded at 18 Mb/s prior being JPEG 2000 encoded. Reference for PSNR computation is the HEVC decoded video (the PSNR due to HEVC encoding/decoding from the original video was 35,4115 dB). For this test, JPEG 2000 has been set with 6 resolution levels and a code-block size of 64 × 64. The comparison is performed for compression ratio in the range [5,100] on the PLT links described in clause 7.2.3.

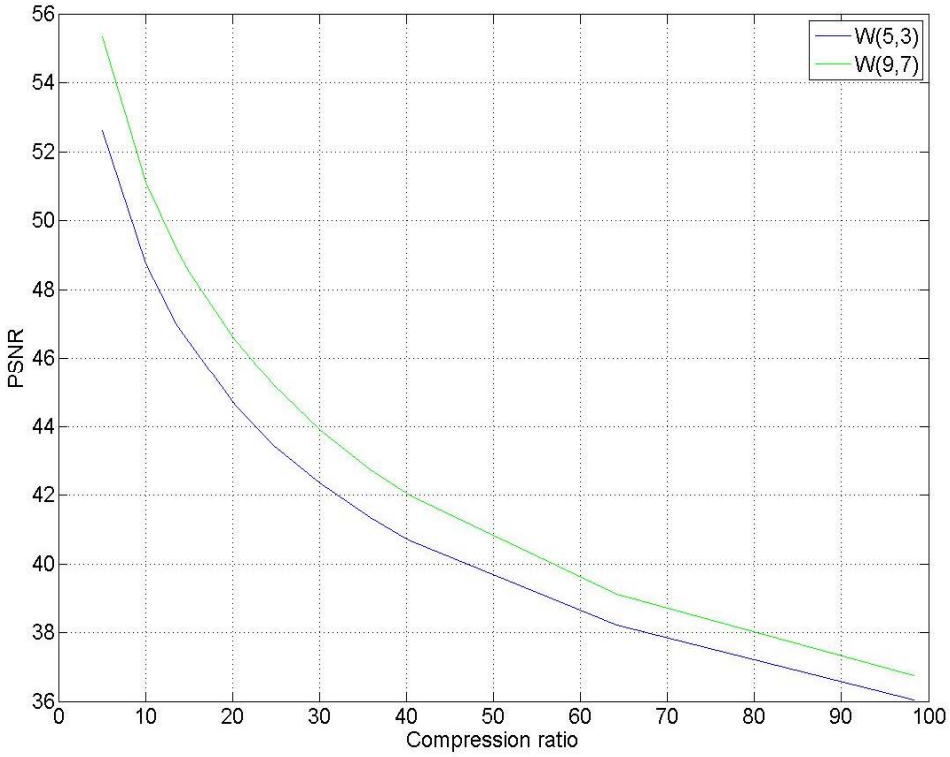


Figure 25: PSNR experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying wavelet parameters of the JPEG 2000 source encoder for the tandem scheme described in clause 6.2.2

Results reported in figure 25 show that wavelet (9,7) outperforms wavelet (5,3) in all the considered scenarios. This trend is also confirmed in terms of mean Y-SSIM metric (see table 13). For three specific PLT links, here numbered 1, 2 and 3, it is also possible to perform a comparison among PLT modems using SISO, MIMO 2×2 and MIMO 2×3, though it is not the main scope of this clause.

Table 13: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying wavelet parameters of the JPEG 2000 source encoder for the tandem scheme described in clause 6.2.2

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
W(5,3) Mean Y-SSIM	0,9832	0,9823	0,9739	0,9687	0,9594	0,9532	0,9466	0,9155	0,8814
W(9,7) Mean Y-SSIM	0,9889	0,9880	0,9815	0,9763	0,9670	0,9612	0,9544	0,9232	0,8843
Link	3	3	3	2	1	2	1	2	1

In figure 26 a comparison of different filters has been performed for Dirac in terms of PSNR: the GoP has been fixed to 48 and MV precision to ½ pixel for these simulations. The scenario is the same considered for JPEG 2000 in this clause.

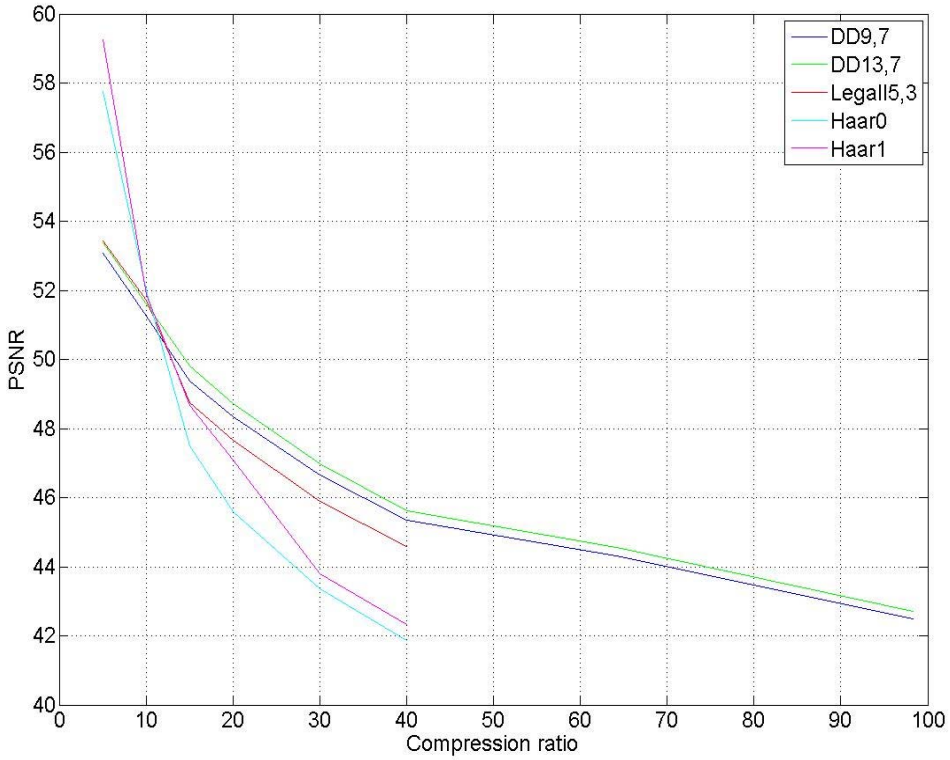


Figure 26: PSNR experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying filter parameters of the Dirac source encoder for the tandem scheme described in clause 6.2.3

It can be observed that DD13,7 and DD9,7 perform quite similar, while the other considered filters experiment more degradation with the increase of the compression ratio (i.e. when the PLT throughput decreases). This is also confirmed looking at the Y-SSIM metric (see table 14).

Table 14: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying filter parameters of the Dirac source encoder for the tandem scheme described in clause 6.2.3

Compression Ratio	5	10	15	20	30	40
DD13,7 Mean Y-SSIM	0,9947	0,9927	0,9901	0,9882	0,9847	0,9801
DD9,7 Mean Y-SSIM	0,9946	0,9925	0,9898	0,9879	0,9839	0,9798
Legall5,3 Mean Y-SSIM	0,9945	0,9923	0,9892	0,9869	0,9820	0,9773
Haar0 Mean Y-SSIM	0,9961	0,9912	0,9837	0,9782	0,9678	0,9585
Haar1 Mean Y-SSIM	0,9955	0,9916	0,9848	0,9796	0,9694	0,9602

For the same specific PLT links considered in this clause for JPEG 2000, a comparison between DD13,7 and DD9,7 in the case of a SISO PLT modem is considered in table 15. Here again the two filters appear quite equivalent with a slight preference for DD13,7.

Table 15: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying filter parameters of the Dirac source encoder for the tandem scheme described in clause 6.2.3: example on the 3 links considered in table 17 with SISO PLT communication with the two selected filters of table 14

Compression Ratio	20,53	64,13	98,3
DD13,7 Mean Y-SSIM	0,9874	0,9724	0,9617
DD9,7 Mean Y-SSIM	0,9871	0,9722	0,9613
Link	3	2	1

In figure 27 a comparison of Dirac with the available motion vectors is performed: half pixel, fourth pixel and eighth pixel. Results for MV1/4 and MV 1/8 are almost identical both in terms of Y-PSNR and Y-SSIM, while MV ½ is slightly worse, but still a good result. Similar results are also obtained in terms of mean Y-SSIM metric (see table 16), which reports results on three specific links using SISO, MIMO 2×2 and MIMO 2×3.

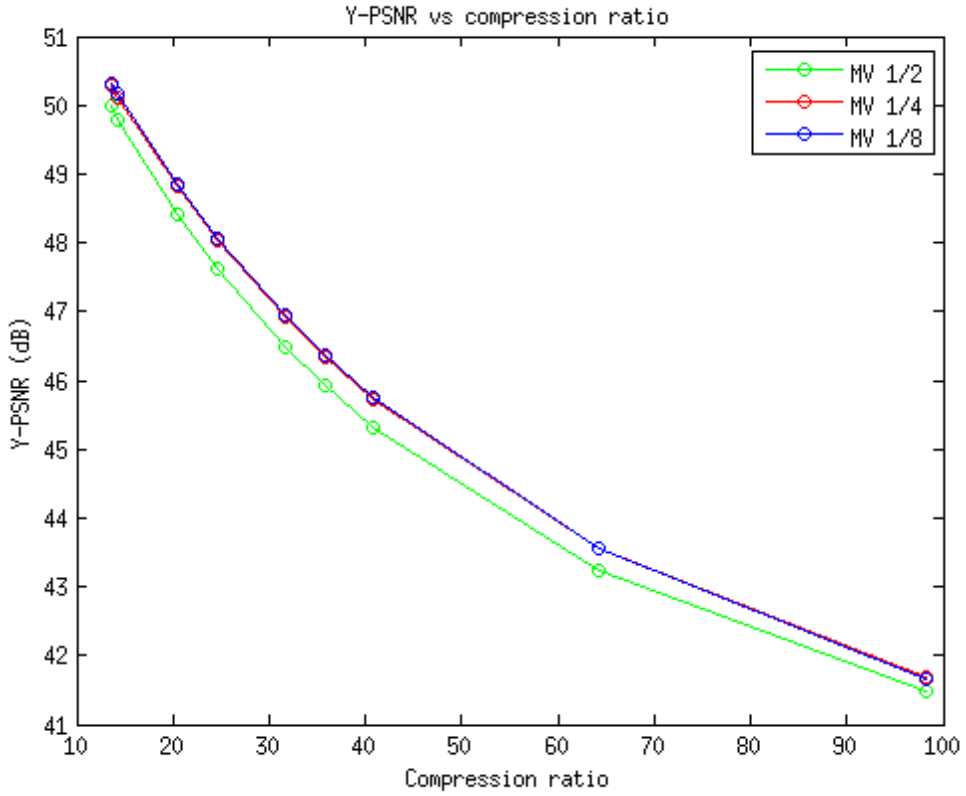


Figure 27: PSNR experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying MV parameters of the Dirac source encoder for the tandem scheme described in clause 6.2.3

Table 16: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) varying MVparameters of the Dirac source encoder for the tandem scheme described in clause 6.2.3 for 3 PLT links

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM Dirac MV 1/2	0,9904	0,9901	0,9876	0,9859	0,983	0,9812	0,979	0,9696	0,9583
Mean Y-SSIM Dirac MV 1/4	0,9909	0,9906	0,9884	0,9868	0,9841	0,9823	0,9803	0,9711	0,9595
Mean Y-SSIM Dirac MV 1/8	0,9909	0,9906	0,9884	0,9868	0,9841	0,9823	0,9802	0,9708	0,9588
Link	3	3	3	2	1	2	1	2	1

9.3 Source encoder resiliency to errors at the PLT level

Results for tandem schemes presented in previous clauses assume that the input to the source decoder is error free. This assumption is reasonable assuming that the retransmission of wrongly received packets occurs at the PLT level. Throughout this clause, the consequences of avoiding retransmission at the PLT level are investigated: in these conditions, the source decoder accepts some wrong bits at its input. The PLT BER assumed for this analysis is 10^{-8}.

In figure 28, a comparison in terms of PSNR has been performed for JPEG 2000. The considered video is the UHD video ParkJoy described in clause 8.2. The video has been HEVC encoded and decoded at 18 Mb/s prior being JPEG 2000 encoded. Reference for PSNR computation is the HEVC decoded video (the PSNR due to HEVC encoding/decoding from the original video was 35,4115 dB). For this test, JPEG 2000 has been set with 6 resolution levels and a code-block size of 64 × 64 and wavelet (9,7). The comparison is performed for compression ratio in the range [5,100] on the PLT links described in clause 7.2.3.

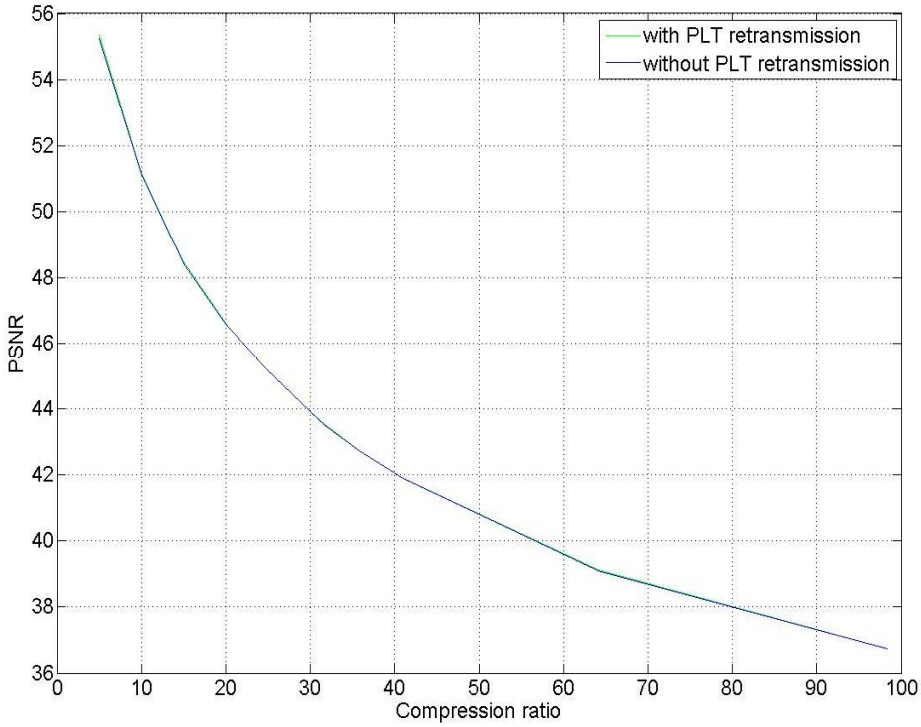


Figure 28: PSNR experimental results for UHD "ParkJoy" (HEVC encoded and decoded) with or without PLT errors at the JPEG 2000 source decoder input for the tandem scheme described in clause 6.2.3

As it can be seen, results of figure 38 show that, as long as the PLT BER is lower than or equal to 10^{-8}, JPEG 2000 is resilient to PLT errors and retransmission is not needed.

