We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

19

Combined Source and Channel Strategies for Optimized Video Communications

François-Xavier Coudoux[1,2,3], Patrick Corlay[1,2,3], Marie Zwingelstein-Colin[1,2,3], Mohamed Gharbi[1,2,3], Charlène Mouton-Goudemand[1,2,3,] and Marc-Georges Gazalet[1,2,3]

[1]*University Lille Nord de France, F-59000 Lille,*
[2]*UVHC, IEMN-DOAE, F-59313Valenciennes,*
[3]*CNRS, UMR 8520, F-59650 Villeneuve d'Ascq,*
France

1. Introduction

Digital video is becoming more and more popular with the wide deployment of multimedia applications and networks. In the actual context of Universal Media Access (UMA), one of the main challenges is to flexibly deliver video content with the best perceived image quality for end-users having different available resources, access technologies and terminal capabilities. In this chapter, we look in detail at the basic source and channel coding techniques for digital video communication systems, and show how they can be combined efficiently in order to fulfill the quality of service (QoS) constraints of video communication applications. The chapter includes several illustrative examples and references on the related topics.

The chapter begins with an overview of digital video compression basics, including MPEG-2 and H.264/AVC. We discuss the most common coding artifacts due to digital compression, and show how the compressed bitstream is made more sensitive to channel errors (Section 2). Since both compression and channel distortions affect the final perceived video quality, image quality metrics are needed in order to estimate the resulting visual quality. Both subjective and objective metrics are presented briefly and discussed in section 3.

The second half of the chapter concerns channel coding and error control for video communication (Section 4). The most common existing techniques are presented, with a focus on forward error correction (FEC) and the hierarchical modulation, used in the DVB standard, for example. We show that such channel coding schemes can be used to increase the error resilience of compressed video data. For example, scalable video coding can be combined with hierarchical modulation in order to allow the most important information in the compressed video bitstream to be transmitted with better protection against channel distortions. Section 5 explains this concept of scalability, and gives the great benefits of this coding tool for robust video transmission. Since all the various bits of the transmitted video data don't have the same level of importance, unequal error protection (UEP), with its different priority levels can be applied successfully.

Source: Digital Video, Book edited by: Floriano De Rango,
ISBN 978-953-7619-70-1, pp. 500, February 2010, INTECH, Croatia, downloaded from SCIYO.COM

For a given application, it is necessary to determine the best combination of lossy compression and channel encoding schemes in order to offer the optimal quality to the end-user. In section 6, we illustrate such a combined source and channel approach by presenting a quality-oriented study for a broadband video distribution system using digital subscriber lines (DSL). Our system allows the coverage area for a given DSL infrastructure to be extended, thus increasing the number of potential end-users. This system applies an adaptation mechanism that determines the "bottleneck" bit-rate at which the reconstructed video has the best quality. First, we detail the complete system architecture, and then we explain how to jointly determine the optimal source and channel coding parameters. Finally, we report our experimental results in order to demonstrate the effectiveness of our system in terms of extended coverage and optimal quality for a given eligibility level. In section 7, we conclude the chapter with a discussion of the current issues and the perspectives for future research on video communications.

2. Digital video compression basics

Digital video signals generally include some redundant information. For still images, the redundancy is spatial and is due to the important correlation between neighboring pixels. In order to reduce this redundancy, lossless data compression can be used, thus allowing a reconstructed image equal to original image to be obtained. The compression ratio obtained with this method is generally very low, close to 2 or 3. Lossy compression insures that a higher compression ratio will be obtained. The deterioration of the reconstructed image is a function of the compression rate, constituting a rate-quality trade-off. Like transmitting still images, transmitting a video stream also requires compressing the video data. This compression is made possible by the video stream's redundancy, both spatial (intra-image) and temporal (inter-image).

To eliminate the redundancy—or, in other words, the correlation—between two images, each image in a video stream is predicted in terms of the previous and/or following images, with only the prediction error being encoded. The first image in the stream (or the group of pictures described below) is always fully encoded without reference to the other images; this is the "Intra" mode. This encoding without reference facilitates the synchronisation of the receiver (i.e., the decoder). The images that follow can be predicted by motion compensation, and the prediction error can be then encoded before being transmitted with motion vectors.

During the encoding process, a video stream is split into a group of pictures (GOP) of a fixed size. The GOP contains:

- Intra-Picture (I): This is the first frame in each GOP. It is the reference, representing a still image, independent of other pictures.
- Inter-Picture (P or B): These frames contain the motion-compensated differences. A P-frame is the prediction based on a previous image, while a B-frame is the result of the encoding from two images, one previous and one following. An error in a predicted picture will propagate up to the final image in the GOP. Figure 1 shows the typical structure of a MPEG sequence.

In digital video broadcasting, the widely used video coding standards MPEG-2 (Mitchell et al., 1996) and, more recently, H264/MPEG-4 (Wiegang et al., 2003) are both based on a hybrid encoding method using transformation and motion compensation (Tekalp, 1996), as illustrated in Fig. 2.

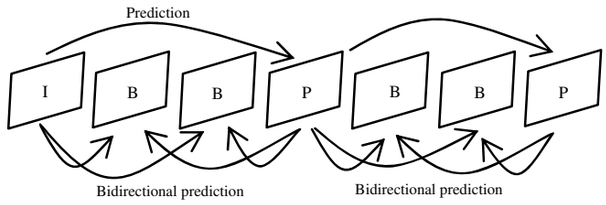


Fig. 1. Structure of MPEG sequence

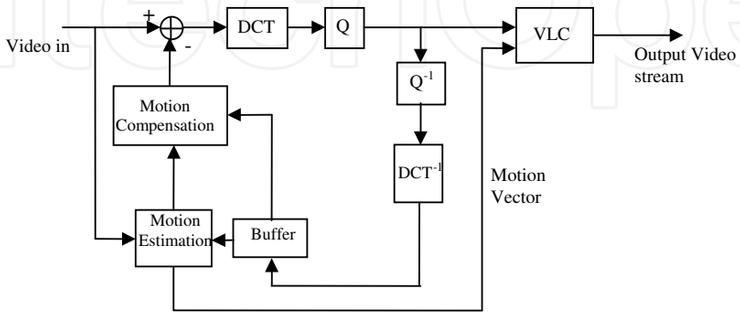


Fig. 2. Hybrid encoding method using transformation and motion compensation

Since the advent of the well-known JPEG algorithms (Pennebaker & Mitchell, 1993), the encoding process has consisted of several steps:

- *Conversion*: The image's color representation is converted into the Y (luminance) and Cr,Cb (chroma) components. The chroma resolution is generally reduced by a factor of 2 both horizontally and vertically.
- *Transformation*: The image is split into blocks, and each block (Intra-Picture or prediction residue) undergoes a discrete cosine transform (DCT). This transform exhibits excellent energy compaction for highly correlated images (Rabbani & Jones, 1991). DCT is independent of the signal to be encoded, and many fast DCT computation algorithms exist. A DCT applied on blocks of 8x8 pixels generates 64 coefficients. The first coefficient represents the constant value (DC), and the others represent waveforms at gradually increasing frequencies. In the most recent H.264 compression standard, the image can be decomposed into blocks of different sizes to adapt to local image statistics, and thus increase encoding efficiency.
- *Quantization*: The amplitudes of the frequency components generated in the previous step are quantized by removing small values. Quantization, which plays a major part in lossy compression, reduces the amount of data needed to represent an image. Since human eye is more sensible to errors in low frequency compared with high frequency (Glenn, 1993), each DCT frequency coefficient is quantized with an adequate step.
- *Scanning*: The quantized DCT coefficients are subjected to zigzag scanning, which arranges the DCT coefficients in order of increasing frequency.
- *Compression*: The resulting data is further compressed using entropic encoding, which is a form of lossless data compression. The entropy encoders compress data by replacing each value, defined by a fixed length, with a corresponding variable length codeword. Since the length of each codeword is an inverse function of its appearance probability, the most common value is represented by the shortest codeword. In JPEG and MPEG2, the entropic encoding is always Huffman encoding.

Typically, in the above encoding process, every 12th frame is an I-frame, and the GOP contains the image string, IBBPBBPBBPBB. The (I) intra-coded frame is split into block of 8 * 8 pixels, and each block is coded independently of the others, using the above encoding process. An Inter-frame (P or B) is divided into 16*16 pixel macro blocks. For each macro block, the motion is estimated in terms of the reference frame, and then the estimation error is compressed.

The H.264/AVC standard allows a compression ratio equal to twice MPEG-2 ratio to be obtained. The H.264/AVC standard is similar to the MPEG-2 standard, but with some rather important differences (Richardson, 2003):

- The intra-coded blocks are predicted in terms of the pixels located above and to the left (causal neighborhood) of those that have previously been encoded and reconstructed. The error prediction is then encoded.
- A deblocking filter is applied to blocks in a decoded video to improve the subjective visual quality.
- The motion is compensated by accounting for different block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4). Using previously-encoded frames as references is much more flexible than past standards, and the precision of the motion compensation is equal to quarter pixel.
- The entropy encoding is enhanced by providing Context-adaptive Binary Arithmetic Coding or Context-adaptive Variable-length Coding for residual data and Exponential-Golomb Coding for many of the syntax elements.

But… what about image quality?

We mentioned in the previous sub-section that digital video compression algorithms use lossy quantization in order to achieve high compression ratios. This quantization results in various kinds of coding artifacts, which may greatly affect the visual quality of the reconstructed video signal, especially for low bit-rate coding. The compression artifacts and their visual significance have been widely studied in the literature. Yuen (1998) provides a comprehensive classification and analysis of most coding artifacts in digital video, compressed using MC/DPCM/block-based DCT hybrid coding methods. In particular, this author shows how the visual impact of coding artifacts is strongly related to the spatial and temporal characteristics, both local and global, of the video sequence, as well as the properties of the human visual system (HVS).

Among the various coding artifacts, *blocking effect* (also called *blockiness*) is the most well-known distortion introduced by video compression algorithms. The blocking effect manifests itself as a discontinuity located at boundaries between adjacent blocks. This spurious phenomenon is due to the fact that common block-based compression algorithms encode adjacent blocks as independent units without taking into account the correlation that exists between them. Hence, at the decoding stage, the quantization error differs from one block to another, resulting in inter-block discontinuities. Figure 3 illustrates this particular coding distortion, which is clearly visible on the face and the building in the background.

This blocking effect is very annoying and mainly affects the visual perception of end-users. The perceptual relevance of this distortion is strongly related to HVS sensitivity: the regular geometric spacing of blocks, the specific horizontal or vertical alignment of block edges, and the spatial frequency of repeated blocks are highly apparent to the human eye. Because of its visual prominence, several methods have been proposed in order to reduce the visibility of the blocking effect in compressed images or sequences. (See the post-processing algorithms

proposed by Ramamurthi (1986), Lee (1998) or Coudoux (2001), for example.) Recently, the H.264/AVC codec has introduced a deblocking filter as a standardized tool to reduce the visibility of this coding artifact.

