This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

MPEG-2 CODED VIDEO TRACES TRANSMITTED OVER A SATELLITE LINK: SCALABLE AND NON-SCALABLE SOLUTIONS IN RAIN FADING CONDITIONS*

Nedo Celandroni[#], Erina Ferro[#], Francesco Potortì [#]
Antonio Chimienti[+], Maurizio Lucenteforte[+], Romualdo Picco[+]

[#]CNUCE, Institute of National Research Council
Via S. Maria 36 - 56126 Pisa - Italy
Phone: +39-50-593111 - Fax: +39-50-904052 - Telex: 500371
E-mail: {N.Celandroni, E.Ferro, F.Potorti}@cnuce.cnr.it

[+]CSTV, Institute of National Research Council
Strada delle Cacce 91 -10135 Torino -Italy
Phone: +39-11-3919242 - Fax: +39-11-341882
E-mail: {chimienti, lucente, picco}@cstv.to.cnr.it

ABSTRACT

The literature is poor in the analyses of the effects produced by corrupted bits in compressed video bitstreams. This paper presents the results of a transmission experiment of MPEG-2 coded video data over a satellite link affected by noise, in order to investigate under which conditions this type of transmission is economically feasible. The signal-to-noise ratio scalability feature of the MPEG-2 encoder was used to produce different bitstreams of the same movie sequence. The scope of the study was to verify which are the best combinations of video and channel codings in the presence of attenuation on the satellite link, in order to optimize the bandwidth utilisation for a requested image quality. The results obtained give indications about the data channel codings to be used to counter the rain fade on the transmission link, which is a non negligible problem especially when satellite transmissions are in the Ka band. Moreover, the results highlight the flexibility of the scalable video coding in the examined scenario.

Keywords: MPEG video, scalability, quality factor, satellite, fading

1. Introduction

The Motion Picture Expert Group (MPEG) video compression schemes have emerged as standards for multimedia applications among the many video coding schemes proposed for the compression of video signals. Compression is necessary to reduce the transmission bandwidth because an uncompressed video source may generate bitstreams at rates in the order of hundred of Mbit/s (about 166 Mbit/s). A variable bit rate (VBR) encoder attempts to keep the quality of the video output constant at the price of changing the bit rate. The resulting traffic is highly bursty, and dependent on the encoding scheme adopted and on the vivacity of the movie's scenes. The

* Work carried out in the framework of the Italian co-ordinated project "*Advanced applications for next generation packet-switching networks*" funded by the National Research Council (C.N.R.)

reduction in the bandwidth needed by VBR video allows data communication networks to agilely support high quality multimedia applications.

The MPEG-2 standard has some embedded error resilience features, and leave the provider free to adopt the appropriate channel protection. For many video applications, such as wired transmission or reproduction from digital storage media (DSM), the channel coding is superfluous, due to the very low bit error rate (BER). When VBR video traffic is transmitted over a satellite link, the noise level, in the link budget design, must be carefully calibrated in order to avoid impairments in the reconstruction of the images received. In fact, the target channel signal-to-noise ratio (CSNR), and thus the resulting BER, is chosen as a compromise between quality and cheapness. Moreover, in satellite transmissions over 10 GHz, the signal attenuation due to rain imposes the adoption of fade countermeasures to ensure an acceptable level of link availability with a reasonable channel quality. The literature for studies on the effects produced by the transmission channel noise on VBR video data is quite poor. This work is a report of an experiment that we carried out, which aims at replying to the following questions.

a) Given an MPEG-2 coded video source, what is the quality degradation of the image in the presence of channel noise, with different levels of BER?
b) Given a required image quality, what is the most suitable combination of video coding (scalable bitstream, non-scalable bitstream) and satellite channel transmission parameters (coding types, coding rates, bit rates) to optimise the channel bandwidth in different link degradation conditions?

In order to provide the answers to the above questions the experiment involved the production of the traces of a film with some scene changes inside, coded according to the MPEG-2 standard with different coding modes. The samples obtained were transmitted between two real satellite stations connected together via a satellite emulator. The received sequences, relative to different transmission parameters necessary to cope with the channel degradation level, were then analysed to evaluate the resulting video quality. The scenario produced showed the actual possibility of MPEG-2 video transmission over a degraded satellite link with limited resource requirements, by adopting a fade countermeasure technique based on both channel coding and bit rate variations.

In Sections 2 and 3 of this paper the environments relevant to the processing of the video signal and the satellite transmission chain involved in the experiment are respectively described. In Section 4 the generation of the MPEG-2 video bitstream is depicted, and the experiment results are presented in section 5. Hints on future work and conclusions are drawn in Section 6.

2. The video environment

The MPEG compression algorithms are intended for the compression of full-motion video. They use inter-frame compression and can achieve compression ratios of 40:1 through exploiting temporal correlation. The MPEG first-phase standard (MPEG-1) [13] is tagged for compression of 320x240 full motion video at rates of 1 to 1.5 Mbit/s in applications such as interactive multimedia and DSM. MPEG-2 standard [14] is intended for higher resolutions, similar to the digital video studio standard ITU-R 601 [15], EDTV, and further leading to HDTV. An MPEG encoder uses two basic techniques: block-based motion compensation for the reduction of temporal redundancy (*inter-frame* coding), and transform domain-based compression for the reduction of residual spatial redundancy (*intra-frame* coding). Three main picture types are defined: intra-pictures (I), predicted pictures (P), and bidirectionally-predicted pictures (B). The first ones are self-contained, since they use only transform coding, and provide access points to the coded sequence where decoding can begin. They are coded without reference to other pictures and with only moderate compression, and are used for predicting P and B pictures in inter-frame coding. They give the lowest compression ratios within MPEG. P pictures are generally used for further prediction and are coded more efficiently, using forward predictive coding, where the actual frame is coded with reference to the previous frame (I or P). The compression ratio of P frames is significantly higher than the I frames. Also, P pictures are used in the inter-frame prediction of other P and B pictures. B pictures provide the highest degree of compression. They are coded using two reference frames, a past and a future frame (I or P frames) for motion compensation. Furthermore they are never used as a temporal reference for other frames.

2.1. MPEG-2 bitstream hierarchy

The MPEG-2 video bitstream structure is a superset of the MPEG-1 structure, obtained using some syntactic extension. One of the major differences between the two standards is the MPEG-2 capability of handling interlaced video sequences such as the ITU-R 601 format.

The MPEG-2 bitstream is a coded representation of I, P and B frames. The highest compression ratio can be achieved by incorporating a large number of B frames. A video data stream is made up of six layers: sequence, group of pictures (GOP), picture, slice, macroblock, and block layer. Each layer consists of the appropriate header and following lower layer. If the transmission is at constant bit rate (CBR) stream, at the beginning of the sequence layer there are two entries: the constant bit rate of a sequence and the storage capacity needed for decoding. These parameters define the data buffering requirements. The transmission can also be done in VBR mode, where the video codec does not provide a buffering system. The sequence header contains information relevant to the image size, the bit and frame rate, and the quantization matrices. A sequence is divided into a series of GOPs. A GOP is a flexible set of pictures, composed of a variable number of I, P and B pictures, according to the distance (indicated by N) between consecutive intra-pictures. Another important parameter is the distance between consecutive P pictures, usually indicated by M. At least one I frame as the first coded frame in the GOP is mandatory. The picture layer contains the whole picture (or frame), that consists of the luminance and two chrominance components. The picture header contains temporal references to the coded image, the image type and information relevant to the source of the image. Figure 1 shows these first three levels. The bits corresponding to the discrete cosine transform (DCT) and the motion vectors are contained in the next three layers: slice, macroblock, and block layers. The slice layer contains the information associated to a portion of 16 rows of pels of the picture and consists of a variable number of macroblocks. Macroblock is the basic unit of coding within a picture (16x16 pixels). For a given macroblock a coding mode is chosen as a function of the picture type. Depending on this coding

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

mode, a compensated motion prediction of the contents of the block, based on past and future reference pictures, is formed. This prediction is subtracted from the actual data in the current macroblock to form a difference signal, which is separated into 4 blocks of luminance and 2 blocks of chrominance, and a DCT is performed on each 8x8 block. The resulting 8x8 block of DCT coefficients is quantized and the two dimensional block is scanned in a zig-zag order to convert it into a one-dimensional string of quantized DCT coefficients. Run-length coding is used for the quantized coefficient data. A consequence of using different picture types and variable length coding is that the overall data rate is variable.

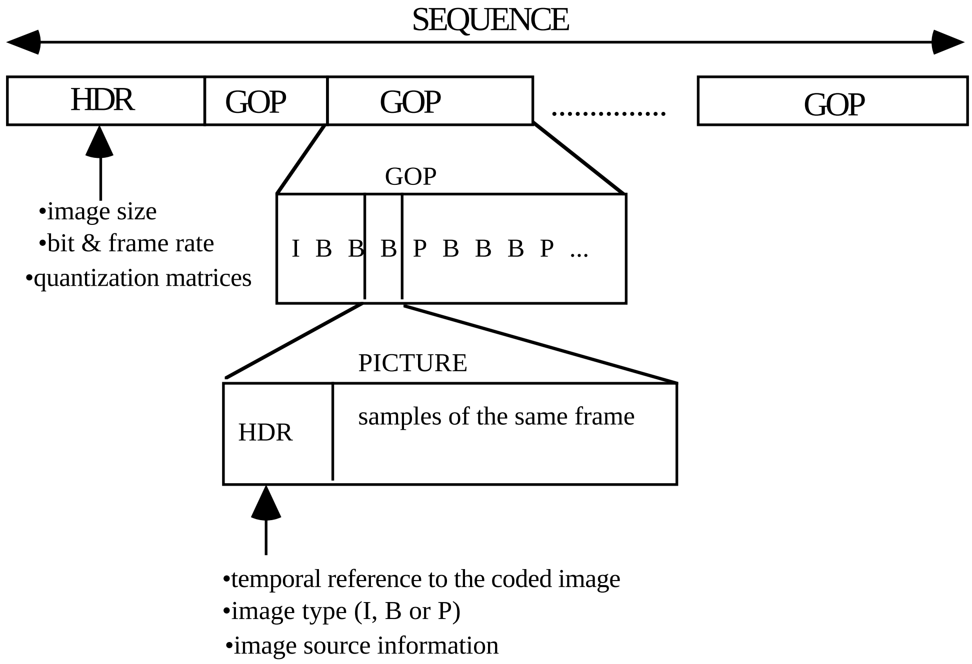

Fig. 1. Sequence, GOP, and picture layers

2.2. MPEG scalability

Scalable video is used for a number of applications where it is necessary to display video at different resolution or quality levels. A first approach to this problem is the *simulcast* technique, where a set of various contents of the same video sequence is transmitted [10]. The disadvantage is a high total bit rate, since the information flows associated to different resolutions or qualities are independent. A lower bit rate, associated to the same picture quality, is possible by using a scalable video coding, where a part of the transmitted information, related to the low quality or resolution, can be reused for coding other scales. In a scalable encoder, two or more layers are generated, coded and transmitted. In the simplest case an enhancement layer encoder utilises information generated by an independent base layer encoder.