JPEG 2000 resiliency to spot PLT errors is also confirmed in terms of mean Y-SSIM metric (see table 17). Table 17 also allows verifying the resiliency for three specific PLT links, here numbered 1, 2 and 3, where SISO, MIMO 2×2 and MIMO 2×3 techniques have been used at the PLT level.

Table 17: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) with or without PLT errors at the JPEG 2000 source decoder input for the tandem scheme described in clause 6.2.3
(the table reports examples on 3 specific links with different PLT communication techniques).

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM (with PLT retransmission)	0,9889	0,9880	0,9815	0,9763	0,9670	0,9612	0,9544	0,9232	0,8843
Mean Y-SSIM (no PLT retransmission)	0,9889	0,9879	0,9814	0,9763	0,9668	0,9610	0,9544	0,9231	0,8843
Link	3	3	3	2	1	2	1	2	1

In figure 29, the impact of PLT errors at the source decoder has been assessed in terms average PSNR for Dirac. Dirac parameters for this figure are DD13,7 filter with GoP = 48 and MV precision of ½ pixel. As it can be seen, Dirac is not resilient to PLT errors.

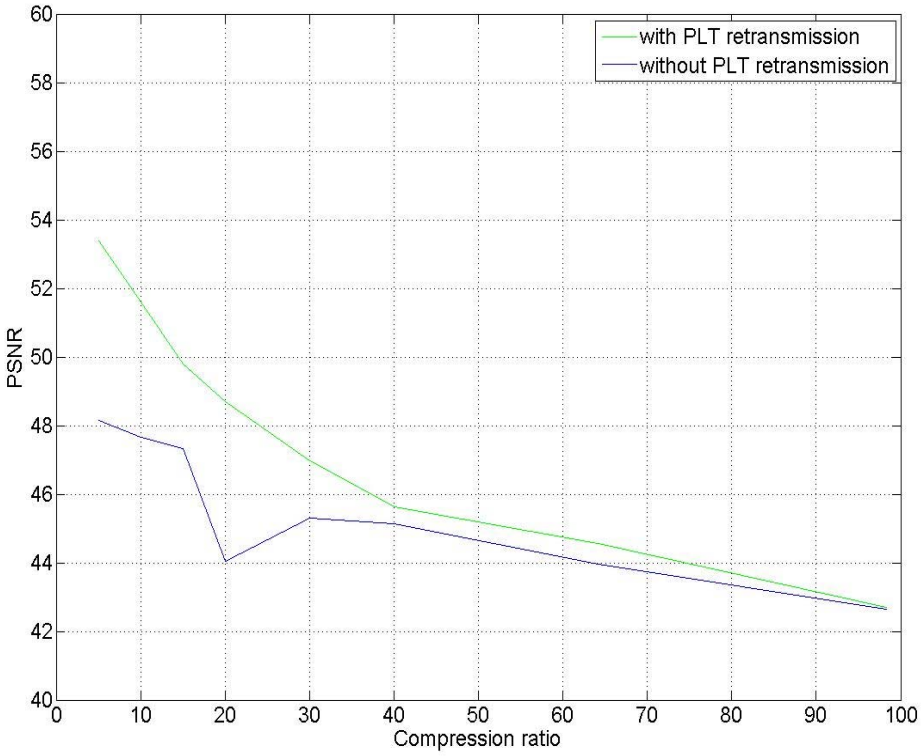


Figure 29: PSNR experimental results for UHD "ParkJoy" (HEVC encoded and decoded) with or without PLT errors at the Dirac source decoder input for the tandem scheme described in clause 6.2.3

The degradation shown in figure 29 is also confirmed in terms of Y-SSIM (see Table 18) on three specific PLT links.

Table 18: Y-SSIM experimental results for UHD "ParkJoy" (HEVC encoded and decoded) with or without PLT errors at the Dirac source decoder input for the tandem scheme described in clause 6.2.3 (SISO, MIMO 2×2 and MIMO 2×3)

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM (with PLT retransmission)	0,9898	0,9895	0,9874	0,9860	0,9834	0,9821	0,9803	0,9724	0,9617
Mean Y-SSIM (no PLT retransmission)	0,9783	0,9825	0,9843	0,9686	0,9742	0,9785	0,9783	0,9689	0,9614
Link	3	3	3	2	1	2	1	2	1

It is worthy to observe that the degradation does not uniformly affect different frames: only the frames that are in the zone where the PLT error occurs are affected: moreover, there is error propagation due to motion estimation. This fact means that locally the degradation is higher compared to the one reported in figure 29 and in table 18 as it can be seen in the following graph (see figure 30). Local degradation can last for a whole GoP.

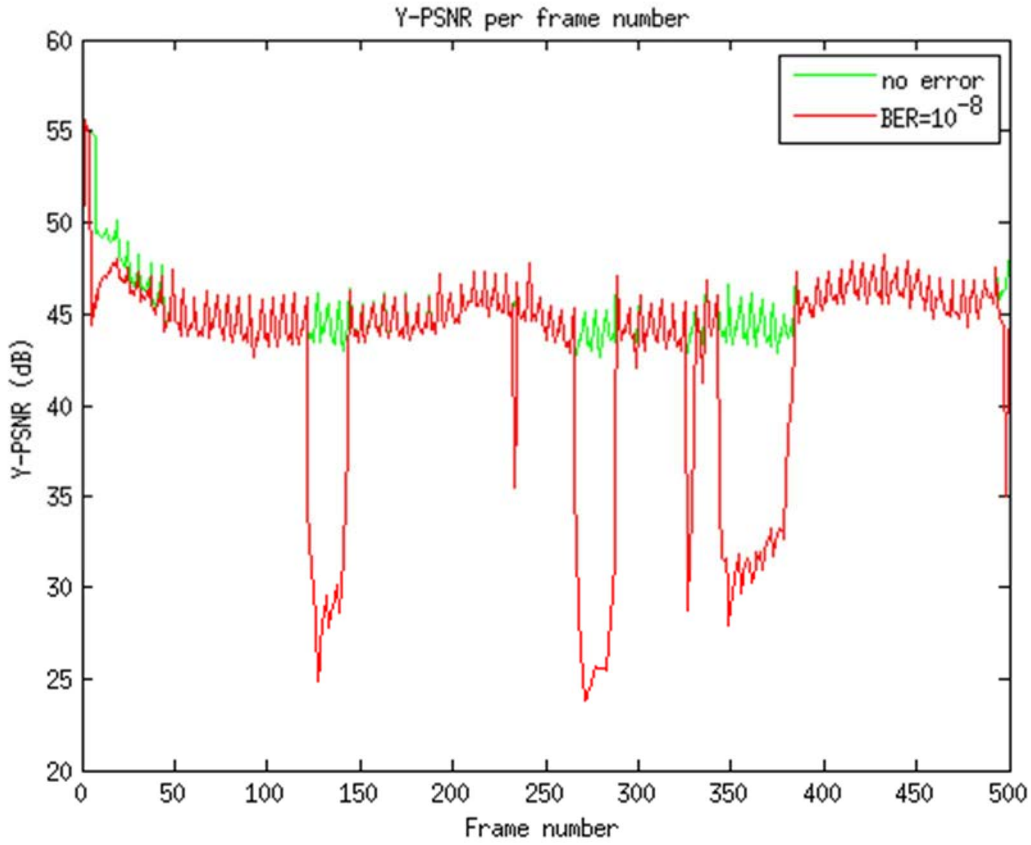


Figure 30: Y-PSNR experimental results per frame for UHD "ParkJoy" (HEVC encoded and decoded) with or without PLT errors at the Dirac source decoder input for the tandem scheme described in clause 6.2.3 on link 1 of table 18 with MIMO 2×2 (CR = 40,77)

9.4 Short summary for clause 9

This clause shows tandem scheme optimization. The JPEG 2000 tandem scheme is optimized with wavelet (9,7). Moreover, as long as the PLT BER is lower than 10^{-8}, the JPEG 2000 based scheme does not require acknowledgements at the PLT level between the PHDMI sink and the PHDMI source. The Dirac based scheme is optimized with the DD13,7 filter though the DD9,7 filter has also reasonable performance. Dirac MV precision of ½ pixel seems sufficient to obtain good results, while practical advantage seem limited in strengthening the precision to MV = 1/4 or MV = 1/8, which can be optional in a specification. Using frame by frame analysis, the clause shows that packet acknowledgement and retransmission at the PLT level is essential for the Dirac tandem scheme, otherwise error propagation would results in locally unacceptable quality drops.

10 Rate controller

10.1 Introduction

In the previous clauses, one has evaluated the performance of tandem and joint schemes assuming that the channel conditions are time invariant and perfectly known. The main issues addressed in this clause are related to the robustness of the HD/UHD PHDMI scheme to variations of the channel characteristics and to impulsive noise, which is unavoidable in the considered context.

10.2 Model of the variation of the channel characteristics

10.2.1 General

Two types of variations of the channel characteristics will be considered in what follows (a permanent decrease of the PLT capacity, see clause 10.2.2 and a temporary decrease of the PLT capacity in clause 10.2.3). Note that the rate controller is supposed to work well when there is an increase of the PLT capacity.

10.2.2 Permanent decrease of the PLT capacity

The first is a permanent decrease of the channel capacity $R^c(t)$, due to a permanent decrease of the SNRs $s_i(t)$, $i = 1,\ldots,N_c$ of the N_c subcarriers at some time instant t_0. When the SNR decrease is larger than the margin considered when the tone mapping $m_i(t)$, $i = 1,\ldots,N_c$ used for $t < t_0$, the bit-error rate at the input of the FEC code (for instance the turbo code in the HomePlug® AV2 case) used at the PHY layer will be too large to allow a good error correction. As a consequence, highly corrupted packets will reach the PLT receiver, which will not be able to exploit them. In absence of acknowledgement mechanisms, this leads to packet losses at the application layer of the receiver, which may have a dramatic impact on the received quality. This is why some acknowledgement mechanism has to be implemented at the MAC layer of the PLT modem to allow retransmission of lost packets and a redesign of the tone map, see figure 31.

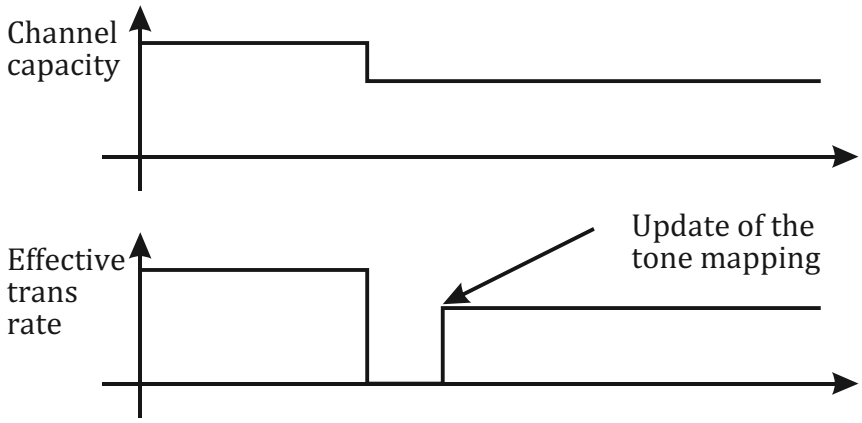


Figure 31: Update of the tone map after a decrease of the channel capacity

A decrease of the value of some $s_i(t)$ at t_0, when redesigning the tone map, results in a reduction of the size of the constellation used for the corresponding subcarriers. This leads to decrease of the effective transmission rate of the PLT link. The permanent decrease of the SNRs is not detected immediately. The delay for receiving successive NACKs and redesigning the tone map is assumed to be upper-bounded by τ_m. In what follows, the PLT capacity is assumed null between t_0 and $t_0 + \tau_m$.

10.2.3 Temporary decrease of the PLT capacity due to impulsive noise

The second source of decrease of the channel capacity is due to impulse noise, which could significantly, but temporarily reduces $s_i(t)$, for all $i = 1,\ldots,N_c$. The consequence is a significant increase of the BER at the input of the FEC decoder. This leads to temporary packet losses. Some models have been considered to described the characteristics of impulse noise, see e.g. [i.5], [i.6], [i.7] and [i.8].

In [i.5], six classes of impulses are identified, depending on their origin:

- Class 1: Electrical switch and thermostat ON event
- Class 2: Electrical switch and thermostat OFF event
- Class 3: Electrical plug plugging
- Class 4: Electrical plug unplugging
- Class 5: Electrical motor start
- Class 6: Diverse weak noise signatures

The class determines the distribution of the amplitude, of the duration of the impulse, and its shape. The duration of the impulse of Class 1 to 4 is described by an exponential distribution

$$f(d) = \lambda \exp(-\lambda d) \quad (10)$$

where d is the duration of the impulse and λ is between $4{,}10^4\ s^{-1}$ and $10^5\ s^{-1}$. Typical impulse durations at the PLT receiver are of the order of tens of microseconds.

In what follows it is assumed that an impulse occuring at time t_0 has an impact on the PLT capacity between t_0 and $t_0 + \tau_i$. During this time interval, the capacity of the PLT is assumed null in order to deal with a worst case situation.

10.3 Mitigating the effect of PLT capacity variations

10.3.1 General

Two mechanisms are employed in what follows to mitigate the effect of variations of the PLT capacity, namely *encoding rate adaptation* (see clause 10.3.2) and *layer filtering* (see clause 10.3.3). Both require the presence of a *transmission buffer* to store the packets containing the encoded video frames prior to transmission and a *reception buffer* at receiver side. This buffer consists of two parts: the *packet buffer*, where received packets are stored before being decoded and a *frame buffer* in which decoded frames are stored before being displayed.

10.3.2 Encoding rate adaptation

This first mechanism involves an observation of the number of encoded frames in the transmission buffer. When this number increases, due to a permanent decrease of the PLT capacity or to an impulse, the number of frames in the frame buffer decreases. In order to restore $n_F(t)$ above M_F, an adaptation of the encoding rate is performed using the scheme described in figure 32.

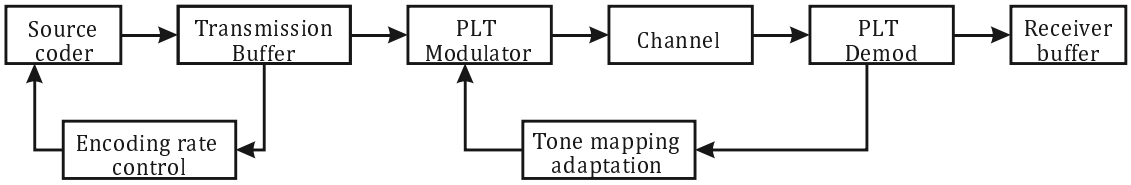


Figure 32: Encoding rate adaptation scheme

Assume that initially, the delay between the outcome from the HDMI® cable of the first frame of the first GoP and the display of this frame at receiver is Δ. Since the frame rate at the output of the HDMI® cable is constant and equal to F, the number of encoded frames within the transmission buffer can provide an estimate of the number of frames in the frame buffer at the receiver.