Fig. 3. Illustrative example of the blocking effect (*Foreman* sequence, MPEG-2, QCIF@128Kbps)

Unfortunately, digital image and video impairments are not restricted to coding artifacts, since errors may also occur when transmitting compressed video bitstreams over a noisy channel. Once the video sequence has been compressed by the encoder, the resulting bitstream is typically packetized in the network adaptation layer using transport protocols, such as ATM or TCP/IP, and then the packets are sent over the transmission network.

Different types of impairments can occur in transmissions over noisy channels: packets may be corrupted, or they may be affected by extensive delays that are incompatible with video applications. In all cases, erroneous packets are considered to be lost and are not available for decoding. Depending on the packet size, this loss may corrupt a limited part of a decoded picture, the entire picture or, in the worst case, a complete group of pictures. In the latter case, error concealment techniques can be used at the decoding stage to limit the visual impact of channel errors. In addition, error control mechanisms may fail at the transport level. In this case, errors occur when decoding at the application level. The wide variety of configurations leads to very different visual distortions depending on the kind of corrupted data. Figure 4 shows examples of localized false macroblocks, erroneous or misaligned slices, and frozen parts of pictures. For example, an erroneous VLC may be decoded as a false value and will subsequently result in a localized distortion in the display.

Fig. 4. Examples of visual impairments due to transmission errors (*Foreman* sequence, MPEG-2, QCIF@128Kbps, BER = 10^{-4})

Visual impairments due to transmission errors generally have a much more severe effect on end-user quality, compared to compression artifacts. In particular, the use of compression techniques based on spatial (e.g., differential encoding of DC coefficient) and temporal (e.g., MC) prediction makes the compressed bitstream very sensitive to channel errors, due to

spatial or temporal loss propagation of the corrupted data up to the next synchronization point (e.g., end of slice/intra-frame).

3. Image quality metrics

Digital video quality obviously constitutes one of the key points for any video service to insure a satisfying level of quality for the end-user. For this reason, it is crucial for researchers, broadcasters and network providers to be able to reliably assess and control the perceived video quality. There are two main approaches to image and video quality assessment: the first one relies on subjective evaluation metrics, and the second one is based on objective quality metrics. These two approaches are described in the following sub-sections.

3.1 Subjective metrics

The best way to measure image quality as perceived by a human observer is to use the subjective viewing experience of human observers themselves. The International Telecommunication Union (ITU) has developed and standardized several subjective methods that provide reliable test conditions and measurements. The ITU recommendations, ITU-R Rec. BT.500-11 and ITU-T Rec. P.910, specify:

- the test conditions (e.g., viewing distance, observer selection process, test material),
- the evaluation procedure (e.g., single vs. double stimuli, type of rating scale), and
- the methods for exploiting the data collected (e.g., statistical tools used for accurate analysis of viewers scores).

For example, ITU-R Rec. BT500-11 specifies the Double Stimulus Impairment Scale (DSIS) method shown in Figure 5. With this method, the reference and the test sequence are shown only once, and the observers have to evaluate the corresponding impairment using a five-level impairment scale. A mean opinion score (MOS) is obtained by averaging the evaluations/scores of all observers. More details on subjective video quality assessment are available in the reference books by Wu (2005) and Winkler (2005).

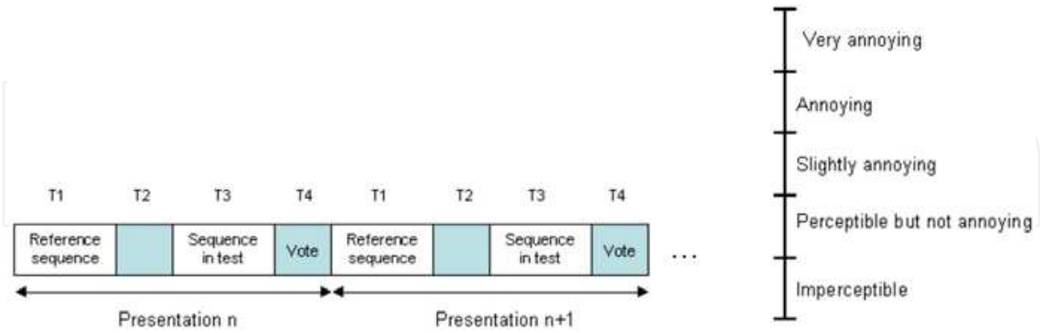


Fig. 5. DSIS method: a) presentation sequence; b) five-level impairment scale

Without any doubt, subjective visual quality assessment is the most correlated with human perception. However, this approach has several drawbacks: it is complex, time-consuming and expensive. Thus, alternative solutions relying on objective metrics are often preferred to such subjective methods. In practice, objective metrics can be used at different places in the broadcasting chain to assess and monitor video quality. They can also be used to optimize the different components of a video communication system, as illustrated later in Section 6.

Nonetheless, in order to be meaningful, objective metrics should be well correlated with the results obtained from the subjective methods.

3.2 Objective metrics

Objective metrics for assessing image and video quality can be divided into three distinct groups:

- signal-based methods,
- structural methods, and
- HVS-based methods.

The most common signal-based method is the Peak Signal-to-Noise Ratio (PSNR), expressed in decibels (dB), and defined as:

$$PSNR(dB) = 10.\log_{10}(d^2/MSE) \tag{1}$$

where d = 255 for 8-bit encoded picture, and MSE is the mean squared error, defined as:

$$MSE = \frac{1}{M.N.T}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{T}\left(f_0(i,j,k) - f_d(i,j,k)\right)^2 \tag{2}$$

where f_0 and f_d are the pictures from the original and distorted sequence, respectively. Parameters M and N represent the size (in pixels) of each picture, the video sequence being made up of T pictures. Unfortunately, this objective metric has been shown to be poorly correlated with human visual assessment (Girod, 1993). PSNR-like metrics do not translate the visibility of the distortions, which depend on picture content, the location of the distortions and the HVS masking properties. For example, the two images presented below have the same PSNR numerical value, although the one located on the left side clearly exhibits worst visual quality (especially due to strong blocking artifacts).

Fig. 6. Image compressed by the JPEG algorithm using high-frequency emphasis (left) and Standard Y (right) quantization tables. Both images have the same PSNR: 30.7 dB.

Recently, the Structural SIMilarity (SSIM) method was proposed by Wang et al. (Wang, 2004a). Instead of measuring a simple pixel-by-pixel difference like *MSE* does, SSIM relies on the fact that human perception depends greatly on the presence of structures inside the video scene: any change in these structures will affect the way that a human observer will perceive the scene. Practically, SSIM is locally computed as a combination of three similarity indicators, one for luminance (based on local means), one for contrast (based on local variance), and one for structural information (based on local covariance). The SSIM values are averaged over the entire image in order to give a single SSIM index by picture. In the case of video sequences, a global note is computed by appropriate weighted averaging of all

the pictures in the sequence, as described by Wang et al. (Wang, 2004b). These authors claimed that the SSIM index gives very satisfying correlation results with respect to subjective assessment methods.

Finally, the most preferred methods model Human Visual System (HVS) properties and then integrate them into the objective metric. There has been plenty of research done on HVS-based objective quality metrics since the beginning of 1970s (e.g., Mannos, 1974; Faugeras, 1979; Lukas, 1982). Such methods can be divided into 2 groups:

- *single-channel methods* -- in single-channel methods, the HVS is modelled as a single spatial filter and is characterized by its Contrast Sensitivity Function (CSF). The final metric is generally obtained by weighting the error signal based on HVS sensitivity.
- *multi-channel methods* -- multi-channel methods are more complex because they assume that the visual signal is processed by separate spatial frequency channels. In the past, several multi-channel vision models have been developed by researchers, including Daly (1993) and Watson (2001), for example. These models generally integrate a more complete representation of significant visual phenomena, such as spatiotemporal masking or orientation selectivity. The global score is typically obtained by error pooling using Minkowski summation. Though more complex, these quality metrics provide better prediction accuracy.

Created in 1997, the Video Quality Expert Group (VQEG) has for the last ten years been evaluating the capacity of various video quality metrics to predict subjective quality ratings, as measured by MOS. Many vision-based video quality metrics have been shown to achieve very good correlation scores with MOS, outperforming the PSNR method. In particular, the so-called Video Quality Metric (VQM) has been recently included in two international Recommendations (Pinson, 2004). Nevertheless, for the moment, no method has ever been judged optimal and subjective rating remains the only reference method for accurately evaluating video quality.

4. Channel coding and error control for video communications

Once a digital video has been appropriately coded and compressed (as described in the previous sections), the concern becomes how to reliably transmit this video over a transmission medium—the channel—that by its nature deteriorates the signal quality. This section presents and analyzes the techniques used to improve digital video transmission quality.

For a given compression algorithm, compressed video quality is clearly a monotone increasing function of the bit-rate: the higher the bit-rate, the higher the video quality. In digital communications, transmission quality is measured as the probability of one bit being flipped during the transmission process – a received “1” corresponding to a transmitted “0” and vice-versa. This probability is referred to as the Bit Error Rate (BER). The BER is a monotone decreasing function of the Signal-to-Noise Ratio (SNR) of the transmission, and a monotone increasing function of the bit-rate. Thus, the higher the video bit-rate, the higher the video quality before transmission but also the higher the BER of the transmission (which results in poor quality of the received video). Thus, a good compromise has to be found for the bit-rate so that the overall video quality deterioration (i.e., the deterioration inherent to the compression algorithm and degradation inherent to the transmission) is acceptable.

In the context of digital video transmission—not to be confused with digital communications in general—the main question is " What are the properties that are specific

to video in terms of its transmission?". One answer is that all the bits in a digital video bitstream do not have the same importance in terms of video quality, which means that the consequences of transmission failure with respect to the video quality can change dramatically depending on which bit(s) in the bitstream fail.

A few examples

- In Pulse Code Modulation (PCM) source coding, each video sample is quantified and binary coded. Clearly, the loss of the least significant bits (LSB) will only slightly degrade its quality, whereas the loss of the most significant bits (MSB) will lead to totally erroneous reconstructed video signal. Thus, the most significant bits are of greater importance than the least significant bits in terms of video transmission.
- In the Discrete Cosine Transform (DCT) coding used for MPEG formats, high-frequency coefficients generally correspond to fine granularity details in the images, whereas low-frequency coefficients correspond to the structure of the images (Richardson, 2003). Thus, in terms of video transmission, low-frequency coefficients can be considered to be of greater importance than high-frequency coefficients.
- In the packetization and framing process inherent to every digital system, headers are added to the video information in order to insure its proper progress in the network, as well as its decoding at the destination. These headers are of crucial importance, since their loss would involve a transmission or decoding failure of a complete packet of video data (corresponding to from a part of an image to several consecutive images). Thus, in terms of video transmission, headers are of greater importance than the other bits transmitted.