Scalable video can be applied in the spatial, temporal and frequency domains. *Temporal* and *spatial scalability* deal with different temporal and spatial resolutions, respectively. These two scalability modes are included in the MPEG-2 standard. The *SNR* (signal-to-noise ratio) *scalability*, a different picture quality scalability mode, has been included in the MPEG-2 standard. This latter type of scalability has been used in our experiment.

2.2.1. SNR scalability

The SNR scalability scheme is illustrated in Fig. 2.

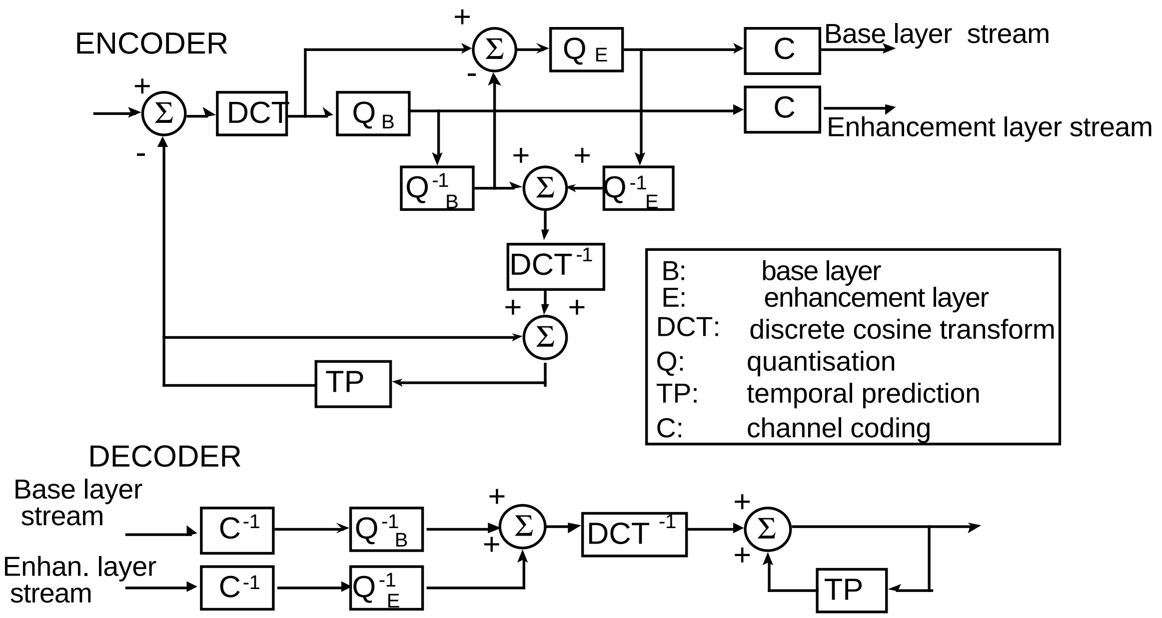

Fig. 2. SNR scalability encoder-decoder schemes

It provides a base layer and an enhancement layer, which contains only coded refinement data for the DCT coefficients of the base layer. In the receiver, DCT coefficients from the two layers are decoded and summed after the inverse quantization process. After DCT summation, the decoding process is the same as in the non scalable decoder. Typically, in an SNR scalable encoder, a re-quantization of the base layer coding error is done: the difference between the original DCT block and the low layer quantized block is re-quantized using a finer quantizer step and then coded and transmitted in the enhancement layer stream. The high level side information is the same as the base layer, and no inter-frame information is needed. Therefore, the enhancement layer mainly contains coded DCT coefficients of the re-quantized base layer error and a small overhead. As a result, the increase in the total bit rate is not large, so the enhancement layer is associated to a higher quality video representation that is impossible to reconstruct without decoding the base layer in parallel. A drift error between the motion compensated image in the encoder and in the lower layer decoder can be observed. This happens because in the encoder the motion compensation for the base layer is performed from the decoded picture of the enhancement layer. This error drift is acceptable for N=12 and M=3.

2.3. MPEG-2 video parameters

The MPEG-2 coding used in our experiment was run on a 1 minute long video sample (1500 frames) extracted from the movie "The sheltering sky"[1]. Both non-scalable and SNR-scalable traces were generated. In order to obtain a constant and comparable error-free video quality, the non-scalable trace and the scalable high layer traces were coded in VBR mode, using a reference quantization parameter set to 3. The other video parameters we used are summarised below:

- Input picture format: ITU-R 601 (720x576 pixels) converted in 4:2:0 chroma format.
- GOP size: N = 12; distance of two consecutive P pictures: M=3.
- Structure of pictures: coded as frame picture.
- Vector search range: ±15 pels/frame horizontal, and ±7 pels/frame vertical with scaled window according to the frame distance (i.e. for P-frames the vector search window is [±45 ¥ ±21]).
- Motion vector estimation: full pel exhaustive motion estimation with the previous defined search window based on the reference original picture; half pel refinement on all the 9 adjacent positions, based on the reference coded picture.
- Quantization of DCT coefficients: linear relation between the reference quantization parameter and the quantizer step.

2.4. Objective video quality assessment

From the user point of view, the main metric of evaluation of video service quality is the perceptual quality, called the *mean opinion score* (MOS). The MOS is calculated from the ratings given by a sample of human observers, under controlled conditions, who judge the image quality. The reference measures used for an objective video quality assessment are the *mean squared error* (MSE) calculated on the difference signal between the original and the coded sequence, and the *peak signal to noise ratio* (PSNR), defined as:

$$PSNR = 10 x \log_{10} \frac{255^2}{MSE}$$

These measures are not necessarily a good quality index of the subjective assessment due to a human perception. This is expecially true when considering a error-prone environment, where the channel error effects can be limited to a relatively small portion of the image, causing some very annoying artefacts not highlighted from the mean MSE.

For this reason in our study we used another objective video quality assessment system [11] that emulates the HVS (Human Visual System). In this method, a linear combination is calculated of three complementary video quality measurements, based on spatial and temporal distortion. The resulting *quality factor* (QF) is strictly correlated with the subjective mean opinion score: imperceptible (5), perceptible but not annoying (4), slightly annoying (3), annoying (2), very annoying (1).

3. The satellite network environment

3.1 Overview on digital satellite communications

[1] The authors would like to thank Videotime for providing the movie.

Telecommunication satellites are mainly classified according to the height of their orbit. GEO (geostationary earth orbit) satellites assume a fixed position at an altitude of about 36,000 Km, while MEO (medium earth orbit: 10,000-20,000 Km) and LEO (low earth orbit: 500-2,000 Km) move with respect to the earth. The lower the altitude, the lower the power required to transmit, and the lower the possible coverage of each spacecraft unit; therefore more units are needed for the coverage of wide areas. The transmission round trip delay is in the order of a quarter of second for GEO, less than 100 ms for MEO, and 8-20 ms for LEO satellites.

Each of the receiving/transmitting chains (payload) on board the spacecraft is called *transponder*. The device which allows the access to the satellite channel is called *earth station* or *earth terminal.* Any earth station is composed of an antenna (its most visible part), a receiver and a transmitter. *Multiple access* is the ability of a number of earth stations to simultaneously share the same satellite transponder for interchanging data. Classically there are three multiple access techniques [2, 3]: *frequency division multiple access* (FDMA), *time division multiple access* (TDMA), and *code division multiple access* (CDMA). The basic problem involved is how to permit a group of earth stations to share a satellite link avoiding interference or collisions. The multiple access technique used in this experiment is TDMA, because of the available satellite equipment [4]. Here only one user at a time transmits and modulates the channel carrier with the maximum base-band signal speed. The link budget must therefore consider this speed, even if the average bit rate of each terminal is generally much lower. On the other hand, TDMA offers many advantages, such as the need for a single modem per user, and flexibility.

The figure of merit of a digital link is the BER, also denoted as P_e, which is the probability that a bit sent over a link is incorrectly received. The relationship between P_e and the CSNR depends on the modulation and coding schemes. Such dependencies can be found in the literature for a variety of modulation and coding combinations when we are in the presence of AWGN (additive white Gaussian noise), i.e. thermal noise [2, 3]. Another important cause of impairment in the transmitting signal is interference. This is mainly due to non-linearities in the amplification devices which produce intermodulation noise, and to unavoidable imperfections in radiofrequency devices such as antennas' side lobes and polarisation discrimination. However most satellite links are designed in such a way as to limit interference, and then thermal noise is predominant.