A possible implementation of the Encoding rate control is via a Proportional-Integral control of the number of frames in the transmission buffer.

The main advantage of this technique is an accurate adaptation of the encoding rate to the PLT capacity. Its main disadvantage is the delay with which the number of frames at the receiver will reach again M_F, since the transmission buffer has to be flushed first. This is illustrated by figure 33. The level and number of frames in the reception buffer increases relatively slowly after some delay, due to the frames that have to be flushed first from the encoder buffer.

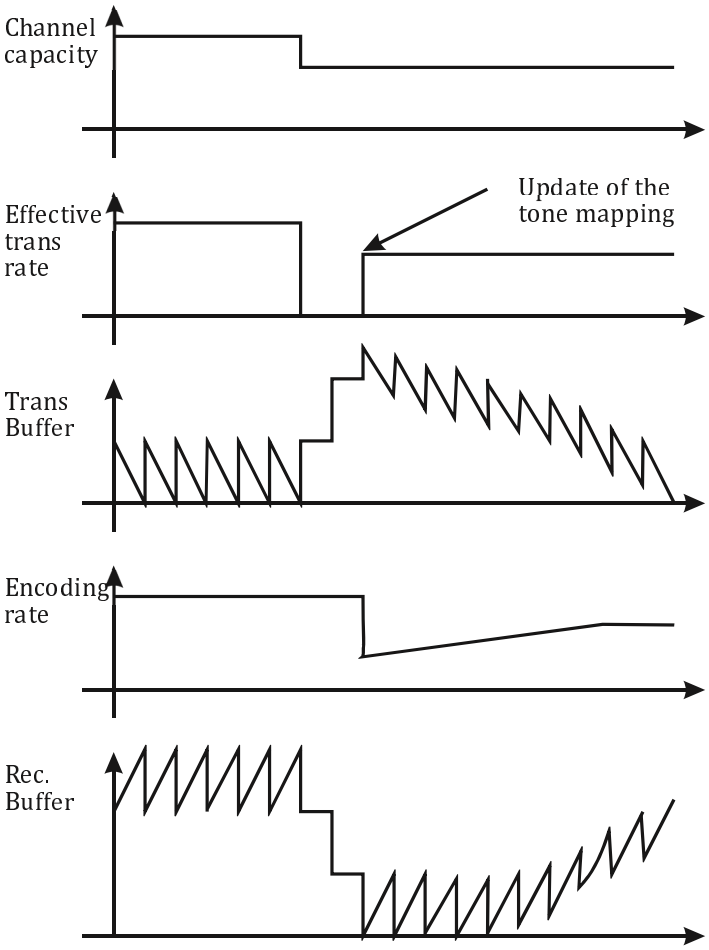


Figure 33: Illustration of the encoding rate adaptation scheme

10.3.3 Layer filtering

Packets generated by JPEG 2000 and Dirac are organized in n_L quality layers. The first (low-index) quality layers contain the most important information, allowing decoding the frames with the minimum quality. The next (high-index) quality layers provide improved quality. A way to efficiently drive $n_F(t)$ above M_F after a temporary decrease of the PLT capacity is to remove packets corresponding to high-index quality layers from the transmission buffer, see figure 34. The main advantage of this approach is a quick adaptation. Its main disadvantage is the coarse granularity at which the control can be performed. Moreover, this technique requires tagging the encoded packets with their corresponding quality layer. This is illustrated by figure 35. Filtering the layer in green at the transmitter, allows increasing the number of frames buffered at the receiver. The price to be paid is a quality decrease.

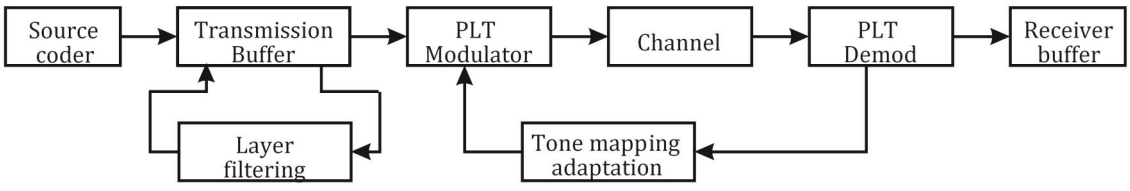


Figure 34: Layer filtering scheme

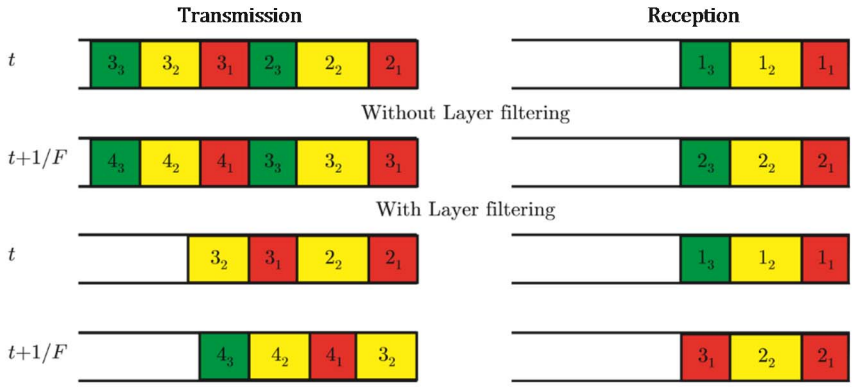


Figure 35: Ilustration of the layer filtering mechanism

10.3.4 Determining the buffers size

The frame buffer should always contain enough decoded frames to be able to compensate for permanent or impulsive PLT capacity decreases. Let $n_F(t)$ be the number of video frames in the frame buffer at time t. The encoding rate control and the layer filtering mechanism have to ensure that $n_F(t)$ remains above some threshold M_F that depends on the frame rate F, the number of frames in a GoP N_F, the values of τ_m and τ_i.

To determine M_F, assume that some capacity drop occurs at time t_0 (due to impulse noise or to a decrease of the carrier SNRs), leading to a null capacity of the PLT channel between t_0 and $t_0 + \tau$. The number of frames in the frame buffer at $t_0 + \tau$ is:

$$n_F(t_0+\tau) = n_F(t_0) - \tau F. \tag{11}$$

After $t_0+\tau$, encoded packets arrive again at the reception buffer and may be decoded once all packets corresponding to a GoP have been received. One considers the worst case, in which (i) the reception buffer is empty at t_0, and thus remains empty at $t_0 + \tau$ and (ii) the last frame in the frame buffer corresponds to the end of a GoP. Assume that packets have been encoded before t_0 at some rate R_1 and that the PLT rate after $t_0 + \tau$ is R_2. In absence of variation of the PLT channel characteristics after $t_0 + \tau$, and neglecting the transmission delay, the time at which the decoding of a new GoP may be performed from packets in the packet buffer is:

$$t_1 = t_0 + \tau + \frac{R_1}{R_2} N_F/F, \tag{12}$$

where $R_1 N_F/F$ is the number of bits of a previously encoded GOP at a rate R_1. If the decoding delay is T_d and neglecting the decoded frame bufferization, a new GoP is available in the frame buffer at time:

$$t_2 = t_0 + \tau + \frac{R_1}{R_2} N_F/F + T_d \tag{13}$$

Between $t = t_0 + \tau$ and $t = t_2$, the frame buffer is drained at a rate F frames per second. The number of frames at t_2, before the arrival of the decoded GOP is:

$$n_F(t_2) = n_F(t_0) - \left(\tau + \frac{R_1}{R_2} N_F/F + T_d\right) F \tag{14}$$

and has to be positive. In general, a given number K of GoPs have been encoded at a rate R_1 and are stored in the transmission buffer. The time at which the last frame of these K GoPs has reached the frame buffer is:

$$t_3 = t_0 + \tau + K\frac{R_1}{R_2} N_F/F + T_d \tag{15}$$

At t_3, one has:

$$\begin{aligned} n_F(t_3) &= n_F(t_0) + KN_F - \left(\tau + K\frac{R_1}{R_2} N_F/F + T_d\right) F \\ &= n_F(t_0) + K\left(1 - \frac{R_1}{R_2}\right) N_F - (\tau + T_d) F. \end{aligned} \tag{16}$$

For JPEG 2000, encoding and decoding is performed frame-by-frame. In this case, $N_F = 1$. Once all packets corresponding to an encoded frame are available in the packet buffer, they are processed by the decoder and the decoded frame is put in the frame buffer. For SoftCast and Dirac, encoding is performed at a GoP level, with $N_F > 1$. The decoder has to receive all packets of the encoded GoP before being able to put the decoded frames in the frame buffer. Several cases have to be considered, depending of the type of capacity change.

In case of an impulse noise, $R_1 = R_2$. In this case, (16) boils down to:

$$n_F(t_3) = n_F(t_0) - (\tau_i + T_d)F, \tag{17}$$

which imposes that:

$$MF > MF,i \tag{18}$$

with

$$MF,i = (\tau i + Td)F. \tag{19}$$

In case of a permanent PLT capacity decrease, R2 < R1 and (16) becomes:

$$n_F(t_3) = n_F(t_0) + K\left(1 - \frac{R_1}{R_2}\right)N_F - (\tau_m + T_d)F, \tag{20}$$

which imposes:

$$MF > MF,m \tag{21}$$

with

$$M_{F,m} = K\left(\frac{R_1}{R_2} - 1\right)N_F + (\tau_m + T_d)F. \tag{22}$$

To illustrate the previous results, assume that $\tau_i = 1$ ms, $\tau_m = 40$ ms, $T_d = 0$, $F = 50$ fps, $K = 2$, and $R_2 = 0{,}7R_1$. In this case, (18) becomes:

$$MF > 0{,}05 \text{ frames}. \tag{23}$$

Table 19 provides the evolution of $M_{F,m}$ as a function of the number of frames N_F in a GoP.

Table 19: Number of decoded frames to store in the frame buffer and corresponding buffering delay (ms) to avoid any freeze due to the decrease of the PLT capacity

N_F	1	2	4	8	16
$M_{F,m}$	2,85	3,71	5,43	8,46	15,71
$M_{F,m}/F$ (ms)	57	74	109	169	314

The effect of the encoding rate adaptation mechanisms is to encode frames at a rate $R'_2 < R_2$ during a certain amount of time after t_2. Frames are then encoded at a rate R_2 corresponding to that of the channel. Assume that K^0 GoPs are encoded at a rate R'_2, then the number of frames in the frame buffer at time $t_4 = t_3 + K^0 N_F/F$ will be:

$$n_F^{RA}(t_4) = n_F(t_0) + K\left(1 - \frac{R_1}{R_2}\right)N_F - (\tau_m + T_d)F + K'\left(1 - \frac{R'_2}{R_2}\right)N_F \tag{24}$$

Since $R'_2 < R_2$, the last term in (24) allows restoring the number of frames to M_F. This restoration may be fast when R'_2 is much less than R_2, the price to be paid is a more significant decrease of the quality of the encoded video. So, a compromise has to be found between quality decrease and speed of the buffer level restoration. One observes, however, that this restoration starts only when the frames encoded at a rate R_1 are flushed from the transmission buffer.

The effect of the layer filtering mechanisms is to drop packets containing high quality layers in the transmission buffer. This is more or less equivalent to assume that the frames contained in the encoded packets stored in the transmission buffer have been encoded at a rate $R"_2$, where $R"_2$ depends on the number of filtered frames. With this approach, (20) becomes:

$$n_F^{LF}(t_3) = n_F(t_0) + K\left(1 - \frac{R"_2}{R_2}\right)N_F - (\tau_m + T_d)F. \tag{25}$$

An appropriate selection of the number of filtered layers with $R"_2 < R_2$ allows to compensate for the drain of $(\tau_m + T_d)F$ frames that has occurred during the capacity drop. The main advantage of this approach is the buffer level restoration takes less time, but an encoding rate adaptation has still to be performed.

10.4 Illustration

10.4.1 General

In all the following simulations, the first 2000 frames of the FTV video sequence have been encoded with JPEG 2000 with $n_L = 6$ quality layers. A buffering delay $\Delta = 5$ frames is considered.

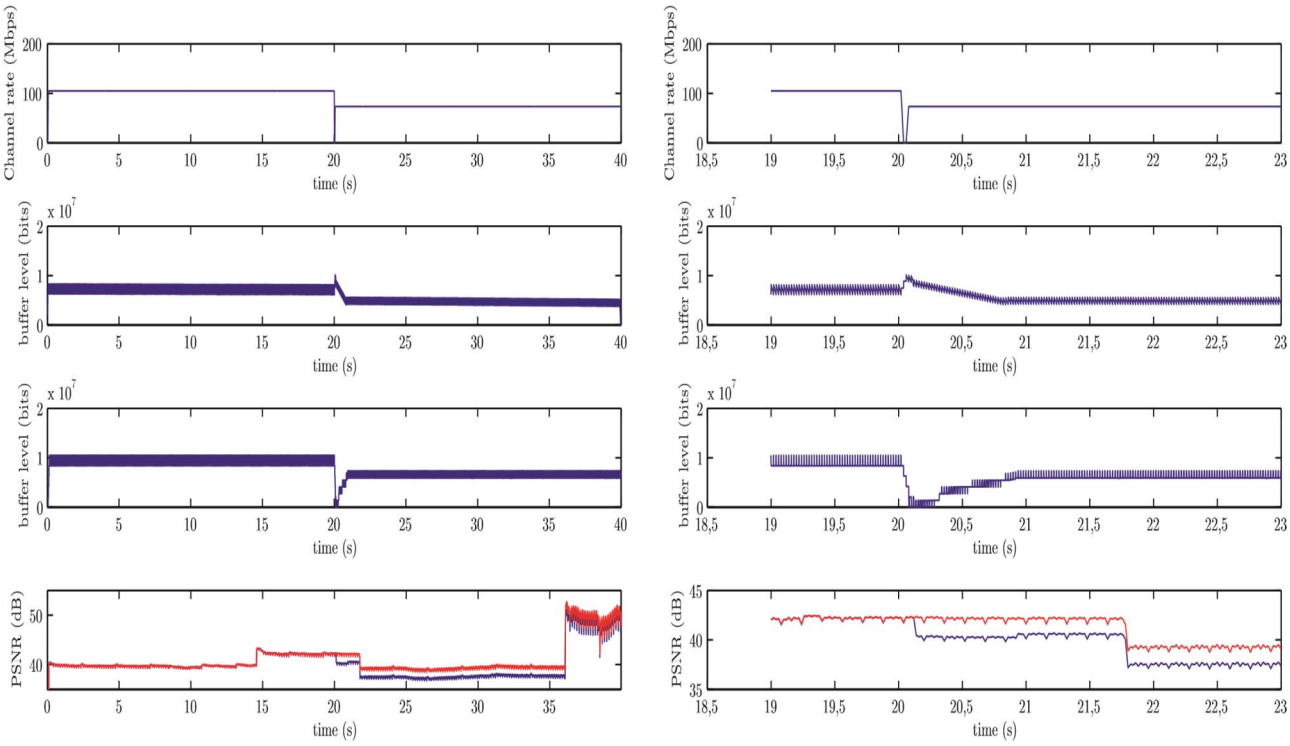
10.4.2 Rate adaptation mechanism

In this first study, the capacity of the PLT channel decreases at $t = 20$ s from $R_1 = 100$ Mb/s to $R_2 = 70$ Mb/s. The delay to update the tone map is taken as $\tau_m = 40$ ms. Just after the update of the tone map, the transmitter buffer has accumulated $\tau_m R_1 = 4$ Mb which have to be flushed. Let's assume that a tolerated flush time τ_F is considered. Then, the encoder has to encode at a rate:

$$R_e(t) = R_2 - \frac{\tau_m R_1}{\tau} \tag{26}$$

between $t + \tau_m$ and $t + \tau_m + \tau_F$. Then, the new encoding rate can be $R_e(t) = R_2$.