Aware of that some bits are more important than others, recent video compression standards (Richardson, 2003) generally make it possible to partition these bits into several different bitstreams, generally 2 or 3. This means that, from the transmission system design perspective, which is the one covered in this section, the idea is to globally optimize the quality of a video transmission by:

- Partitioning the bits transmitted into several bitstreams of different importance, if this has not already been done at the source coding level (possible with the MPEG-2 or H.264 formats – See Section 5) , then
- Providing a way to give different transmission BERs to these bitstreams, where the values of the BERs shall be adapted to the relative importance of the bitstreams. In the following, we use the term *Unequal Error Protection* (UEP) to speak of the transmission techniques that provide different BERs to different bitstreams in a single video sequence.

UEP transmission techniques can have several different effects. For example, they can either improve the video quality of a given user or, for a fixed video quality, extend the range of users eligible for a video service. For example, in a Asymmetric Digital Subscriber Line (ADSL) environment, the UEP techniques would allow users with a poor-quality subscriber line to benefit video services, which would have been impossible in a traditional transmission environment in which all bits are transmitted with the same BER.

Another desirable feature of UEP digital video transmission systems are their ability to provide a simulcast of different video formats—for example, the Standard Definition (SD) and High Definition (HD) formats. This allows a graceful degradation of the video quality as the channel quality degrades, which is frequently the case in mobile communications. It is also an elegant means of adapting the video quality to the receiver's screen definition level.

Although UEP techniques for the different video bitstreams can be considered at the different layers of the Open System Interconnection (OSI) model, we focus on the techniques that do it at the physical layer. This section is organized as follows. In section 4.1, we review the basic modulation principles, which are used to map digital information onto an analog waveform constituted of a sequence of successive symbols. We also examine how the BER is related to the transmission link's signal-to-noise ratio, as well as to the number of information bits conveyed by each symbol. In section 4.2, we focus on UEP modulation techniques. Two independent types of modulation are presented: hierarchical modulation (e.g., Hierarchical Quadrature Amplitude Modulation (HQAM)) and multi-carrier modulation (e.g., Orthogonal frequency-division multiplexing (OFDM)). We also evoke the possibility of using the two types in combination (e.g., HQAM-OFDM). Since the BER performance achieved by any modulation scheme is generally insufficient in terms of the application requirements, forward error correction (FEC) techniques are usually necessary to insure an acceptable quality of service. In section 4.3, we introduce the FEC principles, as well as the UEP techniques that operate at the FEC layer.

4.1 Modulation basics

This subsection does not aim to cover the subject of modulation exhaustively like (Proakis, 1995) did, but rather to provide the essential information needed to understand the techniques developed in the subsequent subsections.

Modulation is the operation that makes the link between digital information – a sequence of bits at a given bit-rate $1/T_b$, where T_b is the duration of one bit – and an analog waveform appropriate for transmission over a channel. Generally, bits are not transmitted one at a time, but are rather grouped in small quantities (typically 2 to 10 bits), called symbols, which are transmitted sequentially. For a modulation scheme that maps b bits into one symbol, the symbol duration will be $T=b\times T_b$, resulting in a symbol rate of $1/T$ bauds.

As an example, let us consider the rectangular Quadrature Amplitude Modulation with eight modulation levels (8-QAM). A temporal representation of the modulated signal appears in Figure 7a. In this figure, the symbols are formed of $\log_2(8) = 3$ bits. These bits are used to code the amplitudes of two orthogonal waves of the same frequency, or to code the amplitude A and the phase π of a single wave, since: $x\cos(2\pi f_0 t)+ y\cos(2\pi f_0 t)$ = $A\cos(2\pi f_0 t + \Theta)$, where: $A = \sqrt{x^2 + y^2}$, and: $\Theta = \arctan(y/x)$.

A signal-space representation of the signal, called the *constellation*, is a useful visualization tool that facilitates the analytical evaluation of the modulation scheme's performance (Figure 7b). A constellation represents the set of possible values for the amplitude and phase of the sinusoid in the complex plan - 8 points for the 8-QAM constellation. In Figure 7b, the point of the constellation corresponding to each symbol of the transmitted waveform is highlighted in red.

This notion of *symbol* leads to defining a second transmission quality parameter, which is of great practical importance: Symbol Error Rate (SER), or the probability that a symbol will be received erroneously. The relationship between the SER and the BER depends on the way the bits are mapped on each symbol. Generally, this mapping is optimized in order that an erroneous symbol results in no more than one erroneous bit (Gray mapping). Such general cases yield the relation BER = SER/b, where b is the number of bits per symbol. Figure 8 illustrates a Karnaugh-style Gray map for the 8-QAM constellation. If this map was used to

construct the modulated signal of Figure 7a, the corresponding transmitted information binary sequence would be …011110000…

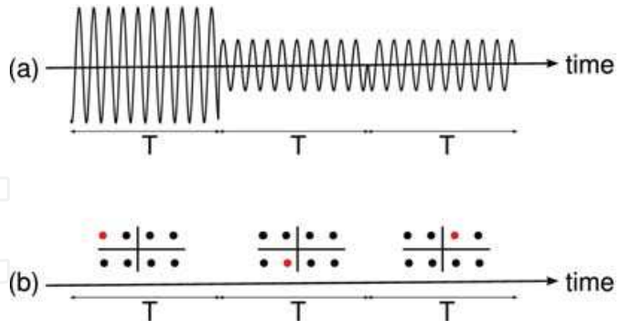


Fig. 7. Different representations of a 8-QAM modulated signal: a) temporal representation; b) signal-space representation (i.e., a constellation)

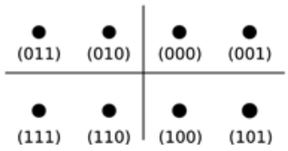


Fig. 8. Karnaugh-style Gray map of a 8-QAM constellation.

For general QAM transmissions over the additive white Gaussian noise channel, the SER can be shown to be related to the distance *d* between nearest points in the constellation to the noise standard deviation ratio *sigma*, or to the SNR channel and to the number of bits per symbol. This relationship is expressed by the equation (Proakis, 1995):

$SER = 4\beta Q(\alpha)(1 - Q(\alpha))$, where: $\alpha = \sqrt{3SNR/(2^b - 1)}$ and: $\beta = 1 - 2^{-b/2}$. In this equation, the β term represents the average number or nearest neighbors of a constellation point. Figure 9 presents the SER performance of several classical QAM modulations.

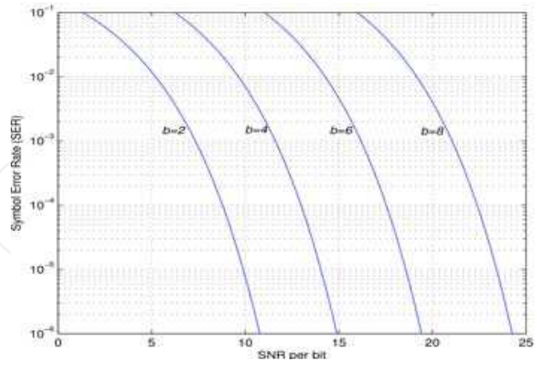


Fig. 9. SER performance of several classic QAM modulations

4.2 Achieving unequal error protection at the modulation level: examples of HQAM, OFDM and HQAM-OFDM combinations

Several possible methods for achieving unequal error protection during modulation are described in this subsection. One possibility is to use hierarchical modulation methods, such as HQAM. The idea of hierarchical modulation comes from the fact that the SER is directly

related to the distance between nearest points in the constellation. Figure 10 illustrates the principle, using the example of a two-level 4/8-HQAM and its associated Karnaugh-style Gray map of two highly important bits and two less important bits. As the figure shows, the points are grouped into four clouds, where the distance between two nearest points in a cloud is d_2 and the distance between clouds is d_1>d_2. The map indicates that the least significant bits serve to differentiate between the four points within each cloud and that the two most significant bits serve to differentiate between the clouds. Hence, the BER for the least significant bits is directly related to the distance d_2. For the two most significant bits, the constellation can also be seen as a 4-QAM constellation with distance between points equal to d_1+d_2/2, where d_2/2 can be seen as a noise component since it contains no information about these most significant bits.

The ratio $\lambda=d_2/d_1$ controls the bits' relative priority. When λ=0, the result is a uniform 4-QAM (the two least significant bits are merely discarded); when λ=1, the result is a uniform 8-QAM (equal priority for all bits). When 0<λ<1, the most significant bits have a greater priority.

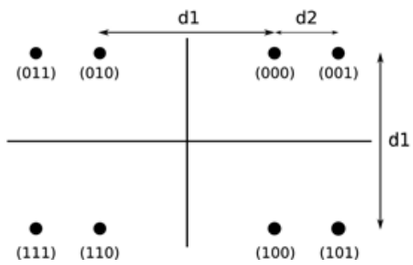


Fig. 10. 8-HQAM constellation and Karnaugh-style Gray map for two most significant bits and one least significant bit

BER analytical expressions have been formulated by Vitthaladevuni (2001) for the class of 4/M-HQAM constellations, in which the number of clouds is 4 (i.e., $\log_2(4)$=2 most significant bits) and the number of points within each cloud is M (i.e., $\log_2$(M) least significant bits). Clearly, the BER expressions are related to the ratio d_2/d_1, which, in practice, must be chosen appropriately according to the requirements of the video bitstreams. For the 4/16-HQAM constellations, these BERs can be approximated as:

$$BER_{most\_significant} = \frac{1}{4}\left(erfc\left(\frac{d_1}{\sqrt{N_0}} \right) + erfc\left(\frac{d_1 + 2d_2}{\sqrt{N_0}} \right) \right)$$
$$BER_{less\_significant} = \frac{1}{4}\left(erfc\left(\frac{d_2}{\sqrt{N_0}} \right) + erfc\left(\frac{2d_1 + d_2}{\sqrt{N_0}} \right) - erfc\left(\frac{2d_1 + 3d_2}{\sqrt{N_0}} \right) \right) \quad (3)$$

Figure 11 illustrates these BERs for a 4/16-HQAM constellation as a function of the SNR, for a ratio of lambda=d_2/d_1=0.8.