The transmitted signal is attenuated by the spreading factor (scintillation), atmospheric losses, and other losses. The main factors causing atmospheric absorption are: uncondensed water vapour, rain, fog and clouds, snow and hail, free electrons in the atmosphere, and molecular oxygen. At most frequencies of commercial interest (up to about 6 GHz) the atmospheric absorption is relatively unimportant (a few tenths of dB). The *attenuation* is defined as the dB difference between the currently-received power and the power received under clear sky conditions. The attenuation increases to large values during unfavourable propagation conditions (*fades*). Rain fades are a major problem in transmissions above 10 GHz in that they attenuate the signal and increase the noise level so worsening the CSNR in the receiving earth station. CSNR may be expressed in terms of C/N, i.e. carrier power-to-noise power ratio, or in terms of E_b/N_0 (bit energy to one-sided noise spectral density ratio). In the rest of the paper we will consider QPSK (quaternary phase shift keying) modulated signals, and denote by *link degradation* the difference between the reference (12 dB) and the current value of E_b / N_0. The reference value, relative to clear sky conditions, allows a BER of 10^{-8}.

3.2. The satellite network emulated in the experiment

Figure 3 represents the test environment. The equipment used was previously employed in experiments on the Olympus and Italsat geostationary satellites [9]. This time, however, instead of using a real geostationary satellite, we used a satellite emulator which introduces the correct round trip time and amount of noise to emulate fading situations.

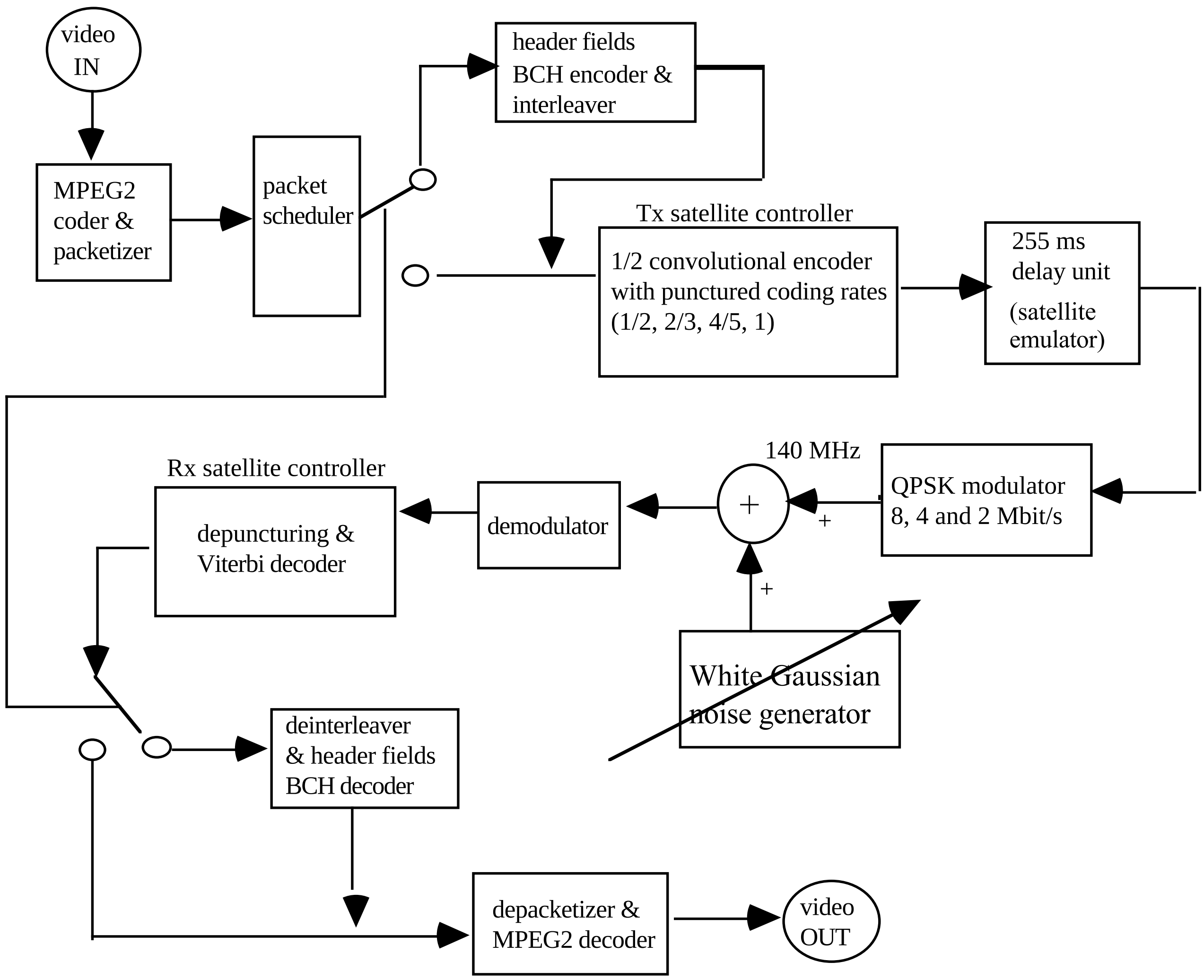

Fig. 3. The satellite network environment

Taking the traces from the MPEG-2 encoder, the packet scheduler generates packets at given time instants so as to emulate the real time data produced by a hardware encoder. The transmission (Tx) satellite controller adjusts the packet sending times according to the TDMA channel access scheme, and applies the channel coding to the base band data stream. The 1/2 convolutional encoder with the puncturing feature allows 1/2, 2/3 and 4/5 coding rates on the satellite channel.

Optionally, an outer BCH[2] encoder with data interleaving can be applied on the header fields. Also no coding at all is possible. A base band delay unit introduces the satellite round trip time of 250 ms, then the QPSK modulation of the 140 MHz intermediate frequency carrier allows the bit rates of 8, 4, and 2 Mbit/s. Each packet can be sent at individual bit and coding rates. Additive white Gaussian noise is introduced at intermediate frequency level, so as to give the required value of E_b / N_0 at the demodulator input. The receiving (Rx) controller decodes and packetizes data for sending to the MPEG decoder which, after data serialisation, reproduces the resulting traces of the video sequence. At this point we analysed the quality of the corrupted video stream.

We assume the following scenario.

•The video encoder can operate in only one coding mode, i.e. it can produce only scalable or non-scalable bitstreams. A typical example is a data base of a video-on-demand service in which all the bitstreams are produced with the same encoding parameters.

•The sending earth station knows:

- the channel degradation of the receiving station [5, 6];
- the target quality factor of the video service;
- the amount of channel bandwidth available for the video transmission.

All this information is used by the earth station:

- to properly select the suitable channel coding and bit rates, in order to compensate for the different fade conditions;
- in case of scalable transmission, to decide on the convenience of transmitting the enhancement layer other than the base layer.

The bandwidth allocation algorithm for video data is beyond the aim of this paper; one of the policies proposed in References [7, 8] is assumed to have been adopted. The residual bandwidth, not actually used by video, can be exploited to send low priority traffic.

After convolutional-encoding/Viterbi-decoding, the residual errors are distributed in bursts of various length, rather than uniformly. In order to make a comparison between the quality of the video sequences obtained with burst and random error distributions, data interleaving has been used as well. Moreover, an outer encoder of BCH type has been applied to the header fields to try to eliminate all the errors, in order to consider the effect of headers corruption on the video quality. This further protection implied the use of the data interleaver.

4. The MPEG-2 encoder simulation

In our experiments three different MPEG-2 coding modes were considered, with and without using the SNR scalability. Without scalability, we produced a VBR single-stream coded with an average rate of 2.37 Mbit/s (here on referred to as *non-scalable*). In each of the other two simulations the SNR scalability was used to produce a base layer coded at a CBR rate refined by a VBR enhancement layer. The two combinations have the following average rates: 1.5 (base)+1.17 (enhancement) Mbit/s (*scalable1)*, and 1.066 (base)+1.482 (enhancement)) Mbit/s *(scalable2)*, respectively. The VBR traces were generated with the quantization step set to 6 in

[2] Bose-Chauduri-Hocquehghem

order to maintain the total channel bit rate within the limits of the hardware at our disposal. In Table 1 the mean PSNR, QF, and mean bit rate for the 3 different experiments are shown.

	MEAN PSNR (QF)	MEAN BIT RATE
non-scalable	39.83 (4.66)	2.37 Mbit/s
scalable1 (base only)	35.57 (4.36)	1.5 Mbit/s
scalable1 (base+enh)	39.81 (4.68)	1.5+1.17 Mbit/s
scalable2 (base only)	32.87 (4.09)	1.066 Mbit/s
scalable2 (base+enh)	39.76 (4.68)	1.066+1.482 Mbit/s

Table 1. PSNR, QF and mean bit rate used in the simulations

In Figures 4, 5, and 6 (where B stands for base flow, and E for enhancement) the QF and the PSNR values for the overall video sequence are plotted versus GOP number. The final video quality of all the coding traces is the same. The mean bit rate is higher in the scalable coding modes, due to the redundancy of the overhead in the enhancement layer.