Figures 36 and 37 illustrates the evolution of the buffering levels and of the PSNR and SSIM at the receiver when this control strategy is put at work with $\tau_F = 0{,}8$ s. The quality does not decrease immediately after $t = 20$ s, due to the frames remaining in the transmission buffer that have to be flushed first. At $t = 20{,}84$ s, the quality increases slightly, due to the encoding rate increase to R_2. The quality decrease at time $t = 21{,}8$ s is due to a change in the characteristics of the encoded video.

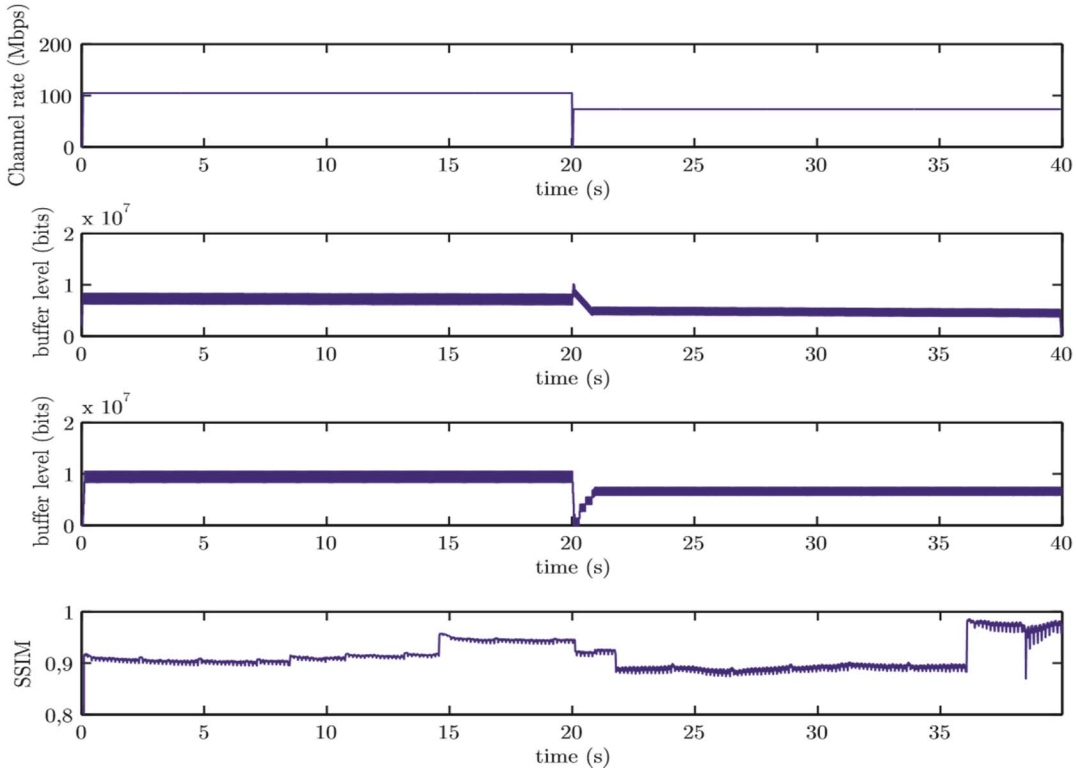


NOTE: On the left panels: the whole sequence. On the right panels: zoom. PSNRmin = 36,7 dB, PSNRmax = 51,2 dB for the whole sequence. Reference PSNR at constant rate is depicted in red.

Figure 36: Illustration of the encoding rate adaptation mechanism

In presence of random impulses of duration $\tau_i = 1$ ms, the previous buffer level control mechanism is relatively inefficient, as illustrated by figure 38.

Due to the impulse noise, the number of frames of the frame buffer progressively decreases. After the fourth impulse, the frame buffer being empty, a freeze occurs. Moreover, the frame buffer is not able to mitigate the effect of the drop of the PLT capacity at $t = 20$ s. This shows that a continuous adjustment of the encoding rate is necessary. Alternatively, the rate adaptation technique can be combined with layer filtering techniques.



NOTE: On the left panels: the whole sequence. On the right panels: zoom. SSIMmin = 0,869 and SSIMmax = 0,987 for the whole sequence.

Figure 37: Illustration of the encoding rate adaptation mechanism

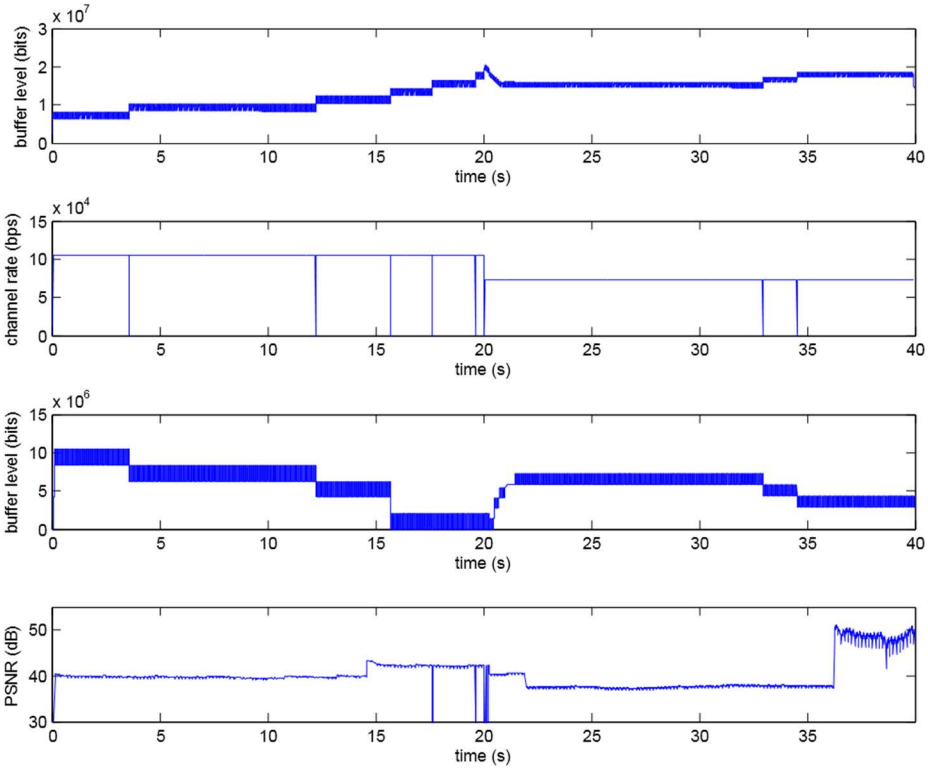


Figure 38: Illustration of the layer filtering mechanism in presence of impulse noise, when the three last quality layers are filtered when the number of encoded frames in the transmission buffer is above 4

10.4.3 Illustration of the layer filtering approach

The layer filtering approach is put at work when the number of frames in the transmission buffer is larger than 4. Figures 39 and 40 show that the layer filtering mechanism with a filtering of the two less important quality layers leads to a decrease of the PSNR when used (here around t = 12,5 s). The good point is that the permanent rate decrease is perfectly mitigated.

Figures 41 and 42 shows that the layer filtering mechanism with a filtering of the less important quality layer leads to a smaller decrease of the PSNR when used (here around t = 12,5 s). Since the margin is less important, when the permanent decrease is observed, a small drop of the PSNR is observed when the permanent capacity decrease is observed.

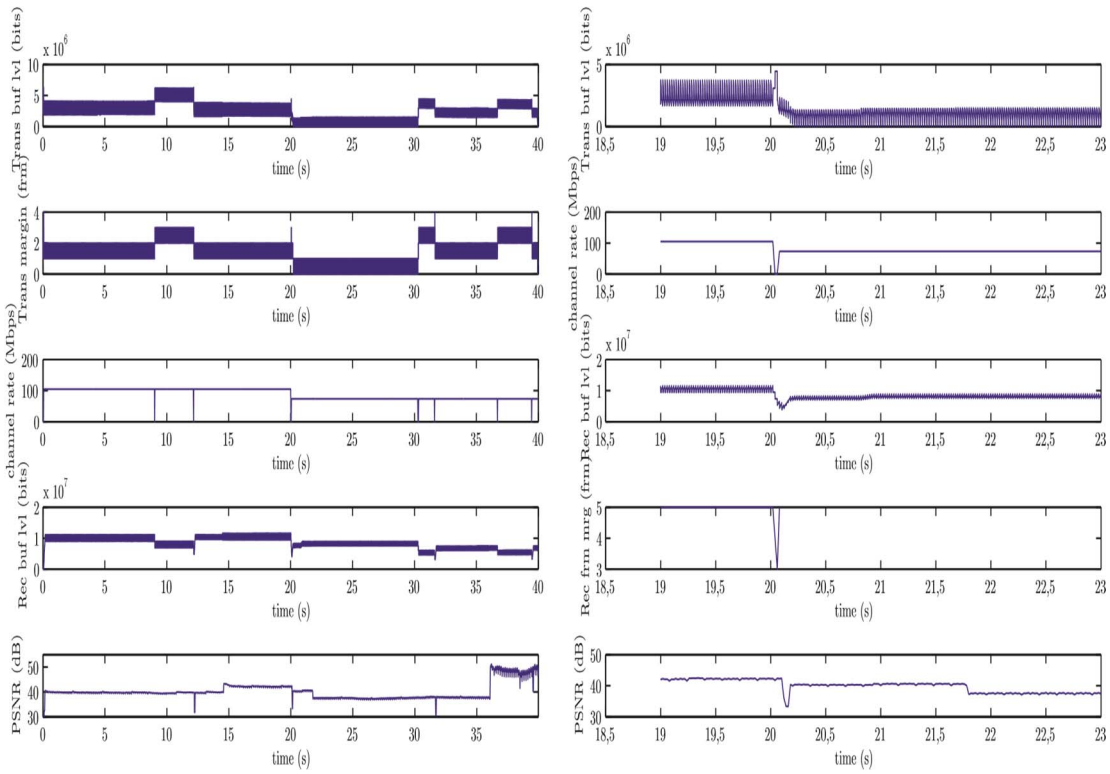


Figure 39: Illustration of the layer filtering mechanism in presence of impulsive noise, when two last quality layers are filtered when the number of encoded frames in the transmission buffer is above 2; $PSNR_{min}$ = 30,5 dB, $PSNR_{max}$ = 51,2 dB for the whole sequence

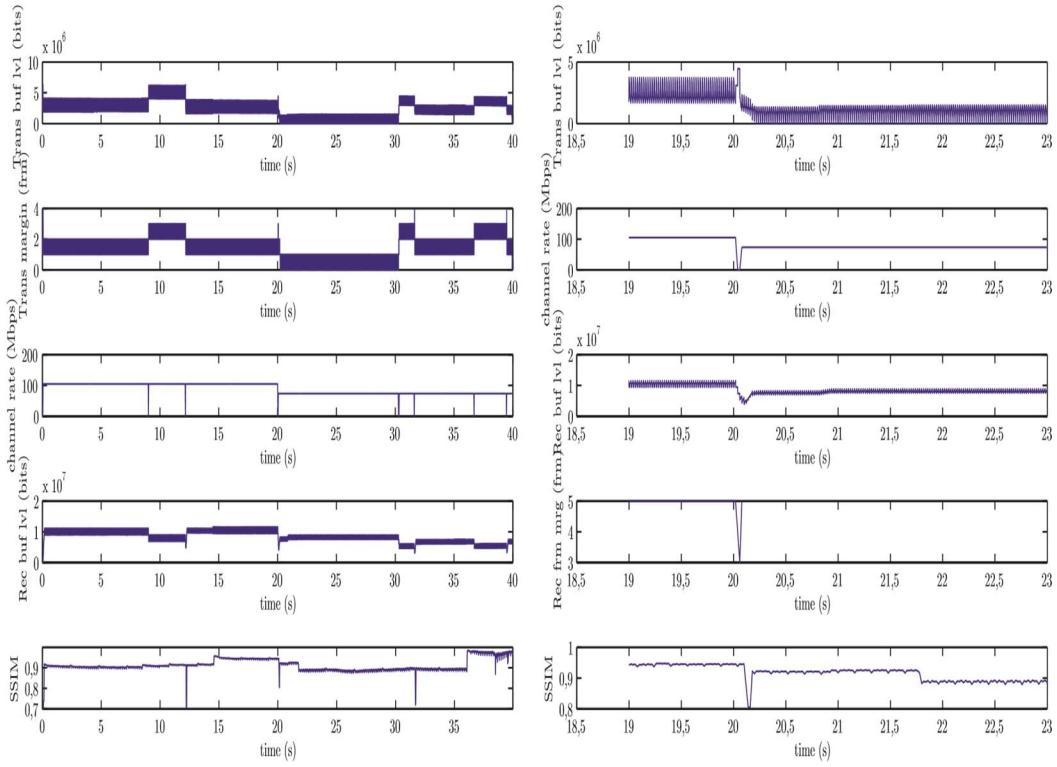


Figure 40: Illustration of the layer filtering mechanism in presence of impulsive noise, when the two last quality layers are filtered when the number of encoded frames in the transmission buffer is above 2; $SSIM_{min}$ = 0,7 and $SSIM_{max}$ = 0,988 for the whole sequence