A second possibility for achieving unequal error protection during modulation is based on multi-carrier modulation methods, such as DMT or OFDM. These methods are used extensively in current digital transmission systems (e.g., DVB, ADSL, Wimax, LTE). In DMT and OFDM, a Quadrature Amplitude Modulation (QAM) is carried out on a sub-channel, using IFFT/FFT processing (Starr, 1999). For a multi-carrier symbol rate 1/T, the spacing between the sub-channels is also 1/T, and the overall bandwidth is N/T, where N is half the

size of the IFFT/FFT used for processing. One of the key issues in designing efficient DMT systems is the bit-loading algorithm that optimizes the bit and power allocation over the QAM sub-channels, based on their power gains and noise levels.

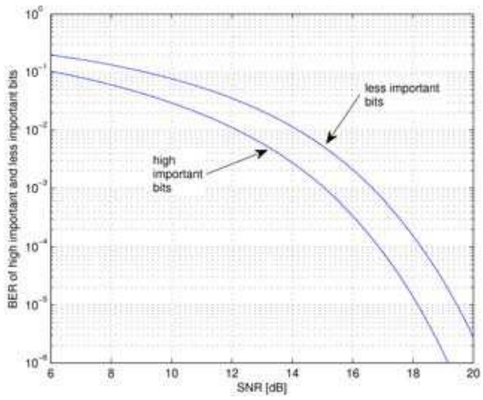


Fig. 11. Approximate BERs of a 4/16-HQAM constellation as a function of the SNR ($\lambda=d_2/d_1=0.8$)

Due to their multi-carrier structure, DMT and OFDM are inherently adapted to provide Frequency Division Multiplexing (FDM) of the different bitstreams. Thus, a natural approach for providing UEP is to multiplex the different streams using FDM, adapting the bit-loading algorithm so that it will provide different BERs for the different bitstreams. Generally, these algorithms assign the "best" sub-carriers to the bits of higher importance, and the other ones to the less important bits, as is suggested by the theoretical water-filling analysis proposed by Starr (1999). Such approaches have been studied by Zheng (2000) and Goudemand (2006), who point out different quality/complexity trade-offs. Figure 12 shows the bit-allocation produced by Zheng's algorithm (2000) for unequal error protection in an ADSL environment, using the European Telecommunications Standards Institute's loop 2 channel model (ETSI, 1996) and two video bitstreams from a data-partitioning MPEG-2 video coder at 6 Mbps (3.6 Mbps for the highly important bitstream and 2.4 Mbps for the less important one). The BERs are 10^{-7} and 10^{-4} respectively.

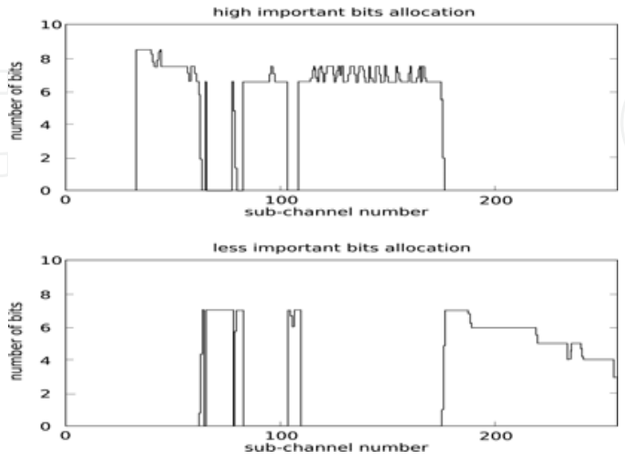


Fig. 12. Bit-allocation result for FDM unequal error protection based on DMT modulation (Goudemand, 2006)

Another possibility for achieving unequal error protection on multi-carrier transmission systems using DMT and OFDM is to combine the two methods described above. This method, called *hierarchical multi-carrier modulation*, involves modulating each sub-carrier using HQAM so that, unlike the pure FDM approach, each sub-channel carries both important and less important bits with embedded unequal error protection. The bit and power allocation over hierarchical multi-carrier modulation minimizes the total transmitted power while maintaining a constant bit-rate and BER for each bitstream. In practice, the bit and power allocation is calculated in two steps, providing two-level unequal error protection. The first step allocates higher priority bits and their corresponding power, and the second one allocates the remaining less important bits (Goudemand1, 2006).
As usual, the bit-loading algorithm must be computationally efficient and must reflect the water-filling principle, in that more bits must be allocated to the sub-channels with the lowest noise levels. Figure 13 shows a typical bit allocation in a hierarchical multi-carrier modulation system.

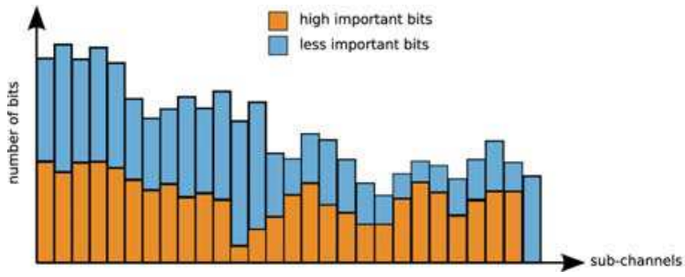


Fig. 13. A typical bit allocation for a hierarchical multi-carrier system

4.3 Achieving unequal error protection at the FEC level

Subsection 4.1 presented our analysis of SER and BER performance in general QAM modulation. This performance is generally insufficient for the needs of video communication. It is thus interesting to compare the bit-rate of a QAM modulation (b bits/symbol) to the maximum bit-rate that could theoretically be transmitted reliably (i.e., at a BER as close to zero as desired) at the same SNR. This maximum bit-rate is known as the channel capacity: c=log2(1+SNR) bits/symbol (Shannon, 1948). Figure 14 facilitates this comparison. In the figure, the SER is plotted as a function of the ratio bit-rate/channel capacity=b/c for different channel capacity values.
Clearly, satisfactory values of the BER for video transmission (below 10^{-4}) can be achieved at a rate that is very far from the channel capacity. In order to improve transmission efficiency (i.e., in order to achieve sufficiently low BER at rates close to the channel capacity), error correcting codes have to be used. These error correcting codes operate by adding controlled redundancy to the information bits so that only certain bit patterns can be transmitted. The error correcting decoder, which is aware of these patterns, will then be able to detect and even correct some erroneous bits.
Among the class of error correcting codes, linear block codes, also known as Forward Error Correction (FEC) codes, are of practical importance (Costello, 1998). One of them, the Reed-Solomon codes, are used extensively in recent popular communication systems, such as DVB or ADSL. Each Reed-Solomon block code is formed of N bytes, of which K bytes are information bytes and (N-K) bytes are redundancy bytes. Figure 15 illustrates the error correcting capabilities of such codes.

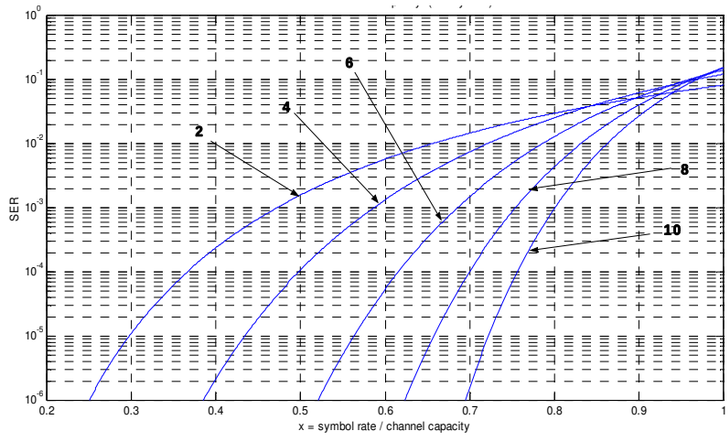


Fig. 14. QAM SER without FEC, as a function of the ratio bit-rate/channel capacity. Parameter: channel capacity (bit/symbol).

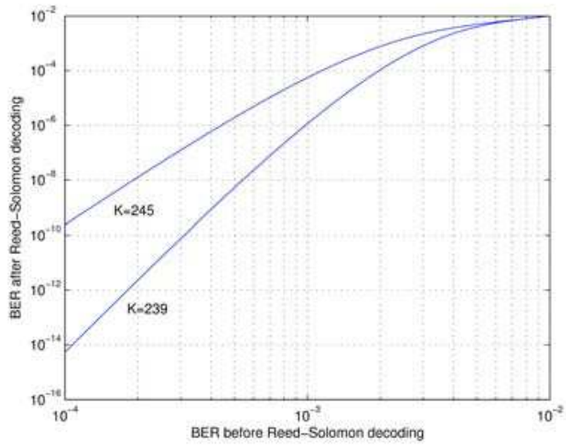


Fig. 15. Error correcting capabilities of some Reed-Solomon codes (N=255,K)

Including FEC in video transmission systems offers new possibilities for achieving UEP. Instead of, or in addition to, providing UEP at the modulation level, UEP can easily be provided at the FEC level by adding different FEC redundancies (and thus different error correction capabilities) to the bits, according to their perceptual importance.

Figure 16 provides a synoptic version of such a UEP approach in the context of a DMT + Reed-Solomon FEC coding, comparing it to a classic DMT + Reed-Solomon FEC coding without UEP.

In this figure, RH and RL, respectively, represent the bit-rates of the bitstreams of higher and lesser importance, and (N_H,K_H) and (N_L,K_L) represent the Reed-Solomon parameters of these two bitstreams. For the classic transmission scheme, R_0=R_H+R_L is the bit-rate, and (N_0,K_0) are the Reed-Solomon parameters. For the two UEP approaches, the DMT modulators are the same, with the same bit and power loading, which makes the BER before Reed-Solomon decoding the same in both approaches.

In order to evaluate the possibilities of this approach to UEP, the cumulative redundancy of the two Reed-Solomon codes, (N_H,K_H) and (N_L,K_L), shall equal the redundancy of (N_0,K_0) (i.e., $R_0.N_0/K_0$=$R_H.N_H/K_H$+$R_L.N_L/K_L$). Figure 17 illustrates the relative variations of the BER of the highly important bits (BER_H) versus the relative variations of the BER of the less

important bits (BER_L), in terms of the BER of the classic method without UEP (BER_0) for an ADSL environment (ANSI CSA5 test loop) in which (N_0,K_0)=(255,239) and for different proportions of less important bits (10% to 50%).