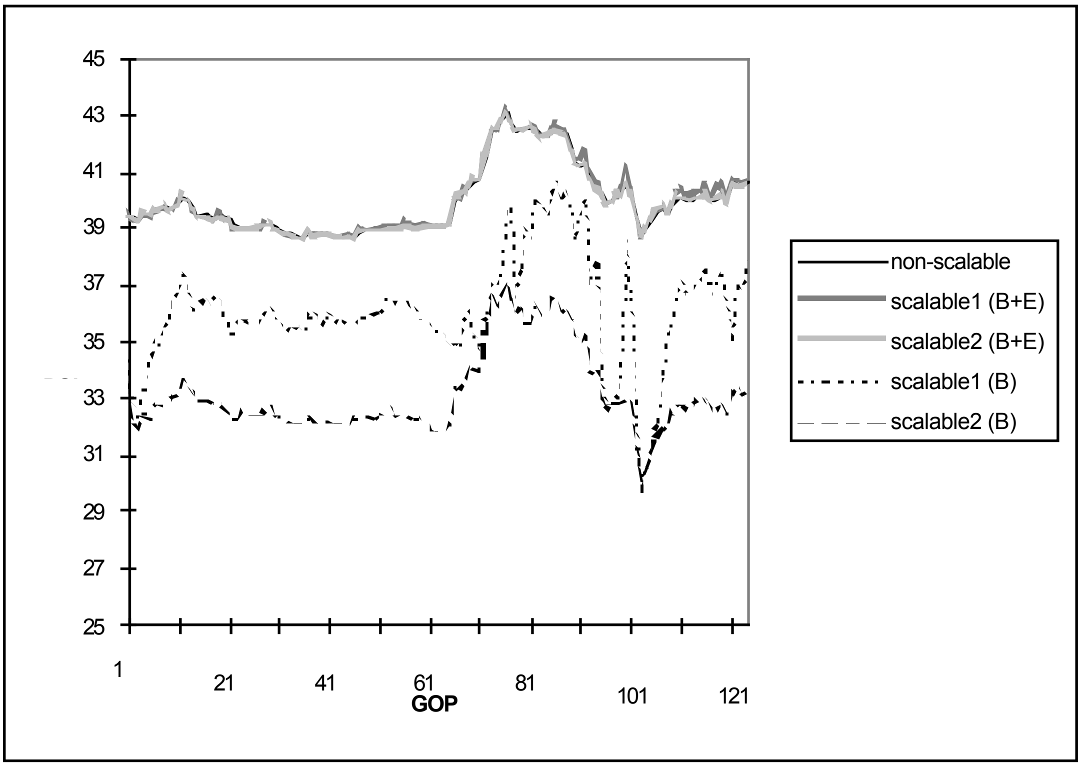

Fig. 4. PSNR vs GOP

The advantage of the scalable coding is generally related to the possibility of receiving and decoding moving video with a different resolution or quality, having a better channel utilisation of an equivalent simulcast transmission, since the low quality information is used for the high quality video reconstruction. As to error-prone transmission environments, a scalable encoder can give a better performance than an equivalent non scalable one, because it allows the organization of information hierarchically into different layers and a better protection of the most important ones. Channel errors have different effects on the decoded video, if associated to the base or to the enhancement layer.

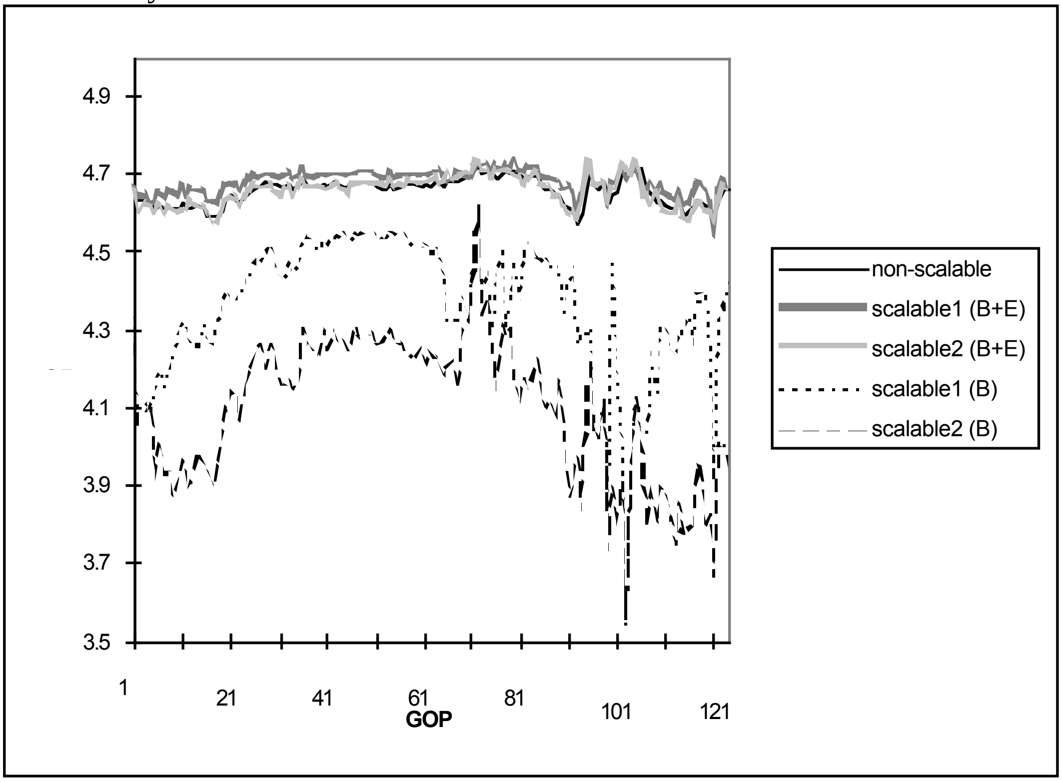

Fig. 5. Quality factor vs GOP

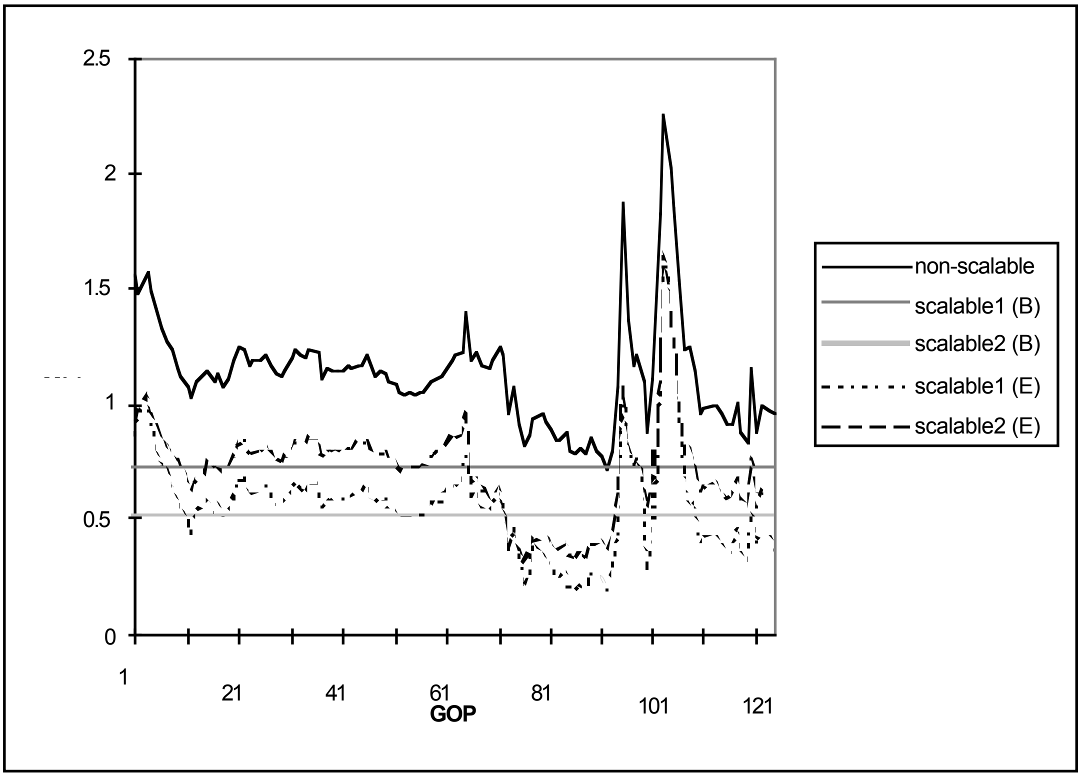

Fig. 6. Bit rate (in Mbit/GOP) vs GOP

In fact, errors in the enhancement layer never have destructive effects because they simply prevent the refinement of the base layer. Another interesting characteristic of an SNR scalability is the possibility of the decoder discarding the high layer flow, when the bit error rate of this layer is too high, without provoking a dramatic reduction in the video quality.

In our experiments an *unsupervised* MPEG-2 decoder [12] is considered at the satellite receiver, that is to say, the MPEG-2 decoder does not acquire error information from any external device and errors are detected solely by the decoder. The MPEG family standards are flexible in the sense that a lot of parameter values are embedded in the bitstream. This generality has the drawback of making it impossible to detect errors when they produce valid syntax and parameter values.

4.1. Error detection

The syntactic rules of MPEG-2 make it possible to detect some channel errors by just parsing the bitstream. The detectable error types can be summarised into the following four classes:

- Semantic error: some header fields can assume values not consistent with the semantic or the specific profile/level mode.
- Variable length coding (VLC) error: the variable length or entropy coding is mainly used for the quantized DCT coefficients. The MPEG-2 VLC tables satisfy the prefix rule, that is to say each codeword is not a prefix of any other codeword. The VLC decoding usually uses a tree structure, where each path from the root node to a terminal node identifies a VLC codeword. The VLC decoding starts from the root node and each input bit individuates a binary decision about which a child node must be selected. A VLC decoding error occurs when the next input bit is not associated to a child node and the current node is not terminal.
- DCT coefficients number: the DCT coefficients are entropy coded using a combination of VLC and run-length code, i.e. each VLC codeword individuates a couple run (number of consecutive zero DCT coefficients) length (value of the non zero DCT coefficient). The decoder detects an error when the coefficients count has reached 64 and an end-of-block has not been encountered.
- Motion vectors: in addition to the above VLC decoding error (motion vectors are VLC coded) a typical detectable semantic error occurs when the macro-block prediction area falls out of the picture area.