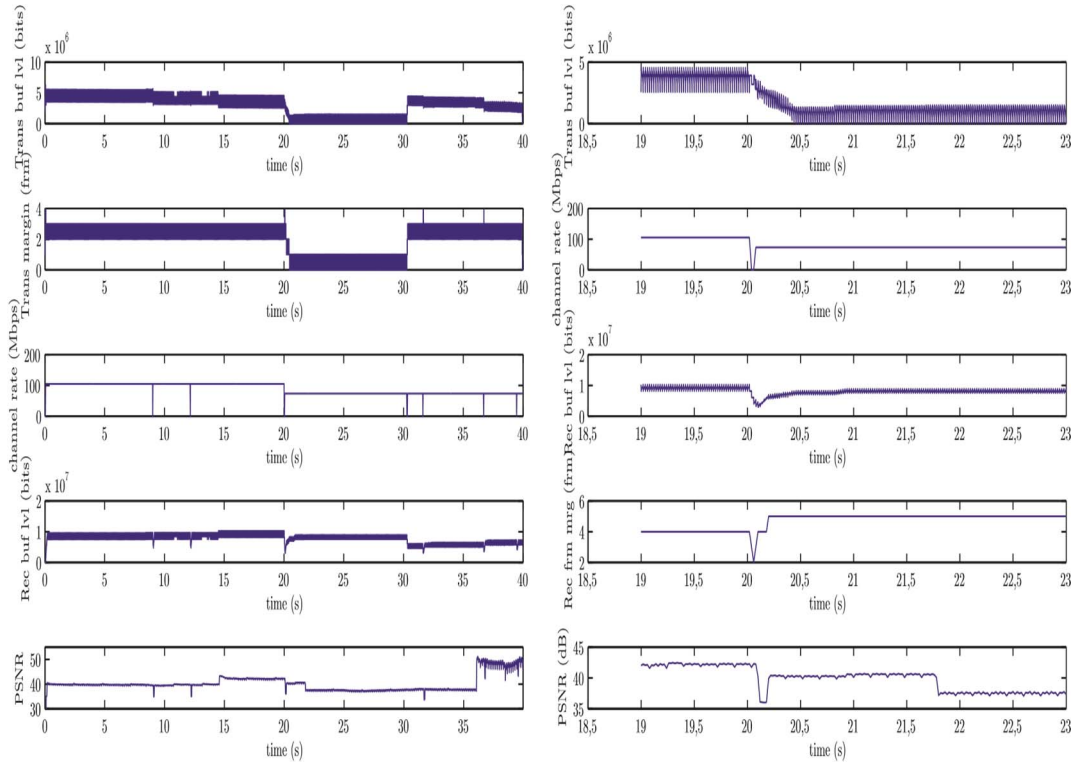


Figure 41: Illustration of the layer filtering mechanism in presence of impulsive noise, when the last quality layer is filtered when the number of encoded frames in the transmission buffer is above 2; $PSNR_{min}$ = 33,5 dB, $PSNR_{max}$ = 51,2 dB for the whole sequence

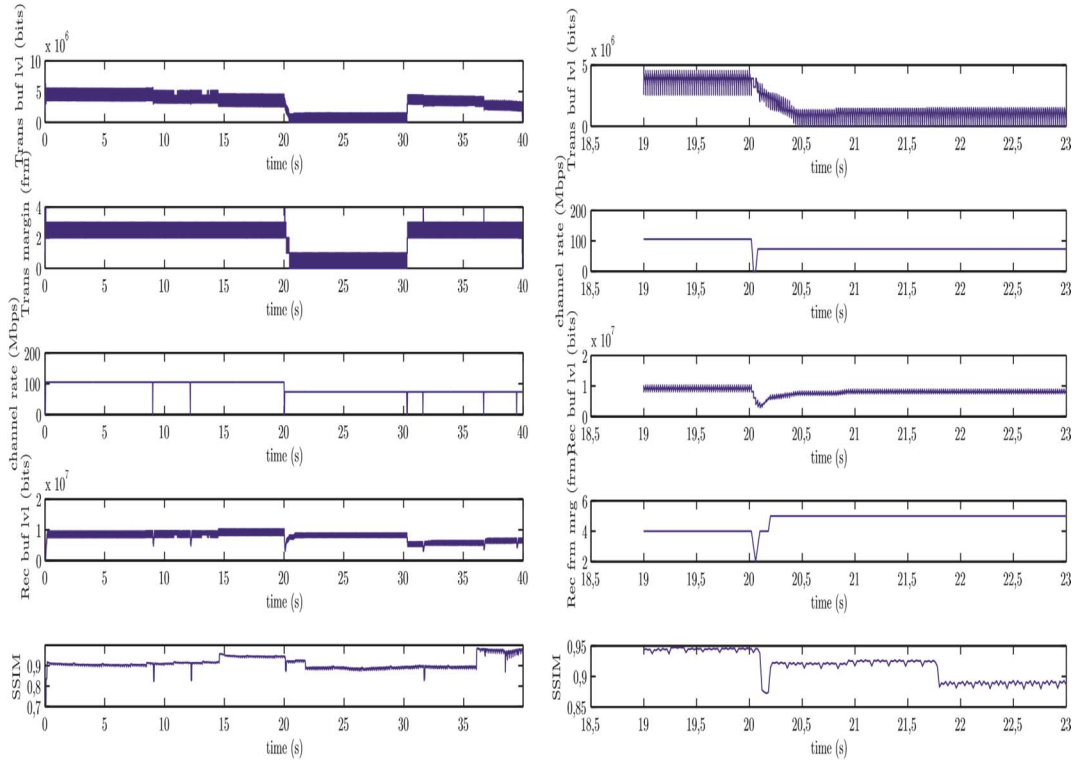


Figure 42: Illustration of the layer filtering mechanism in presence of impulsive noise, when the last quality layer is filtered when the number of encoded frames in the transmission buffer is above 2; $SSIM_{min}$ = 0,82, $SSIM_{max}$ = 0,988 for the whole sequence

10.5 Short summary of clause 10

Rate adaptation and layer filtering are efficient mechanisms to mitigate the effects of impulse noise and of decrease of the PLT capacity. Both require a buffer to store compressed frames at the transmitter and the decoded frames at the receiver. The rate adaptation mechanism provides a fine adaptation the channel capacity, but its effects are delayed. The layer filtering technique is efficient, low-delay, but provides a relatively coarse adjustment of the buffering level.

Combining both schemes provides the best results: fast restoration of the buffering level at receiver and fine adaptation in permanent regime.

11 Conclusions

From this technical report, one can draw the following recommendations for a future PHDMI specification. Tandem schemes are more mature than joint schemes since they use more consolidated capabilities both at the source encoder (compression codecs) and at the channel encoder level (SISO or MIMO PLT on 2 MHz to 100 MHz. In the present document PLT rates between 50 Mbit/s and 400 Mbit/s have been considered at the application level). Considering that the video bit rate will increase in the future (8K ?!), the Dirac-based scheme, with the optimized parameters of clause 9, appears offering better perspectives since it has more compression capabilities than the JPEG 2000-based scheme. This was particularly apparent when considering the real UHD video provided by France Télévision (see clause 8). To insure QoS, at the PLT level a TDMA scheme is recommended, while packet acknowledgement and retransmission have been proven essential for this tandem scheme works (see clause 9). The rate controller principles described in the present document (see clause 10) are a possibility in order to interface the source and channel encoding parts and to face even abrupt changes in the PLT rate. Joint schemes have some nice properties such as being more compact, and not requiring link adaptation or rate controlling, but cannot be recommended as principal solution. Since nobody can exclude that joint schemes will improve in the future to reach similar or better performance than tandem schemes, one possibility that can be considered is specifying two profiles: a Dirac tandem scheme profile and a joint scheme profile.

Annex A:
Additional results on HD and UHD video sequences

A.1 General

In the main text, the most significant results are reported. In particular, the most important video that has been analysed is the "FTV" sequence which is a real very long UHD sequence. In this annex, additional results derived on other typical test video sequences are reported, for the sake of completeness. It is worthy to notice that the short summary at the end of clause 7 and clause 8 also takes into account the hereinafter information.

A.2 Additional results on HD sequences

Results in figures A.1 and A.12 and in tables A.1 and A.2 are obtained with homogeneous PLT SNRs.

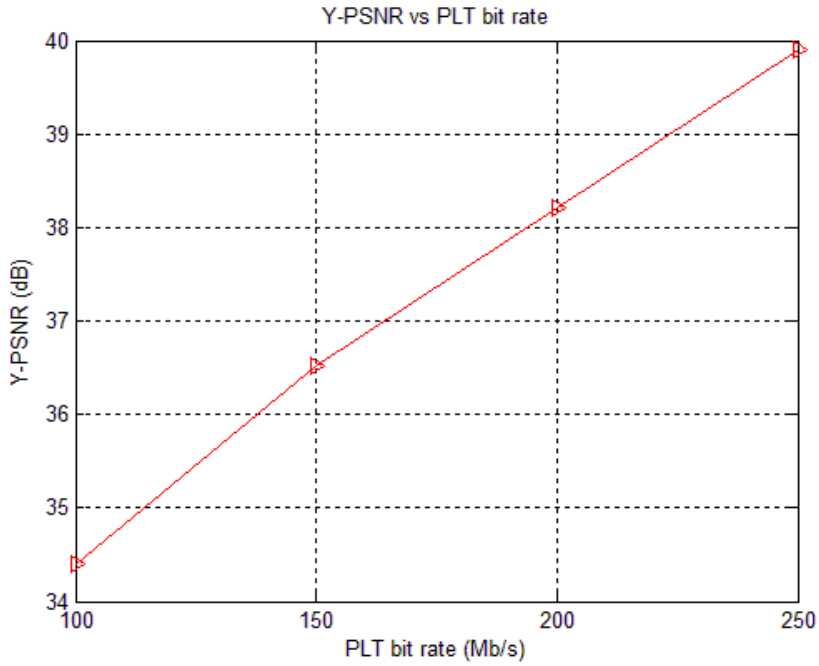


Figure A.1: Experimental results for "DucksTakeOff" using the tandem scheme described in clause 6.2.2

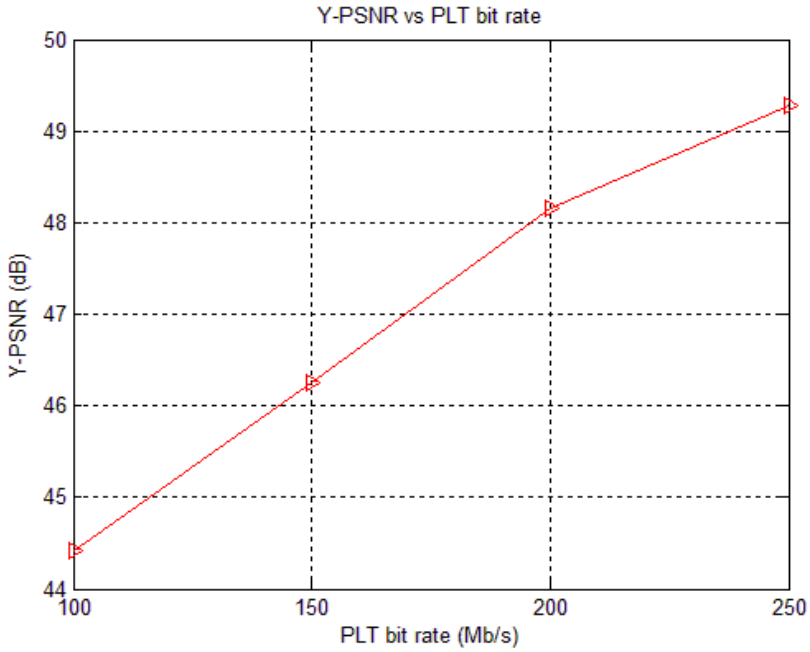


Figure A.2: Experimental results for "YachtRide" using the tandem scheme described in clause 6.2.2

At values of the channel SNR corresponding to a target bitrate of 150Mb/s (8,5 dB – 9,5 dB), one sees that: for "DucksTakeOff" (see figure A.3), the performance of the JPEG 2000 (here and in what follows, JPEG 2000 is sometimes indicated OpenJPEG to highlight the software used to emulate it) tandem scheme is about 0,7 dB better in PSNR than the Dirac tandem scheme and about 1,5 dB better in PSNR than SoftCast; for "YachtRide" (see figure A.4), the considered tandem schemes are about 7 dB better than the considered joint scheme. In the last case, the larger difference comes certainly from the high motion of the video, which may require other choices of the coding parameters for SoftCast.

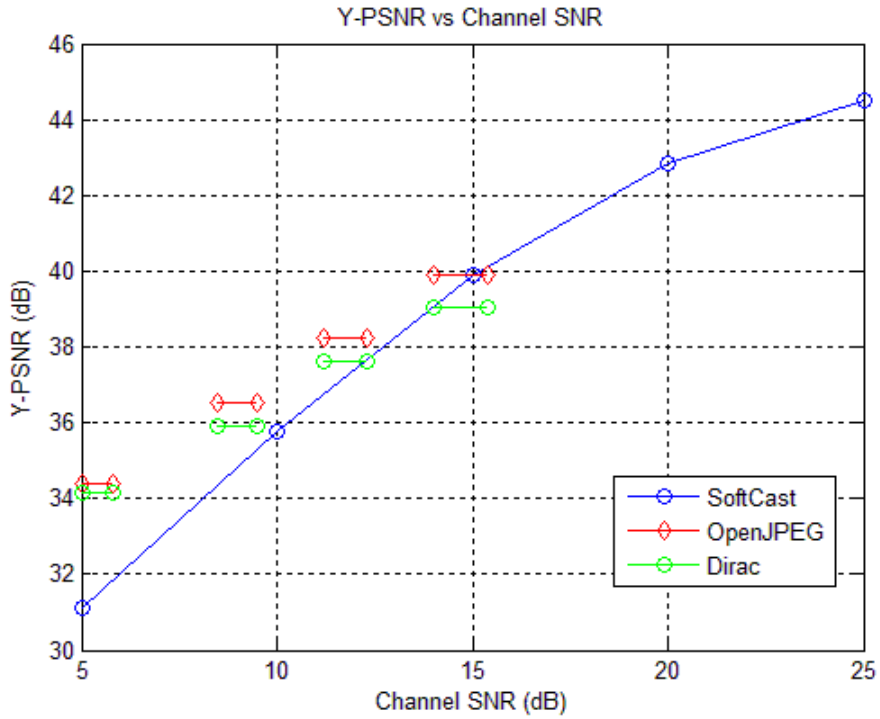


Figure A.3: Experimental results for "DucksTakeOff" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.1

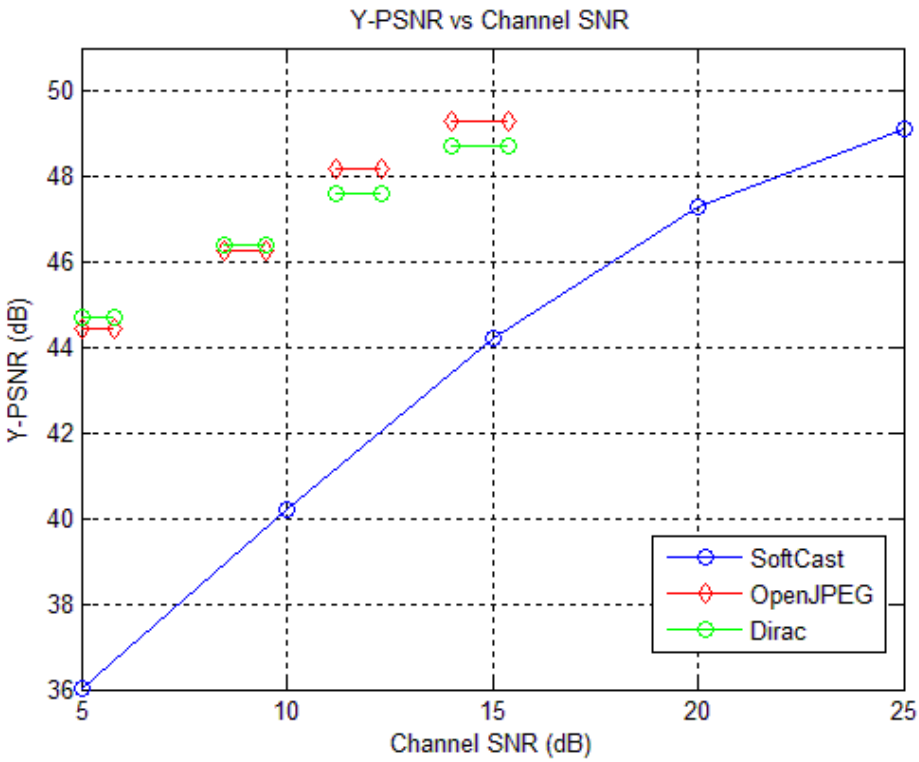


Figure A.4: Experimental results for "YachtRide" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.1

In figure A.5, the Y-PSNR obtained by the SoftCast scheme is compared for various rates at the PHY layer equal to 1,2 Gb/s, 1 Gb/s, or 800 Mb/s.