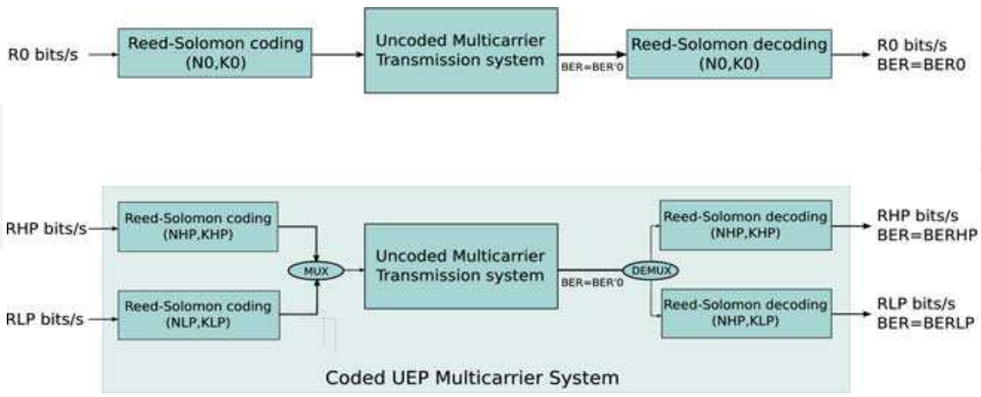


Fig. 16. Synoptic version of a DMT system using FEC with UEP and a classic DMT system using FEC without UEP.

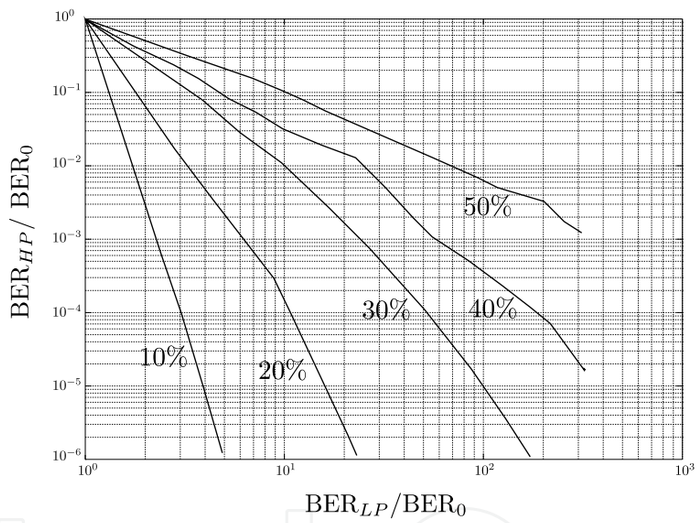


Fig. 17. Evaluation of the possibilities offered by providing unequal error protection at the FEC level.

More sophisticated UEP methods based on error control codes have been explored in the literature. In particular, the UEP methods using multilevel codes and multistage decoding accomplish Unequal Error Protection quite satisfactorily (Wachsmann, 1999; Chui, 2008).

5. Scalability tools

In the previous section, we underlined the fact that UEP with different priority levels can be applied successfully in modern video transmission systems to provide flexibility as well as the QoS level required by end-users. This is true because not all the various bits of the transmitted video data streams have the same importance with respect to reconstructed video quality. Recently, digital video compression standards have introduced the concept of *scalability* to allow video content to be encoded with different levels of resolution and

different levels of quality. In this section, we provide a brief overview of the existing scalability modes, as well as the corresponding coding tools.

Typically, video coding standards propose four main scalability modes: SNR scalability, spatial scalability, temporal scalability and data partioning. Each of the modes is presented briefly below, with more detail given for the first and the last modes.

SNR scalability allows the delivery of a video bitstream compressed into several separate layers with the same spatio-temporal resolution but different quality levels. The base layer, which contains a reduced quality version of the encoded video signal, is typically transmitted with a high protection level—for example, using one of the techniques described in Section 4 —to guarantee that video will be decoded even with high error rates. Then, each additional layer enhances the quality of the reconstructed video, but is transmitted with a reduced protection level. Hence, SNR scalability provides graceful degradation of decoded video according to the transmission quality.

A block diagram of a SNR scalable video encoder is shown in Figure 18. First, the base layer is encoded using coarse quantization. Then, this base layer is decoded, and the residual error between the reconstructed base layer and the original video signal, which constitutes the enhancement layer, is computed and then re-encoded during a second encoding stage, using finer quantization.

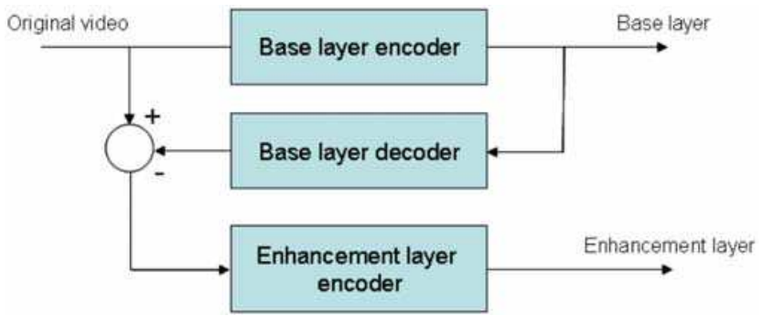


Fig. 18. Block diagram of a SNR scalable video encoder.

Other scalability tools rely on the same basic scheme. With *spatial scalability*, however, the video signal is encoded into separate streams corresponding to different spatial resolutions. The base layer consists of an encoded version of the video signal at lowest spatial resolution. Then, each enhancement layer contains additional data related to higher resolutions. *Temporal scalability* allows the simultaneous delivery of video signals with different frame rates.

The last mode, *data partitioning* (DP) (Mathew & Arnold, 1999), is quite similar to SNR scalability. The video signal is encoded into two separate bitstreams, with the additional layer allowing increased video quality if received at the decoding stage. However, data partitioning has the great advantage of being easier to implement. In fact, the DP mode is applied directly to the encoded bitstream and does not require any significant modification of the codec. In MPEG-2, the compressed video bitstream is split into two separate streams: the first one, corresponding to the base layer, contains the most important data for each of the 8x8 encoded blocks, such as low frequencies DCT coefficients or motion vectors. The remaining high frequency DCT coefficients, corresponding to details of each block, constitute the enhancement layer. The priority breakpoint (PBP) parameter marks the place where the bitstream is split into two parts. In the next section, we demonstrate that data partitioning is a great advantage for increasing error resilience, if unequal error protection is applied before transmission.

Since 2003, the scalability concept has been put forward through the design of the scalable video coding (SVC) amendment of the H.264/AVC standard. The SVC technology is based on the coding tools available in H.264/AVC, but includes also specific hierarchical and inter-layer predictive coding techniques. For further details on SVC, see (Schwarz et al., 2007) and (Huang et al., 2007). SVC provides a way to encode a video sequence into a single compressed bit-stream composed of a H.264/AVC compatible base layer and multiple additional layers that enhance spatial resolution, frame rate and quality. It is therefore possible to extract on-the-fly (inside the network or at the terminal) a limited number of layers in order to decode video with given resolution and quality level. Another obvious advantage of SVC is that UEP can be applied efficiently to the different parts of the bit stream having different importance in terms of reconstructed video quality. Hence, using SVC allows supporting a broad range of devices with different capabilities (display resolution, battery power) and access networks. Several papers on SVC applications have been proposed recently in the literature (Schierl et al., 2006) (Kouadio et al., 2008) (Thang et al., 2008). For example, Hellge et al. propose to combine of the spatial and SNR scalability features of SVC with the hierarchical modulation of DVB-H (Hellge et al., 2009). The authors show the benefits of jointly using SVC with hierarchical modulation in terms of error robustness or increased number of services.

6. Application to digital video delivery over digital subscriber lines

In the following Section, we present a complete quality-oriented transmission system for video distribution over DSL, optimized by applying techniques previously described.

The diffusion of audiovisual content, or Video on Demand (VoD), was one of the major objectives in the development of ADSL technology. This technology was initially a huge success for high-speed internet access and gaming. With increased bit-rates, in recent years, there has been a return to one of the original objectives, which was audio and video diffusion through ADSL.

ADSL involves transmission over the physical link portion of the telephone network, called subscriber loop, which connects DSL Multiplexer (DSLAM) located at the central office (CO) and Customer Premises Equipment (CPE) on the Subscriber side. The subscriber loop consists of a twisted-pair copper line. ADSL coexists with voiceband service by using the high frequency band above the one allocated to Plain Old Telephony Services (POTS). This relatively inexpensive technology benefits from the existence of a reliable widespread copper-wire infrastructure. Typically, several customer lines from the same CO ends at the same DSLAM, and the DSLAM outputs are connected to a high-speed Internet backbone line. The modulation used by International Telecommunications Union (ITU) in various standardized versions of ADSL is the DMT (see paragraph 4).

ADSL experiments and knowledge have resulted in several developments (e.g., ADSL1, ADSL2, ADSL2+). Initially, ADSL had a theoretical maximum bit-rate of about 6 Mb/s, which was rapidly multiplied by a factor of almost five in ADSL2+. For a fixed error rate and a fixed transmission power, throughput is strongly dependent on the line characteristics:

- attenuation of the copper line, which depends on line length; and
- noise, including Near End CrossTalk and Far End CrossTalk interference from lines in a same bundle, and impulsive noise.

Unfortunately, high bit-rates are available only to users located near the DSLAM. For users more than 2.5 kilometers from the CO, line attenuation restricts the bit-rate so that it rarely exceeds 4 Mbps. As a result, the subscribers cannot all receive the same bit-rate with the same error rate. If a video is sent with the same bit-rate for all the subscribers, then each subscriber receives the video at a different error rate, which may cause unacceptable visual degradations for a lot of users if the error rate is high. This usually happens for users with limited link capacity below the required bit-rate. In the same way, if the same error rate is set for all subscribers, then the bit-rate must be adapted for each subscriber. Generally, the latter solution is preferable to insure reliability of the received data.
However, in the context of video transmission, it is also necessary to take the real time constraint into account in the video broadcast. All subscribers must be able to receive video simultaneously with an acceptable quality level. To allow this, transcoding is usually done in platform somewhere between the server and the customer. Generally, using transcoding makes it possible, for example, to multicast the video over heterogeneous links. Transcoding is an all-embracing term, which can involve adapting the standard, the format, the bit-rate or the frequency. In this Section, we deal only with transrating, which consists of adapting the video bit-rate to each subscriber loop by removing part of the transmitted video information. Of course, this information removal procedure must be carried out in such a manner that the video received is the least degraded version possible.
In the literature, there are many transrating algorithms that are more, or less, complex. Some are based, for example, on the suppression of high-frequency spatial information (i.e., removal of high-frequency DCT coefficients) (Werner, 1999; Assuncao & Ghanbari, 1997a; Sun et al., 1996; Celandroni et al., 2000). Others are based on the suppression of certain images in order to reduce the display frequency, as in temporal scalability (Horn et al., 1999; Lagendijk et al., 2000). Still others are based on a stronger quantization of predefined image areas (Tudor & Verner, 1997; Assuncao & Ghanbari, 1997a; Assuncao & Ghanbari, 1997b; Sun et al., 1996; Celandroni et al., 2000). These algorithms can be divided into two groups: *closed-loop transrating*, which require a complete decoding and re-encoding of the compressed video source, and *open-loop transrating*, which generally do not resort to a complete decoding, making them much less complex (Assuncao & Ghanbari, 1997a).
In this section, we present a Joint Source and Channel Coding (JSCC) approach for broadband MPEG-2 video distribution over DSL. Originally published by Coudoux et al. (2008), this approach combines a layered hierarchical video transrating scheme with an unequal error protection (UEP) technique and multi-carrier modulation for DSL video distribution in order to optimize the end-to-end video quality. We first highlight the existence of a "bottleneck bandwidth" of the video source after transrating; at this bandwidth, the quality of the video received by the subscriber is optimal. Following that, we present the JSCC system, and then explain the transrating optimization procedure. Finally, we report our simulation results.