4.2. Error handling

When a channel error has been detected during the MPEG-2 decoding, it can be handled by skipping the received data until the next sync-symbol is found. The layered MPEG-2 stream structure makes this easy, since each layer has a different 32 bits start-code which is never present in other parts of the bitstream, except if an error occurs. All the parts of the video sequence associated to the lost information are marked as *faulty*, and must be concealed. Usually a channel error corrupts a slice, that is the area starting at the macroblock where the error occurs up to the slice end. This error is handled by forcing a re-synchronization to the next slice start-code. This is the less destructive and more frequent error type. The situation is more serious if an error occurs in a header field of a high level syntactic layer like sequence, GOP or picture. In these cases the

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

error consequences can be temporarily extended to several consecutive pictures, and the effect is a picture freezing of the received sequence.

The goal of error concealing is to mitigate the effects of transmission errors in order to obtain an acceptable subjective quality of the areas associated with a channel error de-synchronization. This is possible by exploiting available redundancy in the decoded picture. Different strategies have been adopted [16], depending on the corrupted picture type. One is based on spatial interpolation for the intra-coded pictures while the other uses temporal replacement with motion compensated concealment, if motion information is available.

When errors are found in the I frames, a macroblock which has been lost is reconstructed by filling the 8x8 sub-blocks with values synthesised by interpolation from the nearest blocks in the top and bottom macroblocks. Since a channel error generally causes a loss of data in a series of macroblocks, the left and right neighbours are not used for synthesis.

When errors are found in the P and B frames, the first step is to mark the macroblock type of the concealed macroblock, depending on adjacent macroblock types. For P-frames, if either the top or bottom macroblock is coded as forward prediction mode, the damaged macroblock is assigned the forward mode. If both the neighbours are intra-coded then the damaged macroblock is assigned the intra-code, and the strategies applied for intra-pictures is used. Similar strategies apply for B frames.

		Top MB	
	MB type	forw	back
Bottom MB	forw	forw	forw
	back	forw	intra

Table 2. Macroblock type of concealed MB for P picture

		Top MB			
	MB type	forw	back	inter	intra
Bottom MB	forw	forw	inter	inter	forw
	back	inter	back	inter	back
	inter	inter	inter	inter	inter
	intra	forw	back	inter	intra

Table 3. Macroblock type of concealed MB for B picture

These strategies are summarised in Tables 2 and 3, respectively, where “MB” stands for macroblock, “forw” for forward prediction, “back” for backward prediction, “inter” for bi-directional interpolation, and “intra” for intra-coded mode.

The motion vector synthesis follows a similar philosophy: if the top and bottom vectors are defined, then the average of motion vectors is used for the synthesised macroblock. If only one of the neighbours has valid motion vector(s) defined, then this vector(s) is used; if no motion vectors are available then the macroblock is synthesised as specified in the intra-frame technique.

5. The results

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

In Fig. 7 the average QF, for non-scalable and scalable2 codings with several satellite channel coding rates, is reported as a function of the VBR streams' BER.

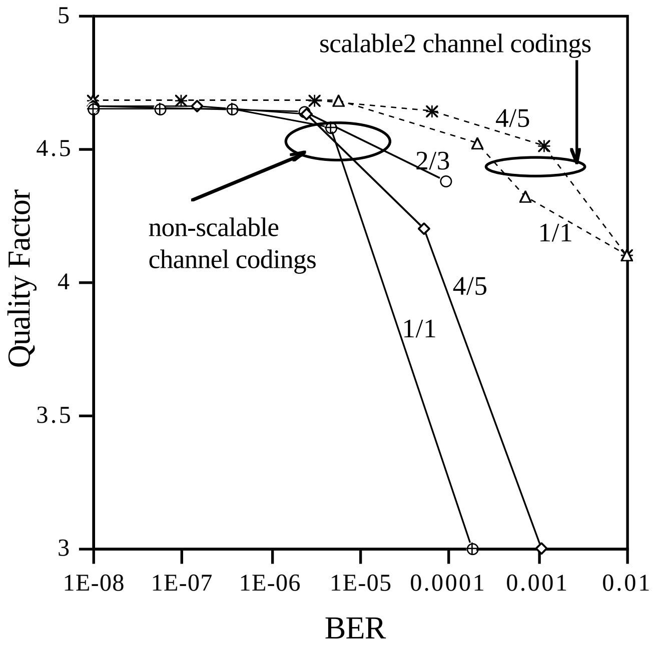

Fig. 7. Average QF versus BER for non-scalable and scalable2 with different channel coding rates of the VBR sequences.

In the graph, we highlight how the error distributions of different FEC[(3)] rates influence the QF of MPEG-2 bitstreams. The scalable coding base stream is error-free (thanks to sufficient protection), and the enhancement stream has different FEC rates according to the labels in the graph. As mentioned above the residual errors, after convolutional encoding and Viterbi decoding, are distributed in bursts whose length increases when the coding rate decreases. Thus, for a given BER, the number of the error bursts decreases with the coding rate. The picture shows that, for a given BER, any MPEG-2 coded stream has a greater tolerance for clustered errors, so it is preferable to have fewer, even if longer, erroneous sequences. The same result has been achieved with the experiment of the header fields protection with a BCH encoder. In this case, the better protection of headers is compensated by the spreading of erroneous bits produced by the data interleaver, which is needed by the BCH decoder to recover all the errors. In fact, the headers' protection did not produce any appreciable difference in the resulting average QF.

In the following we have considered the need to find out which is the best video coding type to be used over a noisy satellite link (among non-scalable, scalable1 and scalable2). Indeed, the base-only versions of the two scalable codings should be considered as particular cases of the base+enhancement codings. For example, let us assume that one of the two scalable codings is chosen as the best one for this type of transmission. Then the MPEG-2 encoder at the source will

(3) Forward Error Correction. It is also used to indicate the channel coding.

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

generate two flows, one for the base and one for the enhancement data. The transmitting earth station is then responsible for choosing the best channel coding and data rate for the two flows, and possibly for discarding the enhancement flow. These operations are based on real-time measurements of the attenuation on the satellite link, made by the receiving station. The decisions of the transmitting station influence both the QF of the received MPEG-2 data and the satellite average bandwidth occupancy. In general, a higher channel occupancy leads to better quality, so the choice is a matter of trading off small bandwidth with high video quality.

Bit rate	Coding rate	Redundancy
8 Mb/s	1:1	1
8 Mb/s	5:4	1.25
8 Mb/s	3:2	1.5
8 Mb/s	2:1	2
4 Mb/s	5:4	2.5
4 Mb/s	2:1	4

Table 4. Bit and coding rate combinations used for the experiment.

In order to make such a trade off, a rationale must be chosen. We used the criterion of maintaining a given QF while using the smallest possible bandwidth on the satellite channel. Thus, the transmitting station continuously monitors the link attenuation and selects, for both the base and the enhancement flows, bit and coding rates such that the QF at the receiver is not smaller than the target value, while trying to keep the average channel occupancy as small as possible. Other criteria are possible, and are the subject of current research.

We started by analysing the QF of the received flow as a function of the satellite channel bit and coding rates, and of the channel attenuation.

The channel bit and coding rate combinations used for the experiment (summarised in Table 4) are selected from those allowed by the burst modem available to us, which dictated most of the parameters chosen for the experiment. In particular the raw throughput of the video stream after channel coding does not exceed 7.2 Mb/s, which is the maximum we could afford given our modem and the overhead needed by the available satellite access scheme. With this constraint, we chose the parameters of the MPEG-2 encoder, and those values that afforded us the maximum number of different channel coding choices. Each combination of channel bit and coding rates for both the base and the enhancement flows corresponds to an average bandwidth occupancy, which thus assumes discrete values. Nevertheless, representing the QF on a bandwidth-attenuation plane gives a thorough insight into the behaviour of the system as the attenuation changes. Figures 8, 9, and 10 show regions of the bandwidth-attenuation plane where the QF is comprised of some given threshold values. In these figures, M indicates the non-scalable (mono), B the base, and E the enhancement flows.

Figure 8 shows the simplest case. The usable average bandwidth values are those corresponding to the horizontal lines, which are labelled with the relative bit and coding rates. For example, the line at 2.37 Mb/s is labelled M(8, 1/1), which means 8 Mbit/s channel bit rate and 1:1 coding rate, that is, no FEC is applied. This combination is the one with the smallest bandwidth

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

occupancy. The other horizontal lines represent other channel bit and coding rates, with increasing redundancy and, consequently, increasing data protection from corruption.