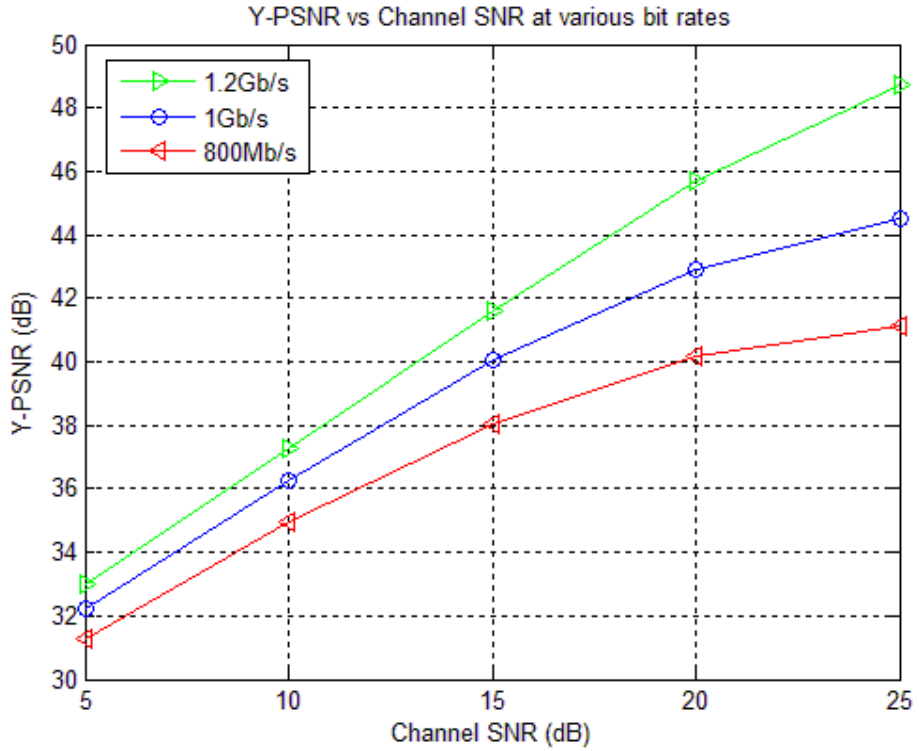


Figure A.5: Experimental results for "YachtRide" using the joint scheme described in clause 6.3.2 and assuming different PHY PLT bit rates

Results in terms of the Y-SSIM metric related to "DucksTakeOff" and "YachtRide" are reported in table A.1 and table A.2.

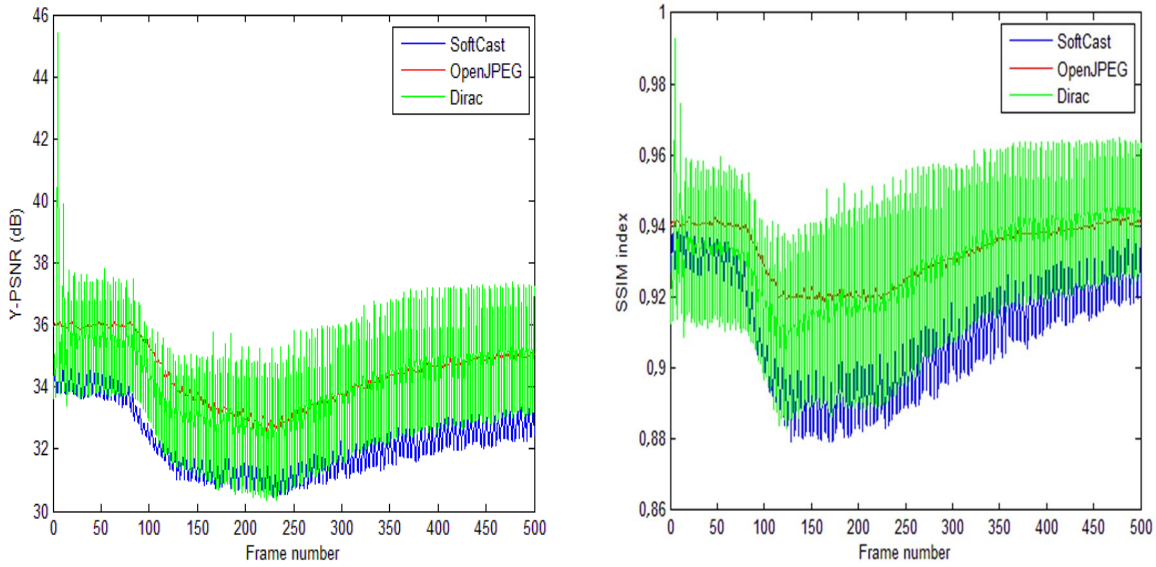
Table A.1: Y-SSIM experimental results for "DucksTakeOff" using the schemes described in clause 6

SNR	5	10	12	15	20	25
Mean Y-SSIM-OpenJPEG	0,9322	0,9550	0,9671	0,9766		
Mean Y-SSIM-Dirac	0,9297	0,9469	0,9596	0,9683		
Mean Y-SSIM-SoftCast	0,9108	0,9599		0,9827	0,9912	0,9941

Table A.2: Y-SSIM experimental results for "YachtRide" using the schemes described in clause 6

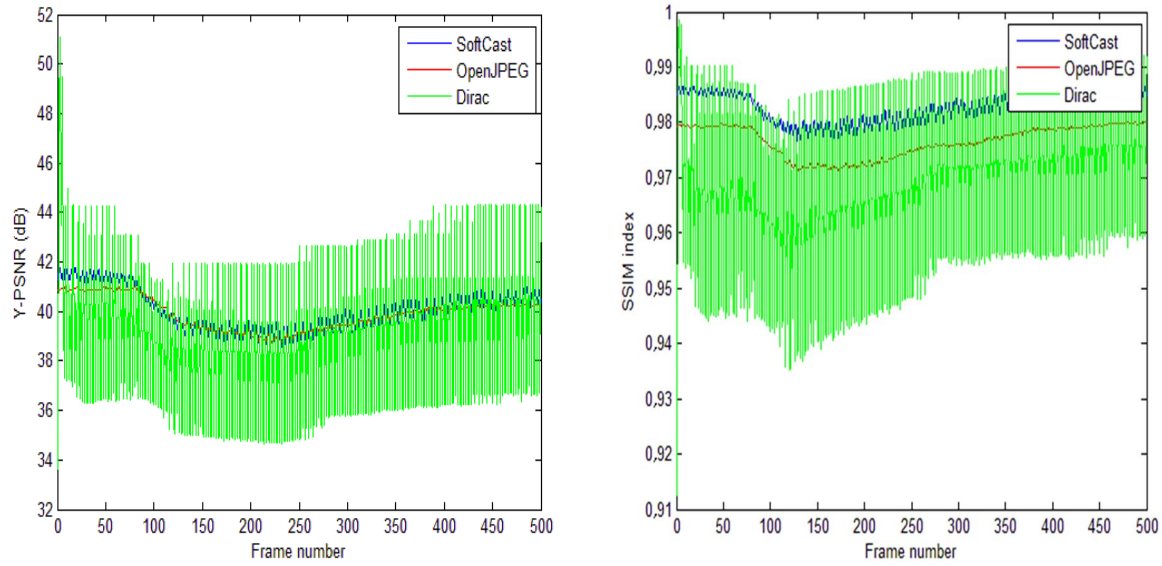
SNR	5	10	12	15	20	25
Mean Y-SSIM-OpenJPEG	0,9826	0,9872	0,9907	0,9926		
Mean Y-SSIM-Dirac	0,9847	0,9889	0,9913	0,9929		
Mean Y-SSIM-SoftCast	0,9240	0,9653		0,9854	0,9933	0,9959

Frame by frame details on the obtained PSNR and Y-SSIM on all the considered video sequences can be found in figures A.6 and A.7 ("DucksTakeOff"), figures A.8 and A.9 ("ParkJoy"), figures A.10 and A.11 ("YachtRide") and figure A.12 ("FTV").



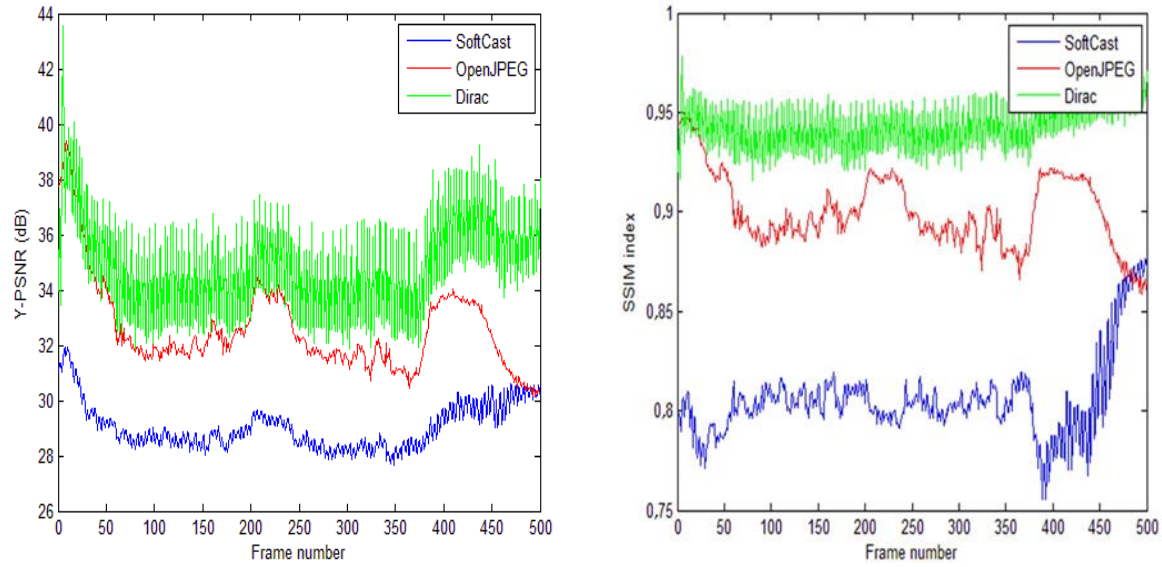
NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 13 and table 8 for average results.

Figure A.6: Frame by frame results for "DucksTakeOff" at SNR = 5 dB using the schemes of clause 6



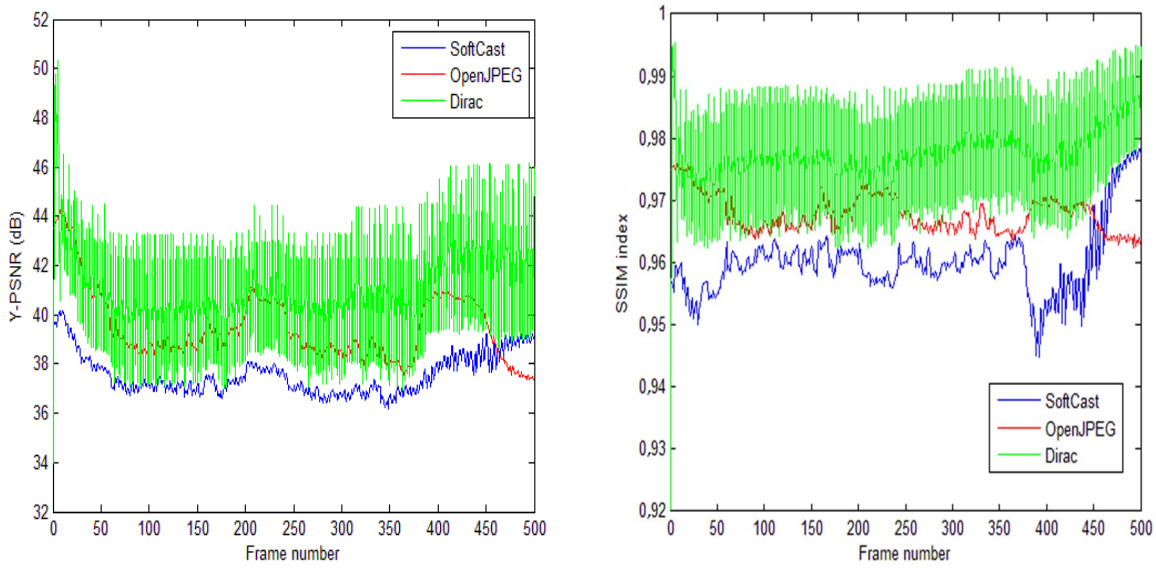
NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 13 and table 8 for average results.

Figure A.7: Frame by frame results for "DucksTakeOff" at SNR = 15 dB using the schemes of clause 6



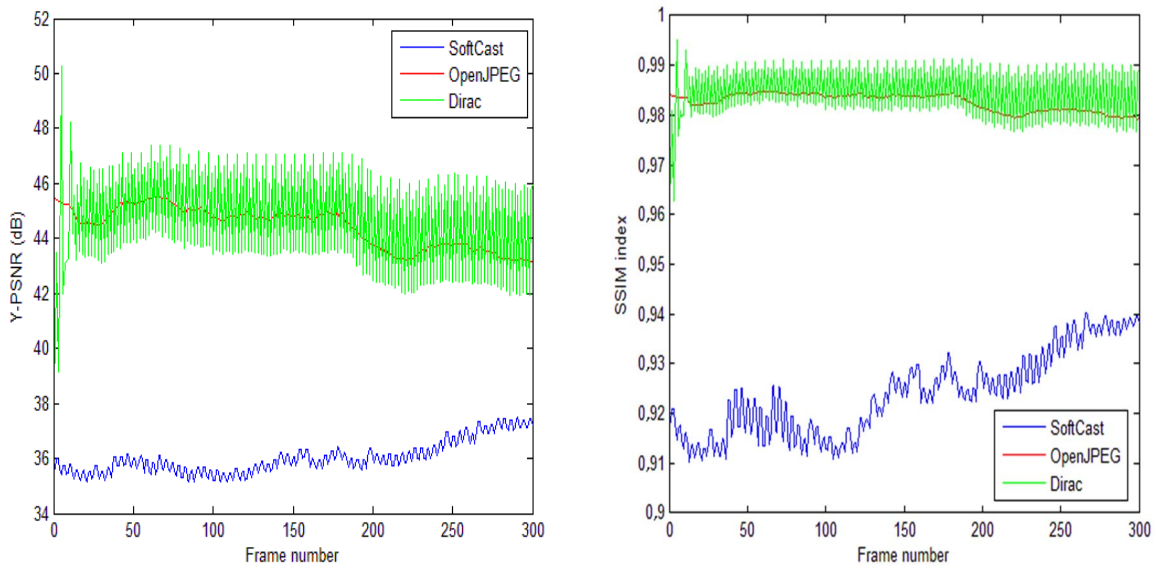
NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 14 and table 9 for average results.

Figure A.8: Frame by frame results for "ParkJoy" at SNR = 5 dB using the schemes of clause 6



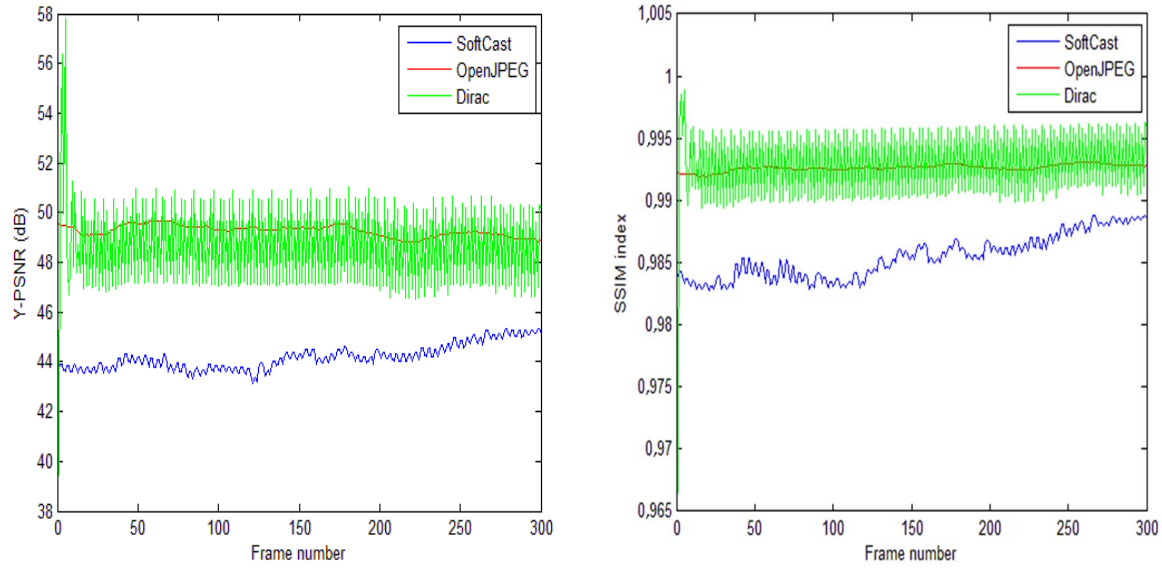
NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 14 and table 9 for average results.

Figure A.9: Frame by frame results for "ParkJoy" at SNR = 15 dB using the schemes of clause 6



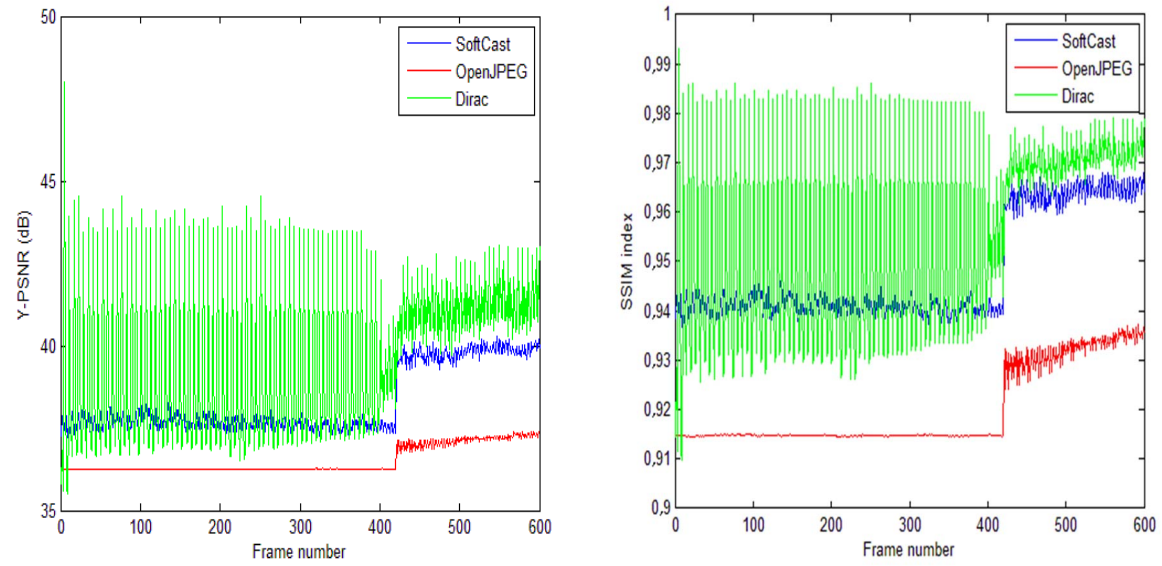
NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 15 and table 10 for average results.

Figure A.10: Frame by frame results for "YachtRide" at SNR = 5 dB using the schemes of clause 6



NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 15 and table 10 for average results.

Figure A.11: Frame by frame results for "YachtRide" at SNR = 15 dB using the schemes of clause 6



NOTE: On the left: PSNR evolution. On the right: SSIM evolution. See figure 28 and table 12 for average results.

Figure A.12: Frame by frame results for "FTV" at SNR = 15 dB using the schemes of clause 6

A.3 Additional results on UHD sequences

Results in figures A.13 and A.12 and in tables A.1 and A.2 are obtained using the 500 PLT link SNRs of the database.

Figures A.13 and A.14 show respective results using UHD sequences "DucksTakeOff" and "Crowd_Run" in terms of PSNR. Tandem schemes perform better than the joint scheme. Between them, Dirac seems generally better with JPEG 2000 only competing at the lower compression ratios. This same trend can be spotted on three specific links (SISO, MIMO 2×2 and MIMO 2×3) in terms of Y-SSIM.

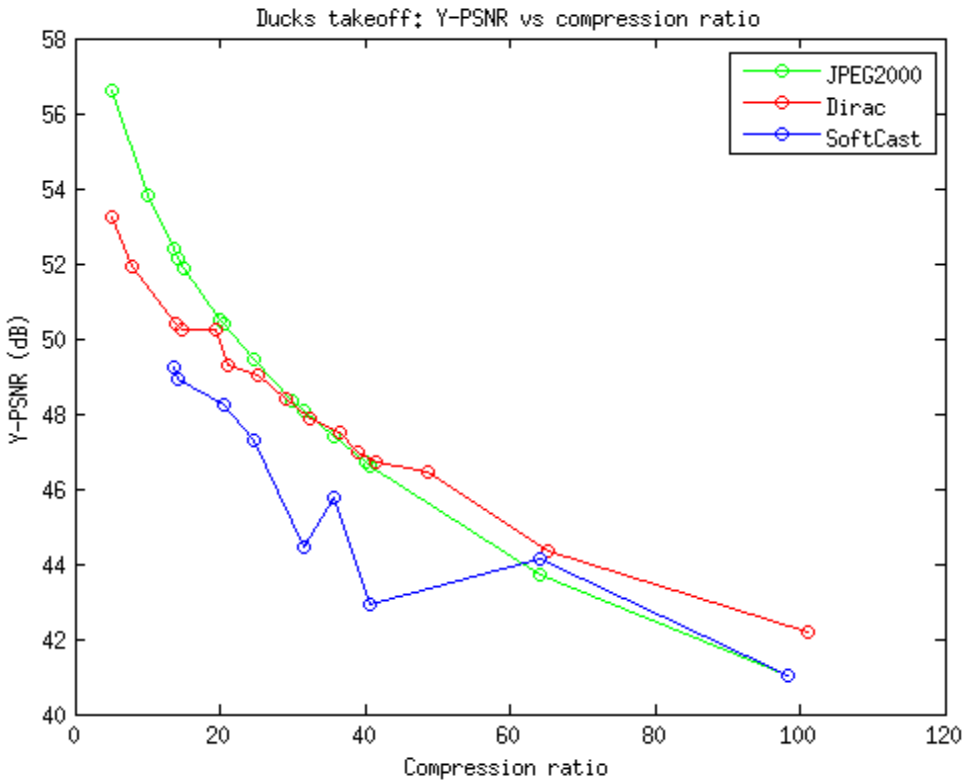


Figure A.13: PSNR experimental results for UHD "DucksTakeOff" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

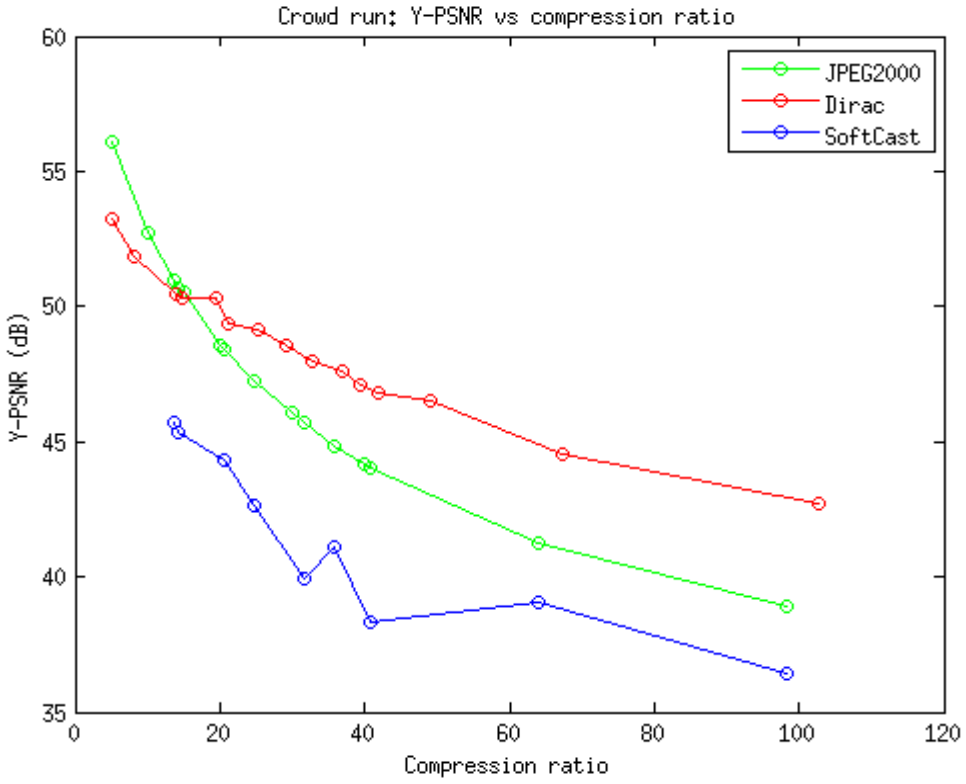


Figure A.14: PSNR experimental results for UHD "Crowd_run" using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Table A.3: Y-SSIM experimental results for UHD "DucksTakeOff" (HEVC encoded and decoded) using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM JPEG 2000	0,9940	0,9936	0,9913	0,9894	0,9861	0,9841	0,9810	0,9658	0,9433
Mean Y-SSIM Dirac	0,9914	0,9911	0,9888	0,9872	0,9846	0,9829	0,9807	0,9702	0,9547
Mean Y-SSIM SoftCast	0,9905	0,9899	0,9883	0,9861	0,9742	0,9806	0,9651	0,9732	0,9498
Link	3	3	3	2	1	2	1	2	1

Table A.4: Y-SSIM experimental results for UHD "Crowd_run" (HEVC encoded and decoded) using the joint scheme described in clause 6.3.2 and the tandem schemes described in clause 6.2

Compression Ratio	13,55	14,33	20,53	24,7	31,71	35,84	40,77	64,13	98,3
Mean Y-SSIM JPEG 2000	0,9889	0,9880	0,9815	0,9763	0,9670	0,9612	0,9544	0,9232	0,8843
Mean Y-SSIM Dirac	0,9898	0,9895	0,9874	0,9860	0,9834	0,9821	0,9803	0,9724	0,9617
Mean Y-SSIM SoftCast	0,9758	0,9738	0,9692	0,9629	0,9321	0,9481	0,9107	0,9182	0,8764
Link	3	3	3	2	1	2	1	2	1

Annex B:
State of the art

PHDMI is a completely new approach. It tries putting together the best of PLT technology knowledge and compression coding/decoding techniques.

PLT technology has recently added MIMO communication to SISO. SISO is typically based upon the differential use of the L-N (Line or Phase to Neutral) wire pair. In a MIMO-PLT channel, differential signaling between any 2 of the 3 wires leads to 3 different feeding possibilities: L-N, L-PE (Line to Protective Earth), and N-PE. According to Kirchhoff's rule the sum of the 3 input signals has to be 0. Therefore only 2 out of the 3 independent input ports can be used. On receiving side all 3 wires are available as reception ports. In the frequency domain, the MIMO-PLT channel can be described by a $N_R \times N_T$ channel matrix for each frequency f:

$$\mathbf{H}(\mathrm{f}) = \begin{pmatrix} h_{11}(f) & \cdots & h_{1N_T}(f) \\ \vdots & \ddots & \vdots \\ h_{N_R1}(f) & \cdots & h_{N_RNT}(f) \end{pmatrix} \quad \text{(B.1)}$$

For orthogonal frequency division multiplexing (OFDM) based communication systems, the above formula describes the MIMO PLT channel for each subcarrier f = c used. Hence, the MIMOPLT channel can be regarded as a set of $N_R \times N_T$ SISO correlated channels: for instance, coefficient h_{11} represents the channel between L-N and L. In order to characterize the MIMO-PLT channel, a previous project [i.3] undertook a comprehensive measurement campaign and measured channel transfer functions in the [0, 100] MHz band, noise, reflection parameters, and EMI characteristics in many homes in six European countries. The project analysed a 4×2 MIMO channel matrix, the fourth path at the receiver given by the common mode (according to the project, common mode reception could be possible and advantageous in scenarios where a large ground plane is available). Similar measurement campaigns have been performed in USA by members of the HomePlug Alliance [i.9] for the 3×2 MIMO-PLT channel matrix. In [i.9] the measurement campaign has been used to derive SISO and MIMO in home PLT channel models. Further investigations on the channel spatial correlation are reported in [i.10]. Using matrix singular value decomposition, the MIMO channel can be decomposed into 2 independent SISO streams and the singular values describe the channel gains for the logical streams of the MIMO-PLT channel. ETSI TR 101 562-3 [i.3] uses the following formula to compute the MIMO-PLT channel capacity:

$$C = \frac{B}{N} \sum_{i=1}^{N} \sum_{j=1}^{2} \log_2 \left(1 + \frac{\rho}{N_T} \lambda_{i,j}\right) \quad \text{(B.2)}$$

where $\lambda_{i,j}$ is the eigenvalue of stream j on carrier i, N is the number of frequency points and B is the bandwidth. Finally ρ is the ratio of the transmit power spectral density (PSD) to the noise PSD and is assumed to be constant over the frequency. Clearly, this assumption is a simplification because the transmit power spectral density is not necessarily constant, while the noise PSD cannot be assumed flat as also shown by ETSI TR 101 562-3 [i.3]. Nevertheless, it made easy to understand the average capacity MIMO gain of 2 over SISO when the same total power is used. When it comes to characterize the noise, besides the analysis reported in ETSI TR 101 562-3 [i.3], it is shown in [i.11] that the noise is correlated between the receive ports. The strongest correlation is measured for the (L,PE) and (N,PE) pairs of receive ports. The correlation is stronger on lower frequencies when compared to higher frequencies. The effect of the noise correlation on the capacity of a MIMO-PLT system with two transmit ports and three receive ports is studied [i.11], and it is observed that noise correlation indeed helps to increase the MIMO channel capacity (compared to uncorrelated noise of the same power). On poor channels, the correlation in the channel can increase capacity by as much as 40 %.

OFDM solutions that have added MIMO to SISO are HomePlug® AV2 (HomePlug Alliance) and G.hn MIMO from (ITU-T). An extensive overview of HomePlug® AV2 can be found in [i.4].

Video compression coding/decoding techniques are fundamental engineering tools for image/video processing. Fundamental of compression techniques are based on tailored mathematical functions, such as wavelet. A nonstationary signal, such as an image, is processed through DWT allowing to get a representation of both time and frequency domain variations. For easier implementation of the DWT, the notion of multiresolution analysis comes into place. A scaling function has the property that, if it can be used to represent a function, then dilated versions of itself can also be used to represent the given function. If a function can be exactly represented by the scaling function at resolution j + 1, it can be decomposed into a sum of functions consisting in a lower-resolution approximation j and a sequence of functions generated by dilations of the wavelet that represent the remaining details. This is similar to subband coding.

The difference is that, while the subband decomposition is in terms of sines and cosines, the decomposition in this case uses scaling functions and wavelets. In subband coding schemes, the signal is decomposed using filter banks. The outputs of the filter banks are downsampled, quantized, and encoded. The decoder decodes the coded representations, upsamples, and recomposes the signal using a synthesis filter bank. In order to obtain a one-dimensional (1-D) wavelet transform, a signal is passed through a lowpass and highpass filter, h and g, respectively, then down-sampled by a factor of two, constituting one level of transform. Multiple levels or scales of the wavelet transform are obtained by repeating the process on the lowpass outputs only. The process is carried out for a finite number of levels K, and the resulting coefficients are the wavelet coefficients. The 1-D wavelet transform can be extended to a two-dimensional (2-D) wavelet transform using separable wavelet filters. The 2-D transform can thus be computed by applying a 1-D transform to all the rows of the input, and then repeating on all of the columns. Detailed description of wavelet-based image and video coding techniques can be found in [i.12]. One of the properties of the wavelet transform that can be used for data compression is that it tends to compact the energy of the input into a relatively small number of wavelet coefficients, equivalent to reducing the correlation among the coefficients. In naturally occurring images, given a decomposition with K levels, much of the energy in the wavelet transform is concentrated into the low frequency subband. The low frequency subband is a coarse version of the original data. The energy in the high frequency bands is also concentrated into a relatively small number of coefficients. The high frequency subbands contain the horizontal, vertical, and diagonal details which cannot be represented in the low frequency band. This interpretation is true for any number of decomposition levels. When it comes to the coding of coefficients for transmission, by the design of the subband filters, the coefficients in each subband are (approximately) uncorrelated from coefficients in other subbands. As a result, the coefficients in each subband can be quantized independently of coefficients in other subbands with no significant loss in performance. The variance of the coefficients in each of the subbands is also different, and thus each subband requires a different amount of bits to obtain best coding performance. The issue to be resolved is that of bit allocation, or the number of bits to be assigned to each individual subband to minimize the distortion of the reconstruction. Nevertheless, in subband coding, the optimal bit allocation changes as the overall bit rate changes, which requires the coding process to be repeated entirely for each new target bit rate. It is also difficult to code an input given an exact target bit rate due to the entropy coding of the quantizer outputs and the approximation inherent in the quantizer model. One solution to this problem, if the coded output is too large, is to truncate the coded data. Truncation removes entire subband coefficients, which can result in visual artifacts. If the coded output data is too small, more bits may be allocated, but the subband coding procedure has to be repeated.

Embedded Zero-Tree Wavelet Coding (EZW) introduced in 1993 by Shapiro [i.13] uses the multiresolution decomposition to improve performance at low bit rates relative to the existing JPEG standard. The two key concepts of EZW are the significance map coding using zero-trees and the successive approximation quantization. The EZW algorithm exploits the fact that wavelet coefficients in different subbands represent the same spatial location in the image. The relationships of these coefficients can be described by a tree, with coefficients in higher subbands considered as descendants of coefficients in lower subbands. The role of zerotree coding is to avoid transmitting all the zeros. Once a zerotree symbol is transmitted, all the descendant coefficients are zero, so no information is transmitted for them. Zerotrees are a form of run-length coding where the coefficients are ordered in a way to generate longer run lengths as well as making the runs self-terminating, so the length of the runs need not be transmitted. Further details can be found in [i.13]. An enhanced implementation of the zerotree algorithm, known as Set Partitioning in Hierarchical Trees (SPHIT), has been done by Said and Pearlman [i.14]. Their method is similar to the Shapiro algorithm, but with more attention to detail. The public domain version of this coder is very fast, and improves the performance of EZW by 0,3 dB to 0,6 dB. While in the original zerotree algorithms special symbols are given only to single zerotrees, in SPHIT other sets of zeros that appear with sufficient frequency are represented by special symbols. The Said-Pearlman coder provides symbols for combinations of parallel zerotrees.

Wavelet transforms have been used for Scalable Video Coding (SVC), for multi-resolution spatiotemporal representation of video sequences. In SVC, from a single bit-stream, a number of decodable streams can be extracted corresponding to various operating points (on the rate-distortion cuve) in terms of spatial resolution, temporal frame rate or quality of the reconstruction. The wavelet-based SVC architectures (WSVC) differ depending on the space-time order in which the wavelet transform is applied. A comprehensive report can be found in [i.15] with the main points presented in the following. For a scaled decoding, in terms of spatial, temporal and/or quality resolution, the decoder only works on a portion of the originally coded bit stream according to the specification of a desired working point.

A functional block, called extractor, extracts a decodable bit stream matching or almost matching the working point specification. Usually a low complexity extractor is used, which does not require coding/decoding operations and consists of simple parsing operations on the coded bitstream. This means that in order to properly decode higher resolution one needs to also use the portions of the bitstream which represent the lower spatial resolutions, while for each spatial resolution level a temporal segmentation of the bit-stream can be further performed. Since most successful video compression schemes require the use of motion compensation (MC), each temporal resolution level bitstream portion should include the relative motion information (usually consisting of coded motion vector fields).

With such a bit-stream, every allowed combination in terms of spatial, temporal and quality scalability can be generally decided and achieved according to the decoding capabilities.

Spatial scalability can be obtained by multirate filter banks, as presented earlier, or by lifting schemes. In a lifting scheme, the signal is split in two polyphase components, the even and the odd samples, on which a pair of linear operators is recursively applied. The first operator performs the prediction step, where the current odd sample is predicted from a linear combination of even samples. The prediction step outputs the prediction error. A suitable combination of prediction errors is used to update the current even value. This stage is called update step. A couple of prediction and update stages is a lifting step. By combining a suitable number of lifting steps, it is possible to obtain any wavelet filter. The output of the last prediction stage constitutes the high-pass band of the corresponding WT filter; the output of the last update step is the lowpass band. Daubechies and Sweldens [i.16] proved that every DWT can be factorized in a chain of lifting steps.

The WT can be extended to three-dimensional signals by performing a further wavelet filtering along time dimension. However, in this direction, the video signal is characterized by abrupt changes in luminance, often due to objects and camera motion, which would prevent an efficient de-correlation, reducing the effectiveness of subsequent encoding. In order to avoid this problem, Motion Compensated Temporal Filtering (MCTF) is needed. The idea behind motion compensated WT is that the low frequency subband should represent a coarse version of the original video sequence; motion data should inform about object and global displacements; and higher frequency subbands should give all the details not present in the low frequency subband and not caught by the chosen motion model as, for example, luminance changes in a (moving) object. Taubman and Zakhor [i.17] proposed to apply an invertible warping (deformation) operator to each frame, in order to align objects. Then, they perform a three-dimensional WT on the warped frames, achieving a temporal filtering which is able to operate along the motion trajectory defined by the warping operator. However, this motion model was able to effectively catch only a limited set of object and camera movements. Another approach was proposed by Ohm [i.18] and later improved by Choi and Woods [i.19]. They adopt a block-based method in order to perform temporal filtering. In this method, each spatial block is treated as an independent video sequence. In the regions where motion is uniform, this approach gives the same results as the frame-warping technique, as corresponding regions are aligned and then undergo temporal filtering. If neighbouring blocks have different motion vectors, pixels belonging to different frames can no longer be correctly aligned. These pixels need a special processing, which no longer corresponds to the subband temporal filtering along motion trajectories. Another limitation of this method is that motion model is restricted to integer-valued vectors, while it has been recognized that sub-pixel motion vectors precision is beneficial. A different approach was proposed using motion compensated lifting schemes (MC-ed LS) [i.20]. This approach proved to be equivalent to applying the subband filters along motion trajectories corresponding to the considered motion model, and proved to have better performance than previous WT-based video compression methods.

Among the best image compression schemes, wavelet-based ones currently provide for high rate-distortion (R-D) performance while preserving a limited computational complexity. They usually do not interfere with spatial scalability requirements and allow a high degree of quality scalability which, in many cases, consists in the possibility of optimally truncating the coded bit-stream at arbitrary points (bit-stream embedding). Most techniques are inspired from the zerotree technique EZW and then reformulated with the set partitioning in hierarchical trees (SPHIT) algorithm. A higher performance zerotree inspired technique is the embedded zero-block coding (EZBC) algorithm [i.21], where quad-tree partitioning and context modeling of wavelet coefficients are well combined. Another technique which does not use the zerotree hypothesis is the embedded block coding with optimized truncation (EBCOT) algorithm [i.22], adopted in the JPEG 2000 standard, which combines layered block coding, fractional bit-planes, block based R-D optimization, and context based arithmetic coding to obtain good scalability properties and high coding efficiency.

The JPEG 2000 scheme has been used in this TR as a constituent block in a tandem scheme (see clause 6.2). Tandem scheme philosophy is based on the independence between the source and the channel coder, also known as the separation theorem, and simplifies the construction of communication systems since the coders can be optimized separately. Also, when the source or the channel characteristics change, the affected coder can be changed while leaving the other unchanged. There are nevertheless some drawbacks with tandem schemes. Namely, separation is proven to be an optimal strategy only for stationary and ergodic channels, and in a point-to-point communication. Failing these conditions, and in particular in the case of time-varying channels under delay constraints, separation is generally no longer optimal and joint optimization of the source and channel coders can be considered. An overview of the joint source-channel coding/decoding methods is given in [i.23]. The main points are presented in what follows. Joint schemes exploit the residual redundancy left by the source coding as an a priori to assist the decision process of the channel decoder in case of errors. Notice that this strategy implies no modification at the transmitter side.

The redundancy mainly comes from constraints imposed on the source coder on the bitstreams it can generate, and it can be of three types:

1) Redundancy due to the syntax of source coders. Usually, entropy coding (e.g. Huffman codes, arithmetic codes) is one of the last steps performed by the source coder before packetization of compressed data. There are sequences of bits that cannot be generated by the entropy code. One can thus distinguish valid bitstreams from invalid bitstreams.

2) Redundancy due to semantic of the source and of the source coder. Comes from the statistical properties of the source (sources with memory) and from constraints on the compressed bitstream resulting from the compression standard. A parser at the decoder may be employed to determine whether a bitstream, or a part of a bitstream, is compliant with the source coding standard.

3) Redundancy due to the packetization of compressed data. Packets generated by the Network Abstraction Layer (NAL), referred to as NAL Units, should be more or less independently decodable. A packet should thus contain an integer number of coding units (pictures, slices, or macroblocks). Some information on the packet content is usually introduced in the header, like the number of bits of the encoded sequence.

At the decoder side, when the set of sequences compliant with all constraints imposed by the source coder and the packetization process can be described by a trellis, optimal maximum-likelihood or maximum a posteriori decoders can be employed such as the Viterbi or the BCJR decoders. If modifications at the transmitter side are acceptable, the possibility of deliberately adding redundancy is addressed, either by re-designing some blocks of the source encoder, or by changing its global structure. Rather than re-designing from scratch a totally new encoder implementing joint source-channel coding concepts, with the risk of obtaining a system with decreased performance, many authors chose to add robustness features to existing blocks found in actual, efficient, source encoders.

One possible way of increasing the robustness of source encoders to channel impairments is to introduce structured redundancy in one of the existing blocks, rather than relying on the one naturally found in the standards. Some example are:

- Variable length error-correcting codes: the aim is to build low-complexity codes simultaneously providing good data compression and error correction capabilities. The goal is to obtain joint codes outperforming separate codes, in terms of error rate for a given complexity or in terms of complexity for a given error rate. The compression efficiency of a VLC is measured by the ratio of the average codeword length to the source entropy. In addition to this, error correction capabilities can be obtained by ensuring that the VLC has a well-defined minimum distance. In the cases where Arithmetic Coding is used, e.g. H.264 AVC and JPEG 2000, the robustness against errors is usually achieved by introducing a forbidden symbol in the source alphabet and using it as an error-detection device at the decoder side.

- Redundant signal representation, or multiple description coding: this technique represents source signal with a set of M independently decodable streams, called descriptions. These descriptions are transmitted on different channels, each with given loss and error probability. The system is designed in such a way that the decoder should be able to reconstruct the source with acceptable distortion from any subset of these descriptions.

- Hierarchical modulations: In this approach, proposed for broadcast/multicast communication, the main idea is to cluster a high-dimensional constellation to efficiently transmit bitstreams produced by a scalable source coder. Data from the base layer are associated to the cluster index, data associated to the refinement layer are mapped to points in the constellation of the considered cluster. This allows users with a very good channel to receive to both layers while users with a degraded channel may still receive the base layer. Thus, a graceful degradation instead of complete signal loss can be obtained when the channel worsens.

Annex C: Bibliography

- HomePlug Alliance: "HPAV specification version 2.1".
- Recommendation ITU-T G.9963: "Unified high speed wire-line based home networking transceivers - Multiple input/multiple output (MIMO)".
- M. Marcellin, M. Gormish, A. Bilgin and M. Boliek: "An overview of JPEG-2000" in Data Compression Conference Proceedings, pp. 523-541.
- S. Jakubczak and D. Katabi: "Softcast: one-size fits all wireless video, "ACM SIGCOMM Computer Communication Review", volume 41, number 4, pp.449-450, 2011.

History

Document history		
V1.1.1	December 2015	Publication