6.1 Motivations

Let us consider an example of a MPEG-2 MP@ML single layer video source encoded at 6 Mbps. For the lines not reaching 6 Mbps with a sufficiently low error rate under practical power constraint, decreasing the bit-rate will decrease the transmission error rate. However, this improvement is counterbalanced by the fact that decreasing the bit-rate of the source video (and, as a result, removing some useful video information), increases the deterioration

of the quality of the video received. Thus, lowering the bitrate clearly has two contradictory effects on the quality of the video available to the end-user.
Consequently, it can be assumed that, for each subscriber loop that cannot achieve 6 Mbps with appropriate QoS level, there is an ideal bit-rate, called the "bottleneck" bit-rate, at which the overall deterioration is minimal. This minimal deterioration is the result of a compromise between the distortion generated by the bit-rate reduction and the distortion generated by the transmission. Figure 19 illustrates this basic idea.

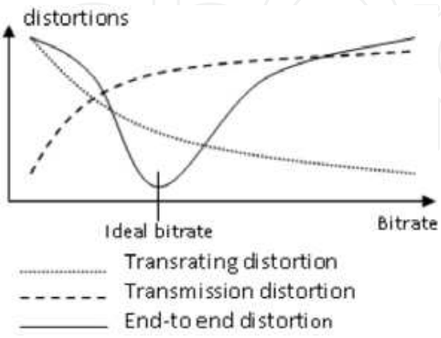


Fig. 19. End-to-end distortion in terms of the overall bit-rate

To insure a minimum level of deterioration, and thus a received video quality threshold, we used a hierarchical video source coding system, coupled with UEP at channel coding level as described in the preceding sections.

6.2 System architecture

Our JSCC system is based on the Data Partitioning (DP) mode used in the MPEG-2 standard, which splits the MPEG-2 bitstream into two hierarchical bitstreams (see Section 5). For simplicity, DP was chosen in order to conform to real-time constraint: indeed, it makes possible to split on-the-fly the incoming video bitstream into two separate ones. The most important bitstream is the base layer containing the most important data (e.g., headers, motion vectors and low-frequency DCT coefficients). The second bitstream is the enhancement layer (EL), containing all the other data, typically high-frequency DCT coefficients. Thus, these two bitstreams can be protected differently against channel errors, using UEP as explained in section 4.
The proposed JSCC system can be decomposed in a four-step adaptation process, as shown in Figure 20:

1. *Estimating the channel* -- The channel-to-noise ratios of the considered line are estimated for each carrier. These ratios are used by the bit- and power-loading algorithm (see Section 4).
2. *Defining the transrating parameters* -- The single layer input video bitstream is transcoded into two video bitstreams. Starting with the estimated channel parameters, the bit-rates of both the base layer bitstream and the enhanced layer bitstream must be determined.
3. *Transrating using data partitioning* -- The MPEG-2 bitstream coming from the very high bit-rate network (from a video server) and received at the CO is transcoded based on the transrating parameters defined in step 2.
4. *Coding and transmission* -- UEP is applied to the two bitstreams. Reed Solomon (RS) codes are used as in the ADSL standard in this step. As explained in section 4, two different RS codes can be applied to the two bitstreams before transmission. The two

bitstreams are then transmitted at fixed BER (with BER1<BER2). The RS code is applied to the base layer, so that the decoded base bitstream can be considered to be error-free at the receiver.

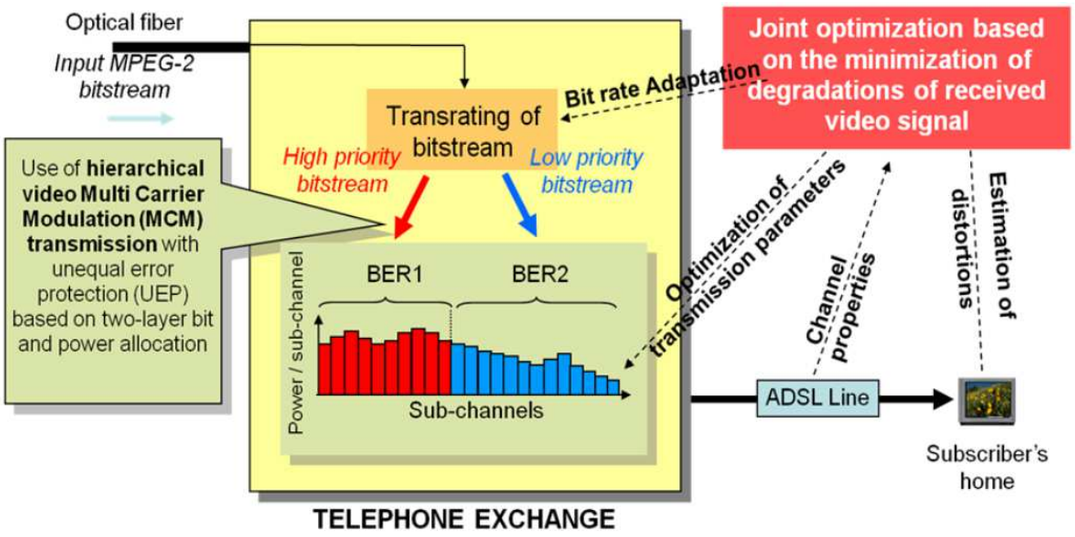


Fig. 20. Overview of the complete JSCC architecture

The first two steps are performed by the ADSL modem during initialization, assuming quasi-static transmission conditions; the last two are specific to our transrating system.

The Data partitioning step uses the Priority Break Point (denoted PBP1 in this chapter) defined in the MPEG-2 standard. PBP1 defines the separation point of the two bitstreams, and its value is related to the number of non-zero DCT coefficients preserved in the base layer (i.e., only the first low-frequency DCT coefficients). PBP1 depends on the number of VLC codes preserved in the base layer after the zigzag serialization MPEG procedure. The JSCC system maintains PBP1 constant in order to insure a minimal video quality level at the receiver side.

As the goal is to reduce the total video bit-rate, we introduce a second Priority Break Point, denoted PBP2 (PBP1≤ PBP2≤127), which defines the point at which the remaining VLC codes are discarded. The extreme values of PBP2 correspond to the situation in which the base layer only is transmitted (PBP2=PBP1) and to the situation in which all the bit-rates are preserved before transrating (PBP2=127). The other intermediate values make it possible to adjust the video bit-rate of the enhanced layer. This adjustment can propagate a quantization error (i.e., drift) and thus lead to a reduction in video quality. However, the quality metric used for determining transrating parameters takes into account of this error (C. Goudemand et al., 2007).

The simulations performed by Goudemand et al. (2006) showed that the optimal value of the PBP2 determines the best compromise between the perceived distortion and the bit-rate reduction obtained. Simulations on different MPEG-2 video sequences transmitted at 6 Mbps were performed with different PBP2 values. From one sequence to another, the bit-rate curves obtained after transrating versus the PBP2 value are very similar. For this reason, the average of these curves was used to determine the video bit-rate after transrating based on PBP2 (Figure 21). As illustrated, the bit-rate increases rapidly from 2.6 to 5.95 Mbps when the PBP2 goes from 64 to 85. Then, from PBP2=85 to 127, the bit-rate increases slowly. This

evolution of the bit-rate with PBP2 can be easily explained by the fact that the DCT coefficient's energy decreases as the spatial frequency increases.

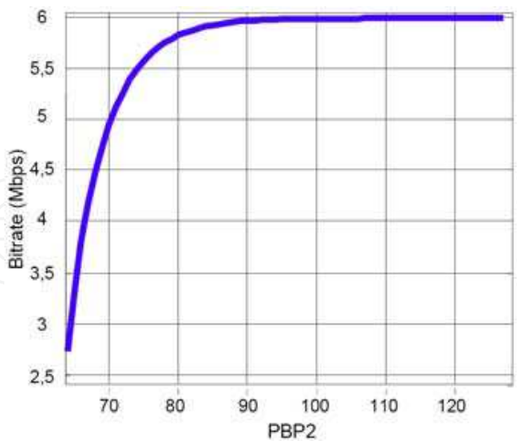


Fig. 21. Evolution of total bit rate as a function of the PBP2 parameter.

We also verified that the bit-rate overhead introduced by data partitioning does not exceed 1.5% of the aggregate video rate and therefore can be neglected. Let us consider our example of a coded sequence at 6 Mbps, transrated with the PBP1 and PBP2 equal to 67 and 85, respectively. According to Figure 21, the bit-rate at PBP1=PBP2=67 is about 4.3 Mbps, which is the base layer bit-rate. With PBP2=85, the bit-rate is 5.95 Mbps, which corresponds to the total bit-rate after transrating (i.e., the total bit-rate of the base and enhanced layers). The bit-rate corresponding to the enhancement layer is therefore equal to 1.65 Mbps.

In order to optimize the overall video quality, it is necessary to have a quality metric which permits to evaluate practically the visual impact of eliminating high-frequency DCT coefficients by transrating. First, let us introduce some features of the quality metric used in our system for the case of still images (i.e., the (I) intra-coded pictures of an MPEG-2 bitstream). Remember that, for the intra-pictures, the DCT coefficients are uniformly quantized. A quantization table defines the quantization step values that were obtained from subjective measurements.

The Normalized Weighted Mean Square Error (NWMSE) quality metric, proposed by Goudemand (2007), is based on the quantization property of the DCT coefficients involved in the MPEG-2 coder. As each quantization step value is related to the importance of the visual impact of the corresponding DCT coefficient, the weights used for this metric are related to these quantization steps. Such WMSE metric has been first introduced by Vandendorpe (1991) in the context of sub-band coding.