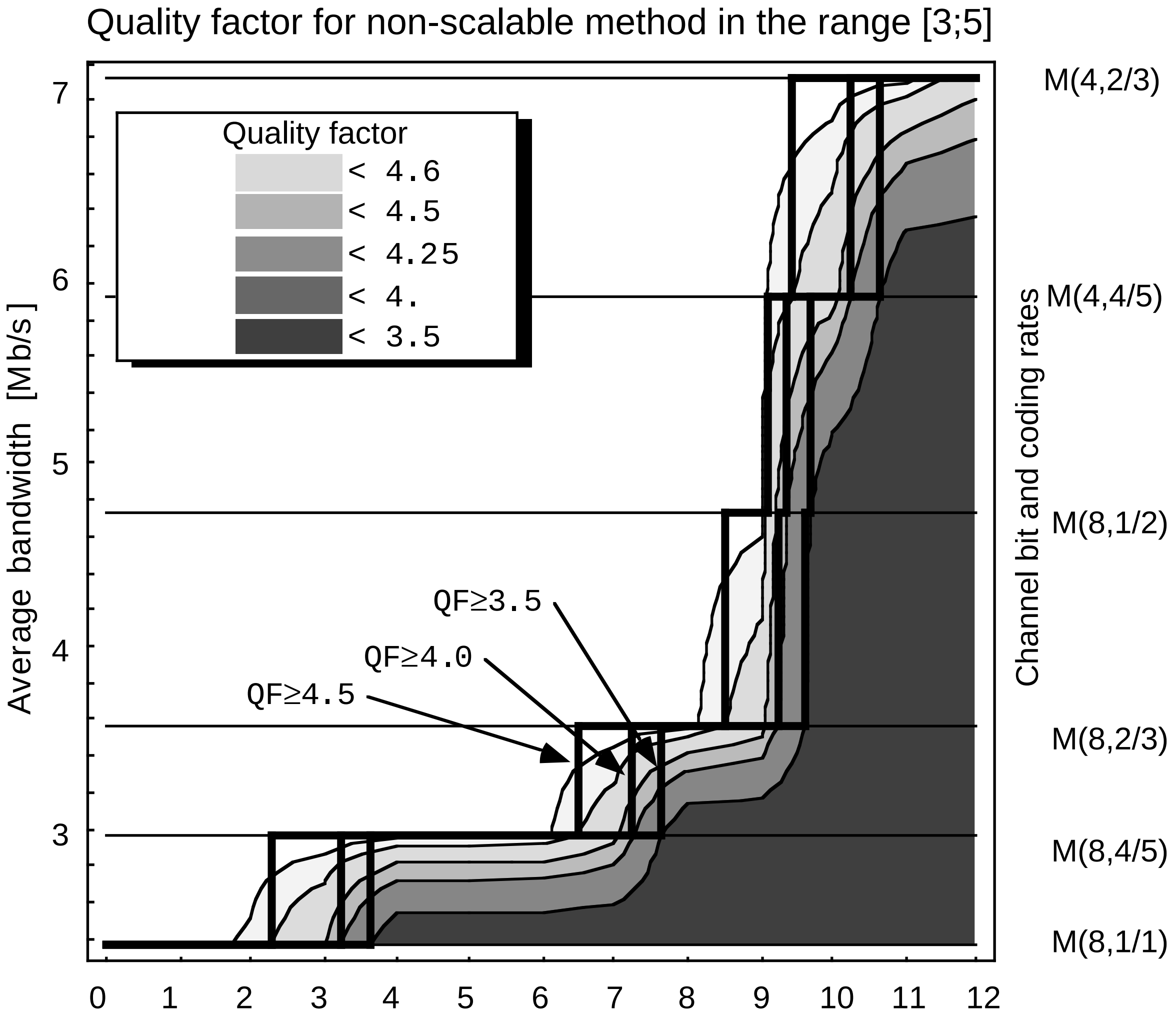

Figure 8. Isoquality regions in the attenuation/bandwidth plane, and paths of constant QF for non-scalable MPEG-2 coding.

The white region is the set of points (attenuation, bandwidth) whose corresponding QF is greater than 4.6, out of an attainable QF of 4.66 for the used MPEG-2 coding. This region is practically error-free, and the quality of the received data is the same as that of an uncorrupted video stream. The darkest region is where the quality of the received data is between 3.5 and 3, which we considered as the floor QF value. As one would expect, a higher bandwidth (higher data redundancy) gives a better video quality for a given value of the satellite channel attenuation, while a higher attenuation gives a worse video quality for a given average bandwidth.

The three thick lines are the paths that the transmitting station follows in order to maintain a quasi-constant QF, jumping from one channel coding and bit rate combination to the next, as the channel attenuation changes. The line for QF not less than 4.5 is the nearest one to the top-left corner of the graph, with respect to the lines for 4.0 and 3.5. This visual indication shows the increasing cost, in terms of average bandwidth, needed for maintaining a QF not smaller than 4.5, with respect to maintaining a QF not smaller than 4.0 or 3.5.

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

Figures 9 and 10 are more complex to read, because of the non monotonic behaviour of scalable coding coupled with FEC coding, and because of the possible elimination of the enhancement data flow.

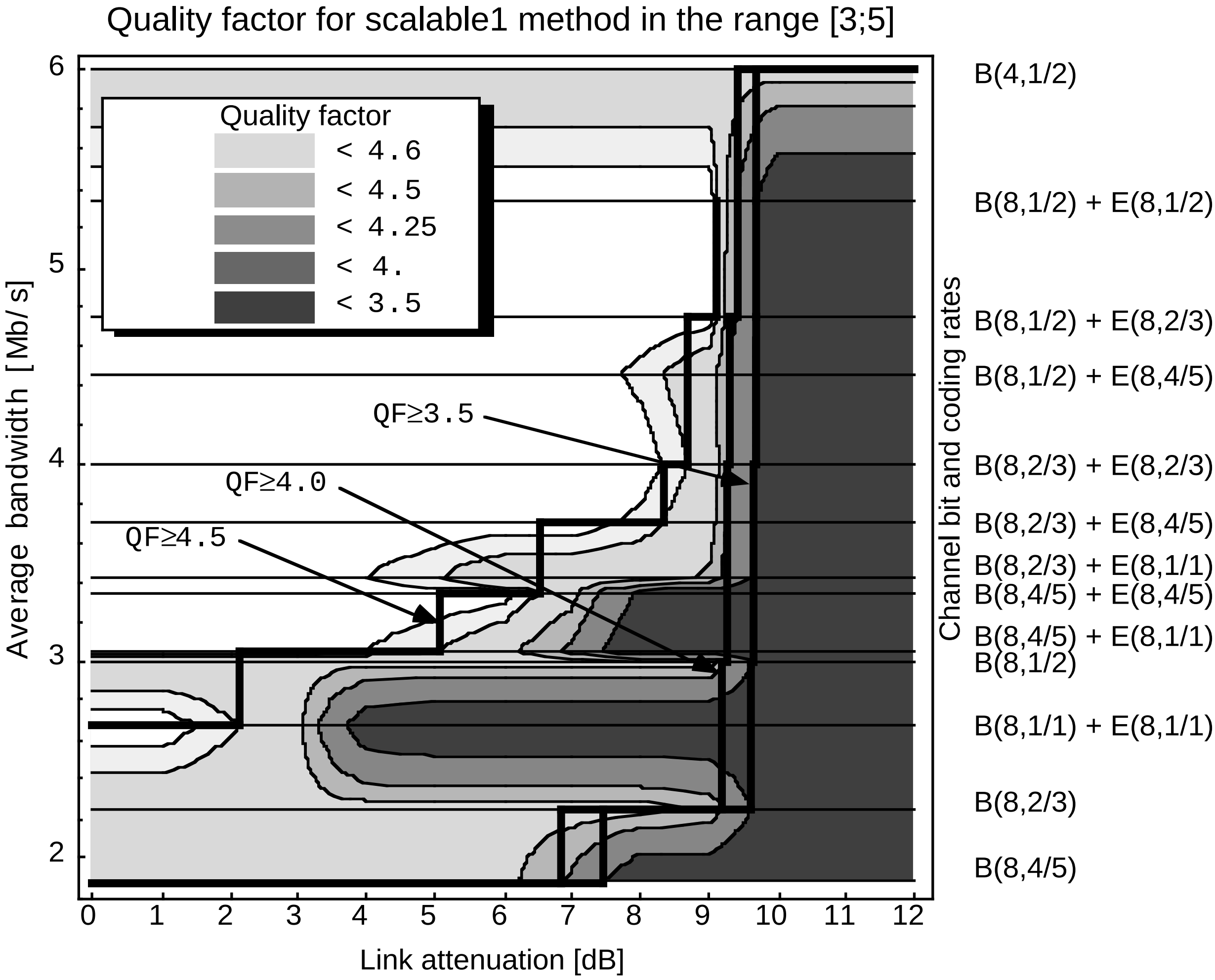

Figure 9. Isoquality regions in the attenuation/bandwidth plane, and paths of constant QF for scalable1 MPEG-2 coding (see Table 1).

As in Figure 8, the white region is the set of points where QF is not smaller than 4.6, and the darkest region is the floor of the video quality we chose to consider (QF=3). The number of possible combinations of channel bit and coding rates for the scalable video codings is higher than in the case of non-scalable one (13 versus 5), which translates into a greater flexibility. Having at its disposal a higher number of possible choices for bandwidth and channel codings, the transmitting station can make a better job of choosing the channel bit and coding rates, while keeping the average bandwidth occupation as small as possible and maintaining the requested video quality. This is particularly true when low qualities are requested, that is when the target QF is 4.0 or 3.5. Indeed, in this case, the base flow alone is able to deliver a quality better than 4 with both the scalable codings considered. Therefore the average bandwidth occupancy is generally smaller with scalable codings than with the non-scalable one, where all the video information is sent to the receiver no matter what the minimum tolerated video quality is.

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

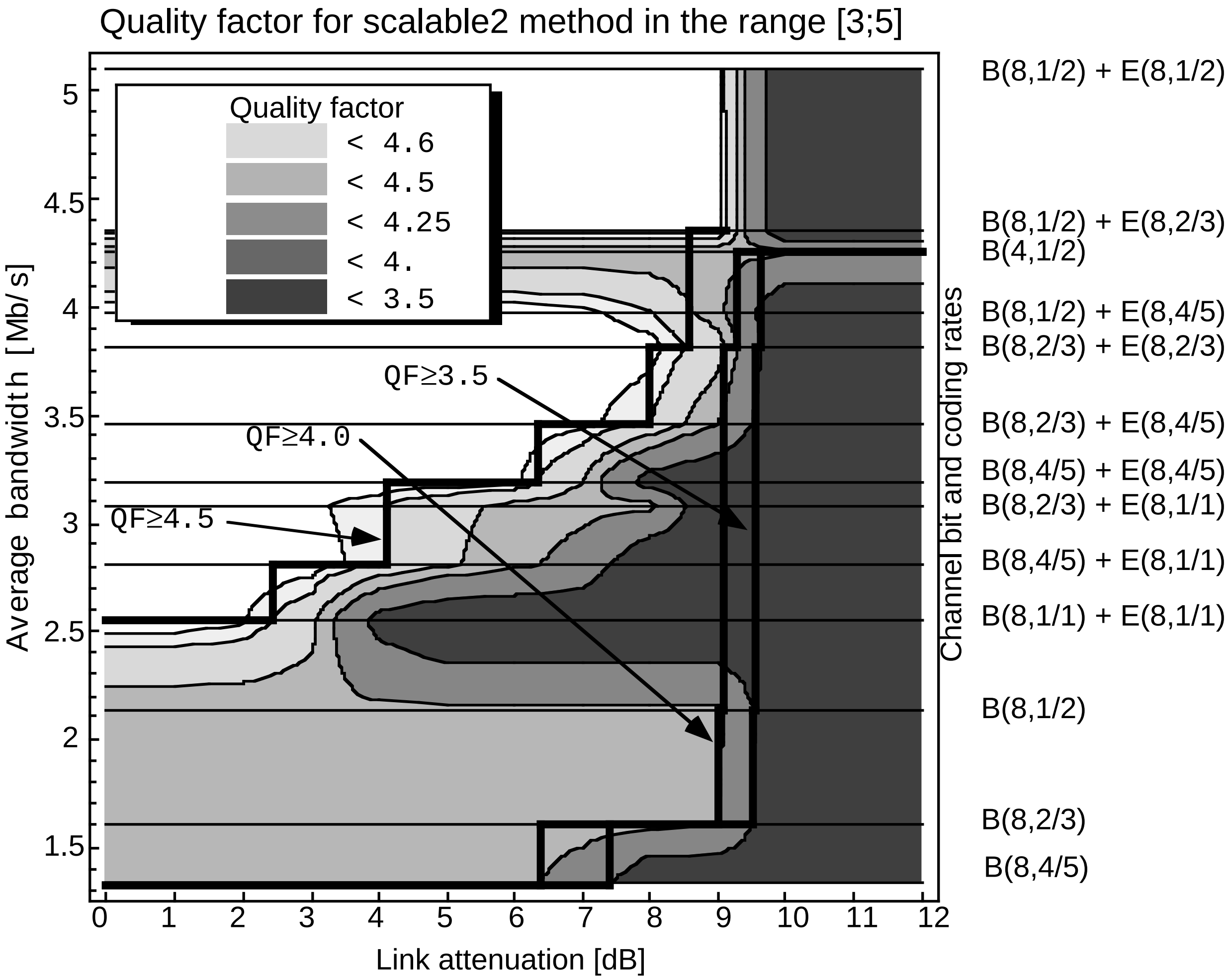

Figure 10. Isoquality regions in the attenuation/bandwidth plane, and paths of constant QF for scalable2 MPEG-2 coding (see Table 1).

Figures 11, 12, and 13 use the paths computed in Figures 8, 9, and 10 to compare the performance of the three coding methods used (see Table 1). For each video coding, given a single target QF, the bandwidth occupied by the transmitting earth station is depicted for a given attenuation. An MPEG-2 coding method has a better performance than another if the path of the former is lower than the other, that is, the former uses less bandwidth for a given attenuation value. Even if it depends on the attenuation value, the graphs nonetheless give a clear idea of what the best video coding method is in most conditions. In Figures 11 and 12, in fact, the scalable2 method is clearly the winner. Apart from a small range of attenuation values, the scalable2 method uses significantly less average bandwidth than the other methods to obtain the same QF. The main reason why the scalable coding performs better than the non-scalable one — for target QF of 3.5 and 4.0 — is that the former has the possibility of dropping the enhancement flow. Looking at Figures 9 and 10, one can notice that the paths for target QF of 3.5 and 4.0 use base-only video codings for almost the entire attenuation range. This advantage over the non-scalable coding is more significant when the ratio between the base rate and the total flow rate is lower.

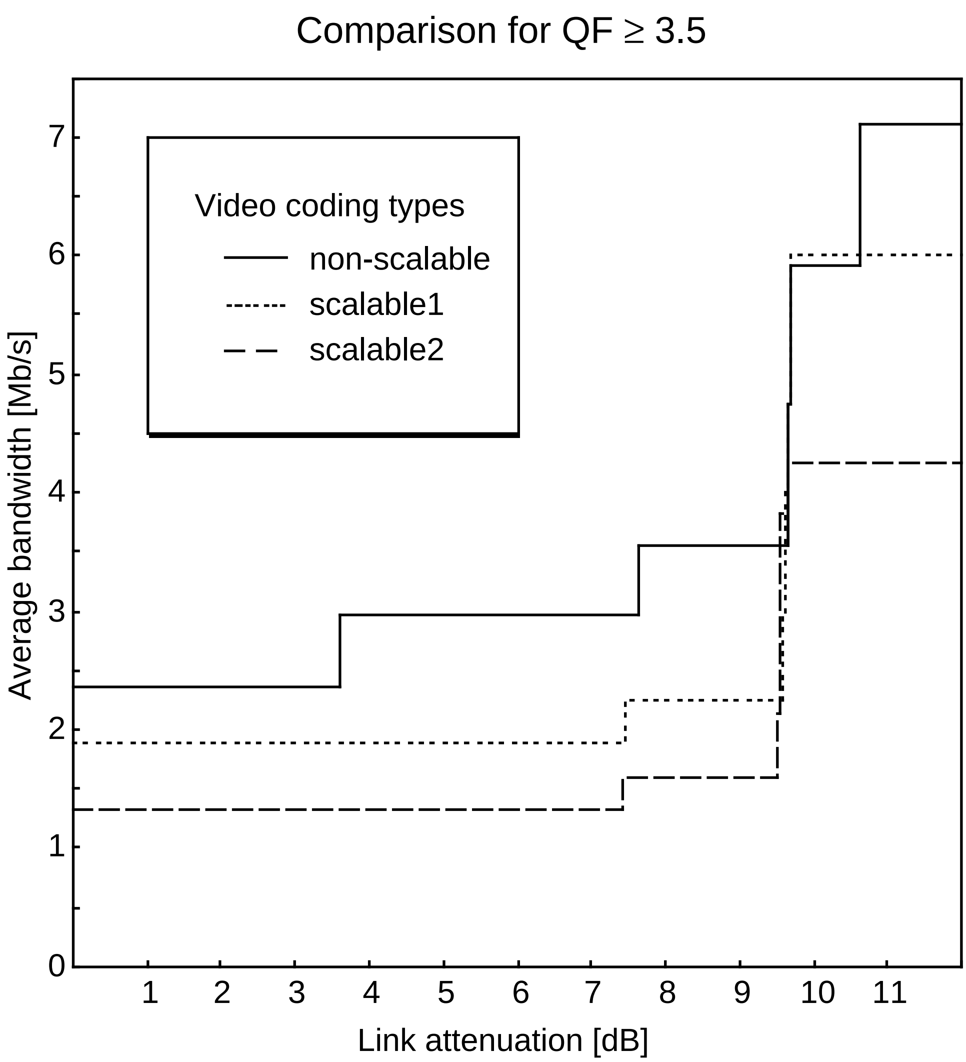

Figure 11. Bandwidth occupation by the transmitting earth station when maintaining a QF not less than 3.5, for the three video codings considered.

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

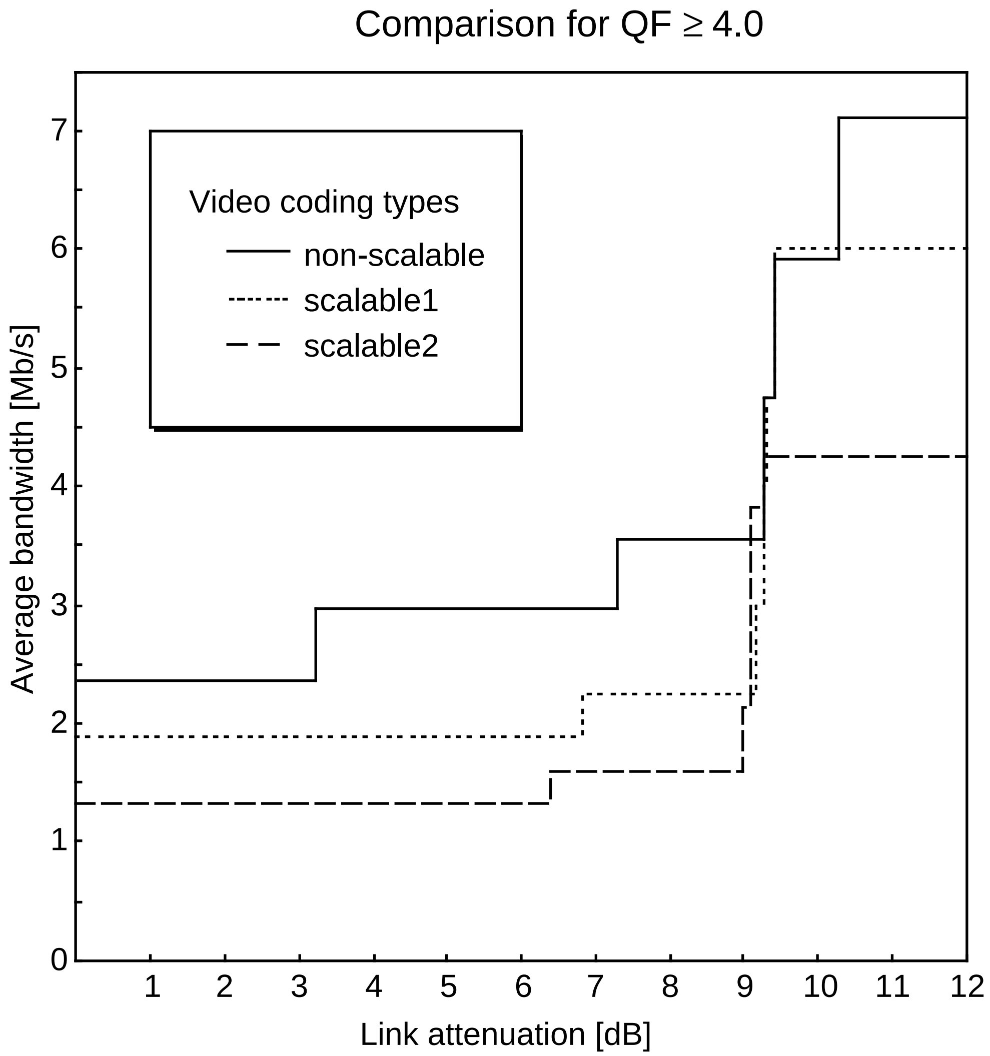

Figure 12. Bandwidth occupation by the transmitting earth station when maintaining a QF not less than 4.0, for the three video codings considered.

The situation is less definite when we consider high video quality. Figure 13 illustrates the performance of the three MPEG-2 codings when the target quality is 4.5. In this case, none of the two scalable video codings can drop the enhancement flow, so this advantage with respect to the non-scalable flow is lost. The non-scalable coding method takes advantage of a smaller overhead (hence a smaller average bandwidth occupancy) required, but on the other hand the transmitting station can use greater flexibility in handling the scalable codings, because it can attribute different channel bit/coding rates to the base and the enhancement flows. In our experiment, the two effects compensate, and indeed the three MPEG-2 coding methods have a very similar performance for attenuations less than about 9.5 dB. The only coding usable for attenuations greater than this value is the non-scalable one, but this result is not particularly significant as it is a consequence of our experiment constraints, which limit the average bandwidth to 7.2 Mbit/s.

This work has been published on Multimedia Tools and Applications, Vol. 10, No. 1, pp. 73-97, January 2000

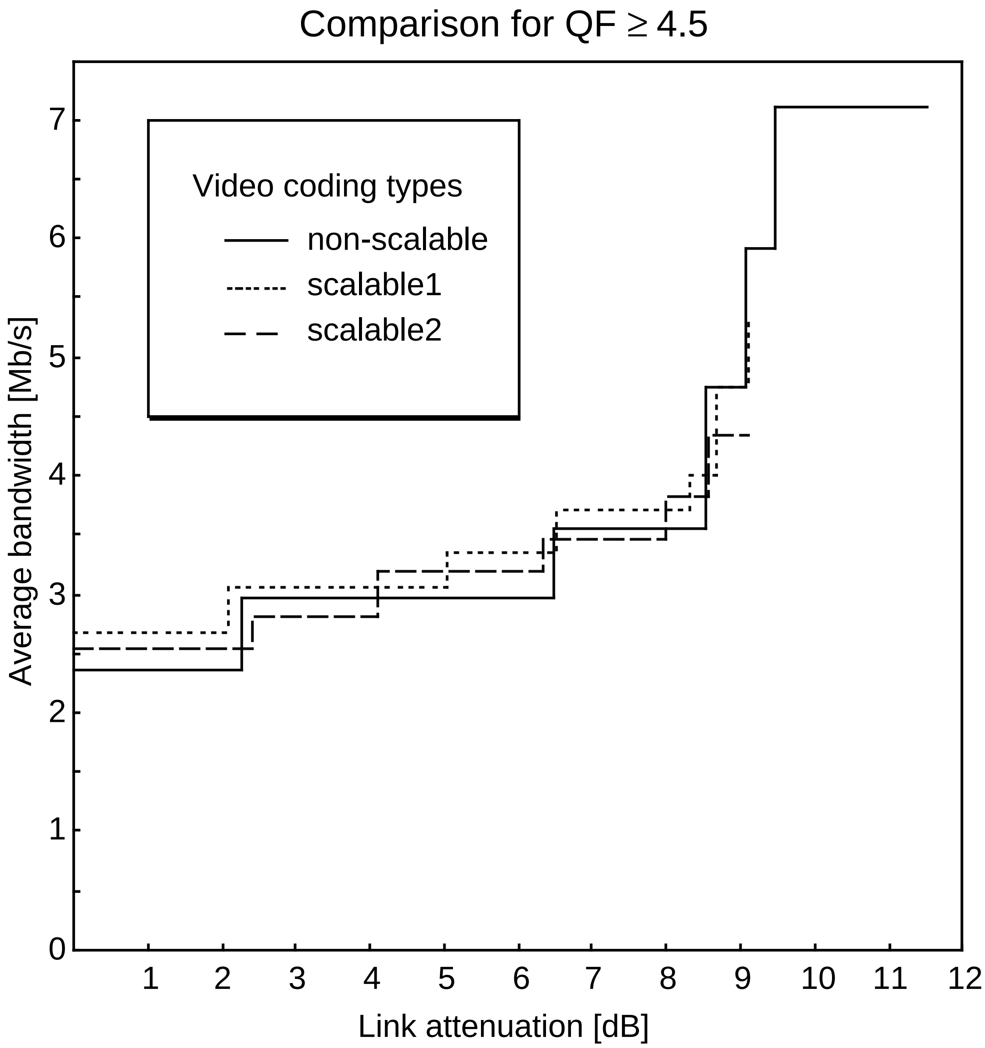

Figure 13. Bandwidth occupation by the transmitting earth station when maintaining a QF not less than 4.5, for the three video codings considered.

6. Conclusions and Future Challenges

The experiment has given some interesting results. First of all, when the transmission policy is to keep a medium quality (quality factor not lower than 3.5 or 4) at the MPEG-2 receiver, a scalable coding results advantageous. The advantage increases with the ratio of the average total bit rate over the base flow bit rate. When a quality factor not lower than 4.5 is required, only the non-scalable coding allows the countering of more than 10 dB of link attenuation without exceeding the maximum channel bandwidth available in our system. A higher bandwidth occupation allows a higher data redundancy, so it gives a better video quality for a given value of the satellite channel attenuation. Scalable methods take advantage of dropping the enhancement flow, when necessary, and obtain a quality factor not lower than 4 even at deep fade levels. When higher quality factor values are required, the intrinsic overhead of the scalable methods is balanced by the possibility of suitably protecting the two flows, and results are comparable with the non-scalable method. The further protection of the headers' field using a BCH code did not produce any appreciable difference in the resulting average quality factor, because the better protection of the headers is compensated by the spreading of erroneous bits produced by the data interleaver, which is needed by the BCH decoder to recover all the errors.

The results of this study do not have a general validity, because they are strictly related to the traces of the video sequence we examined and, moreover, the choice of the video encoder parameters to produce the MPEG-2 traces was obliged by the limits of the satellite channel we had at our disposal. Nevertheless, we think that the scenario presented can be helpful for the design of the earth stations and the payload required for VBR video transmissions. Moreover, the results obtained encourage us to continue the investigation of the transmission of MPEG-2 video codings on a noisy satellite link. Our future work in this field is oriented towards studying a feedback mechanism between the earth station and the MPEG-2 encoder, in order to choose dynamically a scalable or a non-scalable video coding according to the fade levels detected in real time. Of course, some problems must be addressed among others, such as the delay introduced by the feedback mechanism, and how the probability of being at a certain fade level influences the choosing of the MPEG-2 video coding.

References

[1] C.E. Shannon: “A mathematic theory of communications”, Bell Syst. Tech. J., part I, 379-423; Part II, pp. 623-656 (1948).

[2] J. G. Proakis, Digital Communications. McGraw-Hill, 1995.

[3] K. Feher, Digital Communications - Satellite earth station engineering. Prentice Hall, 1983.

[4] N. Celandroni, E. Ferro, N. James and F. Potortì: “FODA/IBEA: a flexible fade countermeasure system in user oriented networks,” Intern. Journal of Satellite Communications, Vol. 10, No. 6, pp. 309-323, 1992.

[5] N. Celandroni, F. Potortì, S. T. Rizzo: "An inexpensive rain fade countermeasure technique for DA-TDMA satellite systems", proceedings of IEEE GLOBECOM' 96, London, Nov. 18-22 1996, pp. 1000-1004.

[6] N. Celandroni, E. Ferro, F. Potortì: "Quality estimation of PSK modulated signals", IEEE Communications Magazine, July 1997.

[7] N. Celandroni, E. Ferro, F. Potortì: “Satellite bandwidth allocation schemes for VBR applications”, CNUCE Report C94-24, December 1994.

[8] N. Celandroni, M. Conti, E. Ferro, E. Gregori, F. Potortì: “A bandwidth assignment algorithm on satellite channel for VBR traffic”, International. Journal of Satellite Communications, Vol. 15, No. 6, pp. 237-246, November-December 1997.

[9] N. Celandroni, E. Ferro, F. Potortì , A. Bellini, F. Pirri: “Practical experiences in Interconnecting LANs via Satellite", ACM SIGCOMM Computer Communication Review", Vol. 25, No. 5, pp. 56-68, October 1995.

[10] Y. Ho, C. Basile, A. Miron, “MPEG-based video coding for digital simulcasting”, International Workshop on HDTV, Kawasaky, Japan, 1992.

[11] A. Webster, C. Jones, M. Pinson, S. Voran, S. Wolf, "An objective video quality assessment system based on human perception", Proceedings of SPIE Human Vision, Visual Processing and Digital Display IV, 1-4 February 1993, San Jose, California, Vol. 1913.

[12] Cheng-Tie Chen, "Error detection and concealment with an unsupervised MPEG2 Video decoder", Journal of visual communication and image representation, Vol. 6, No. 3, September 1995, pp. 265-279.

[13] ISO/IEC 11172 Information Technology, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s", Part 1: system; Part 2: video; Part 3: audio; Part 4: conformance testing, 1993.

[14] ISO/IEC 13818-2, ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio information: video", 1995.

[15] ITU-T Recommandation 601, "Encoding parameters of digital television for studios", 1982.

[16] H. Sun, K. Challapali, and J. Zdepski, "Error concealment in digital simulcast AD-HDTV decoder", IEEE Transaction on Consumer Electronics, August 1992, Vol. 18, N. 3.