In the present case, the reference for calculating the quality metric is the input video sequence compressed at 6 Mbps. First considering the intra-coded pictures, WMSE is typically calculated from the DCT coefficients perceptually weighted according to the contrast sensitivity function (CSF) of the human eye. The weights used in the WMSE computation are inversely proportional to the square of the quantization step values. Therefore, the WMSE of the DCT coefficients can be seen for each block as the MSE of the corresponding quantized values. The perceptual distortion depends only on the error magnitude of the quantized DCT coefficients, independent of the spatial frequencies. The quantized coefficients thus have the same visual importance in the DCT domain. Any additional transrating or transmission distortion can be taken into account in the WMSE of the quantized coefficients.

The previous approach has been extended to (P) and (B) inter-coded pictures, such that WMSE can be computed from quantized DCT coefficients whatever the image type. Normalization is also introduced for each picture type, which depends on the different macroblock types as well as the rate control. However, the suppression of DCT coefficients due to transrating or transmission errors typically results in *drift error* among successive pictures. This phenomenon is accounted for by introducing weighting factors noted W_I, W_P, W_B defined as the average number of frames affected by an error in I, P, B frame, respectively. Finally, the average normalized perceptually weighted MSE of a sequence $NWMSE_{sequence}$ is the weighted sum of the NWMSE averaged for I, P, B frames defined as:

$$NWMSE_{sequence} = W_I E\{NWMSE_{I\ frame}\} + W_P E\{NWMSE_{P\ frame}\} + W_B E\{NWMSE_{B\ frame}\} \quad (4)$$

6.3 System parameter optimization

The optimization process consists in determining the bit-rate of the two bitstreams, and therefore the values of the two Priority Break Points: PBP1 and PBP2, such that the received video quality (measured thanks to the above metric) is the best. The end-to-end perceptual quality of the video is considered to depend on the total bit-rate. In our JSCC architecture, the two-layer bitstreams received undergoes two independent types of degradations: transrating degradation and transmission degradation. Thus, the total degradation of the system is the sum of these two types, which can be evaluated using the NWMSE quality metric. Figure 22 shows two curves. For a given and fixed value of PBP1, the left side of the figure illustrates the evolution of the total transmission power P with respect to the PBP2.

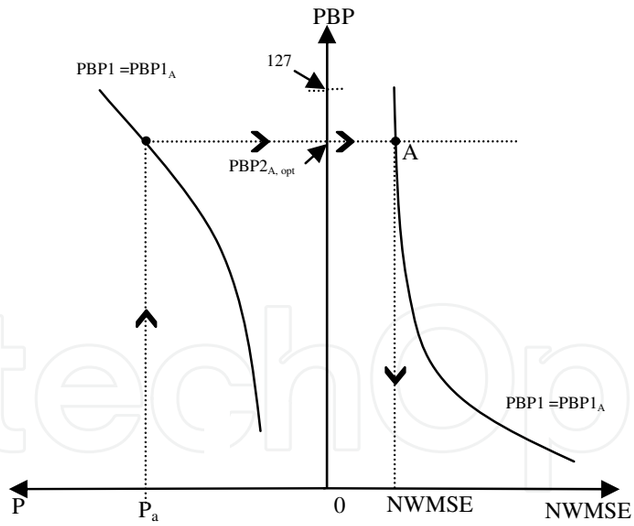


Fig. 22. Proposed approach for the determination of the end-to-end NWMSE

Let us consider the case of a transmission at power P, where P=Pa. This is a case for example in which all the total available power P_a is used for the transmission, making it possible to have the maximum transmission bit-rate. For any value of the PBP1=$PBP1_A$ given, the optimal value of the PBP2=$PBP2_{A,opt}$ is immediately obtained, as shown on the left curve in Figure 22. Thus, the operating point of the system given by ($PBP1_A$, $PBP2_{A,opt}$) can be determined. In addition, the NWMSE is a decreasing function of PBP2, as illustrated on the

curve shown on the right side part of Figure 22. Thus, the NWMSE= $NWMSE_A$ can be therefore determined from the value of $PBP2_{opt}$.
If another similar but different value of PBP1 is considered, other curves must be used, such as those in Figure 22. Since the value of PBP1 is unknown, the direct determination of $PBP2_{opt}$ is compromised. If all the values of PBP1 from 64 to 127 are considered, $PBP2_{opt}$ and the NWMSE can be determined for each one of these values. A series of NWMSE values will be obtained. The operating point of the optimal system is thus the couple ($PBP1_{opt}$, $PBP2_{opt}$), which will minimize the global end-to-end NWMSE.
In short, by varying the PBP1 from 64 to 127, determining the optimal value of the PBP2 and the NWMSE for each value of the PBP1, and selecting the couple ($PBP1_{opt}$, $PBP2_{opt}$) that produces the minimal NWMSE, the locus C of all the operating points for the PBP1 values between 64 and 127 can be obtained, as shown in Figure 23.

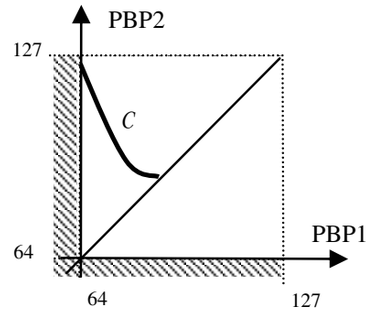


Fig. 23. Evolution of the PBP2 parameter as a function of PBP1, under transmit power constraint (Coudoux et al., 2008).

6.4 Simulation results

Experiments were conducted in two steps: 1) the determination of the optimal transrating parameters and 2) the simulation of the transmission of transrated video sequences over ADSL lines.
The optimal parameters were determined as follows: for each subscriber loop, the transrating couples (PBP1; PBP2) were estimated, which allowed the base and enhancement layers to be transmitted. These parameter couples were obtained using the bit and power allocations at a fixed bit-rate for the two bitstreams, based on the values of PBP1 and PBP2 at a maximum total power of 110 mW as stated in the ADSL standard. The optimal couple was then the one corresponding to the minimum total end-to-end NWMSE.
The simulations were carried out at 6 Mbps initial bit-rate using MPEG-2 (MP@ML) video sequences. ADSL Reed Solomon RS (255, 239) were applied to the base layer, which generated a bit-rate overcost of about 6.7%. The BER of the base layer transmission was fixed at 10^{-6} before RS decoding ($BER_{BL} = 10^{-6}$), which corresponds to about 6.4 10^{-33} after RS decoding. For this reason, the BL transmission was assumed to be error-free. The BER of the enhanced layer transmission was fixed at 10^{-4} ($BER_{HL} = 10^{-4}$) so that visual degradations would be minimized whatever the values of PBP1 and PBP2. The bit and power allocation algorithm used frequency division multiplexing of the two layers, as explained in Section 4. Once the optimal PBP1 and PBP2 parameters for the transmission were determined, the transmission of the video processed with these values was simulated.
For purposes of comparison, the ADSL MPEG-2 transmission was used in our experiments as the reference transmission. Figure 24 represents the evolution of the end-to end NWMSE

as a function of the subscriber line length. The European Telecommunications Standards Institute (ETSI) test loop 1 was used (ETSI, 1996). Remember the combined source/channel transrating system described in this section was not designed to improve the quality of the received image on "a good" line, but rather to increase the range of ADSL video transmission. Given this reminder, a single layer ADSL transmission at a BER of 10^{-6} over the 1800 m-long ETSI loop 1 required 2.5 dB more than the authorized ADSL maximum transmission power of 110 mW. In order to reduce the total power to 110 mW without changing all other transmission parameters, it was necessary to reduce the line length by approximately 110 m. The traditional ADSL transmission on the same line was thus limited to 1700 m around the telephone exchange. The solution proposed by the authors led to a slightly degraded transmission in an area with an 1800 m radius, which can possibly be extended if progressive degradations, increasing with the line length, are accepted.
The two-layer JSCC used in these experiments was compared to a single-layer transrating system based on the same principle, but using only one priority break point PBP2=PBP1. With the same average system quality, the two-layer JSCC system led to an increase in the zone covered compared to the single-layer transmission system.

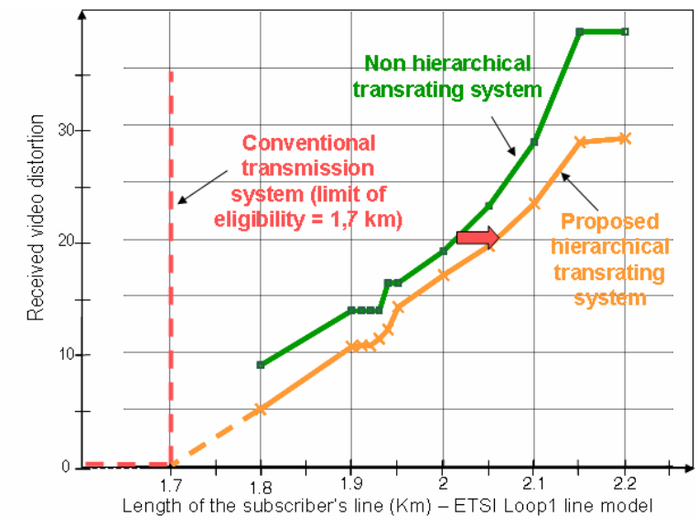


Fig. 24. Evolution of end-to-end perceptual distortions as a function of subscriber's line length (from Coudoux et al., 2008).

7. Conclusion

We demonstrated in this chapter how source and channel coding can be efficiently combined in order to fulfil QoS requirements of modern video communication systems. The underlying philosophy relies on the fact that not all the bits of a compressed video bitstream have the same visual importance. Thus, both digital video compression and transmission parameters should be jointly designed in an end-to-end approach, aiming at delivering the best video quality to the final end-user. We also demonstrated that digital video signals can be encoded into several layers of varying perceptual relevance using scalability tools. Subsequently, if unequal error protection is applied to these layers, the most visually important layers will be the best protected. Such video transmission methods typically increase the flexibility and reliability of video communications in heterogeneous multimedia

environments, and produce a better quality of experience for the end-user. The effectiveness of a combined source/channel coding strategy was illustrated for the particular case of DSL video distribution. We showed that the eligibility level of video services can be optimally extended, by transrating the compressed video bitstream available on a distant server for the optimal bit-rate at which the end-user's visual quality is the best. The extension of these results to the case of the scalable extension of the H.264/AVC standard as well as wireless residential networks is currently under consideration in the TOSCANE project (see: http://www.lien.uhp-nancy.fr/lien/index.php?p=toscane).

Combined source and channel coding strategies constitute a promising research topic in the field of video communications. This is particularly true when considering the emergence of new services, such as mobile video streaming on small display devices or 3D-TV, as well as new transmission architectures, such as wireless vision sensor networks, for example. Future work will try to develop more efficient solutions to respond to the new challenges of video communications in terms of error resilience, accessibility, interactivity and universality.

8. References

Chui, J.; Calderbank, A.R. (2008). *Multilevel diversity-embedded space-time codes for video broadcasting over WiMAX*, Proceedings of the IEEE International Conference on Information Theory, pp. 1068-1072, Nice, France, June 2008.

Coudoux, F.-X.; Gazalet, M.G.; Mouton-Goudemand, C.; Corlay & P.; Gharbi, M. (2008). Extended Coverage for DSL Video Distribution Using a Quality-Oriented JSCC Architecture, *IEEE Transactions on Broadcasting*, Volume 54, Issue 3, Sept. 2008, pp. 525 – 531, ISSN: 0018-9316.

Coudoux F.X., Gazalet M., Derviaux C., Corlay P. (2001). Picture quality measurement based on block visibility in discrete cosine transform video sequences, *Journal of Electronic Imaging*, Vol.10(2), pp.498-510.

Daly S. (1993). The visible differences predictor: an algorithm for the assessment of image fidelity. In Watson A.B. (ed.), *Digital Images and Human Vision*, pp. 179-206, MIT Press.

ETSI Technical Report ETR 328 (1996). *Transmission and Multiplexing (TM ; Asymmetric Digital Subscriber Line (ADSL) ; Requirements and Performance*, Reference DTR/TM-06001, November 1996.

Faugeras O.D. (1979). Digital color image processing within the framework of a human visual model, *IEEE Trans. On Acoustics, Speech and Signal Processing*, vol. 22, no. 4, pp. 380-393.

Girod B. (1993). What's wrong with mean-squared error? , in *Visual Factors of Electronic Image Communications*. Cambridge, MA: MIT Press.

Glenn W.E. (1993), *Digital image compression based on visual perception and scene properties, video processing*, SMPTE Journal, pp. 395-397, May 1993

Goudemand, C., Coudoux, F.X., Gazalet, M. (2006). A study on Power and Bit Assignment of Embedded Multi-Carrier Modulation Schemes for Hierarchical Image Transmission over Digital Subscriber Line, *IEICE Transactions on Communications*, Vol. E89, No.7 (July 2006), page numbers (2071-2073).

Goudemand, C.; Gazalet, M.; Coudoux, F.X.; Gharbi, M. (2006). Reduced complexity power minimization algorithm for DMT transmission – Application to layered multimedia

services over DSL, *Proceedings of the 13th International Conference of Electronics, Circuits and Systems, ICECS 2006*, pp. 728-731, Nice, France, December 2006.

Goudemand C., Coudoux F.X., Gazalet M.G. (2006). Optimal bit rate adaptation for layered video transmission over spectrally shaped channels using multicarrier modulation, *IEEE ICIP-2006*, pp.13-16, Atlanta, USA.

Goudemand C., Gazalet M.G., Coudoux F.X., Corlay P., Gharbi M. (2007). A Low Complexity Image Quality Metric for Real-Time Open-Loop Transcoding Architectures, *IEEE ICC' 07*, Glasgow.

Huang, H-C., Peng, W-H., Chiang, T., Hang, H-M., "Advances in the Scalable Amendment of H.264/AVC", *IEEE Commun. Mag.*, pp. 68-76, Jan. 2007.

Hellge, C., Mirta, S., Schierl, T., Wiegand, T., „Mobile TV with SVC and Hierarchical Modulation for DVB-H Broadcast Services", *IEEE BMSB*, Bilbao, 13-15 May 2009.

Kouadio, A., Clare, M., Noblet, L., Bottreau, V., "SVC – a highly scalable version of H.264/AVC', EBU Technical Review, 2008.

Lagendijk R.L., Frimout E.D., Biemond J. (2000). Low-Complexity Rate-Distortion Optimal Transcoding of MPEG I-frames, *Signal Processing: Image Communication*, no. 15, p531 – 544.

Lee Y.L., Kim H.C., Park H.W. (1998). Blocking effect reduction of JPEG images by signal adaptive filtering, *IEEE Trans. Image Processing*, vol. 7, no. 2, pp. 229-234.

Lin, S., Costello, D. (1983), *Error Control Coding: Fundamentals and Applications*, Prentice-Hall Series in Computer Applications in Electrical Engineering, ISBN 0-13-283796-X, N.J. 07632.

Lukas F.X., Budrikis Z.L. (1982). Picture quality prediction based on a visual model, *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1679-1692.

Mannos J.L., Sakrison D.J. (1974). The effects of a visual fidelity criterion on the encoding of images, *IEEE Trans. On Information Theory*, vol. 20;, no. 4, pp. 525-535.

Mathew R., Arnold J.F. (1999), Efficient Layered Video coding using Data Partionning, *Signal Processing: Image Communication*, n°14, 1999.

Mitchell J., Pennebacker W., Fogg C., Le Gall D. (1995). *MPEG Video Standard Compression*, Van Nostrand Reinhold, New York.

Pennebaker W.B., Mitchell J.L. (1993). *JPEG Still Image Data Compression Standard,* Van Nostrand Reinhold, New York.

Pinson M., Wolf S. (2004). A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp 312-322.

Proakis, G.P. (1995). *Digital Communications*, third edition, McGraw-Hill Electrical Engineering Series, ISBN 0-07-051726-6, International Edition.

Rabbani M., Jones P.W. (1991),. *Digital Image Compression Techniques*, SPIE Press, Bellingham

Ramamurthi B., Gersho A. (1986). Nonlinear space variant postprocessing of block coded images, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1258-1267.

Richardson, I. (2003). H.264 *and* MPEG-4 *Video Compression: Video Coding for Next Generation Multimedia*, John Wiley & Sons, ISBN 0-470-84837-5.

Schierl, T., Gänger, K., Wiegand, T., Stockhammer, T., "SVC-based multisource streaming for robust video transmission in mobile ad hoc networks", *IEEE Wireless Commun. Mag.*, Special Issue on Multimedia in Wireless/Mobile Ad Hoc Networks, Vol. 13, N°. 5, pp. 96-103, Oct. 2006.

Schwarz, H., Marpe, D., Wiegand, T., "Overview of the Scalable Video Coding extension of the H.264/AVC standard", IEEE Trans. Circuits Syst. Video Technol., Vol. 17, N°. 9, pp. 1103-1120, Sept. 2007.

Shannon, C.E. (1948). A mathematical Theory of Communications, *Bell System Technical Journal*, Vol. 27 (October 1948), page numbers (379-423,623-656).

Starr, T., Cioffi, J.M., Silverman P.J. (1999). *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, ISBN 0-13-780545-4, N.J. 07458.

Sun H., Kwok W., Zdepski J. (1996). Architectures for MPEG Compressed Bitstream Scaling, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 2.

Talluri R., Moccagatta I., Nag Y., Gene C. (1999). « Error Concealment by Data Partitioning », *Signal Processing : Image Communication*, No. 14.

Tekalp, M. (1996). *Digital video processing*, Prentice Hall: Signal processing, Chapman&Hall, International Thomson Publishing.

Thang, T-C., Kang, J-W., Yoo, J-J., Ro, Y-M., "Optimal Multilayer Adaptation of SVC Video over Heterogeneous Environments", *Advances in Multimedia*, Volume 2008, Article ID 739192, 8 pages, doi:10.1155/2008/739192

Vandendorpe L. (1991). Optimized quantization for image subband coding, *Signal Processing: Image Commun.*, 4, pp. 65-79.

Vitthaladevuni, P.K.; Alouini, M.S. (2001). BER Computation of 4/M-QAM Hierarchical Constellations, *IEEE Transactions on Broadcasting*, Vol.47, No.3 (September 2001), page numbers (228-239).

Wachsmann, U.; Fischer, F.H; Huber, J.B; (1999). Multilevel Codes: Theoretical Concepts and Practical Design Rules, *IEEE Transactions of Information Theory*, Vol.45, No.5 (July 1999), page numbers (1361-1391).

Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. (2004). Image quality assessment; from error visibility to structural similarity, *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 300-612.

Wang Z., Lu L., Bovik A. (2004). Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, special issue on objective video quality metrics, vol. 19, pp. 121-132.

Watson A.B., Hu J., McGowan, J.F. (2001). Digital video quality metric based on human vision. *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20-29.

Werner O. (1999). Requantization for Transcoding of MPEG-2 Intra frames, *IEEE Trans. Image Processing*, Vol. 8, No. 2.

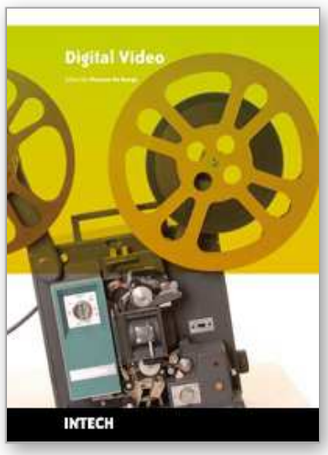
Wiegand T., Sullivan G. J., Bjntegaard G., Luthra A. (2003), Overview of the H.264/AVC video coding standard,. *IEEE Trans. Circuits Syst. Video Techn.* 13(7), pp 560-576 (2003)

Winkler S., *Digital Video Quality: Vision Models and Metrics*, Wiley Press, 2005.

Wu H.R. (Ed.), Rao K.R. (Ed.), *Digital Video Image Quality and Perceptual Coding*, CRC Press, Nov. 2005.

Yuen M., "Coding Artifacts and Visual Distortions", in *Digital Video Image Quality and Perceptual Coding*, H.R. Wu (Ed.), K.R. Rao (Ed.), CRC Press, 2005.

Zheng, H. ; Liu, K.J.R. (2000). Power Minimization for Delivering Integrated Multimedia Services over Digital Subscriber Line, *IEEE Journal of Selected Areas on Communications*, Vol.18, No.6, (June 2000), page numbers 841-849.



Digital Video

Edited by Floriano De Rango

ISBN 978-953-7619-70-1

Hard cover, 500 pages

Publisher InTech

Published online 01, February, 2010

Published in print edition February, 2010

This book tries to address different aspects and issues related to video and multimedia distribution over the heterogeneous environment considering broadband satellite networks and general wireless systems where wireless communications and conditions can pose serious problems to the efficient and reliable delivery of content. Specific chapters of the book relate to different research topics covering the architectural aspects of the most famous DVB standard (DVB-T, DVB-S/S2, DVB-H etc.), the protocol aspects and the transmission techniques making use of MIMO, hierarchical modulation and lossy compression. In addition, research issues related to the application layer and to the content semantic, organization and research on the web have also been addressed in order to give a complete view of the problems. The network technologies used in the book are mainly broadband wireless and satellite networks. The book can be read by intermediate students, researchers, engineers or people with some knowledge or specialization in network topics.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen