Title: Future Performance of Video Codecs (SES2006-7-13)

Research Team

Prof. M. Ghanbari
Prof. D. Crawford
Dr. M. Fleury
Dr. E. Khan
Dr. J. Woods
Mr. H. Lu
Mr. R. Razavi

E-mail: ghan@essex.ac.uk
Tel: +44 1206 872434
Fax: +44 1206 872900

Video Networking Laboratory, founded 1969
Department of Electronic Systems Engineering
University of Essex
Wivenhoe Park
Colchester, CO4 3SQ
United Kingdom

November 2006

Executive Summary

Introduction

The intention of this report is to summarize the impact of new COder-DECoders (Codecs) for video compression on output video bit rates. The expertise of our research group is in video codecs themselves and, therefore, extensions to video and television applications are meant as illustrations only of the likely impact over the next two decades. There is no doubt that without codecs such as the MPEG and the H.26x series the present developments of digital television and storage would not have taken place, as bandwidth capacity for transmission and storage would be insufficient. Codec standardisation by the standard authorities, the ISO and ITU, provide a stable environment for broadcasters and manufacturers to develop their systems and services. At the same time, codecs allow the regulation of delivered video quality by variation of the compression ratio. Therefore, the issues of capacity and video quality become intertwined. Indirectly, future codecs also impact on spectrum allocation, especially in Europe, where in 1991 channel 22 had almost as many television transmitters (over 1400) than for all channels in the USA.

Digital television was introduced as an improvement over analogue television, offering High Definition Television (HDTV); an increased number of programmes offered with existing channels; broadcast to low-power, portable devices; reception in moving vehicles; and distribution over alternative telecommunications networks, including the Internet. Though the compression achieved by the existing standard codec, MPEG-2, has increased the number of programmes offered, many of the other goals are yet to be satisfied. A new codec, H.264, not currently widely deployed, has the ability in the fullness of time to radically reduce output bit rates, though that ability will be moderated by the need to offer higher quality reception than already exists. This report summarises the technical features about to be rolled out in H.264 and the future evolution of that codec or some further replacement within the next decade. There is now sufficient experience to anticipate the evolutionary behaviour of codecs from their inception to replacement by a future codec. This allows predictions of future gains, which are summarized in this Executive Summary. Reference is made below to Sections which continue the analysis and discussion in more detail within the body of the main report.

Background

Studio-quality digital video before compression requires a transmission rate of between 166 Mb/s – 830 Mb/s or equivalently a storage capacity of between 150 GB -- 745 GB per 2 hr movie. Practical channel capacities and storage are respectively limited to about 13-30 Mb/s for terrestrial transmission, 40 Mbit/s for satellite and cable, and 4-9GB for storage devices. For digital television in the UK, rates of 2--5 Mbit/s are the norm, allowing some four or five programmes to be broadcast within an 8 MHz UHF channel. Clearly, to achieve this reduction in bit rate between studio and broadcast quality video requires compression prior to transmission and recompression at the receiver or set-top box. This role is performed by a codec, and currently a standard codec, MPEG-2, is employed to achieve this process. Fortunately, video frames essentially repeat much of their visual information whilst creating the illusion of a moving image. It is this redundancy and others (Section 1.1) that are exploited to achieve the overall compression.

Table E1 summarizes production compression ratios (applied in practice) for television applications (see also Section 1.1), illustrating the vital role played by codecs in reducing bit rates and by extension bandwidths, to more manageable proportions. These rates are content and format dependent and it is possible to increase the number of programmes broadcast on one television channel by reducing the average video bit rate. MPEG-2 is unsuitable for rates below 1 Mbit/s. More recent and future codecs will enable a further reduction in output rates depending on quality expectations, which vary over time and by application.

Table E1: Illustrative compression ratios for production television applications with the MPEG-2 codec (25 frame/s, 4:2:2 colour sampling, 8-bit samples) at recommended qualities.

Pixel resolution (horiz.× vert.)	Format	Input rate from studio (Mbit/s)	Output rate (Mbit/s)	Compression ratio
1920×1080	HDTV	829	20	41:1
720×576	SDTV	166	3-6	33:1
360×288	SIF	31	1	31:1

In the last fifteen years, there has been a continuous evolution of video codecs (see Section 1.2 and "Evolution of Codecs" below) and there is no sign that this evolution is at an end. MPEG-2 is now an "elderly" codec, having been standardized by international authorities in the period 1994/5, and been overtaken in respect to the degree of achievable compression by H.263, MPEG-4 and in 2003 by H.264. Along with H.261, MPEG-1 and MPEG-2 were the first codecs to combine multiple ways (algorithms) of removing redundancy in one codec. Essentially, each video frame is split into blocks and matching blocks between successive frames are sought. Only the difference after it has been further encoded is then transmitted or stored. Each of the contributory algorithms has been refined over the years as intensive and competitive global research has taken place. Much of this report is taken up with the impact of those algorithms in the next two decades.

The latest codec to emerge, H.264, has taken advantage of the hardware bonus, as achievable computational complexity has increased in line with Moore's law (a doubling of processing power every 18 months). In particular, the size of the blocks that are compared has been reduced and made more flexible, which reduces the difference data that remains to be encoded. Of course, improved compression allows either a reduction in the spectrum required to transmit the same programmes or improved video quality using the existing spectrum or a combination of both. With the advent of second generation High Definition Television (HDTV) in Europe there will be an inevitable demand for greater bandwidths, as bit-rates of about three times those of Standard Definition Television (SDTV) are expected. It is likely that H.264 will provide many times the compression that was once achieved by MPEG-2 but that the gain will vary according to the size, resolution and quality of the image, either HDTV, SDTV or one of the smaller formats for handheld devices.

Evolution of codecs

To understand the evolution of codecs it is necessary to understand the standardization process. Due to engineering research there is a continual invention or refinement of compression algorithms, which is reported in journals such as those of the IEEE in the US. These innovations, after competitive assessment, are encapsulated by one of the two standards bodies, the ISO and the ITU, in standard codec specifications such as the MPEG and the H.26x series. However, the standard body standardizes only the format of the bit stream arriving at the decoder end, though obviously it is aware of algorithms that can exploit the information in the bit stream. The advantage of this procedure is that successive refinements can be made to the algorithms at the encoder or sender side, without occasioning the replacement of end-users' equipment, that is the myriads of set-top boxes, digital televisions, and so on, or the need to transmit multiple encoded formats (simulcast). A codec can be implemented as software and to the surprise of some it has become possible to run MPEG-2 software on a PC at video rates. This is not the case for the latest codec in its current form (H.264) but already the first all hardware solutions have been produced that will form future set-top boxes. Simplification of the codec holds out the prospect of an eventual software version appearing.

Therefore, successive refinements can and have taken place to a codec such as MPEG-2 during its life-cycle. These changes, for the same quality of video, have reduced the bit rate required for broadcast quality SDTV from 8.5 Mbit/s to 3 Mbit/s. Without entering into details, these changes can be chronicled as enhanced motion estimation in 1997-8, film noise reduction pre-processing in late 1999, and advanced pre-processing in early 2002. In Section 3.4 of the report, a similar set of changes are identified that are likely to lead to coding gains for H.264. Though we estimate the rate of improvement arising from changes in the pipeline, for HDTV particularly, this is still a matter for research.

In MPEG-2's case the rate of improvement has declined according to an exponential rule. That is the most gains were made in the first 3-4 years, whereas later improvements have not reduced the bit rate by as much. In fact, from about 2002 improvements have bottomed out or rather have approached an asymptote. Well before reaching the asymptote a new codec (MPEG-4) was introduced, see Figure E1. At this point, there is said to be a Leap Gain (Section 3.3) in compression efficiency, when a new codec by its structure is able to take advantage not only of all the existing improvements from a previous codec but also those innovations that have been stored up, as it were, in the research literature. In fact, Figure E1 is a simplification, as it only shows members of the MPEG family of codecs, whereas the alternative coding scheme H.263 predates MPEG-4 and has a similar performance. In H.264 the two standards tracks have merged. While the new and old codec continue to improve over time (not shown in Figure E1), the new codec continues beyond the effective life-time of the old, which has reached its asymptote.

New codecs are introduced to service new applications. MPEG-2 was developed for video broadcast, whereas its predecessor was intended for video storage on CD-ROM. H.263 was intended for video conferencing. From MPEG-4, the MPEG series have diverged towards compression services, including video animation and video database construction. H.264 aims to serve a variety of applications (Section 2) from very low bit rates of less than 20 kbit/s to HDTV quality video at around 20 Mbit/s.

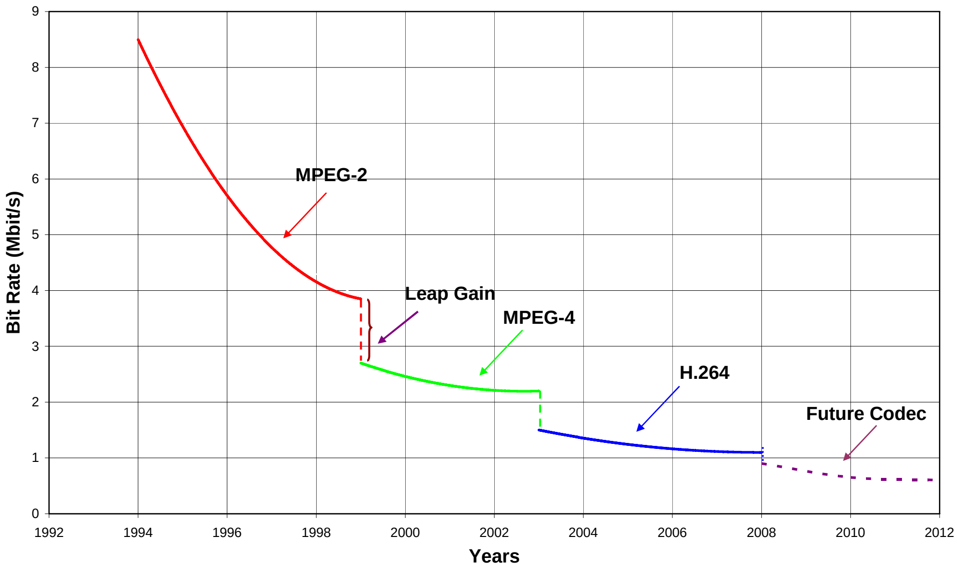


Figure E1: Stylized evolution of the standard codecs over time.

In fact, the introduction of improvements does not necessarily follow the rate suggested by Moore's law (see earlier Background) but behaves in a more conservative fashion. Another well-known commentator, Ken McCann of ZetaCast consultancy (see Section 3.1), has aligned his prediction more closely with Moore's law, opting for a year-on-year reduction in bit rates of about 15%. However, our prediction is closer to 7%, based on a more considered estimation of codec life-cycles, an expectation of higher video quality in the future, and higher performance from video decoders and displays. The principle reason for the retention of a codec is the need to sustain the economic life of consumer devices and video services. In the case of national television broadcasting, as socio-political factors come into play, this retentive effect tends to be masked. However, an early codec such as H.261 is still retained for some services for reasons of backwards compatibility and to maintain video conferencing across legacy ISDN circuits (a precursor of 'broadband'). Another restraining factor on bit rate reduction is the need to improve video quality at the receiver and Ken McCann has argued that bit rate reduction should not come at the cost of quality reduction.

Quality expectations

It is important to realise that for the sake of comparison, Figure E1 compares video of the *same* quality. The degree of compression influences the quality of the video as measured at the receiver device after decoding. MPEG-2 was designed to output good-quality video at medium bit rates, while very low bit-rate video (i.e. highly compressed video) is better achieved with H.263/4. However, lower quality is only acceptable for some applications such as video conferencing, and is heavily dependent on viewer expectations. At higher compression ratios and lower bit rates artefacts such as temporal flicker (awareness of picture/frame changes) or blockiness become apparent.

Therefore, another restraining force that reduces the bit rate reductions that might otherwise be expected is the continual rise in viewer expectations of broadcast and entertainment video services. In other words, for some applications, the gains that might have been expected from freeing up spectrum by improvements in compression will not become available because of the need to improve the quality of the delivered video (after decoding). The impact of flat panel displays in that respect are analysed in Section 2.1. The most likely area where there will be a growth in viewer expectations is in HDTV, for the reasons developed in Section 6.4.

Owing to expectations of higher quality services, the new generation codec achieves a relatively lower rate of coding gain when applied to high quality HDTV than they do when required to compress medium to low quality. Published results resulting from tests made when H.264 was first introduced tended to concentrate on lower pixel resolutions when quality defects are less apparent. However, it would be wrong to extrapolate from these results to HDTV. This is because H.264 largely improves over earlier codecs by more precisely identifying or addressing matching areas in successive video frames. This means that the residual information that is actually transmitted is reduced. In fact, for lower quality video much of the residual information is simply not transmitted. However, for higher quality video it becomes essential to transmit that information albeit in encoded form. However, H.264 has relatively little to offer over MPEG-2 in that respect, as is further shown in Section 2.2.

Therefore, while other commentators have identified the gains in compression leading to a reduction in spectral consumption based on a technical analysis, it is the view of the authors of this report that these gains will not be as significant as might be thought. Indeed rising quality expectations could lead to *less* gain in bit-rate reduction for high-quality HDTV over MPEG-2, than is expected from improving upon SDTV to lower quality video encoded with MPEG-2. Hence the bit rate of a future codec in Figure E1 may be increased slightly..

Services

We examine three new areas for which emergent codecs are set to have an impact, and look at how SDTV will be affected.

DVB-H (Section 6.3) is an exciting emergent standard for digital reception of television on handheld devices, which is already implemented (2006) in some Nordic countries. DVB-H is capable of reception on cars and trains, where it would be attractive to commuters. Transmitting to simple rod aerials rather than roof-top aerials raises issues such as transmitter power and error control, while battery powered devices have limited processing power. Future adoption of DVB-H is predicated on adopting a newer codec such as H.264 or VC-1, rather than MPEG-2, in view of the likely lower available bit-rates. The display format for mobile TV is likely to be either 352×288 pixels (CIF) or 176×144 pixels (QCIF). At these resolutions and in general for mobile applications, viewers are known to be more tolerant to weaker quality video than for SD or HD video. As remarked earlier, the greatest coding gain in new codecs comes at lower resolutions and, therefore, in Figure E2, year-on-year evolution of the codecs is predicted to reduce CIF bit rates to 64 kbit/s in two decades time. However, we do not expect the same gains for QCIF because rather than reducing the bit rate the quality of video, as well as its robustness against transmission errors will be increased, hence QCIF rates do not change in Figure E2. It should also be borne in mind that any further reduction in rate implies increased processing latency at the codec, which also may be a limiting factor on further reductions in bit rate. Figure E2 also applies to other wireless systems such as wireless LANs, cellular phone networks, and wireless access networks.

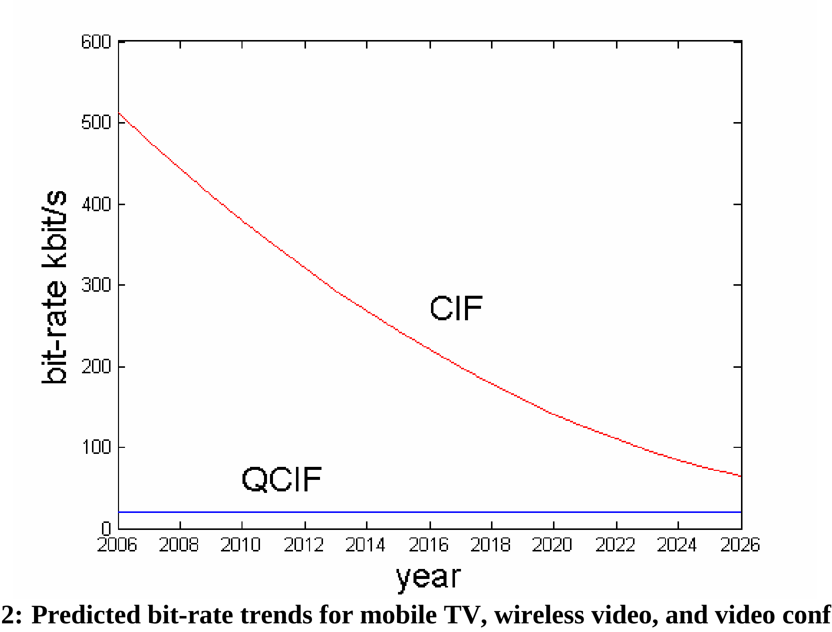


Figure E2: Predicted bit-rate trends for mobile TV, wireless video, and video conferencing arising from the evolution of future codecs.

In the UK, BT Movio piloted mobile TV in 2005, but for an alternative system based on an extended Digital Audio Broadcasting (DAB) standard, DAB-IP, which uses already allocated spectrum in the UK. The BT Movio platform is actually agnostic to the broadcasting standard. DVB-H is backwardly compatible with the DVB-T system, which the UK was an earlier adopter of in 1998. DVB-T was originally intended for portable devices (with simple aerials) and it was known that high-speed mobile reception was possible. DVB-H could co-exist with existing television programmes broadcast to fixed or portable receivers in the same multiplex on an 8 MHz television channel. However, though theoretically possible, both the differences in required field strengths between fixed/portable versus mobile broadcast and also the likely choice of the 4K variant of frequency multiplexing for mobile, rather than the current 2K or

8K variants respectively in the UK and Europe (Section 6.3) for fixed/portable broadcast, make it unlikely that mobile will co-exist in the same multiplex as fixed/portable broadcast. DVB-H transmission adds: power saving features; enhanced error protection (necessary for mobile reception); and a new mode (4K) that is a compromise between greater resilience against interference and mobile reception at higher speeds (at UHF). In an urban environment, such as Oxford, where DVB-H has been trialed, a typical bit-rate at the receiver for DVB-H would be between 250 kbit/s and 500 kbit/s, selecting a TV programme from a 10 Mbit/s multiplex. Actual bit rates are dependent on the pixel resolution of the picture and its quality.

SDTV (Section 6.2) Currently, broadcast quality video under MPEG-2 may be coded at 3-6 Mbit/s for fixed rate channels, depending on the content of the scene and expected quality, as lower quality commercial programmes may dip as low as 1.8 Mbit/s. Refer to Section 6.2 for current terrestrial bit rates specific to the UK. We predict a bit rate reduction of 7% per year for such a service and as Figure E3 shows, the illustrative fixed rate SDTV rate of 3-6 Mbit/s is most likely to be reduced to 1.5-3 Mbit/s in the next decade and to 1-1.5 Mbit/s in the following decade. These figures take into account an increase in quality expectations for SDTV. Section 3.4 details a number of algorithmic innovations that are likely to contribute to the future evolution of H.264 or a future codec, contributing to the overall bit rate reduction. The SDTV encoded range in Figure E3 is illustrative of a fixed rate broadcast quality and *not* lower quality commercial programmes. The SDTV plot implies that if statistical multiplexing is used more programmes per channel or equivalently some channels will become available for other purposes.

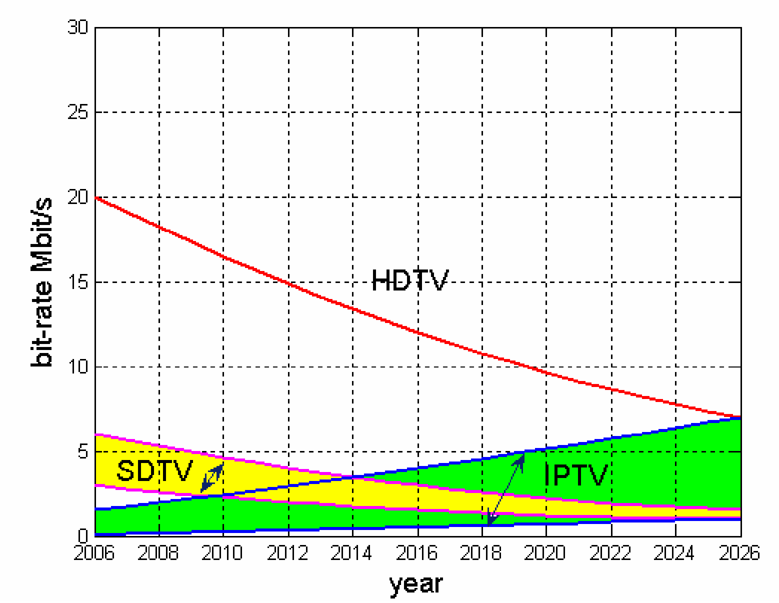


Figure E3: Predicted bit-rate trends for higher quality broadcast TV arising from the evolution of future codecs.

HDTV (Section 6.4) Japanese HDTV, originally in a pioneering satellite analogue form, now (2003) employs a DVB like digital system, Integrated Services Digital Broadcasting (ISDB), and a recent demonstration was of HDTV reception in a moving car with a window aerial. In ISDB, a 6 MHz channel is split into 13 segments, with all 13 segments needed for HDTV broadcast. In the USA, the FCC has mandated that all TV stations should be capable of broadcasting HDTV by the end of 2006 (with the ATSC system which is based on single carrier rather than the multi-carrier DVB system). When the USA introduced digital HDTV services in 1997/1998 with a resolution four times that of SDTV, Europe's experiment was not successful, possibly as a result of 'format wars'. Renewed interest in Europe coincides with growing consumer familiarity with flat panel displays (Section 2.1), spurred on by recent sporting events. A number of pixel resolutions are being considered by the European Broadcasting Union, including 1280×720 pixels with progressive (single frame) scanning at

50 frame/s; and 1920×1080 pixels at 25 frame/s, with either interlaced (two fields) or progressive scanning. HDTV aspect ratio is 16:9, which is closer to the behaviour of human peripheral vision than the 4:3 ratio of SDTV. From Table E1, the current encoded bit-rate for HDTV is about 20 Mbit/s. However, these figures assume the standard reduced rate sampling of colour components. If, as may arise for reasons of quality, the colour resolution in a European standard is made the same as the luminance (light intensity) resolution, then a factor of 1.7 increase in the input bit-rate would result (and a doubling compared to the lower quality 4:2:0 colour sampling system). 10-bit sampling, already provided for in the H.264 codec for high quality HDTV, may also impact. These illustrate the trend already mentioned to move to meeting the expected higher quality aspirations of viewers. In addition, the reduced coding gain available for high-quality video needs to be considered. There will also be a loss of statistical-multiplexing gain for multiplexes carrying only HDTV channels (caused by the low number of channels in the multiplex). Taking these effects into account, in Figure E3, for HDTV, we estimate the reduction rate to be less than 7% per year (rather than 7% as for SDTV). Figure E3 does not assume a fixed quality throughout but tries to take into account the joint increase in compression ratios and the increase in expected quality.

IPTV (Section 6.5) A number of market surveys report an increasing take-up of IPTV within Europe and in Italy IPTV has an entrenched position, while UK demand lags that of continental Europe. There is strong interest in a triple play option including IPTV that would extend available services for broadband users. Peter Cochrane, former head of research at BT’s Martlesham Labs., Ipswich has gone so far as to predict that 85% of Internet (or converged IP network) bandwidth might be taken up by visual services. We predict IPTV services will become the second most important visual service of the future. Currently, the average access network bit-rate in the UK is about 2.8 Mbit/s with rates of 8 Mbit/s possible. However, in Japan rates of 54 Mbit/s over ADSL 2 and experimental enhancements of ADSL exist, made more possible by the shorter average cable lengths.

In IPTV, current quality expectations are low and encoded bit-rates are much reduced compared to their broadcast counterparts. We believe that this situation will not continue and that quality expectations for some IPTV services will aspire upwardly. The improvement is most likely to occur for SDTV and possibly for an HDTV service, and not for lower resolutions. At the same time, various wireless LAN access networks may co-exist with ‘broadband’ access. These considerations explain the range of bit-rates and implied range of resolutions charted in Figure E3 for IPTV. In Figure E3, the region boundary lines represent upper and lower bounds on the bit rate. At the top of the range after two decades, IPTV is distributed at HDTV resolution and quality, made possible by the coding gains from evolved and future codecs. At the lower end of the range, for low resolution distribution, there will be minimal change, with if anything an increase in overall bit rates at the end of two decades. Currently, with an H.264 codec, reasonable video quality can be achieved with low resolution pictures of 128 × 96 pixels at 20 kbit/s. We believe future codecs, rather than trying to reduce this rate further, will try to improve the quality at the expense of keeping it at this rate, as there is little to be gained from further reductions. Just as in respect to QCIF video, there is a delay constraint, with an added delay from the need to packetize that sets a lower bound on the bit rate. IPTV will be delivered by simulcast (multiple formats at the same time) through scalable coding (Section 3.4.2) for a very wide range of access network bandwidths and types of users with differing satisfaction levels.

Hardware issues

Opinions differ on how long gains in computation will continue in line with Moore’s law, but even the originator of the law predicts the end within the next two decades. The failure of Intel and other manufacturers to increase clock speeds for the latest generation of general-purpose processors implies that gains will no longer come in reduction of chip feature size, which was the main driver of Moore’s law. Software versions of H.264 on PCs are not predicted to work in real-time now or later, but special-purpose integrated circuits for set-top boxes and camcorders are already available. In general, as codecs such as H.264 depend largely on intensive computation to match repeated areas in successive frames, this will have a drag on the rate of coding gain. As computation is directly related to resolution, this conclusion reinforces the reduced reductions in bit-rates for HDTV. Otherwise, we predict the emergence of low-power codecs to fall in line with the European power directives. There will also be research into conserving battery life by modification of codec implementations. Other hardware and environmental issues are considered in an Appendix to this report, as they do not impact directly on bit rates.

Conclusion

We have predicted a year-to-year reduction in bit rate of about 7% for SD quality video. This is a reduced estimate compared to predictions by other commentators. Our reasons for putting forward a more conservative estimate lie with: several parameters influencing the life-cycle of video services; viewers’ expectations of higher quality video combined with higher fidelity displays in the future. At this reduction rate, it is expected that the current rate of 3-6 Mbit/s for good quality SD video under MPEG-2 will be reduced to 1.5-3 Mbit/s in the next decade and then to 1-1.5 Mbit/s in the following decade. Equally, the year-to-year reduction rate for HD video will be less than that of SD video, where the current rate of 20-25 Mbit/s will be reduced on a decade by decade basis to about 12-15 and then 7-10 Mbit/s.

We believe along with market analysts that IPTV will be an important visual service of the future and, to satisfy a wider range of users, rather than a reduction in its bit rate, its bit rates will be broadened. We predict its current rate 64 -512 kbit/s will span over a large range of between 0.5-10 Mbit/s over the coming two decades.

For mobile TV, medium sized CIF pictures, its year-to-year reduction in bit rate may be larger than that of SD video, resulting in reducing the current rate of 256-512 kbit/s to 128 and 64 kbit/s over the two decades respectively. In order to improve quality, video with the smaller QCIF picture size is more likely to sustain its current rate of 20 kbit/s.

Contents

1 Introduction to video encoding 11
1.1 Standard video codecs 15
1.2 Motion estimation 17
1.3 Long term memory prediction 20
1.4 Transform coding 22
1.5 Statistical coding 23
1.6 Summary 17
2 Efficiency and computational complexity 25
2.1 Displays and Viewers' expectations 29
2.2 Summary 30
3 Trends of future video codecs 31
3.1 McCann Law versus Moore's Law 31
3.2 Russian steps 33
3.3 Leap Gain 34
3.4 Rate of coding gain 35
3.5 Factors adversely affecting compression 39
3.6 Summary 41
4 Other codecs affect on the Leap Gain 43
4.1 Fractal coding 43
4.2 Vector quantisation 43
4.3 Model based coding 44
4.4 Object based coding 45
4.5 Summary 45
5 Hardware issues affect on the Leap Gain 46
5.1 Summary 46
6 Applications 48
6.1 MBMS 48
6.2 SDTV 48
6.3 DVB-H 49
6.4 HDTV 52
6.5 IPTV 53
6.6 Summary 53
References 56
Appendix A Hardware issues 58
Glossary of acronyms 66

1. Introduction to video encoding

Raw video, when transmitted directly in at video rates, is notorious for its consumption of bandwidth. This is a result of digitally sampling each video frame both for luminance (light intensity), and chrominance (relative colour components) and subsequently sending at (say) 25 frame/s. The Source Input Format (SIF) and the Common Intermediate Format (CIF) are the starting points for a set of frame formats. While the SIF family apply to broadcast TV, the members of CIF (Table 1) are used in video conferencing and mobile applications. Their difference for the European standard is in their frame rate: SIF uses 25 frame/s and CIF 30 frame/s. Transmitting Quarter CIF (QCIF) frames at 25 frame/s would result in a bit rate of 8.7 Mb/s. Yet, QCIF is hardly the largest of formats, and Table 1 shows the range of CIF sizes with matching per frame bits, assuming eight-bit sampling[1]. These are shown for a 4:2:0 sampling pattern [1], as commonly used on DVDs. The eye's insensitivity to the spatial resolution of colour allows colour components to be sampled at quarter the rate of luminance ones, resulting in an average of 12 bits per colour pixel. If 4:4:4 sampling was used, giving equal weight to chrominance and luminance, an average of 24 bits would be needed. This obviously implies a doubling of the bits per frame in Table 1. In Table 1, 4CIF is appropriate for Standard Definition Television (SDTV), CIF and QCIF may be applied to videoconferencing and QCIF and below have been applied to mobile multimedia applications. There are other formats, notably CR.601, which for PAL/SECAM signals has a spatial resolution of 720 × 625 pixels, giving a total bit rate of 216 Mb/s[2]. Though there is debate over the most fitting High Definition Television (HDTV) format (Section 6.4) the implications in terms of raw bit rate for just one of these (say) at a frame resolution of 1280 ×720 pixels at 50 or 60 frame/s, if 4:4:4 sampling pattern is used, as is now suggested, with possibly 10 bits per sample, are easy to calculate.

Table 1: Video frame formats.

Format	**Luminance resolution (horiz. ×vertical)**	**Bits per frame (4:2:0, 8 bits per sample)**
Sub-QCIF	128 ×96	147456
QCIF	176 ×144	304128
CIF	352 ×288	1216512
4CIF	704 ×576	4866048

To allow raw video to be transmitted with realistic bandwidths (or stored on DVDs) it is clearly necessary to compress raw video. For most natural scenes[3], the eye can tolerate the loss or distortion of some pictorial information. This report is mostly concerned with 'lossy' compression, with lossless compression finding applications in still image, document and medical imagery. To compress a video sequence it is necessary to remove various redundancies from the information conveyed by the sequence. Removal of one of these redundancies, correlation between components of the red, green and blue (RGB) colour space, has already been implicitly assumed in the foregoing discussion. This is achieved by a simple

[1] The number of bits required for interlaced video is the same as for progressive video.

[2] This figure includes blanking intervals. Without blanking intervals the rate is 166 Mbit/s, as quoted in the Executive Summary.

[3] Technically, a natural scene forms an order-one Markov field.

linear transformation to form Y (luminance) C_r (red colour difference --- chrominance) and C_b (for blue) samples.[4]

The job of removing other redundancies is the task of a CODEC (enCOder DECoder) the most well-known of which are the MPEG series from the International Standard Organisation (ISO), which were actually awarded Emmy awards. However, there is another series of codecs from the International Telecommunications Union (ITU), the H.26x series, where x can be 1, 2, 3, or 4. The two series have now merged to form the MPEG-4 Part 10/H.264 standard. Further details of the evolution of the standard codecs are given in Section 1.1. Other codecs exist but information on some is proprietary, such as that from RealVideo Inc., and others have not had widespread take-up by hardware codec manufacturers, such as VC-9 from Microsoft (MS) Corp. for HDTV, possibly due to later progress towards standardization. Tests by the BBC in 2003 comparing H.264 with VC-9 for 720p/50 TV concluded that edge preservation and blockiness removal by VC-9 were less than by H.264, though there were differences in the type of artefact occurring [50]. Still image codecs, especially JPEG and its successor JPEG2000, do not remove motion redundancy, which is the most notable difference between images and documents compared to video. For high-quality (telemedicine) or high-spatial resolution or both (digital cinema --- Section 6.4) employ motion versions of these codecs.

An interesting codec operating at various spatial resolutions for progressive and interlaced television is the BBC’s Dirac codec as its code is open source [49] (though of course, the source code for reference implementations for the standard codecs discussed below is available at the standard body’s web pages and elsewhere). At the 2006 International Broadcasting Convention, there was a demonstration of 1080 line HDTV compressed with Dirac at broadcast bit rates, i.e 18 Mbit/s, which is similar to MPEG-2 rates for HDTV. Dirac may also have a future in low latency applications, typically conversational services but also video streaming, and there are other applications for studio quality video. As at the time of writing, the Dirac specification was not complete and the software was at release zero, further detailed analysis must await the passage of time.

The standard codecs consist of a series of algorithms to remove various forms of redundancy, giving rise to the term hybrid encoder. It is worth remarking that a decoder only receives a quantized difference signal from what it could predict from reconstructing preceding frames. Therefore, an encoder includes a decoder in its encoding loop in order to form the difference between the present and previous frame. This explains the asymmetry in hardware complexity between encoder and decoder (Section 5 and Appendix A). There now follows a summary of the various forms of redundancy and the algorithms for their removal. Because of the complexity of these algorithms it is completely impossible to fully describe them in a short space and the reader is referred for further details to an account such as [1], which itself is a compact narrative.

A hybrid video encoder, Figure 3, firstly removes temporal redundancy, then removes spatial redundancy and, finally, removes statistical or entropic redundancy. There is a qualification: because when transmitting set of pictures it is possible for data to be lost or corrupted, leading to accumulative errors in prediction, temporal redundancy is not removed from anchor, Intra- or I-pictures. Every group of pictures (commonly every 12) another I-picture is formed, while intermediate pictures are predicted (P-pictures) or bi-predicted (B-pictures) from previous and succeeding pictures.

[4] The green chrominance component can be re-created from the other three. The separation of the luminance allowed backwards compatibility with monochrome TV sets.

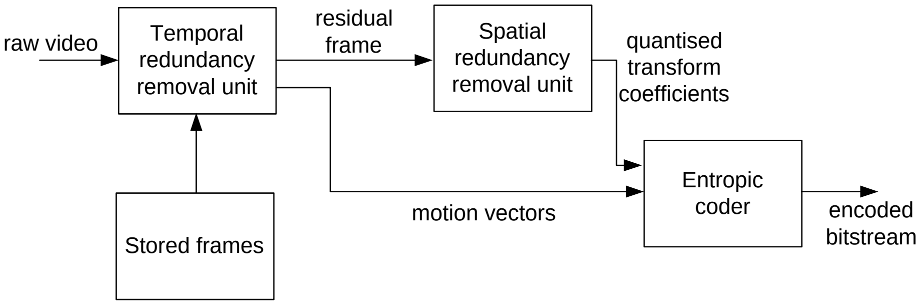

raw video
Temporal redundancy removal unit
residual frame
Spatial redundancy removal unit
quantised transform coefficients
Entropic coder
encoded bitstream
motion vectors
Stored frames

Figure 3: Simplified block diagram of a generic video encoder.

From Figure 3, raw video frame enters the encoder in raster scan order. The purpose of the temporal redundancy removal unit is to remove information that has already been sent to the decoder in the previous frame(s). It should be noted that, in the literature, information is often referred to as energy because of an analogy with statistical physics, in which disorder requires more information to describe it, while disordered particles are more energetic. If a simple difference frame was formed, the luminance difference image might still contain very dark and light areas, as well as considerable spatial variation in the distribution of the light and dark areas (energy). Leaving aside scene changes/cuts, the most likely cause of the differences is due to motion of the objects within the scene or camera motion. Therefore, video coders employ motion compensation to compact the information or energy in the difference frame.

For computational convenience, motion estimation (finding the position of a feature in the current frame in a previous frame already sent to the decoder) is commonly based on blocks and until recently on 16 × 16 luminance macroblocks (MBs). The best match, in terms of some form of pixel-wise mean error, is sought. Because of the nature of natural images, this search need only take place in the local vicinity. Nevertheless, there are many variations to the search pattern and recent codecs reconstruct intermediate search positions between the sampling grid by forming sub-pixel MBs through interpolation. Once a partial match is found, a 2-D motion vector (MV) describing the relative position of the vector is sent to the entropic coder for processing (refer to Figure 3). The residual MB (after differencing) is sent to the spatial redundancy unit. The main difference between an encoder and a decoder is that there is no need for the decoder to estimate motion, thus removing the most computationally intense part of the processing loop.

It is still possible to predict an MB even if no reference is made to a previous frame. In Intra-picture encoding this is possible through comparison with the pixels of adjacent macroblocks. Whatever the means of formation of the residual MB, the next stage is to decorrelate the inter-dependence in value of nearby pixels in the spatial domain by transforming to some type of spatial frequency domain. For computational tractability the transform is normally performed on the four 8 × 8 pixel sub-blocks of an MB. The Discrete Cosine Transform (DCT) is reversible (preserves energy or information), has a fast algorithm, and for natural images approaches optimality in terms of decorrelation and compactness (the ability to represent the energy of multiple samples in the spatial domain in a few samples in the transform domain). The DCT in most codecs before H.264 takes place through floating point operations, while in H.264 an integer-based version avoids calculation discrepancies at the decoder when the reverse transform is applied. The DCT is by no means the only reversible transform or more precisely orthogonal transform. For example, the Hadamard transform may be applied in the H.264 codec to the zero spatial frequency (DC) coefficients across already DCT-transformed blocks (Section 1.4). The Hadamard transform is restricted to binary operations on the spatial intensity values.

The wavelet transform has risen to prominence and has been widely applied especially to still images. The wavelet transform is most appropriately applied to a complete image rather than a block. This is because a wavelet's spatial frequency filters can be tuned to the characteristics of the local spatial region. In fact, the same wavelet filter acts rather like a subband encoder, as it is applied on successively down-sampled versions of the original image. As Section 1.4 explains, though the wavelet may outperform the DCT in compactness that is not its principal attraction but rather the natural way that its method of transformation lends itself to scalable coding at the entropic coder. In H.264, it has been used as such for scalable coding between pictures (not for residual MBs) through the Haar transform, which has been rediscovered as a wavelet transform though it also has the nice processing characteristics of the orthogonal transforms.

If block based processing occurs then discontinuities occur at block edges. In H.264, but not in the MPEG series, a de-blocking filter is applied at the decoder to reconcile discontinuities. Removal of blocking artefacts is an example of the exploitation of psychovisual redundancy.

Assuming now that a DCT has been applied to a residual MB, at this point no loss of information has been occurred. However, the transform coefficients are now quantised removing the ability for perfect reconstruction. If linear or quasi-linear (with a dead zone at zero level) quantization is applied then the step size can be tuned to change the ultimate bit rate (against a coarsening of the output). Another technique exploits psychovisual redundancy through a quantization scaling matrix but is not carried over into H.264 from the MPEG series, except when the codec is applied to the highest quality video.

On arrival at the entropic coder in Figure 3, the quantised transform coefficients are reordered. This step is needed so that runs of coefficients with similar values are placed in adjacent positions. In particular, many of the values will have been reduced to zero because of the compaction achieved by the DCT. As the largest value is invariably the DC coefficient, the reordering starts here and proceeds in zig-zag fashion (for frame blocks), selecting the low spatial frequency coefficients before higher spatial frequency ones. Runs of zeros are now coded by their number and the next non-zero value, as well as indicating, optionally, whether the final run of zeros has occurred.

An entropic encoder weights the number of bits applied to arriving symbols according to their statistical distribution, hence the term variable length encoding (VLC). In the case of a video encoder the symbols are the run lengths or the MVs and other miscellaneous data. Huffman coders have been applied but compared to arithmetic coders, a Huffman coder can only code to an integral number of bits per symbol. Therefore, arithmetic coders first appeared in H.263 and in more computationally efficient form in H.264. Both Huffman and arithmetic coders can be applied based on a static predetermined distribution or can adaptively be formed according to the evolving context of the video sequence being encoded. The gain from entropic coding is limited (about 3-4 times reduction in JPEG Lossless (LS)), which is why information loss from quantization is required.

Finally, it should be noted that an encoder such as MPEG-4 or MPEG-2 can be used for numerous applications and bit rates. Implementation of a codec to operate over a wide range of applications can be very expensive. To reduce the cost, codecs are designed for a limited range of applications, with their limits being specified by Profile and Level [1]. Profile is a subset of the entire bitstream syntax[5] and the level is a defined set of constraints imposed on the parameters in the bitstream. For example, the well known MPEG-2 MP@ML: main profile main line, which is a widely used pair for broadcast TV, describes a non-scalable

[5] The bitstream syntax is the set of rules for the configuration of the encoder's output bitstream so that a conformant decoder is able to decode the bitstream.

bitstream containing I,P and B pictures of 720 pixels by 576 lines, 25 frames/s with 4:2:0 colour format.

1.1 Standard video codecs

Almost all standard video codecs invariably follow the generic structure consisting of: motion estimation/compensation (to remove temporal or inter-frame redundancies); transform coding (to de-correlate inter/intra pixel redundancies); and entropy coding (to remove statistical redundancies) [1]. Of course, video codecs are intended for lossy compression of natural scenes, whereas lossless dictionary coders such as Lempel-Ziv (LZ) are intended for text. In the last one and half decades, coding efficiency has been significantly improved by evolving the algorithms in each of these areas to improve the performance of codecs, resulting in more advanced standard codecs. However, it is in motion estimation/compensation that the largest compression gains have occurred, though these improvements have impacted considerably on computational complexity. In this report, we will briefly describe each of the component parts of the standard codecs, with an emphasis on factors that are important to future predictions. Of course, improvements in compression allow the same video quality to be achieved, either in storage or video communication, for a reduced bit rate. A convenient measure of video quality is Peak-Signal-to-Noise Ratio (PSNR), which is measured logarithmically in decibels (dB)[6]. Although the reliability of PSNR figures has been debated as to whether or not they equate to the results of subjective testing (mean opinion scores), nevertheless, PSNR figures are still widely quoted in the literature as a measure of performance. However, in comparing video codecs with different type of induced distortions (*e.g.* smearing distortions introduced by wavelet based codecs versus the picture blockiness by DCT based codecs), PNSR may not be valid. Nonetheless, it is generally agreed that, within a given codec, improving PSNR also improves the quality. Since the types of distortions in all standard video codecs are similar (they are all DCT based), then various standard video codecs can be compared on a PSNR basis. It should also be noted that PSNR figures, like mean opinion scores are content dependent, but values in the range of 30-39 dB, depending on the application, are regarded as acceptable [2]. For example, the target PSNR value for prototype video codecs of lower quality video (e.g. mobile TV) is around 34 dB and that of broadcast quality is around 39 dB, though recently the Fraunhofer-Heinrich Hertz Institute (HHI) of Germany has increased its target PSNR to 42 dB. This is, of course, another indication that future video codecs are required to have a better video quality than that which is now acceptable.

Before evaluating how much has been gained by successive refinements, it will be helpful to briefly review the various standard video codecs developed so far.

H.120 was the first video-coding standard developed by CCITT (now ITU-T) in 1984 for video conferencing applications. This codec was based on Differential Pulse Code Modulation (DPCM), scalar quantisation, and conditional replenishment (direct re-use of similar regions from a prior frame). It supported bit rates of 1.544 and 2.048 Mbit/s of the first digital hierarchy of North America and Europe respectively. This codec was abandoned soon afterwards and is not in use anymore, as shortly afterwards a new standard H.261 was developed.

[6] Specifically, PSNR = $10\log_{10}\left[\frac{p^2}{(1/n)\sum_{i,j}(Y_{ref}^{i,j}-Y_{prc}^{i,j})^2}\right]$ dB, where p is the peak value for a given pixel resolution, e.g. for 8-bits p = 255, N is the total number of pixels in a picture, i,j range over every pixel of the picture, and Y_{ref} is the luminance value in the reference picture, while Y_{prc} is pixel value in the processed picture.

H.261, developed in 1990 to replace H.120, is regarded as the basis or originator of all modern video compression standards. The basic structure (the hybrid coding structure) proposed in H.261 is still dominant today. H.261 was based on 16×16 MB motion estimation/compensation, 8×8 DCT, zig-zag scanning of DCT coefficients, scalar quantisation of those coefficients, and subsequent variable length coding (VLC). The other key aspects of this coder were a loop filter (to remove artefacts at block boundaries), integer-pixel motion compensation accuracy (optional) and 2-D VLC for coding of coefficients. H.261 operates at bit rates of k×64 kbit/s, where k is an integer with values in the rage from 1 to 30 and 64 kbit/s is the base rate for ISDN links. H.261 is still in use (mostly as a backward compatibility feature) but it has now been overtaken by H.263.

MPEG-1 (1991) was the first coding standard for motion pictures developed by ISO. It was mainly developed for video storage applications (on CD-ROM). MPEG-1 utilizes the same structure as H.261 but introduces the concept of bi-directional prediction (B-pictures are predicted from anchor I- and P-pictures). MPEG-1 provides superior quality to H.261 when operating at high bit rates (>= 1 Mbit/s for CIF 360 × 288 pixels – spatial resolution). MPEG-1 also adds half-pixel motion estimation to the H.261 design.

MPEG-2 (also known as ITU-T H.262) was developed jointly by ISO and the ITU-T in the period 1994/95. It is one of the most commonly used video coding standards deployed at this time, particularly for DVD and Digital Video Broadcasting (DVB). MPEG-2 supports two new features namely: interlaced scan pictures and scalability. Otherwise, in all other aspects it is essentially the same as MPEG-1. Although MPEG-2 has many applications, it was designed mainly for high quality video at bit rates in the range 2-20 Mbit/s, and is *not* suitable for low-bit rate applications (below 1 Mbit/s). Its various applications are defined under levels and profiles.

H.263 was first developed in 1995 to replace H.261 as the dominant video conferencing codec, owing to its superior performance at all bit rates. In particular, at very low bit-rates it reduces the bit-rate by a factor of two compared to H.261. The basic algorithm in H.263 employs: half-pixel motion compensation; 3-D VLC of DCT coefficients; and median motion vector prediction. In addition, H.263 proposes many optional enhanced modes such as: increased motion vector range; advance prediction mode (with Overlapped Block Motion Compensation (OBMC) to counter blocking, and switching between one and four motion vectors (MVs)); optional arithmetic entropic coding; and coding of PB frames – two P and B-pictures are coded as one unit, reducing overhead at low bit-rates. H.263 went through many refinement phases resulting in H.263+ (1998) and H.263++ (2000). Most of the improvements involved error resilience and scalability aspects to cater for a new range of applications over mobile networks and the Internet.

Frame based MPEG-4 (v.1, early 1999) approximately follows the H.263 design and includes all prior features, including various trick modes (simulation of VCR functions such as fast replay). However, MPEG-4 can also code multiple objects within a video frame, with shape coding of video objects being an important object coding technique. MPEG-4 also includes zero-tree wavelet coding of still pictures, as well as dynamic 2D mesh coding of synthetic objects and facial animation modelling. There are many application profiles and levels in MPEG-4. Some of them are implemented while others are still at a development stage or may never be implemented. Therefore, MPEG-4 is better regarded as a toolset of compression tools, rather than a codec in the mould of MPEG-1/2. It has a high degree of complexity compared to MPEG-2. Version 2 of MPEG-4 introduces quarter-pixel motion compensation and global motion compensation. Despite many fanfares, MPEG-4 has not been as popular with manufacturers as anticipated. The demise of MPEG-4 can be ascribed to the failure of hardware manufacturers to take up its object-based features, as these would require a radical new design compared to the macro-block-based processing streams that manufacturers have been accustomed to.

H.264/ Advanced Video Codec (AVC) (also known as MPEG-4 Part-10 or Joint Video Team (JVT), after the developers) is a state-of-the-art video codec, standardised in 2003. It is suitable for a wide range of applications such as broadcast with set-top-boxes, DVD storage, packet networks, and multimedia telephony systems. H.264 encompasses the full range of bit rates and quality resolutions, unlike some previous codecs. Profiles such as the High profile for HDTV and Blue-ray disc storage support a wide set of applications. The Baseline profile is intended for applications with limited computing resources, such as video-conferencing and mobile applications. The Main profile was intended for broadcast TV and storage but has been overtaken by the High profile. The Extended profile is intended for streaming applications, with robust coding, trick modes, and server switching. H.264 is the first video codec that has been explicitly designed for fixed-point implementation and the first network-friendly coding standard. It has higher computational complexity but better coding efficiency than previous standards. H.264's better coding efficiency is mainly attributed to some of the features arising from H.263++, such as predictive intra-frame coding, multi-frame and variable block size motion compensation, quarter to one-eighth pixel motion estimation precision, integer DCT transform, adaptive in-loop de-blocking filtering, and more efficient/advance entropy coding. It has an increased range of quantisation parameters, and employs Lagrangian optimised rate control. In the latter, the components of the target bit-rate in a rate-distortion function are divided between the individual coding parts in such a manner that maximum reduction in distortion is achieved at the expense of minimal increase in the bit rate.

Figure 4 compares the relative performance of various standard video coders for a Quarter-CIF (QCIF) resolution video (Foreman). It can be seen that by improving the various parameters, the bit rate of standard codecs is reduced significantly, starting from Motion JPEG (JPEG still image codec applied to video without removal of temporal redundancy) to H.264/AVC. In interpreting the performance figures, one has to note the reference point. For example, in Figure 4, for a picture quality of 34 dB, 77% reduction in the bit rate, means that H.264 is 100/(100-77) or nearly four times more efficient that motion JPEG!

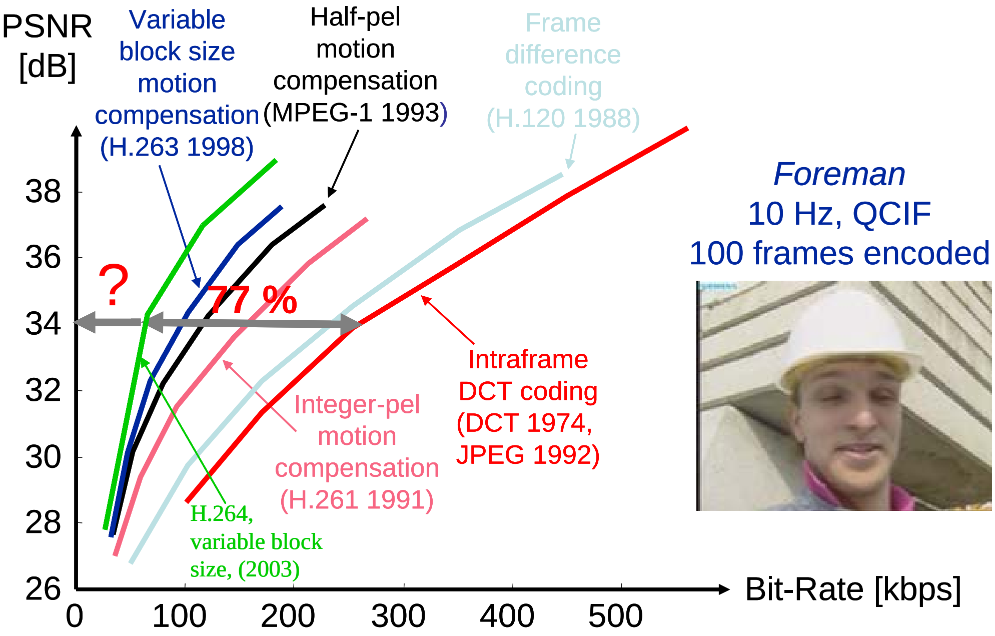


Figure 4: Comparison of various standard video codecs for the Foreman test sequence, after [15].

1.2 Motion Estimation

As stated earlier, almost all codecs avail themselves of simple block based motion estimation and compensation. However, over time many enhancements have been introduced. Although most of these are based on motion prediction (which is difficult), the net result has still been significant compression gains. As noted previously, these improvements come at a cost, as motion estimation dominates the computational complexity of a codec (e.g. in MPEG-1 alone, more than 65% of the processing power of the encoder is assigned to motion estimation, and this figure for H.264 can go beyond 80%) and is also the most memory hungry component, with repercussions in cost, power usage, and form factor. The different improvements in motion estimation and compensation parts can be broadly classified into three groups:

1.2.1 Effect of block-size on coding.

It is found that accurate modelling of motion vectors can improve the performance by up to 15%. This can be performed in many ways: either by means of a rectangular block structure but with variable sizes (the size of a MB can be varied depending upon the motion activity in the picture) or through a non-rectangular MB structures (such as quadrilateral). For example, MPEG-1 has a fixed block size (16×16) for MBs and only one pair of MV per MB, while H.263+ in advanced prediction mode has an option for either one or 4 pairs of MV corresponding to block sizes of 16×16 or 8×8. On the other hand, H.264/AVC has variable and hierarchical block sizes with horizontal and/or vertical divisions of MBs (a 16×16 block may be divided into 16×8, 8×16 or 8×8 and each of 8×8 block can further be divided in to 8×4, 4×8 and 4×4 size blocks). In general, smooth regions/stationary parts of the frame are coded with large block-size and moving parts with relatively smaller block sizes. A comparative study shows that variable (two) block sizes in H.263+ reduces the bit rate by 5-8% over a fixed block size of 16×16 for CIF resolution sequence for a quality of 34 dB [3]. However, the variable block size in H.264/AVC can improve the coding efficiency by up to 15% over the fixed block size of H.263 or MPEG-2 [4]. As Figure 4 shows, larger improvements are achieved at higher PSNR values and higher bit rates. However, for much larger bit rates at SDTV and HDTV quality, the curves converge and differences are reduced. Further, the use of OBMC [5] results in an additional improvement of 0.2-0.5 dB in the quality, which depending on the desired quality, can result in another 5% reduction in bit rate.

1.2.2 Fractional Pixel Accuracy of Motion Vectors

Earlier codecs such as H.261 employed integer pixel accuracy for MVs, or even motion compensation was made optional. However, it was later observed in MPEG-1 that half-pixel resolution MVs can result in a significant gain of 25-30%. Therefore, MPEG-2 and H.263+ were designed with half-pixel accurate MVs. H.264/AVC extended this to either one-quarter or one-eighth (optional) pixel accuracy, achieving a further gain of respectively 15-20% and 20-30% over half-pixel accuracy. In addition, filtering during interpolation of sub-pixels also plays a small role. Figure 5 shows the effect of increasing the motion vector accuracy for a typical Mobile and Calendar sequence [6]. It is noticeable that a significant coding gain can be achieved by representing MVs with either half pixel or quarter pixel accuracy, but further increase in accuracy (beyond one-eighth accuracy) has no significant gain, despite the increased complexity.

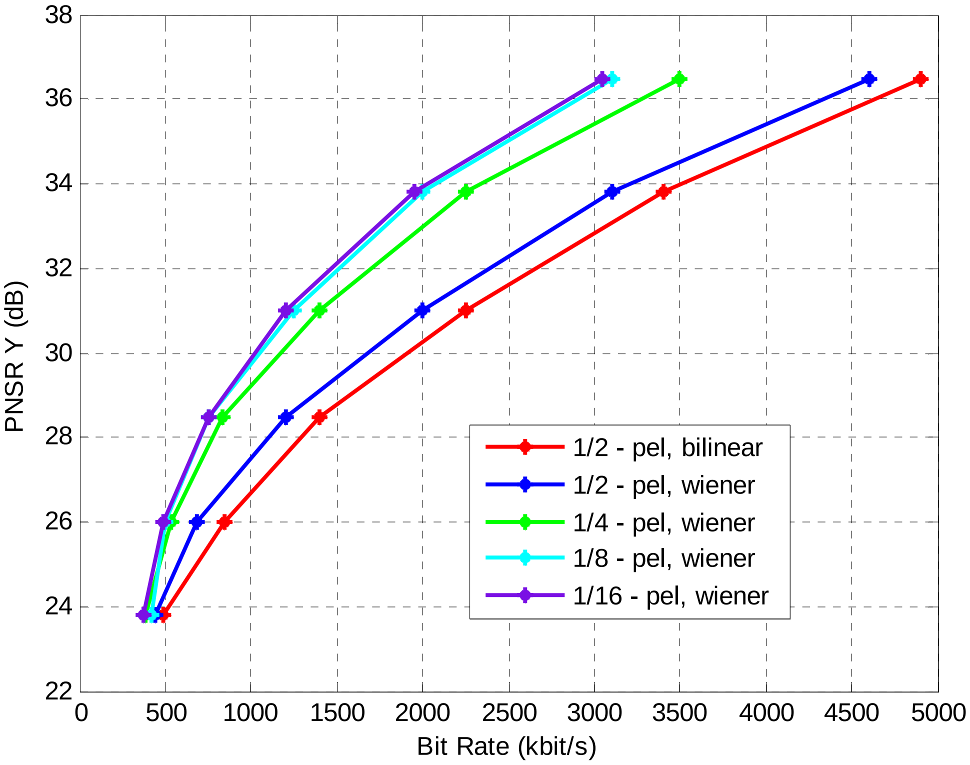


Figure 5: Effect of sub-pixel accurate motion vectors on the performance of video codec, after [6].

1.2.3 New Generation of Motion Estimation

Block-based motion estimation and compensation is very widely practised, but the major drawback is that only translational motion[7] is considered. Although variable block size motion estimation and compensation, to some extent, takes into account the different levels of motion in a frame, it fails to completely model any non-translational motions (such as rotational motion or affine motion[8] in general). Next generation video codecs may incorporate more of these advanced motion models, known as spatial transformations (warping and deformed mesh) to incorporate both translational and non-translational motion alike. In this case, a rectangular block is matched against a quadrilateral through picture interpolation. Figure 6 shows how accurate motional modelling (by warping the block through shifting the vertices of the quadrilateral) can improve the performance of video coders by more than 1.5 dB [7]. Although this technique is very complex, with faster computational resources it will not be difficult to employ these types of motion model in future codecs.

[7] Translational motion in this context is that in which a block moves from one point to another without changing its orientation to fixed points.

[8] In addition to translational motion, affine motion in this context includes changes through rotation in a block's orientation to fixed points and also changes due to shearing.

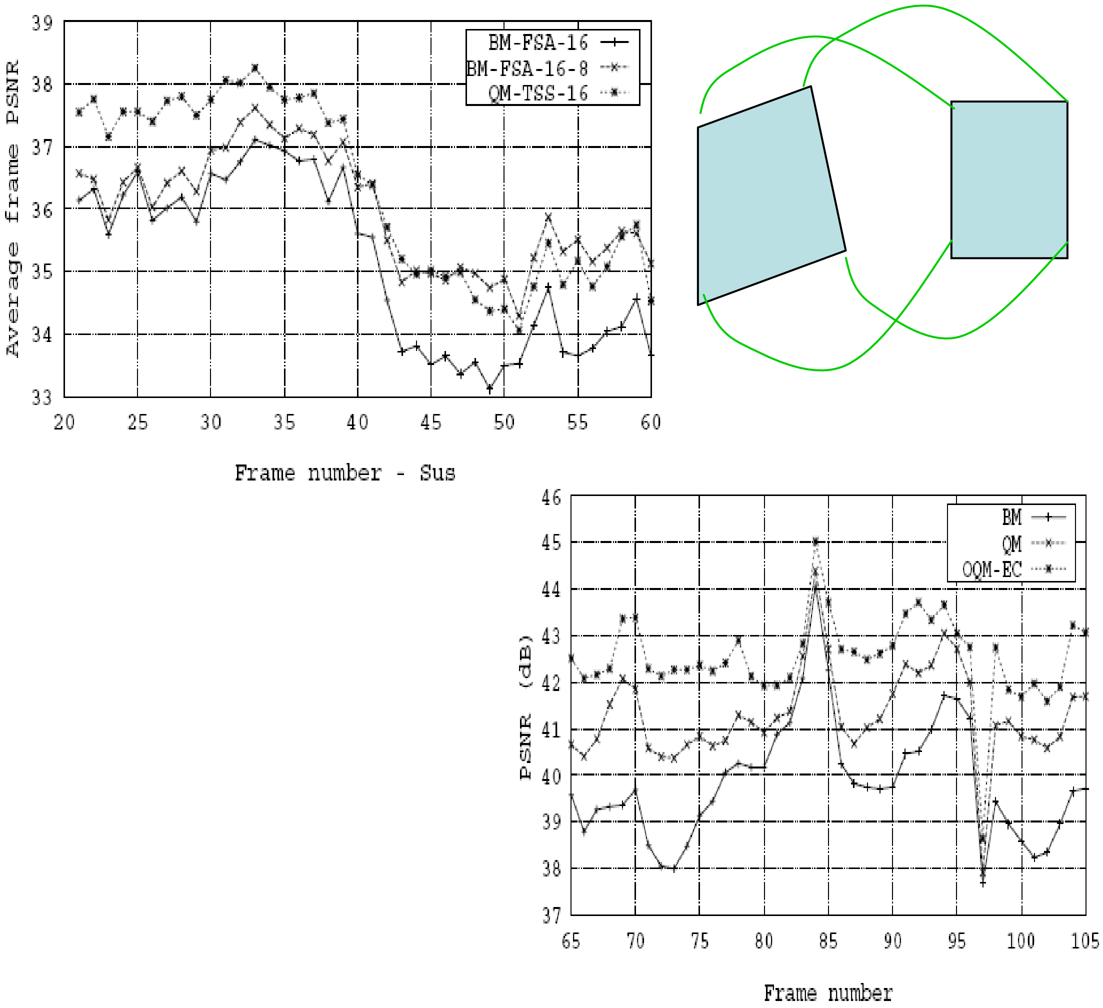


Figure 6: The improvement due to the use of deformed block structures (Spatial transformation in block matching), after [7].

1.3 Long term memory prediction

In all standard codecs prior to H.263++, a P-frame/slice references only one (immediately before) I- or P-picture, whereas B-frames are predicted from both directions (previous as well from future I- or P-pictures). However, in Annex U of H.263++, there is a provision for more than two reference pictures. This was primarily designed to combat against channel errors, as the erroneous reference frame may not be used for the prediction of future frames. However, later on it was discovered that some of the past frames may be more similar to the current frame than its immediate predecessor. Hence, they can also improve compression efficiency. Figure 7 shows the effect of long-term memory prediction in H.263++ for the Foreman sequence and it can be observed that multi-frame prediction can improve the performance by up to 17% [8]. This concept is adopted and made mandatory in H.264/AVC and 10-15% compression improvement may be contributed by long-term memory prediction (up to 15 frames).

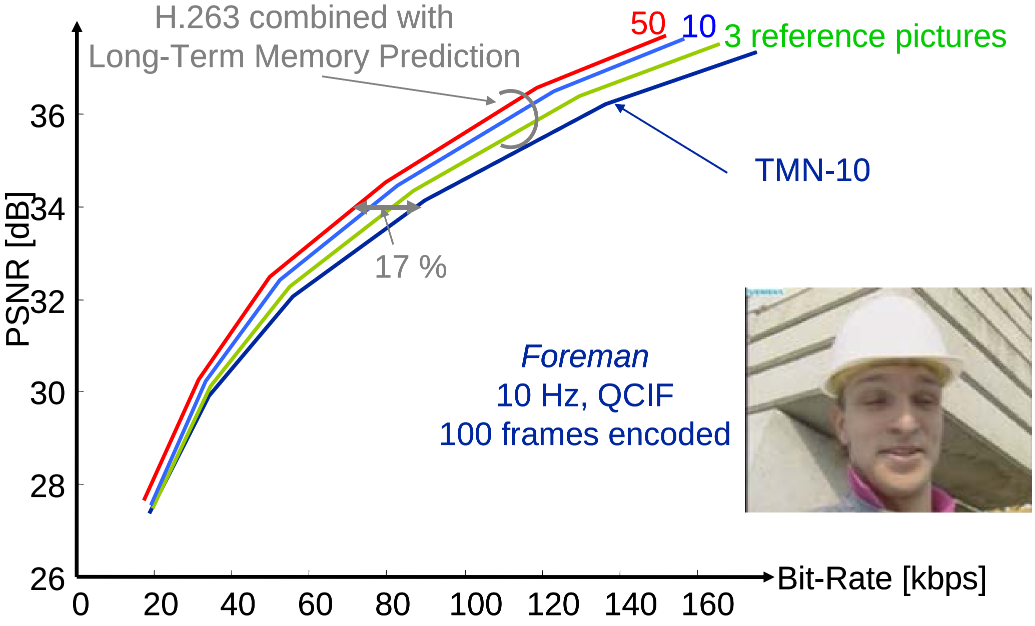


Figure 7: Effect of long-term memory prediction in H.263++ (Annex U), after [9]

However, the number of required reference frames not only depends on the frame rate of the source video (e.g. 10 Hz or 25 Hz video), but is also content dependent, Figure 8. For example, in the Stephen (a tennis player) and Mother&Daughter (a head and shoulders pictures with little movement of eyes and lips) sequences, due to repetition of events, longer frame references are useful. In these 10 Hz sequences, predictions can go back up to 6 seconds! At the other extreme, when there is no repetition in the scene, such as in the Container (a boat moving along a river) sequence, longer duration frame reference is not helpful. Thus, in some TV programmes when shots alternate between two scenes, extended memory prediction can reduce the peak rate significantly.

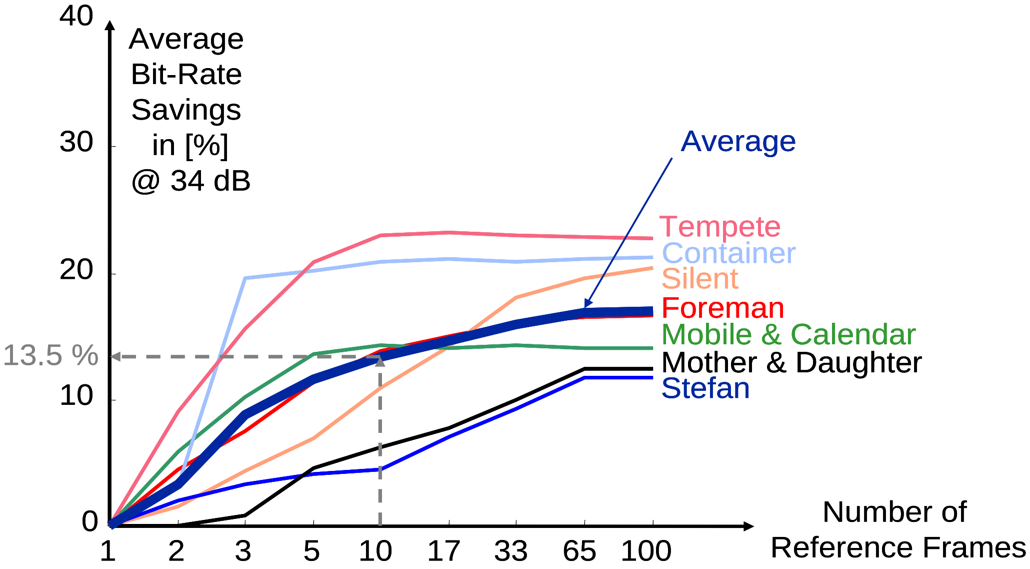


Figure 8: Effect of long-term memory prediction for various sequences after [9]

1.4 Transform coding: (DCT versus Wavelet)

Both Inter- and Intra-frames are transformed from the time-domain into the frequency domain to de-correlate their energy. Inter-frames (B- and P-) incorporate motion estimation/compensation, whereas Intra-frames (I-frames) do not and, hence, are more suited to act as anchor frames in case of transmission error. It is a well established principle that orthogonal transforms have the best energy compaction property. The idea is that, in the transform domain, the energy is concentrated among a few coefficients only, as most of the energy in a natural scene is concentrated in the low spatial frequencies. This allows only the significant energy coefficients to be quantised and coded, while discarding the coefficients with insignificant energy. Starting from JPEG and H.261, almost all video coders are based on an 8×8 DCT. The advantage of the DCT is that it results in real-valued DCT coefficients, and has performance close to the optimal transform (*i.e.* the Karhunen-Loève Transform or KLT). The importance of the DCT can be judged by the fact that during the call for proposals of H.261 (in 1991), there were 14 DCT-based proposals and only one Vector Quantisation (VQ) based proposal. Since then all video coders, including H.263, MPEG-1/2/4, have employed an 8×8 DCT. However, a new 4×4 Integer DCT (IDCT) with the option of an 8×8 IDCT is incorporated into the H.264/AVC standard. The IDCT provides an exact match inverse transform (no rounding errors or what is called DCT-mismatch), which is important when different devices at encoder and decoder are involved. An additional advantage of the IDCT over a conventional 8×8 floating point DCT is the reduced hardware complexity, as an IDCT can be implemented with add/subtract and shift registers only (no multiplier). The 4×4 block size generates less ringing noise as compared to 8×8 block size. The H.264/AVC also includes a secondary Hadamard (orthogonal) transform for DC coefficients (one in each MB) to gain more compression in smooth regions.

In the recent past, due to the arrival of efficient quantisation algorithms and associated coders, wavelet-based video encoders have attracted considerable attention amongst researchers. Due to the wavelet transform’s better energy clustering and because it is a frame-based transform rather than block-based, the wavelet transform was selected in JPEG2000 and for intra-frame coding (optional) of MPEG-4 (though JPEG2000 also has a block-based wavelet version). It is observed that for still images, the Discrete Wavelet Transform (DWT) outperforms a DCT by an average 1 dB [10]. Furthermore, due to bit-plane processing the DWT is naturally suited to fine-grained, progressive scalability. One of the advantages of a wavelet based video codec is that scalability is a natural part of the codec, in that the coding gain increases at higher scalability levels, whereas in DCT-based codecs, larger levels of scalability reduce its compression efficiency. With increased demand for scalability in video coding, wavelet-based video coding has come to prominence and will be part of future codecs (e.g. use of Motion Compensated Temporal Filtering (MCTF) in H.264, see later). In the recent call for proposals of a Scalable Video Coding (SVC) extension of H.264/AVC, a total of 14 proposals were submitted, out of which 12 were wavelet-based and only two considered a scalable extension of the existing hybrid structure of H.264.

H.264/AVC also introduces the concept of predictive intra coding. Coding relies on the fact that neighbourhood prediction of block pixels (spatial prediction) improves the coding efficiency. In fact, the intra-predicted DCT has even better coding efficiency than the wavelet transform. As shown in Figure 9, the performance of Intra-only H.264 with spatial prediction has 1-1.5 dB better performance than Motion JPEG2000 (M-JPEG-2002) and up to 0.8 dB better than an efficient zerotree-based wavelet video coder, known as partitioning, aggregation and conditional coding (PACC) [11]. It is not surprising that Intra-predicted DCT is to replace JPEG-2000 in the future digital cameras. This Figure also shows the importance of inter-frame coding over intra-frame coding to reduce the bit-rate drastically.

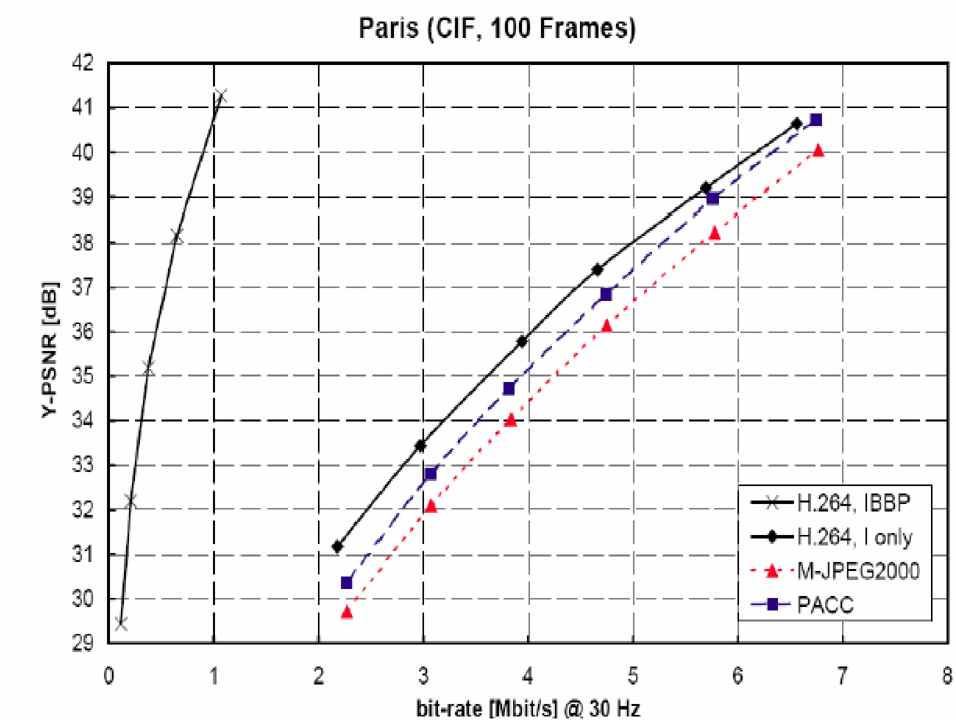


Figure 9: Wavelet versus Intra-predicted DCT for the Paris sequence, after [11].

1.5 Statistical coding

As stated earlier, entropy coding is an integral part of any image/video coder. In image/video coding, after zig-zag scanning of DCT coefficients, the 2D or 3D non-binary symbols are entropy coded using either Huffman or arithmetic coding to generate binary bit-streams. Early codecs mainly employed static Huffman coding (with a predefined LUT). Huffman coders are optimal if and only if symbol probabilities are integral powers of a half, otherwise arithmetic coding has better coding efficiency than Huffman coding. Arithmetic coding may assign a fractional number of bits per symbol, whereas Huffman coding always assign an integer number of bits/symbol, with consequent inefficiency. More recent versions of video codecs such as H.263+ and H.264/AVC have the option of both Huffman and arithmetic coding. For the H.263 codec, its performance with arithmetic coding is 5-8% better than that with Huffman coding. The performance of entropy coding can further be improved by means of context-based adaptive probability models. The new coder for H.264/AVC specifies two types of entropy coding [12][13]. One type is Context-Adaptive Variable Length Coding (CAVLC), which is, in fact, an adaptive version of Huffman coding in which multiple VLC tables are used. The coder adaptively switches between tables, depending on already transmitted syntax elements. It is reported that this version of adaptive Huffman coding has 5-10% better efficiency than single table VLC. The version of arithmetic coding used in H.264 is known as Context-Adaptive Binary Arithmetic Coding (CABAC) [14] and [2], which reduces the bit-rate by 5-15% more than CAVLC, with a range of acceptable video quality from 30-38 dB, as shown in Figure 10. However, a CABAC coder is highly complex [15]. Compared to conventional static Huffman coding, CABAC has a gain of 10-20% in coding efficiency.

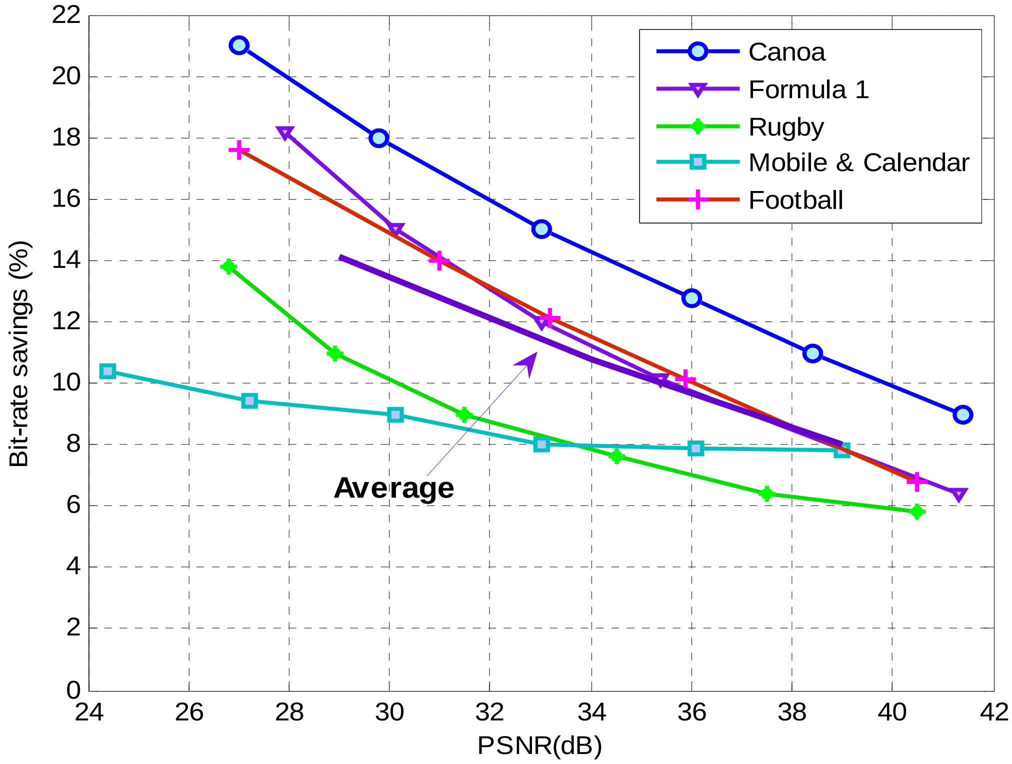


Figure 10: Bit-rate saving provided by CABAC relative to CAVLC in the H.264/AVC codec, for various sequences, after [15].

1.6 *Summary of Section 1*

Video codecs take advantage of similarities among the pixels within a picture (intraframe) and/or between the pictures (interframe) and among the generated symbols (statistical) for bit rate reduction. The degree of compression desired of a particular codec depends on the desired picture quality. The target picture quality for most low bit rate applications is around 34 dB and that of higher quality broadcast television approaches 39 dB.

Interframe coding is the most significant part of the compression chain, whereby through prediction of the motion of objects between two consecutive frames the bit rate can be reduced by about 40%. When the prediction accuracy of objects positions is raised to half or quarter of a pixel, despite the larger addressing overhead, a further 10-15% reduction of the bit rate is achieved. Lastly, by adaptively changing the motion compensated block size, a further 10% reduction in bit rates can be achieved. Temporal prediction can be extended up to 10-15 frames. Depending on the type of motion in scenes, a 10-15% further reduction in bit rate is achievable by temporal prediction. Apart from coding the residual of interframe motion compensated pixels, some parts of the picture are better coded in intraframe mode. In that case, use of intra prediction can reduce the bit/picture by 10-20%. However, since purely intra coded pictures are rare, and in most applications only 8% (e.g. one in 12 frames) of pictures are intra coded, the overall effect on compression due to spatially predicted intra pictures is only a reduction of the bit rate by 1-2%.

Finally, through variable length coding (VLC) of already compressed data, additional bits are saved. The conventional method of VLC is by means of an Huffman coder, but when such a coder is replaced by arithmetic coding, depending on the arithmetic coder type, a 5-10 % further reduction in bit rate is achieved. When all these compression tools are implemented, then bit rate can be reduced my more than 80%, as shown in Figure 4.

2. Relative efficiency and computational complexity of MPEG-2, MPEG-4, and H.264

MPEG-2 is still the most successful and most widely used video coding standard for broadcast applications. MPEG-4 Part-2 has better compression efficiency but is much more complex and has a multiplicity of profiles and levels, some of which, as stated earlier, may never be implemented. H.264/AVC has much better coding efficiency (twice that of MPEG-2) and is targeted to take over MPEG-2 in the near future. The higher efficiency of H.264/AVC (defined in terms of bit-rate reduction while maintaining the same subjective picture quality) is paid for in terms of increased complexity in both the encoder and the decoder. Table 2 shows the increased complexity of H.264/AVC decoder and corresponding coding gains over MPEG-2 [15] Although, the H.264/AVC encoder is approximately 8-10 times more complex than MPEG-2 encoder, computing power has increased by a factor of 100 since the MPEG-2 was initiated.

A profile is a point of conformance with a standard, allowing interoperability between similar applications. While MPEG-4 includes at least 19 profiles, the number of these has been sharply reduced in H.264/AVC. The H.264 profiles [13] in Table 2 are further explained as follows. The Baseline profile supports all features in H.264 except two sets of features: Set 1 includes bi-predictive pictures, CABAC entropic coding (adaptive arithmetic), and interlaced fields; and Set 2 includes support for server swapping and an extra error resilience feature (to protect data on wireless links). Main profile includes all of the features of Set 1 but does not include two other error resilience features that are supported by the Baseline profile. The Extended profile includes all features of the Baseline profile, plus the two sets not supported by the Baseline profile except CABAC. The significance of CABAC is that it is computational intensive with limited coding gain compared to CAVLC coding (adaptive variable length coding) and, therefore, CAVLC may be preferred for real-time applications or mobile devices. There are three variants of the High profile, the principal features added to the Main profile being more flexible (non-linear) quantization. Two of the variants (High 10 and High 4:2:2) support denser sampling of the chrominance signals and allow 10 bits per sample. A further variant, High 4:4:4, as the name suggest, supports equal density sampling of luminance and chrominance, as well as 12 bits per sample. Regions of interest can also be coded without information loss. However, High 4:4:4 is being re-developed at the time of writing, implying that the existing version will be replaced (i.e. is deprecated).

As, from Table 2, the Baseline profile has reduced computation it is more suited to low latency applications such as teleconferences or video-phones, conversational services. The Main profile has moderate latency (0.5 to 2 s) and is suited to entertainment services such as broadcast television in all its forms, DVD and video-on-demand. Streaming services with more than 2 s latency and lower bit rates (less than 1.5 Mb/s) if communicated over the Internet (for example IPTV) may choose the Extended Profile. If a wireless link is involved then the Baseline profile should be preferred. As the name suggests, the High profile is intended for HDTV including storage on DVD and Blu-Ray optical storage discs.

Table-2: Efficiency/Complexity of H.264/AVC with respect to MPEG-2.

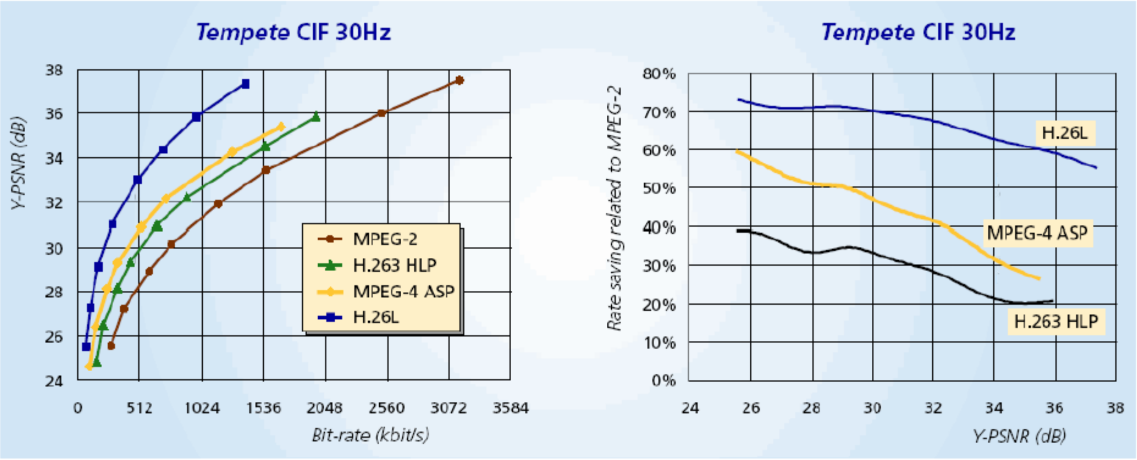
Profile	Efficiency with respect to MPEG-2	Increase in decoder complexity
Baseline	About 1.5 times	About 2.5 times
Main	About 1.75 times	About 3.5 times
Extended	About 2 times	About 4 times
High	About 2.25 times	About 5 times

Table 3 compares the bit-rate reduction of H.264/AVC and MPEG-4 Part-2 with reference to MPEG-2 for various picture resolutions. The performance gain of H.264 is apparent over MPEG-4. It can also be observed that at higher resolutions (HDTV and Standard Definition, SDTV), the percentage gains over MPEG-2 are less than at lower resolutions (CIF and QCIF) [16][17].

Table-3: Saving in bits of H.264/AVC and MPEG-4 Part-2 over MPEG-2.

Resolution	Bit rate saving in H.264/AVC compared to MPEG-2	Bit rate reduction in MPEG-4 Part-2 compared to MPEG-2
HDTV	20-40%	10-15%
SDTV	40-50%	15-20%
CIF	50-60%	20-30%
QCIF	50-70%	30-50%

The rate-distortion profiles of various standard video codecs (MPEG-2, MPEG-4, H.263, H.264) are compared in Figure 11, along with the corresponding bit-rate reductions over MPEG-2. In the Figure, the H.26L codec, the early version of H.264 is used (L for long term codec). It can be seen from this Figure that H.264/AVC has much improved performance compared to other codecs. This is because of the various factors previously described. In comparing the bit-rate reduction, it can be seen that for an acceptable quality of 34 dB, H.263, MPEG-4 and H.264/AVC have gains of 20%, 30%, and 60% respectively over MPEG-2 [15]. Again, these figures should be properly interpreted. For example at 34 dB quality, H.264 (H.26L) requires less than 40% of the bit rate of MPEG-2, or it is more than 2.5 times efficient than MPEG-2.



PSNR of Y-Frames Vs. Bit-rate

Rate Saving Related to MPEG-2

Figure 11: Comparisons of MPEG-2, MPEG-4, H.263 and H.264/AVC in terms of rate-distortion and percentage bit-rate savings over MPEG-2, after [15].

In comparing the compression efficiency (coding gain) of codecs, the target bit rate and quality are important factors. Figure 11 may imply that as the bit rate and desired video quality increases, the gap between H.264 and MPEG-2 widens. This is true at lower bit rates, since at very low bit rates, the various compression overheads, such as addressing motion vectors, MB types, etc. comprise a larger portion of the bit rate, while little is left for efficient coding of the transform coefficients, to improve picture quality. As the bit rate increases, these coefficients are given the chance to be coded, and, hence, the rate of improvement for these types of codec is faster. However, at very large bit rates or higher quality video this growth rate is not sustained. Unfortunately, data for high quality video in the literature is not available to show this. This may indicate a need for further research in our laboratory, to look at the relative performance of high quality video under various codecs. What we could find in the literature is shown in Figure 12, where highly active *Entertainment* and very quiet *News* video sequences are coded with MPEG-2 and H.264 codecs [15].

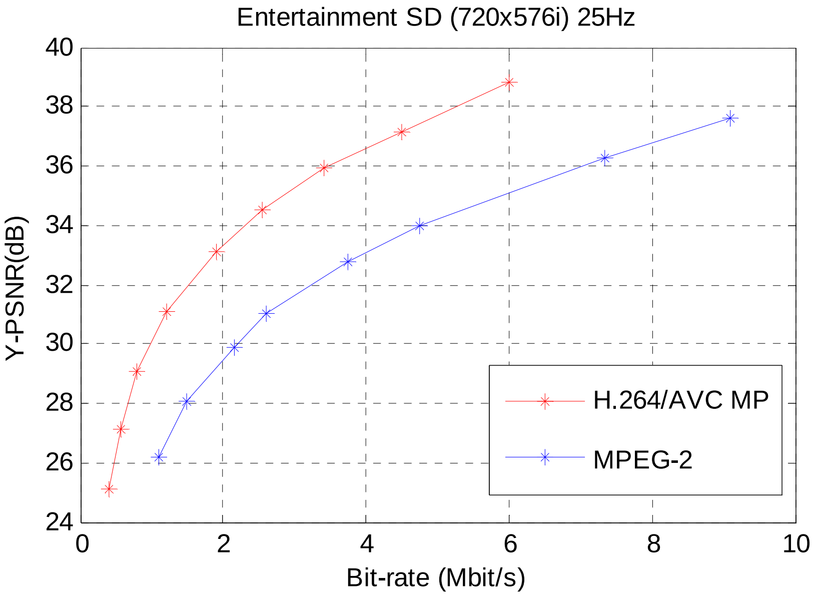


Figure 12-a: Quality of highly active video, after [15].

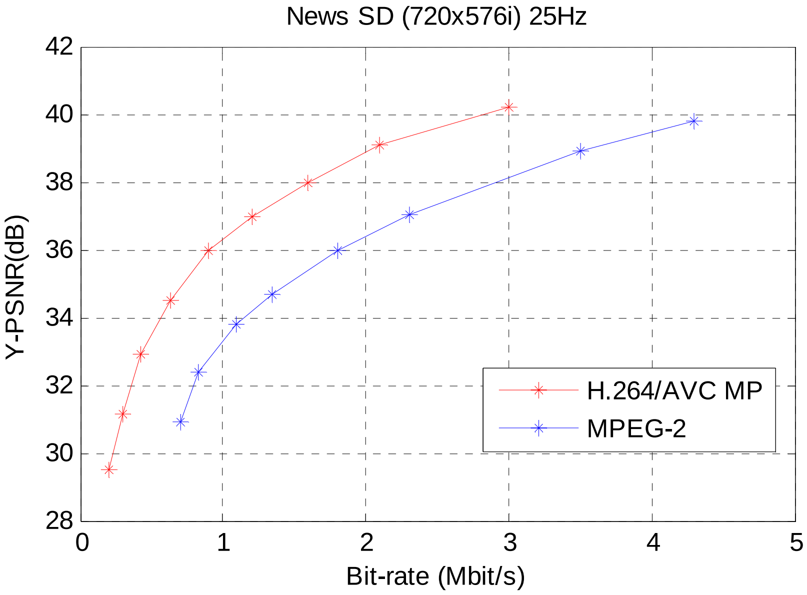


Figure 12-b: Quality of less active video, after [15].

Comparison of Figures 12-a and 12-b reveals two important findings:

1) The difficult *Entertainment* sequence that cannot be coded at quality better than 38 dB at a rate of almost 9 Mbit/s under MPEG-2, is coded almost at half of its MPEG-2 rate (4.5 Mbit/s) with H.264. This is almost a 100% improvement in coding gain. On the other hand, a soft sequence like *News*, which can be easily coded with MPEG-2 at almost 4.5 Mbit/s, and still gives an excellent quality of almost 40 dB, when coded by H.264, requires almost 3 Mbit/s for the same quality. Thus for higher quality coding, the coding gain drops to 50%. It should be noted that, though this difference looks as if it is content dependent, if this sequence is compared at a quality of 38 dB, we see that the relative required bit rates of MPEG-2 and H.264 are reduced to 3 and 1.5 Mbit/s respectively. That is the coding gain for this relatively lower quality becomes again 100%. Thus, it can be concluded that at very high quality video, the coding gain of H.264 over MPEG-2 drops (as can also be understood from Table 3).

2) The other message of this Figure is that, if the quality expectation of future video is to increase (e.g. above 40 dB), which is very likely, then not only the existing 8-10 Mbit/s MPEG-2 video of Fig 12a will be unacceptable (we may need 15 Mbit/s or more) but also the bit rate of future broadcasts under H.264 may just become comparable to the existing 8-10 Mbit/s. Nevertheless for applications involving statistical multiplexing, such as satellite TV, Digital Terrestrial Transmission (DTT), IPTV, the larger bit rate requirement of hard sequences can be compensated by the lower bit rates of soft sequences. Experiments indicate [1] that when video is coded at a fixed quantiser step size (almost constant quality), the peak bit rate of a difficult picture frame may become more than 10 times the mean bit rate. However, when sufficient number of videos (e.g. around 10) are statistically multiplexed, the overall rate is typically only 1.2 times the long term mean bit rate. Higher bandwidth per channel is required for a fewer number of multiplexed videos. Hence, an average rate of 3-6 Mbit/s for the current MPEG-2 video of SDTV quality is a realistic figure and that of the future H.264 at a rate of 1.5-3 Mbit/s is not far from reality.

However, the reason that the relative compression efficiency of H.264 over MPEG-2 is reduced at higher video quality is that, in H.264, due to better prediction of motion compensated blocks, the residues are so small that they do not need be coded by the DCT.

Hence, most of the H.264 blocks are skipped or are just represented by MVs. One of the key components of the compression efficiency of H.264 lies in efficient addressing of these kinds of blocks. Of course, this introduces a small error, which is acceptable at low quality video, but surely is not acceptable for high quality video. For high quality video, if the small residue, no matter how small it may be, must be coded, then the coding gain of H.264 diminishes. The higher the quality (less compression distortion) the smaller will be the coding gain. Ironically, for loss less video H.264, due to larger addressing overhead, can even have higher bit rates than MPEG-2! However, Intra-coded macroblocks in H.264 due to spatial prediction are still better coded than in MPEG-2.

For HDTV video the same reasoning as for SDTV applies, but expectations may be for an even higher quality. Figure 12-c shows how a difficult picture at HDTV resolution is coded by the two codecs. Again, we need much higher quality than 38 dB, which increases the required bit rate of MPEG-2 to around 30-40 Mbit/s and even the future H.264 codec may have to code this type of video at above 20 Mbit/s. Subjective tests [45] confirm the relative weaker performance of H.264/AVC when encoding higher quality video, though it should be noted that Figure 12 and the subjective tests were compiled at an early stage in H.264/AVC's development, whereas MPEG-2 implementations are much more mature. Equally, for real-time video processing, it may not be possible to tune the various H.264 options to the video sequence, as is possible for tests.

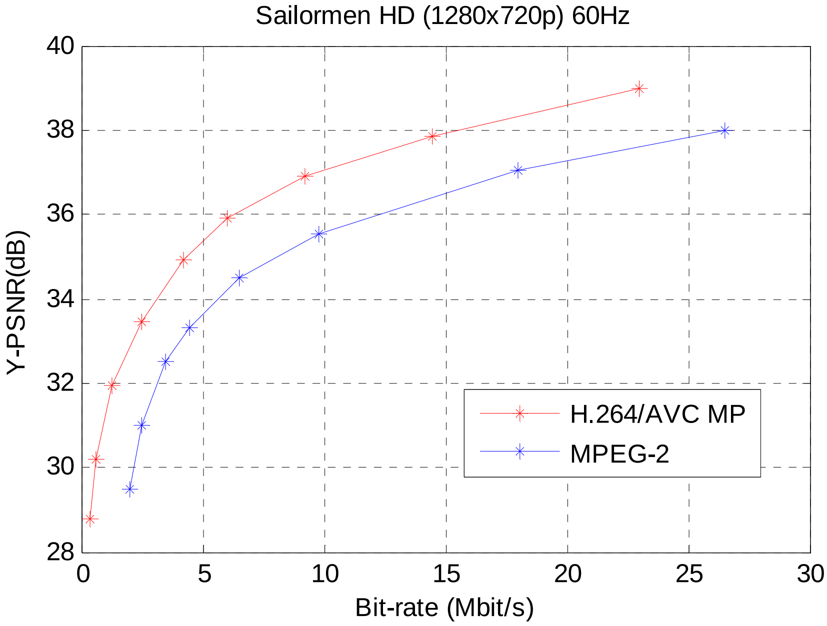


Figure 12-c: Quality of highly active video of HDTV resolution, after [15].

2.1 Displays and Viewer Expectations

It is likely that improvements in display technology will drive up viewers' expectations as viewers' become more aware of their capabilities. Display technology is moving towards Liquid Crystal Display (LCD), Plasma Display Panel (PDP) and projection technology. PDPs mask impairments less well compared to traditional Cathode Ray Tube (CRT) displays in televisions and monitors, and indeed may magnifier impairments [51]. PDPs have increased sizes over CRT, from 32" diagonal size and above. The resolution of the largest type, ALIS panels, is as much as 1024 × 1024 pixels, though the WideXGA (1280 × 720 pixels) format may be more suitable for 720p HDTV. LCDs are difficult to make beyond 25" and motion blur has been a problem, but the relative merits of rival technologies are not static [52]. The availability of second generation DVD storage such as the Blu-Ray Disc, Advanced Optical Disc from Toshiba, and HD9 discs may also raise expectations.

Subjective testing by the BBC, RAI Turin, and Sveriges Television (SVT) in 2002 concluded that for SDTV on WideVGA (725 × 480 pixels) at viewing distances of 4H and 6H (i.e. 4 or 6 × screen height) for moderately difficult to compression sequences, an MPEG-2 rate of 10 Mbit/s would be necessary, as opposed to 6 Mbit/s for an CRT display. (Viewing distances may actually be closer than CRTs for flat-panel displays.) 2003 tests by the BBC [53] compared MPEG-2 at 10 Mbit/s with H.264 at 7.3 Mbit/s, as H.264 was better at this rate than MPEG-2. The tests suggested that the types of artefacts arising from MPEG-2 especially blockiness and ringing were different to those arising from H.264, which retains edges but tends to remove textures, the lower the bit rate. Certainly, were MPEG-2 to be retained then current European rates of 2-5 Mbit/s will be insufficient as PDPs become widely disseminated. Swiss broadcasting companies have already (2003) agreed to limit the number of programmes to three that can be broadcast in a multiplex, rather than a typical five or six (as does the BBC for a 24 Mbit/s Digital Terrestial TV (DTT) multiplex and four for the FreeView 18 Mbit/s DTT multiplex). For HDTV, using MPEG-2 rates of 15-22 Mbit/s are required.

Interestingly, public broadcast companies such as the BBC may include material [54] such as children’s and gardening programmes that are more difficult to compress than sports programmes favoured by commercial broadcasters.

2.2 *Summary of Section2:*

In the past decade, there have been significant advances in the compression of video signals. While 10 years ago broadcast quality video under MPEG-2 could be coded at 6-8 Mbit/s, 5 years later, with MPEG-4, the bit rate could be reduced by 10-50 %, with higher reductions for lower quality video. Now a decade later, with new MPEG-4 (H.264) this bit rate can be reduced to 20-70%. That is for lower quality video, today’s video codecs can be three times more efficient than those of 10 years ago. However, for the higher quality of HD and SD broadcast quality video, this saving is about 30% and 50% respectively. The reduced saving for higher quality video is due to both less tolerance by viewers of compression distortion and the higher fidelity of newer video decoders and displays. Lower compression distortion (higher quality) video reduces the compression efficiency of address efficient coding techniques, like H.264.

3. Trends and tendencies of future video codecs

3.1 McCann Law versus Moore's Law

Unlike the computer industry, broadcasters do not benefit from the effects of Moore's Law, except when they change technologies (e.g. analogue to digital). Indeed, without these exponential changes, the transition to digital broadcasting would simply not be possible. One of the reasons for this phenomenon is the great emphasis put by broadcasters on the compatibility with equipment used by consumers. For example, existing receivers were not made obsolete by the introduction of FM stereo, Radio Data System (RDS), colour TV or stereo sound for TV. Broadcasters should be praised for protecting the interests of the public, whereas the computer industry pays minimal attention to the principle of assured compatibility with previous equipment. These contrasting attitudes demonstrate that broadcasters cannot hope to match the speed of developments in the computer industry.

Recently McCann [17] has investigated the evolution in video coding technology, with the aim of predicting the future performance of video codecs. His own law (McCann's Law) has a significant deviation from that of Moore's Law. In Figure 13 the required bit-rates for compression of video under various codecs are shown. The interesting feature of the Figure is enhancement of the coding gain of a particular codec over time. The green line tracking MPEG-2 shows that continual improvements to its compression tools from 1995 until 2004 have resulted in a reduction in bit-rate from about 6 Mb/s to around 2 Mb/s for the same quality video. At the time of introducing H.263, around 1997, there was a sudden decrease in the bit-rate requirements (indicated by the downward black arrow). This coincided with the progress made with MPEG-4 (yellow line). In 2001, H.263 was 15-20% more efficient than the evolved MPEG-2. However, H.263 and MPEG-4 continued to evolve, with H.263 spawning H.26L (red plot), which eventually transformed into H.264. In 2001, for the same quality video, H.264 was about 50% more efficient than MPEG-2. In summary, McCann estimates bit rate reduction over the last decade improves by almost 15% per year or halves its rate every 5 years.

However, though McCann looks at the compression efficiency of video codecs in the past decade, he does not take into account backward compatibility nor puts too much emphasis on the desired quality of future video. The need for backward compatibility must surely be one of the reasons why Moore's law type rates have not been evidenced in the codec world. Consumers expect an economical life time for each generation of video codecs and do not simply abandon products.

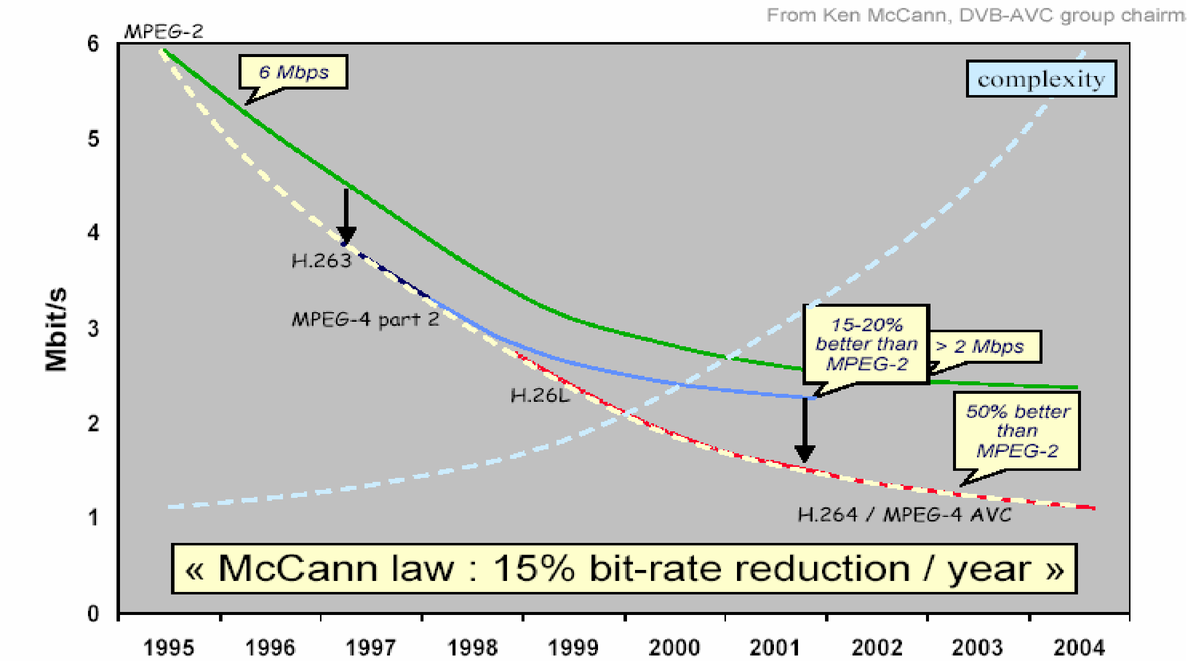


Figure 13: Predicted trends in compression performance over the last decode, after [17].

We believe that the actual rate of improvement is even less than this, even perhaps half of it, *e.g.* 7%, which may be called Ghanbari's Law. There are numerous reasons to believe that the rate of improvement is slower than what McCann has predicted. In fact, McCann in his January 2006 report [18] makes a plea to broadcasters that, "though based on my prediction the bit rate of good quality HDTV will be about 8-10 Mbit/s, for 1080i and 6-8 Mbit/s for 720p formats, broadcasters should avoid the risk of giving HDTV a bad name by prematurely cutting the bit rate". He suggests allocating 15 Mbit/s to such services, rather than those quoted above, i.e. 1.5 to 2 times those figures.

The main driver underlying the slower rate of bit rate improvement is that digital television viewers become more aware of distortions and consequently can no longer tolerate the current level of coding distortions. For example, Table 3 shows that bit rate saving in going from MPEG-2 to MPEG-4 and H.264 (new MPEG-4) is much higher at the lower quality of video of smaller pictures (QCIF and CIF) than the high quality and resolution of SDTV and HDTV pictures. This is despite the fact that larger resolution pictures exhibit larger correlations among the pixels and are easier to code, as the acceptable levels of distortion do not allow higher compression gains. In the previous Section, we also discussed why, at higher video quality, coding gain is reduced. Moreover, higher quality video displays (LCD and plasma) demand even higher quality video. This becomes particularly noticeable when future multimedia displays will display video and text side by side. For example while SD encoded video with MPEG-2 at 2.5 – 5 Mbit/s on current CRT displays gives almost acceptable quality, on flat display panels they should be coded at 8-10 Mbit/s [19]. In particular, as digital modulation is able to broadcast luminance and chrominance (YUV) signals separately, there results a display of higher quality video, in a manner that is simply not possible by analogue transmission. This is because, in composite signals like PAL, through a low pass filter, the luminance and chrominance signals are separated and the effective bandwidth of the luminance is reduced from 5.5 MHz to almost 3 MHz, while in digital TV with separation of colour components, the luminance bandwidth remains at 5.5 MHz. With digital video, it is also possible to increase the chrominance bandwidth, more than what is offered now under analogue transmission. On the quality issue, David Wood [16] also believes that the long term gain for grade 4.5 pictures (picture quality with mean opinion score of 4.5) of professional broadcast quality is about 5-10 % per year.

3.2 *Russian steps*

In predicting the future performance of video codecs, David Wood [16] has taken an interesting approach, which pays greater attention to the compatibility of video codecs. Similarly to McCann's approach of looking at the performance of past video codecs over the time, he assumes a life time of around 5-8 years for each codec, whereas after this period a new codec is introduced, as shown in Figure 14. The interesting point in his approach is that he assumes every codec's performance during its life time is improved steadily (similarly to McCann) but a new codec at the time of introduction is better than the old one. However, an improved version of an old codec in the future can be better than the new non-refined codec. This is very much in line with reality. We believe this is the right approach for prediction of future performance of video codecs and we define, two parameters, rate of improvement and Leap Gain, for our predictions. In fact, if one can estimate the rate of improvement, Leap Gain, and the lifetime, then it is possible to arrive at a crude estimate of the likely future performance of video codecs.

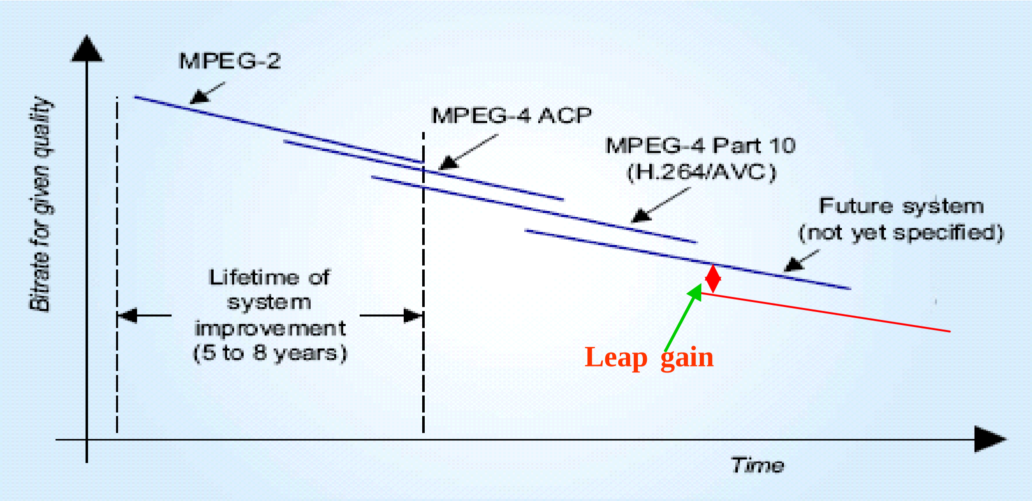


Figure 14: The evolution of standard video codecs, after [16].

3.3 *Leap Gain*

As all standard codecs are DCT-based they appear similar to each other, but there are differences that make a new codec incompatible with an older one. For example, the main difference between MPEG-1 and MPEG-2 is that, the former is used for progressive video but the latter for interlaced video. This small difference may have significant implications, even on the compression efficiency. In this example, for MPEG-1 with progressive video, a macroblock (MB) has 16×16 pixels, while, in interlaced video, the MBs might be taken from the previous odd or even fields (16×8 pixels) or even from the past frames, as in progressive mode. Unfortunately also, the VLC tables are defined differently between MPEG-1 and MPEG-2, which makes them not entirely compatible. In short, the lack of compatibility between seemingly similar codecs and the need to avoid replacement of consumer equipment at a rate faster than economically sustainable results in a distinct lag in compression rate improvements in the standard codecs. However, when the pressure of new innovations becomes over-whelming and the mass market is ready, then a Leap Gain will take place.

Since at the introduction of a new codec all the lessons learnt from past codecs can be exploited, then naturally a new codec will have a better performance than the old one. This coding gain, or we call it Leap Gain, depends on how much compression efficient tools have been incorporated in a codec. For example in moving from H.261 to MPEG-1, two important compression efficient tools, namely bi-directional motion compensation (B-pictures) and half-pixel motion estimation were incorporated into MPEG-1.

The impact of B-pictures on coding is enormous. For example, in MPEG-1, -2, with M=3, there are two B-pictures among the anchor I- and P-pictures, and B-pictures bit rates are almost 30% less than those of P-pictures.[9] To gauge the compression gain due to these pictures, compared to pure P-picture type coding used in H.261, one has to consider that in H.261 P-pictures are consecutive pictures, close to each other, while in MPEG-1,-2 P-pictures are three frames apart. Hence, P-pictures are more easily coded in H.261 than in MPEG type codecs. Thus, this might reduce the nominal coding gain of 30% to around 10-20%.

Another coding gain is due to the half-pixel accuracy of MPEG-1 over H.261. According to Figure 4, this could be in the order of 10-20 %, depending on the content and the bit rate, as can be seen from the Figure.

Note that multiple of these two coding gains, does not give the overall Leap Gain, since use of I-pictures, which normally consume more than twice the P-pictures bit rates, reduces this gain. However, in MPEG-1,-2 there is normally one I-picture in a Group of Pictures (GOP), eight B-pictures and three P-pictures (e.g. M=3 and N=12). Thus, since the majority of pictures are of B-type, and there is only one I-picture every 12 pictures, this only slightly reduces the coding gain. Thus, overall, it is not going too far to assume that MPEG-1 is about 20-30 percent more efficient than H.261.

It should be noted that, to measure the value of Leap Gain in going from one codec to the next, it is required that the new codec is compared with the best representative of the old one. This sort of information is rarely available in the literature, since it is more than likely that a refined new codec is compared against moderate past codecs. This is because researchers normally publish their new work based on the improvements that they have made to a new codec. Thus, depending on what stage of improvements a codec is, its coding gain is in fact its overall gain (Leap Gain plus its rate of improvement so far, see later comments) rather than its Leap Gain. Of course, when a codec has matured over its lifetime of 5-8 years, then we may define its overall gain as its overall Leap Gain.

The Leap Gain from MPEG-1 to MPEG-2, is only due to use of field pictures in the prediction loop, which are closer to each other than frame pictures. However, this is not that significant and to some extent is compensated by the extra overhead of identifying the type of prediction pictures to the receiver decoder. Thus, one may assume a Leap Gain of around 2-5 % for this codec.

The overall Leap Gain from MPEG-2 to MPEG-4 is much more than this. This is because, numerous compression efficient tools, like quarter pixel accuracy motion estimation, predictive coding of motion vectors, use of arithmetic coding, efficient addressing of coded and non-coded MBs, overlapped motion compensation and numerous other tools (of course as options or annexes) can lead to a substantial Leap Gain, which is around 100%. That is, for the same quality, MPEG-4 is twice more efficient than MPEG-2, or its bit rate is 50% of the MPEG-2 bit rate.

H.263 has almost a comparable performance to MPEG-4 and its overall Leap Gain over H.261 can be more than 130-150%. This, of course, represents the basic H.263, as some of its options can increase the compression gain, and other options that improve the robustness of the codec reduce this coding gain.

[9] This, of course, is picture dependent and since B-pictures are not used in the prediction loop, they can be compressed even more heavily (the introduced distortions are not accumulated).

The overall Leap Gain from H.263 to H.264 (or from old MPEG-4 to new MPEG-4) is also around 100%, which may also be called a 50% saving in bit rates. That is at the same quality, H.264 uses half the bit rate of early versions of H.263. This ratio is much less for the H.263+ and H.26++.

The overall Leap Gains over a 5 year period (in saved bits) in compression performance that have already taken place can be summarized as:

- From MPEG-2 to H.263+ (1995-2000) there was a saving of 20-40%.
- From H.263+ to MPEG-4 (ASP) (1998-2003) there was a saving of 8-10%.
- From MPEG-4 to H.264/AVC (2000-2004/5) there has been a saving of 10-30%.
- Thus, from MPEG-2 to H.264/AVC over a 10-year period (1995-2005) there has been a saving in bits of nearly 40-80% (more at low quality video and less at high quality video).

3.4 Rate of coding gain

When a new codec is introduced, researchers round the world try to improve its compression efficiency. This is almost true for all codecs, and specifically evident in H.263, when through the years of improvement it was called H.263+ and two years later H.263++. The Plus (+) indicates an improved version of the old system. Rate of improvement is associated with major/minor changes within a codec to improve its compression efficiency. This can be as simple as extending the motion estimation range to a fraction of pixel, up to a more sophisticated entropy coding scheme. Although it is not difficult to estimate the compression efficiency gained at each stage of coding gain, it is the amalgamation of all these improvements that finally represents a codec’s performance. This is particularly interesting for software based codecs, when at each stage the improvements can be easily incorporated into the older versions and users can benefit of these rate of improvements.

It is important to note that, rate of improvement is steady and new coding methods can improve compression efficiency and eventually contribute to an overall Leap Gain in performance. Currently several improvements are under study for the H.264 state-of-the-art codec, that, when they are matured, may lead to a new codec, perhaps to be called H.265!.

There follows a more detailed analysis of coding gains. The headline figures are summarized at the end of each Sub-Section.

3.4.1 Lagrangian rate-distortion optimisation

In compression of video, like other signals, increasing the bit rate will reduce coding distortion. However, since compression involves numerous sub-compression elements, and each with its own rate-distortion characteristics, then care should be taken to invest the bits where it has the highest coding gain. This can be achieved via Lagrangian optimizer, in which on the rate-distortion curve one tries to find the steepest rate of distortion decay, for a small increase in bit rate. This optimization is currently implemented in the H.264 codec with a compression gain of around 5-8 %, but the Lagrangian parameters for all coding elements, like macroblock type, DCT coefficients, motion vectors, etc are kept the same. Currently we are investigating how different parameters for different coding elements can further improve compression efficiency.

To summarize, if the codec's component algorithms are tuned to emphasize the component that most reduces bit rate, then there is a coding gain of 5-8%.

3.4.2 Motion Compensation Temporal Filtering (MCTF)

Motion Compensated Temporal Filtering (MCTF) causes an input sequence of pictures to be split into two or more sequences. For example, if just two new sequences are created then this is achieved by taking only even pictures (pictures every even interval in time) to form one sequence and only odd pictures to form the other. Each sequence is then downsampled. A sub-sampled picture is predicted by means of motion estimation from an earlier picture in time. Given a motion vector, a prediction picture is formed (motion compensation) by means of a prediction operator, which adjusts the estimate given by the motion vector by interpolating between matching picture elements. In the case of odd and even sequences, an odd picture is predicted from an even one, as these alternate in time relative to the original input sequence. Prediction pictures subtracted from the original to find a residual picture (i.e. the part that was not correctly predicted by motion estimation, given that objects do not just move between pictures, they also change their orientation). The residual picture can be used recursively to update the picture from which the prediction was formed (the even picture in a two sequence operation). Therefore, in the case of odd and even pictures, the even picture is updated by the residual picture. Again the update is based on an update operator which may use matching picture element interpolation.

The purpose of this process is rather than send the original input sequence, to send downsampled versions of one or more sequences and the residuals needed to recreate other intermediate sequences forward in time from them. Hence, the term "temporal filtering" is included in MCTF. In the twosequence case, the downsampled even pictures are sent together with residual pictures from the odd sequence. Because residual pictures contain less bits and are downsampled anyway, there is a reduction in bit rate. Moreover if the prediction and update operators are inverses of each other, then the process described in the preceding paragraph can be reversed to achieve perfect reconstruction of the original input sequence.

The prediction operator may well be chosen to be a short range interpolation filter, such as a spline or a Haar wavelet filter (see Section 1). Application of a wavelet transform in this manner is called lifting in the technical literature.

One further aspect of the MCTF scheme is that separation into multiple sequences, downsampling, and motion compensation can be recursively applied. Thus, in the two sequence case, these two sequences can be further decomposed themselves, with the two new sequences downsampled, etc in relation to the parent sequence.

In H.264, lifting based MCTF with and without update steps (corresponding to open and closed loop predictions) is selected as the replacement for the existing IBBPBBP…. structure of temporal scalability. Initial results suggest that MCTF extension with 5/3 wavelet filter banks have either better or similar performance compared to H.264/AVC IBBP…… (see Figure 15). It is believed that future video codecs will incorporate MCTF, not only for more efficient temporal scalability, but also to improve compression efficiency. Figure 14 shows that MCTF can improve the compression efficiency of the H.264 video codec by more than 1.5 dB. At the picture quality of this Figure, this is equivalent to 20-25% efficiency in the bit rate.

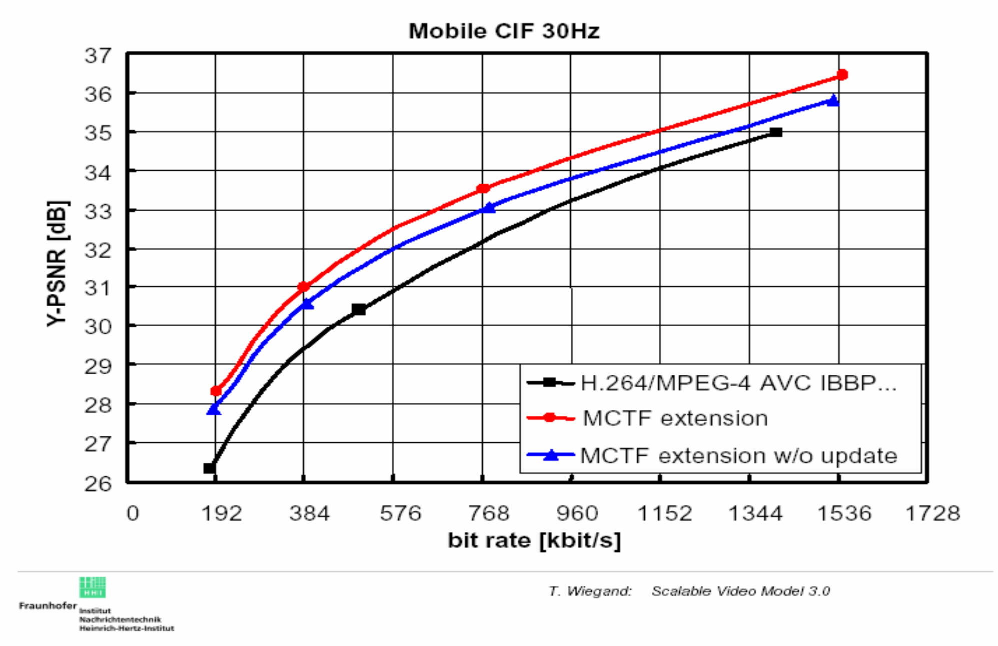


Figure 15: Effect of MCTF on the coding gain of H.264, after [20].

MCTF is in fact a joint temporal decomposition of pictures (separating into two or more sequences with picture members of the sequences taken from pictures alternating in time in the original input sequence) along with motion compensation for better exploitation of inter-frame redundancy. Although normally a 2-tap Haar filter (i.e. taking a weighted selection from just two pictures) is widely used in conventional MCTF schemes [21], recently we have extended the filter length through motion compensated DCT temporal filtering (MCDCT-TF) to take a better advantage of the object linkage (i.e. the presence of the same object in multiple frames) within each group of frames (GOF) [22]. A simple way of achieving this is to make GOF length a multiple of three frames. Figure 16 illustrates how a GOF of 9 frames is decomposed into five temporal bands by 4-stage recursive decomposition of the lower temporal subband frames. It is then divided into three sub-GOFs of 3 frames. MCDCT-TF is applied to each sub-GOF of 3 frames resulting in two low and one high subband frames. After one stage decomposition a new GOF is made of the six low subband frames and the remaining three high subband frames are the first set of generated high band frames. This GOF of 6 frames is then divided into two sub-GOFs of 3 frames, which are recursively decomposed into two sub-GOFs of 2 frames at stage two. MCDCT-TF is then applied to each sub-GOF of 2 frames, generating one low and one high subband frames.

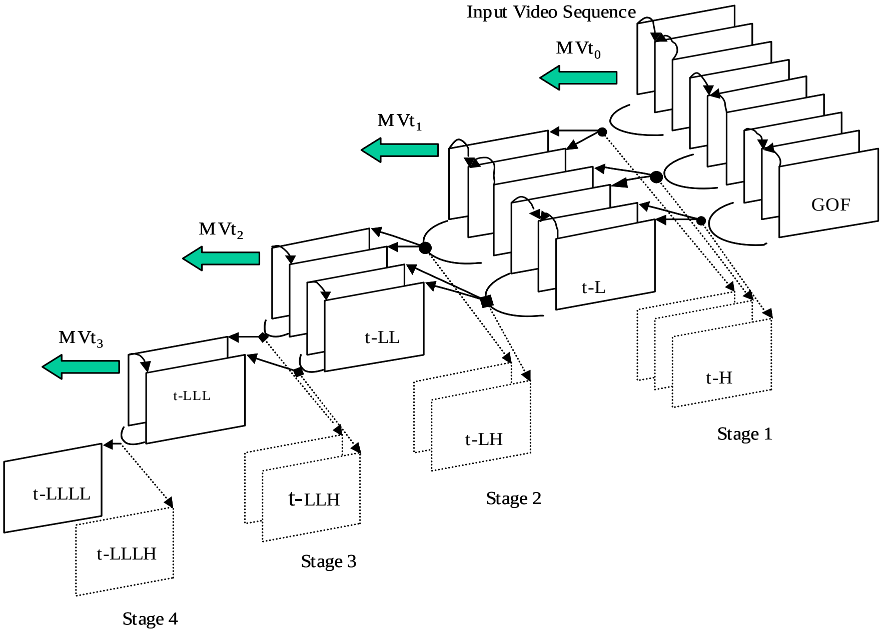


Figure 16: A group of frames with 4 temporal analysis stages

A significant advantage of this method, apart from its superior compression efficiency, is the generation of a scalable bit stream without increasing the overhead, contrary to most scalable coders. The other notable advantage is its very low bit rate base layer, which is again hard to realise with the conventional standard codecs. However, the price paid for these advantages is the long delay (there are 9 frames delay in Figure 16). This, of course, can impose restrictions on applications such as interactivity, editing, pause etc. Its ideal usage would be broadcast of video to users with various degrees of network constraints.

In summary, a 20-25% reduction in bit rate over the existing H.264 codec by a form of layered coding that extends in time.

3.4.3 Use of metadata in coding

It is envisaged that, in the future, broadcast video will be indexed for easy access. In this method, a video program will be divided into segments, and each segment into shots, with the frame as the smallest unit. Indexing video in this form makes it possible to find a desired programme among the several broadcast channels (particularly among the satellite programmes). Video indexing also helps to retrieve video clips on the World Wide Web (WWW), which is the main aim of MPEG-7. The additional data inserted into the video stream is called metadata, which carries all the required information about the content of video.

Metadata indexed video can also improve coding efficiency, since if metadata within the pictures are defined, then these areas can be more efficiently addressed. Figures 17-a and 17-b show the relative improvement in coding gain due to use of two indexing metadata based coding techniques that exploit metadata information in order to improve the efficiency of current hybrid codecs [20]. The first technique uses the MPEG-7 colour layout descriptor and the second technique is based on the MPEG-7 Segment Data Shuffling technique. As the Figure shows the first method results in 3-12% bit-rate savings, and the second one has 3-8% savings over H.264/AVC baseline. These developments are at the early stage of exploring metadata and it is believed that in the future incorporation of more MPEG-7 descriptors will improve the coding gain further.

`

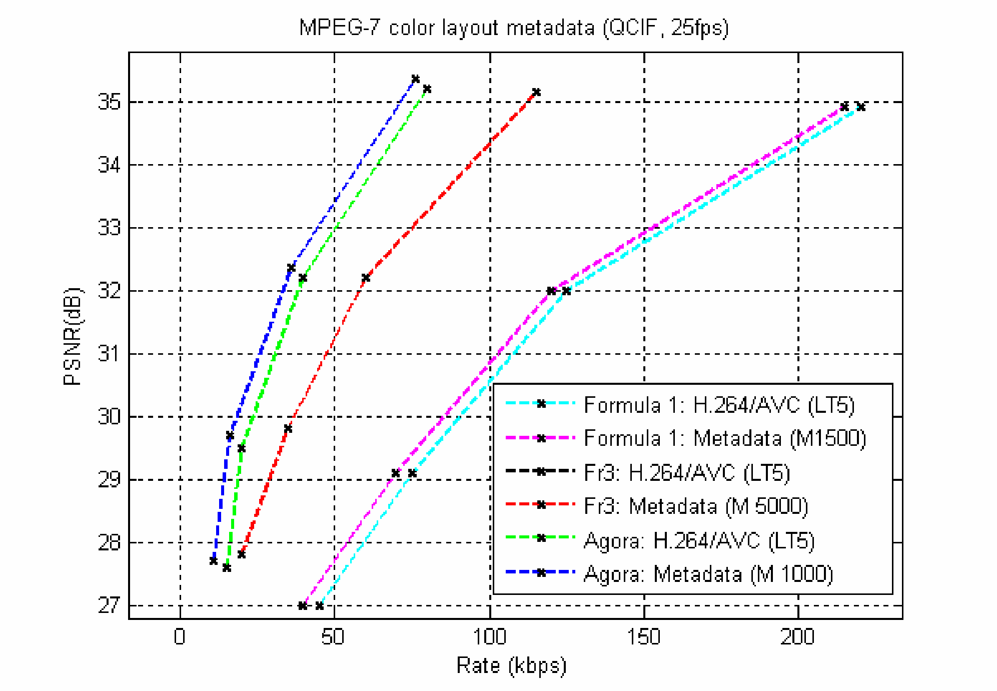


Figure 17-a: Effect of colour metadata on the coding gain of H.264, after [20].

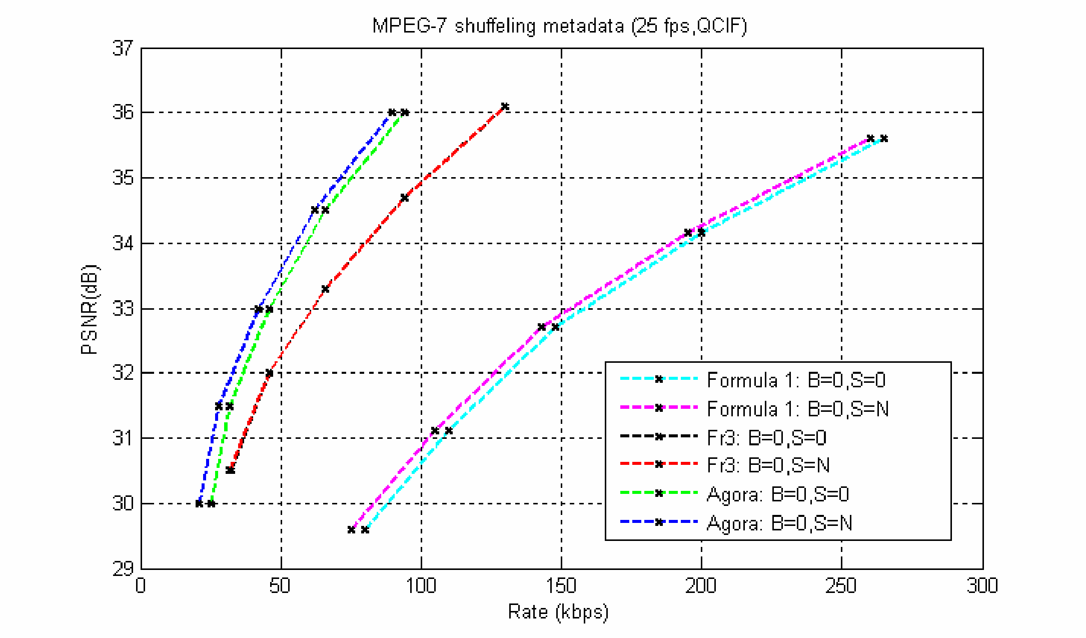


Figure 17-b: Effect of shuffling metadata on the coding gain of H.264, after [20].

In summary, 3-12% bit rate savings may be achieved by sending auxiliary information called metadata.

3.5 Factors adversely affecting compression efficiency

Not all coding tools are for compression purposes and there are other factors that may dissipate these gains. For example the need for stronger Forward Error Control (FEC) in wireless applications, in certain scenarios may double the bit rate. Even simple Group of Blocks or slice headers, for small size images (e.g. Sub-QCIF) can be very costly. In addition, provisions of auxiliary pictures in combining computer generated images with live footage, flexible macroblock ordering, as well as redundant MVs for better error concealment and many more similar tools can increase the bit rate. However, one of the tools that will have greater impact on future video codecs is the scalability or layering technique. Scalability allows users to access the same bit stream at various, spatio-temporal and quality resolutions. In other words, users' end devices may have differing spatial resolutions and/or may only be able to process pictures at a particular frame rate (temporal resolution). Equally, the quality of the received video may be traded against the available bandwidth or capacity. The most likely application of scalability is at the point of distribution to various mobile devices with varying display capabilities and differing wireless interconnects. It is interesting to note that while spatial and quality scalability increases the overall bit rate of the encoder, in fact temporal scalability reduces it [1]. A good example of temporal scalability is the inclusion of B-pictures in the bitstream, which compared to I- and P-pictures reduces the bit rate significantly.

Following on from the pioneering work on layered video coding by Ghanbari in 1989 [23], various types of scalability have been introduced into the standard video codecs, as well as naturally from the wavelet transform and tree-coders employed therein. Spatial scalability refers to the ability to vary the spatial resolution of an image, which is immediately applicable to the various display sizes of portable devices. Moreover, this technique may be applied without the need for transcoding. Signal-to-Noise Ratio (SNR) scalability, also called quality scalability, allows the video quality to be progressively varied alongside spatial scalability. Fine-grained scalability (FGS) was also applied in the MPEG-4 streaming profile [24] to support variable-rate adaptation within the same bit stream. FGS is a type of SNR scalability where each bit plane is coded as a layer. In [25], MCTF in the form of a temporal wavelet transform is applied as temporal scalability. The scalability feature of H.264/SVC is likely to provide all types of scalability in a single bitstream, with slight loss of coding efficiency. In this case, the new temporal scalability under MCTF will compensate for the compression deficiency of other quality and spatial scalabilities. Due to the increased heterogeneity of networks and display devices (from mobile to HDTV), fine scalability is the future solution to provide video services to the customers; in marketing jargon "anytime, anywhere and on any device" along with MCTF for temporal scalability. In Figure 18, JSVM2 refers to the H.264 scalability extension, with the 'SingleLayer' plot showing performance without scalability. The reduction in performance for spatial and SNR (video quality) scalability are shown separately, and then in the 'CombinedScalability' plot, the reduction in performance resulting from combining spatial and SNR scalability in the same bitstream is even greater. Note that this reduction in coding gain can be compensated with the MCTF gain of Figure 15.

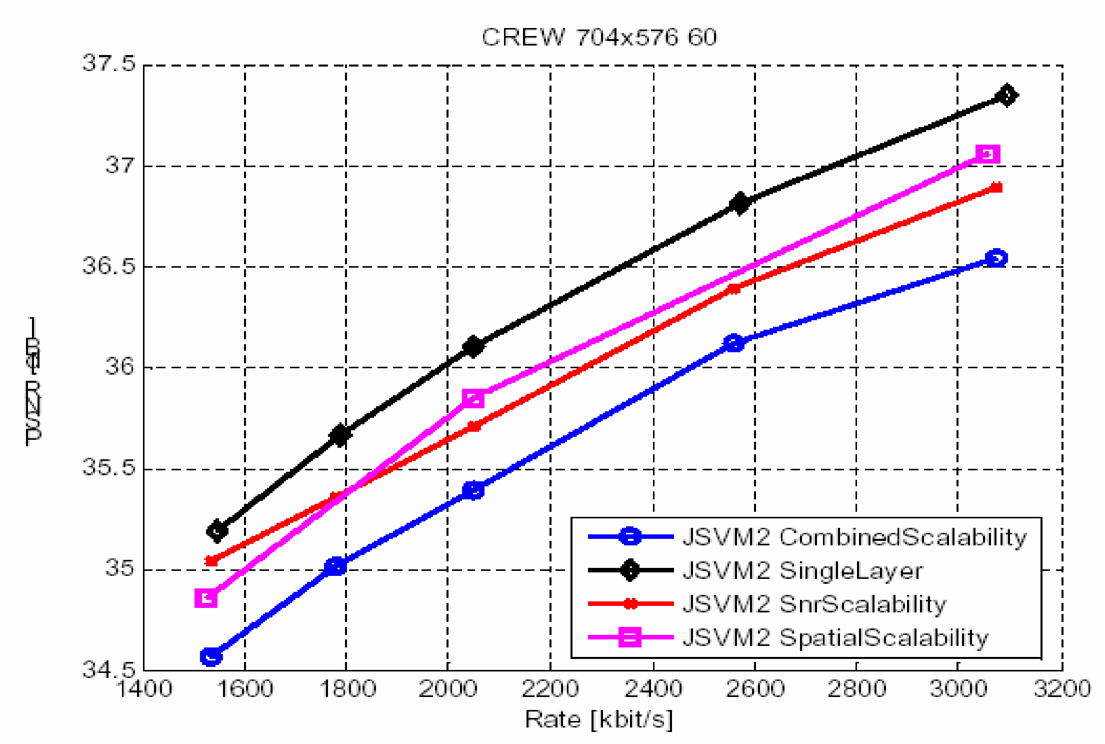


Figure 18: Successive reduction in performance from adding various forms of scalability, after [26].

3.6 *Summary of Section 3*

In the past 15 years, a variety of video codecs have been developed. Although these codecs have been devised for a specific application, they are versatile enough to be used for a wide range of usage. H.264 is a good example, as it can be used at about 20 kbit/s for QCIF-sized video and at about 15 Mbit/s for 16CIF sized HDTV quality video. However, each codec has a life time of 5-8 years, while at the end of each cycle a new codec is introduced. Each codec in its life time goes through numerous enhancements and its coding gain steadily improves, but a new codec at the time of introduction can be better than its predecessor. This codec again undergoes refinement and its compression efficiency increases.

Although currently our best available codec, H.264, is almost 2-3 times better than MPEG-2, efforts to improve its coding gain are an on-going process. Currently, various enhancements in that respect, such as motion compensated temporal filtering, and employing metadata for efficient addressing, will lead to a equivalent improvements for H.264 in the next decade or beyond.

While over a decade, the coding efficiency gain for low quality video can be almost three times, for the higher quality SD and HD video, this value is reduced to almost twice. Due to higher quality expectations for future video and display devices, it is believed by us that a rate of improvement of amounting to a doubling in compression efficiency every decade is a realistic assumption. This is of course far less a rate than Moore's law has predicted over the same period, i.e. a halving in semi-conductor feature size every 18 months --- implying an equivalent growth in available hardware complexity (but see Section 5 for why Moore's law may no longer hold). The main reason for the lag is that broadcast tools need to be backward compatible and have to have a minimum life cycle, which is not the case for many electronic goods following the Moore's law. Therefore, the codec lifecycle resembles that of complex software systems more than it does hardware lifecycles.

Our estimated figure of a doubling in efficiency per decade, or a 7% improvement per year, is also less than that of McCann's estimate of 15% per year for similar applications. Our main reason for advancing a more conservative estimate is the higher quality expectations which will inevitably arise in regard to future video services and display devices. Our estimate is very much in line with that of David Woods' estimate of 5-10% per year.

4. Potential of other codecs to affect the Leap Gain

This report has concentrated on mainstream (standard) macro-block-based codecs (MPEG-1, 2, MPEG-4 Part Part 10, H.263/4), which successively remove spatial, temporal, and statistical redundancy (spectral redundancy is removed prior to the codec by transformation of the colour components). Other principles have been applied but for some of these (fractal coding and vector quantization) over time have been found lacking, at least in terms of video processing. However, some of these codecs do have the potential to upset the predicted compression performance Leap Gains from the macro-block codecs and, hence, they are briefly reviewed next.

It might be imagined that encoders could be combined, for example by encoding high frequency detail, which is harder to code through standard MPEG-4/H.264 HDTV by a different type of encoder. This technique is actually applied in the AAC+ audio codec. However, in audio each frame over time is coded separately, whereas in video there are temporal dependencies that do not make that possible. In fact, SNR scalability already exists in standard codecs through motion JPEG-2000 and the MCTF SVC extension to H.264. It is possible to combine a conventional MPEG-2 encoder with a wavelet encoder by means of virtual coding trees [41].

4.1 Fractal coding

Fractal is best suited to natural scenes rather than computer-generated graphics, as it relies on matching different parts of an image. Natural images are said to be order-one Markov fields, which implies that there is autocorrelation between spatial images in close proximity, though that correlation reduces thereafter. Fractal compression appeared to be innovative and promising in the late 1980s, when it was claimed by Barnsley [27] and associates to outperform JPEG for some images, its main competitor at that time (resulting in an astonishing bit rate of a few hundred bits per picture). Codecs and software libraries are available but fractals have never achieved widespread use, due partly to patenting issues but more significantly due to excessive processing time and lack of quality. Fractal techniques do have an advantage over JPEG still image compression at low image resolutions, but low quality is seldom required. The track record of fractal coding in the video processing domain does not suggest that it has a future.

4.2 Vector quantisation (VQ)

In this technique, an input image is divided into blocks that are arranged as vectors. At the decoder, a look-up-table (LUT) or code book is searched to find a matching pattern for each block. The LUT can be dynamically constructed or is predefined. When dynamic construction occurs there is a marked asymmetry between encoder computation and decoder computation. Compression is achieved through efficient coding of the indices marking the location of the vector within the LUT. Disadvantages emerge as the vector dimension or codebook size increases: searching becomes computationally intense; the memory required to store the codebook becomes large; if the codebook is constructed dynamically it becomes vulnerable to scene changes, while a static codebook is unique to the training regime; and lastly inevitable blocking artefacts arise. Of course, there is also an overhead from communicating the code book. Despite an IEEE Transaction special issue on vector quantization [28], to which this research group contributed [29], there may be room for future improvements. It subsequently emerged that for still image compression, vector quantization was equalled or outperformed by the JPEG standard. In fact in late 1987 vector quantization (VQ) was one the 15 proposals for low bit rate video coding. Though in the subjective tests there was no apparent quality difference between the proposed VQ codec and the other 14 DCT based codecs, eventually

DCT-based codec was chosen as the core codec for H.261 and hence, work on VQ was almost abandoned in the 1990s. There is a possibility that vector quantization may be revisited in the future due to its ability for high quality video (in telemedicine and digital cinema applications, where presently motion JPEG2000 is a leader). Due to its processing asymmetry, it may prove attractive for decoding on portable devices. Vector quantization may become a subsidiary technique [30] within a larger codec. No significant impact is expected from vector quantization within the next five years.

4.3 Model based coding

Model-based coding came to prominence in the 1980s and 90s for videophone applications, and this research group was at the forefront of that research [31]. It was proposed to replace the actual video image with a parameterized computer generated facsimile of it, which could be transmitted at a very low bit-rate, by only communicating motion parameters. Figure 19 shows the technique whereby a wire-framed model (1) is fitted to a video scene (2), which can then be animated by communication of motion parameters (3) and (4). This promised representation of head and shoulders video images at full frame rate for at rates of a few kbit/s. Alas, the human mind is very aware of facial expressions, and is very unforgiving to the types of distortion produced by the early codecs, with their simple LUTs of expressions. Upon realising this, the research community began to progressively increase the complexity, but in doing so lost the original bit rate advantage against the standard block-based coders. This realisation caused the technique to be almost abandoned at the end of the 1990s. All may not be lost though; the research had significant implications for cartoons and personal desktop avatars, and may yet find widespread use in personalising handheld communications, especially combined with laser scanners. Model-based coding may have widespread application in computer animation and as a web browser `friend’. It is unlikely to have an impact within video codecs over the next five years.

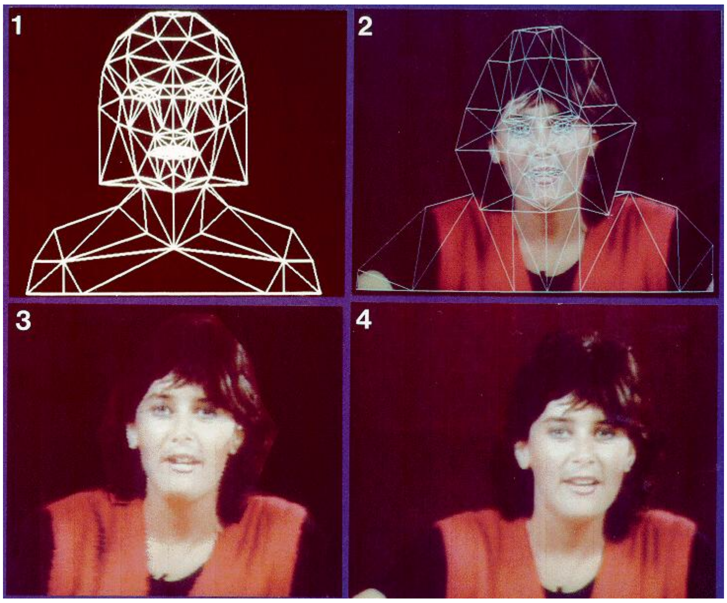


Figure 19: Model based coding employed to animate a head-and-shoulders image.

4.4 *Object based coding*

Object-based coding is closely associated with the MPEG-4 codec [32], which supplies a series of compression tools for manipulation of video objects within a scene. Despite an extensive toolset of coding techniques, video object planes have not been widely adopted. However, there are commercial toolsets that employ sprites (objects subject to perspective projection[10]) within a scene. The advantage of sprites is that they can be reused at later times but that advantage only appears at low video quality and depends on the persistence of the sprite within a video scene. The individual algorithms are complex and the compression rates do not compete with H.264. There is a trend towards coarsening granularity in motion estimation, and it seems likely that object-based motion estimation will emerge as a viable technique in the near future. However, since in object based coding individual objects can be marked as metadata, this can be exploited for compression, as discussed before (see Figure 16).

4.5 *Summary of Section 4*

In the past two decades, along with DCT-based codecs (e.g. the MPEG and H.26x families), a variety of video codecs have been proposed. However, none of them have proved to be a viable competitor to the DCT-based codecs. Nevertheless, the features of these codecs might be employed within DCT codecs to enhance their coding gain. Among these codecs, object-based coding appears to have a better chance of successful implementation than the others, such a vector quantization, model-based, and fractal type codecs.

[10] A perspective projection changes the view of an object to take account of how actual 3-dimensional objects reduce in size as they recede in space. This is the same projection that is applied in many video games, and is the basis of all representational painting since the Renaissance.

5. Potential of hardware issues to affect the Leap Gain

The bit-rate performance of future codecs is based on an ideal premise that future hardware will be able to match computational demands. Complexity analysis studies based on machine instruction counts are likely to be too optimistic by a factor of 3.6 according to a recent case study [33]. The cancellation of Intel's next generation Pentium series processor, because the power consumption was unsupportable at higher clock speeds, is a clear sign that the 18 monthly halving of chip feature sizes predicted by Moore's law have been checked. While there is likely to be a compound annual growth rate (cagr) of 80% [34] in computational demand, for example in digital movie processing, clock speed increase over the next five years is estimated to be 15-20% cagr. In the long term, this implies that complexity issues will gain greater prominence in the design of future codecs. As previously mentioned, the object-based features of MPEG-4 have not been implemented by hardware manufacturers, as these would depart from the two-stage pipeline macro-block processing model that had successfully delivered real-time video processing. It is also the case that real-time video processing of the latest macro-block-based codec, H.264/AVC, for SDTV and HDTV will *not* be possible in software-only implementations, even for decoders. An SDTV/HDTV at 720p encoder requires 2470/3600 giga instructions per second (GIPS) and 3800/5570 giga bytes per second (GBPS) memory access [46], which are far beyond any existing general purpose processors such as those of the Pentium series. The same is true for decoders for HDTV at 720p with 83 GIPS and 70 GBPS required.

Nonetheless, in the short term it is expected that system-on-chip designs, with heterogeneous processors and assorted hardware on a chip will match the immediate needs of H.264/AVC codec HD (1024-interlaced and 720-progressive) video..

In detail, the first SoC chips for the H.264 codec have emerged in the last six months such as the A199 from Ambarbella Inc, aimed at high-end embedded applications such as digital camcorders. The Main and High profiles of H.264 are implemented (both encoder and decoder) with a low-power ARM-9 embedded processor and runs as 216 MHz to encode HD video at 15 Mbit/s dissipating 1 W (manufacturer's figures). Finally, [35] reports at least four single-chip implementations of HD decoding for set-top boxes with satellite service operators including DirectTV, BskyB, Dish Network, Euro1080, Premiere and ProSieben Sat 1 all announcing imminent deployment of 2nd generation HDTV.

Battery capacity and power density are critical for mobile devices such as cellular `phones and PDAs. Lithium-ion (LI-ion) batteries dominate the market, with their high nominal voltage, 3.6V, being suitable for 3G cell `phones. Though nanotechnology has made strides in increasing energy density and reducing charge time in recent years, it is apparent that *no* order of magnitude increase in capacity and charge/density is on the horizon for portable devices. Recent methanol-based fuel cells have a capacity of 3 to 4 times that of LI-ion batteries but will need to be combined with LI-ion batteries, as they are not otherwise suited to burst processing. Despite these unpromising battery prospects, there is considerable potential for low-power and battery-friendly algorithms, especially in motion estimation at encoders, to reduce energy demands. For example, [36] reports up to 60% reduction in power for only 0.3 dB reduction in PSNR. Further details of algorithms, sometimes in association with specialist h/w architectures such as associative processors [37], are given in Appendix A.

5.1 *Summary of Section 5*

The old certainties of exponential increases in clock speeds driven by Moore's law have been recently upset as all microprocessor manufacturers veer towards multi-core devices. In the short term, H.264 ICs for HD set-top boxes will be available but if H.264 goes through further evolutionary stages, hardware may not be able to follow, especially for conversational video services. However, it is unlikely that H.264-based IPTV will run on existing PCs if these are software based. Battery capacity is not predicted to keep pace with the demands from handheld TV devices unless a generation of low-power codecs are developed. However, there is considerable scope for research into low-power codecs.

6. Applications

Current data rate ranges for various applications can be summarized as follows:

- QCIF (176 x 144: for wireless/mobile networks, UMTS handsets) 64-128 kbit/s.
- CIF (352 x 288: for DVB-H, palmtop and hand-held devices) 256-512 kbit/s.
- 4CIF (720 x 576: for SDTV broadcast, IPTV and DVD storage) 3-6 Mbit/s.
- HD (1920 × 1080: for HDTV at 50 fps) 15-25 Mbit/s.

The following summarizes the up and coming applications:

6.1 MBMS (Multimedia Broadcast Multicast Services) in UMTS Networks

The advantage of using mobile telecoms operators' new 3rd generation (3G) networks for uni-/multicasting services is that these networks already have licensed spectrum allocations, and are already planned and rolling out. All the necessary identification, authorisation and billing systems for Universal Mobile Telecommunication System (UMTS) are in place. Thus, operators' new services such as Multimedia Broadcast Multicast Service (MBMS) can offer broadcast-like provision of television content to subscribers within a 3G mobile 'phone cell area. This is achieved by aggregating the content into one or two 'multicast' channels (typically 64 kbit/s each), and all subscribers in the network share access to the channels for live streaming or downloading of the video programmes to their handsets. Video compression for the display of QCIF images is achieved via H264 or VC-1, with Turbo error coding, as specified by the mobile industry consortium 3GPP. The main disadvantage of MBMS is that mobile networks are traditionally optimised for one-to-one, bi-directional communications with short holding times. Multicasting or broadcasting is a one-to-many, uni-directional, continuous service, and as such will tend to reduce available cell capacity (limited to 2 Mbit/s per 3G cell), unbalance the network and reduce the quality of service to other subscribers. As demand for capacity on 3G networks increases in the future, those services with long holding times will become expensive and impractical.

6.2 SDTV

SDTV will continue to be a significant format for video content for some years into the future for delivery to static and nomadic receivers – i.e. for satellite, cable, terrestrial and IPTV services. For terrestrial broadcasting, future developments in coding (as detailed in this report) and efficiency improvements in the statistical multiplexing of the individual data-streams will allow more programmes to be carried within the spectrum channel allocation (8 MHz in the UK supporting 18 Mbit/s or 24 Mbit/s according to the form of modulation). However, over the next 10 years, it can also be expected that the demand for mobile multimedia and HDTV services will reduce the significance of SD in the consumer world. Fortunately, the standards being developed for coding and delivery under the auspices of organisations like DVB can handle a flexible range of picture resolutions and qualities, together with multi-channel sound and associated ancillary services.

Currently, broadcast quality video services can be coded with an MPEG-2 encoder at a fixed bit rate from 3 to 6 Mbit/s, depending on the scene content and the desired quality. In a multiplexed environment, such as satellite or digital terrestrial transmission (DTT) the higher bit rate requirement of a service may be compensated by the lower bit rate of a softer service, or some services may be deliberately coded at a higher quality than others sharing the same channel. This statistical multiplexing reduces the average bit rate of required for a fixed bit rate channel. For example, currently a good quality BBC1 programme is coded at 4.25 Mbit/s but with statistical multiplexing, this rate may come down to 3.3 Mbit/s. There are of course some commercial TV services at a lower quality of 1.5-3.5 Mbit/s, and with statistical multiplexing the average bit rate will be around 2 Mbit/s. We expect that the future bit rate of SD will be reduced by 7% per year. For example, this means that a fixed bit rate of 6 Mbit/s video will be 3 and 1.5 Mbit/s in the coming two decades, albeit at a higher video quality than today. With statistical multiplexing, these rates can be further reduced by a factor of 1.2 to 1.5, depending on the number of services sharing the channel, with a larger reduction for fewer services sharing the multiplex.

6.3 DVB-H

DVB-H (DVB-Handheld) arises from the success of the DVB-T standard and, hence, it is backward compatible with DVB-T. DVB-T was adopted for digital television in the UK with an assumption that roof-top aerials would allow line-of-sight reception (with a Rician noise model). In Germany, reception on portable devices with simple rod aerials (with a Rayleigh noise model) has become popular, and indeed DVB-T was originally designed with such applications in mind. Rod aerials have no gain and are not directional, which without a remedy implies an increase in transmitter power. DVB-H is intended for battery-powered devices with smaller screens (typically CIF-sized) (for those who would like to watch TV on the beach!) and introduces improvements in error control, making mobile reception more practical.

To conserve battery power at the receiver and prolong battery life (see later discussion of this issue), DVB-H transmitters employ time-slicing, which adapts a DVB-T rate, typically MPEG-2 at 10 Mbit/s to a rate more suitable for mobile devices, say 500 kbit/s, and with a new codec such as H.264 or VC-9. Data is transmitted in bursts that can be scaled from a few milliseconds to some seconds. The data-rate at the receiver can be traded against the need to conserve the battery. During a burst the data-rate approaches the peak data rate available over the broadcast channel.

Another addition to DVB-T is time interleaving and FEC (Reed-Solomon) to guard against burst errors arising from fast fading (due to multiple signal paths), collectively known as Multi-Protocol Encapsulation (MPE)-FEC [38]. (Time-interleaving of the transmitted stream converts burst errors to random errors.) DVB-T already has an inner and outer level of interleaving and error control, whereas MPE-FEC introduces a further layer of protection. Typical Doppler shift (the effect of a moving receiver) catered for is 100 MHz and above. Currently, there is no provision for Unequal Error Protection (UEP) and, because of the small size of mobile devices, it is not assumed that multiple aerials will be available (aerial diversity or MIMO) to counteract fading.

DVB-H can be distributed over an IP network from a server to various broadcast transmitters. The UHF band can be utilised for DVB-H services, but this band may not be available for all countries. DVB-H can offer more than 30 TV programs for small displays using 4-QAM (QPSK), 16-, 64-QAM modulation combined with either 2K, 4K, and 8K mode orthogonal frequency division multiplexing (OFDM)[11] to gain flat rather than frequency dependent fading [39]. (The modes refer to the size of the FFT involved in OFDM. The transmitted power is modulation dependent (e.g. 25.5 dB for 16-QAM and 18.5 dB for QPSK). 4K mode is a new mode for DVB-H over DVB-T, and as such the 4K mode has some incompatibilities with DVB-T. 4k mode is suitable for single transmitter operation at multiple frequencies as well as small to medium SFNs (up to 35 km) (provided, Inter Symbol Interference (ISI) is not a problem at 4K, depending on urban conditions). The transmitter power is limited in the 2K

[11] An 8K mode has longer symbol length than 2K giving greater resilience to echos. This in turn allows the creation of single frequency networks (SFNs))

mode, which is unsuitable for SFNs, whereas the 8K mode allows larger SFNs. The lower 2K and 4K modes are respectively four and two times[12] more resilient to Doppler than 8K mode, as the sub-carriers are closer together in the 2K and 4K modes.

DVB carries IP datagrams in an MPEG-2 Transport Stream (TS) using multiprotocol encapsulation (MPE). A stream of MPEs forms an elementary stream (ES). The full MPEG-2 TS may have a bit rate of 10 Mbit/s (as previously stated). However, time slicing allows a handheld device to select the DVB-H ES from the full MPEG-2 TS. Therefore, the average bit rate at the receiver may only be 250 kbit/s. The burst size is scalable and therefore intermediate bit rates up to 10 Mbit/s (peak) are receivable.

Table 4 gives an example of useable (peak) bit rates by modulation scheme, according to error code rate and guard interval between symbols (OFDM mode does not affect bit-rate). The Table assumes the full multiplex is DVB-H, i.e. a set of TV programmes all of which are potentially receivable by a DVB-H device. The guard interval should obviously be larger than the signal propagation time. Normally, an error code rate of 1/2 or 2/3 using convolutional coding at the physical layer would be applied for mobile networks. As previously mentioned, DVB-H also includes additional (compared to DVB-T) link layer Reed-Solomon MPE-FEC, which is applied at a rate of 3/4 in Table 4. (It is also possible to have no MPE-FEC in DVB-H). Table 4 assumes an 8-MHz channel, as current in the UK[13] and Europe. For other channel capacities (5/6/7 MHz channels have also been investigated) a simple scaling can be applied [39]. Live performance tests have taken place, including a pilot in Oxford, confirming the gain from the addition of RS coding and the flexibility gained from adding the 4K mode.

Several competitors to DVB-H have appeared in the marketplace [47], emanating from various developments round the globe. These include DAB-IP, T-DMB, ISDB-T, tdTV, FLO and DMB-TH [27]. Taking into account market size and technical implementation considerations, the FLO system from the USA and DMB-TH from China will be the main competitors to DVB-H. FLO ('Forward Link Only') has a very similar specification to DVB-H being a 4K multi-carrier system, but with a layered coding technique for the video content, and Turbo coding rather than Viterbi in the inner coding processes. DMB-TH is the newly-announced Chinese standard with both single-carrier high power and multi-carrier cell-based options.

The industrial/commercial commitment to DVB technology around the world and the maturity of product and VLSI developments, assures DVB-H of the leadership position for multimedia broadcast platforms. Handsets and terminals will undoubtedly develop into multi-band and multi-standard devices over a 3-5 year period.

[12] The actual speed tolerated is inversely dependent on carrier frequency --- at UHF the lower modes tolerate speeds up to 250 km/h but at VHF 8K mode is also tolerant.

[13] The UK, unlike a number of continental countries, which adopted an 8K mode, adopted a 2K OFDM mode for terrestial DVB, leading to reduced noise immunity but, at the time, cheaper receivers. The UK uses 81 existing main transmitter sites for terrestial TV, making SFNs unnecessary.

Table-4: Useable bit rates (Mbit/s) for an 8 MHz broadcast channel, depending on FEC and guard interval, after [39].

Modulation	Error code rate	Guard interval			
		1/4	1/8	1/16	1/32
QPSK	1/2	3.74	4.15	4.39	4.52
	2/3	4.98	5.53	5.86	6.03
	3/4	5.60	6.22	6.59	6.79
	5/6	6.22	6.92	7.32	7.54
	7/8	6.53	7.26	7.69	7.92
16-QAM	1/2	7.45	8.30	8.78	9.05
	2/3	9.95	11.06	11.71	12.07
	3/4	11.20	12.44	13.17	13.58
	5/6	12.44	13.82	14.64	15.08
	7/8	13.07	14.51	15.37	15.83
64-QAM	1/2	11.20	12.44	13.17	13.58
	2/3	14.93	16.59	17.57	18.10
	3/4	16.79	18.66	19.76	20.36
	5/6	18.66	20.74	21.95	22.62
	7/8	19.60	21.77	23.06	23.75

DVB defines the video/audio coding specifications for its standards in terms of ‘Toolboxes’, Table 5, from which the network implementer can select an appropriate coding ‘tool’ for his service. There are two ‘toolboxes’ of compression tools which can be selected for DVB digital TV services covering both Transport Stream and Internet Protocol transmissions:

- DVB-H IP-based services in ETSI doc TS 102 005 V1.3
- DVB-T transport stream based services in ETSI doc TS 101 154 V1.8

Table 5. Toolbox of allowable compression processes for DVB systems

Medium	Codec	TS 101 154 (TS)	TS 102 005 (IP)
Audio	MPEG-1 Layer II	√	
	AC-3	√	to be included
	Enhanced AC-3	√	to be included
	DTS	√	
	MPEG-4 HE AAC	√	√
	MPEG-4 HE AAC v2	√	√
	AMR-WB+		√
Video	MPEG-2	√	
	H.264 / AVC	√	√
	VC-1	to be included	√

Future market pressures for lower bit rates (more programmes in the multiplex), better spectrum efficiency, and more robust reception over larger coverage areas will catalyse further developments in video compression and transmission. With this in mind, the DVB project is already starting work on developments like DVB-T2 for a new terrestrial standard (specification to be completed in 2009) and DVB-SSP for complementary satellite/terrestrial DVB-H service (specification to be completed in 2007).

Video and audio coding developments over the next 10 years, as described in this report, will be included in the DVB Toolboxes, with DVB-T2 also introducing additional transmission improvements, for example, Multiple Input Multiple Output (MIMO) and spectrum shaping techniques. Similarly, DVB-SSP will specify a new version of DVB-H with fewer carriers and longer interleaving, whilst allowing the operator to choose the best audio and video compression techniques available from the Toolbox.

6.4 HDTV

Similar to standard definition television, HDTV has evolved in line with technological developments for cameras, coding and displays. The original implementations of HDTV in the 1980's were based on analogue technology and included the 'MUSE' system of Japan and the European HD-MAC project 'Eureka-95' [40].

The first generation digital HDTV systems were developed from these analogue roots, and some degree of standardisation was eventually achieved with proposals for production systems based on formats of 1080 scanning lines with 2:1 interlace and 1920 pixel horizontal resolution. MPEG-2 compression was used in these systems for network distribution at around 150 Mbit/s.

Second generation digital HDTV has developed over the last 10 years as camera and digital signal processing speeds and performance have improved. Two basic formats have evolved, with similar technical complexities and subjective qualities:

- 720p/50-60. 1280 horizontal pixel and 720 vertical line resolution, with progressive scanning of 50/60 frames per second
- 1080i/25-30. 1920 horizontal pixel and 1080 vertical line resolution, with interlaced scanning of 50/60 fields resulting in 25/30 frames per second.

There is still some debate about the correct resolution to employ in Europe, in an echo of an earlier controversy, between those who wish to retain analogue TV's interlaced fields and those who favour progressive raster scanning. In a departure from earlier practice, it is no longer expected that reduced chrominance sampling (4:2:0 format) will suffice for HDTV and (4:4:4 format) will be demanded by consumers, at a doubling in the number of samples. This impact on the bit-rate is an example of how increasing consumer expectations will counteract gains made from improved codecs. Typically, the serial digital interface (SDI) in a studio for the real-time exchange of either of these second generation HDTV signals, via a single coaxial cable or fibre link, is approximately 1.5 Gbit/s. State of the art MPEG-4 compression codecs can reduce this to around 18 Mbit/s or less, for distribution to the home.

At high resolutions, especially if the viewing distance is reduced, quality difference is easily visible. For example, if HDTV and SDTV are viewed simultaneously, there is about two quality grade difference between HD and SD in favor of HD. The acceptable quality at HDTV resolution is of the order of 35-40 dB. In the future, the vast majority of large flat panel screens for European homes will be in the range of 30-40''. For these display, the optimum view distance is 2.7 m. At 2.7 m viewing distance, the difference between 720p and 1080p would be unnoticeable, but if watched at closer range, as consumers tend to for Internet viewing, the difference is noticeable.

So, in today's production environment, as perhaps the final development for the next decade, new cameras are available which produce a resolution at the sensor of 1920 x 1080 pixels (or more) with progressively-scanned frame rates of 50/60 Hz (or more). Third generation HDTV can thus be defined as:

- 1080p/50-60. 1920 horizontal and 1080 vertical resolution with progressive scanning of 50/60 frames per second

A disadvantage for the third generation 1080p/50-60 standard is that the studio SDI for a 4:2:2 component sampling structure with 10 bits/sample requires a serial bit rate of 3 Gbit/s, which would limit maximum coaxial cable lengths to some 10s of metres. This is likely to mean that in the short term, because of the high investment in existing studio infra-structure, some mild (near lossless) compression will be used at the camera, so that the 1080p/50-60 signal can be mapped into the existing 1.5 Gbit/s second generation HD-SDI systems. Next generation studios, however, will use 10 Gbit/s optical fibre links to carry uncompressed and future higher resolution formats.

An advantage for the adoption of third generation systems is that the high spatial density of samples and progressive scanning of frames allows greater compression efficiency by schemes such as MPEG-4. Subjective tests have shown superior quality ratings for third generation 1080p/50-60 content compressed at 8 Mbit/s, over ratings obtained for either of the second generation formats at the same bit rate, using equivalent MPEG-4 coding strategies. Notice that if future HDTV is to employ a 4:4:4 colour format, then almost quadrupling the volume of chrominance data over the normal 4:2:0 sampling pattern, resulting in a doubling in the sampled bits per pixel, will push up the overall bit rate. An overview of the evolution of the generations of HDTV and the subjective tests which indicate the superiority of the third generation systems is given in [48].

‘HDTV Ready’ consumer displays will shortly be able to handle both second and third generation formats plus their MPEG-4 decoding requirements, and third generation HDTV will be an affordable option for the ordinary consumer within 3 to 4 years.

Ultra-high resolution systems have appeared in experimental form. NHK of Japan, for example, has already demonstrated a 7680 × 4320 pixel 60 Hz progressively scanned format, employing a 22.2 multi-channel sound system. This has 16 times the raw data rate of third generation HDTV. The resulting SDI and transmission problems in the studio, plus the clock and processing speeds required in any subsequent compression process, make this an unlikely product for the consumer market for the next 10 to 15 years.

Digital Cinema has developed along slightly different lines, because of the even higher resolutions required for display on very large public screens of various sizes. In July 2005, the Digital Cinema Specification was published, in which JPEG-2000 was chosen as the compression and distribution format for digital cinema content. A key advantage of JPEG-2000 is utilisation of wavelet transform coding, which enables multiple resolution images to be extracted from a single coded datastream. This makes it as easy to extract a 2k (2048 horizontal x 1080 lines x 24 frames/second) version, as it is to display the full resolution version of an image from a 4k (4096 x 2160 x 24) compressed code-stream running at 250 Mbit/s. Thus the Digital Cinema standard has a degree of built-in ‘future proofing’ for applications requiring various image resolutions over the next 10 years.

6.5 IPTV (TV over the Internet or a converged IP network)

According to a survey conducted in March 2005 [43], more than 81% of broadband subscribers in Europe are either interested or very much interested in receiving Internet, TV and telephony services (triple play) from a single supplier. This indicates the growth of IPTV in future and is in line with experience in Italy and elsewhere, where IPTV is very popular. Video-on-demand (VoD) services are now maturing and can reliably and cost-effectively deliver MPEG-4 and/or VCI streams at a bandwidth around or just below 2 Mbit/s. A stringent requirement of IPTV is guaranteed Quality-of-Service (QoS). Techniques tuned to combat packet losses are likely to be the essential features at IPTV encoder/decoder.

Currently (year 2006) there are almost 1.2 million IPTV customers in 10 major European countries. Research group OVUM early in this year predicted that by 2009 there will be over 7 million IPTV subscribers in 5 major European countries, with Italy leading the way followed by France, Spain, Germany and UK. Even Diffusion group predicts 34 million subscribers by 2010 (worldwide) and iSuppli predicts 63 million by 2010 [44]. However, broadband users in these countries are much more plentiful than IPTV users, with France Telecom predicting its customer numbers by year 2008 to reach 6 million and those of Deutsch Telekom of Germany predicted to be 11.5m over the same period, with over 8 m in Italy and 4 m in Spain. In the UK, BT expects its customer numbers to reach 2.1 m by the end of 2006. Thus, if IPTV proves to be attractive, then the network infrastructure is in place to increase the number of users.

With the intention of making IPTV a popular future visual communication service, several operators have currently started to roll out services over more than one type of network. In Denmark, TDC, the major cable TV operator, is also offering IPTV, using the same set-tops as for its cable service, due to its limited number of customers. MaLigne TV of France offers 30 TV channels, while Belacom TV currently offers 61 channels with an additional 20 channels for its Classic plus package.

While for the time-being TDC is retaining MPEG-2 codecs, it foresees a transition to MPEG-4 and the Microsoft TV platform, which will improved the delivered video quality. More and more operators such as MaLigne and Belacom are taking advantage of the MPEG-4 codec,
MS's VC-1 codec offers similar compression/quality performance to MPEG-4, but with the additional advantage for IPTV service operators, of an integral Digital Rights Management package.The Dirac codec under development by the BBC, using frame-based wavelet transform and arithmetic coding [49], is an open source codec, which may reduce licensing costs. However, MS VC-9 licensing costs are likely to be less than that of its competitors [50].

One of the attractive features of IPTV, in addition to its interactivity, is user access to a vast number of TV programs and the ability to customize advertising. Thus, the main issue in the IPTV industry is bit rate reduction, in order to be able to accommodate as many TV services as possible, both from the storage point of view and transmission channel. However, in predicting the required bit rate for a service, the main difficulty is the picture format. For example, while, in the USA, already with HDTV services, it is easier to offer IPTV at HD resolution, in Europe SDTV format might be preferred, though through the use of video transcoders smaller resolution pictures at lower quality and lower bit rates are also possible. One has to note that the picture quality of existing IPTV services are much poorer than their broadcast counterparts, despite using a much more sophisticated codec; *e.g.* current HDTV/SDTV are coded with MPEG-2, while those for IPTV are coded with an H.264 type codec. While MPEG-2 HDTV is broadcast at around 20-25 Mbit/s, for IPTV, H.264 achieves around 6 Mbit/s. This does not mean that H.264 is 3-4 times more efficient that MPEG-2 for HDTV, but rather that its target quality is set at lower level. (Note that in Table 3 we have shown that for HDTV, H.264 is around 20-40 per cent more efficient than MPEG-2.) Similarly, the current IPTV rate even with an MPEG-2 codec for SDTV format is around 1.5 Mbit/s, much poorer than the same program in terrestrial and satellite networks at around 4-5 Mbit/s. However, for a smaller SIF-sized picture, which is the main picture format for most IP users, the bit rate is around 256-512 kbit/s.

We believe that bit rate/quality of future IPTV will be a compromise between bit rate and quality. That is, unlike broadcast services, where the bit rates will be significantly reduced with little improvement in quality, except for HDTV, in IPTV the reduction in rate will be gradual but improvement in quality will be more significant. In particular, with the roll-out of broadcasting over IP networks, eventually IPTV services will be at the rate and quality of broadcast services. However, technically to offer IPTV for a broad range of users with various network capabilities, TV programs might be coded in a scalable format, where due to overhead, its overall bit rate can be higher than single layer broadcast video (see Figure 17).

6.6 *Summary of Section 6*

In the coming two decades, visual services will tend to be delivered at a better quality and still lower bit rates by more efficient future video codecs than occurs today. Compression of HD video will be improved by 20-40% per decade, whereby the current rate of 20 Mbit/s HD video will be reduced to 12 Mbit/s over the coming decade and to 7 Mbit/s thereafter. This is about 5% reduction in bits per year. This trend for SD video will likely be at almost 50% per decade or 7% per year, whereby the current rate of 3-6 Mbit/s SD video will be reduced to 1.5-3 Mbit/s in the next decade and to 1-1.5 Mbit/s in the following decade.

For mobile video of QCIF size, rather than reducing the data rate at 50-70% per decade, the quality will be improved, with a small reduction in bit rate. For the higher spatial resolution of CIF video, the rate is more likely to be reduced from the current rate of 512 kbit/s to 128 kbit/s over the next decade and to 64 kbit/s in the following decade.

IPTV is emerging as the most popular broadcasting service with a wide range in bit rates. It appears that at the start, its bit rate will be low, in the range of 0.5-2 Mbit/s with smaller size pictures (CIF size), but the bit rate will gradually increase along with the spatial resolution. At the end of two decades IPTV may end up at a rate and quality close to HD video.

References

[1] M. Ghanbari, “Standard Codecs: Image Compression to Advanced Video Coding”, IEE Telecommunication series, publ. IEE Press, Stevenage, UK, 2003.

[2] D. Marpe, H. Schwarz and T. Wiegand, “Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard”, IEEE Transaction on Circuits and Systems for Video technology, Vol. 13, pp. 620-636, July 2003.

[3] ITU Video Coding Experts Group Document, VCEG-L13, 112th Meeting, Eibsee, Germany, January 2001.

[4] TU-T Recommendation H.263: Video Coding for Low Bit-rate Communication, Version 1, November 1995; Version 2 (H.263+), January 1998; Version 3 (H.263++), November 2000.

[5] “Report on the Formal Verification Tests on AVC”, ISO/IEC JTC1/SC29/WG11, MPEG-2003/N6231, Waikoloa Hawaii, USA, December, 2003.

[6] T. Wedi, H. Musmann, “Motion- and Aliasing-Compensated Prediction for Hybrid Video Coding”, IEEE Trans. Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 577-586, July, 2003.

[7] F. Lopes, “Motion Estimation for Very Low Bitrate Video Coding”, Ph.D thesis, Department of ESE, University of Essex, 2000.

[8] M. Ghanbari, S. de Faria, I. N. Goh and K. T. Tan, “Motion Compensation for Very Low Bit-rate Video”, Signal Processing: Image Communication, Vol. 7, pp. 567-580, November 1995.

[9] T. Wiegand, X. Zhang and B. Girod, “Long-term Memory Motion-compensated Prediction”, IEEE Transaction on Circuits and Systems for Video Technology, Vol. 9, pp. 70-84, February 1999.

[10] X. Zixiang, K. Ramchandran, M. T. Orchard and Y. –Q. Zhang, “A Comparative Study of DCT and Wavelet Based Image Coding”, IEEE Transaction on Circuits and Systems for Video Technology, Vol. 9, pp. 692-695, August 1999.

[11] D. Marpe and H. L. Cycon, “Very Low Bit-rate Video Coding using Wavelet-based Techniques”, IEEE Transaction on Circuits and Systems for Video Technology, Vol. 9, pp. 85-94, February 1999.

[12] “Draft ITU-T Recommendation H.264 and Draft ISO/IEC 14496-10 AVC”, in Joint Video Team of ISO/IEC JTC1/SC29/WG11 & ITU-T SG16/Q.6 Doc. JVT-G050, T. Wiegand Ed., Pattaya, Thailand, March 2003.

[13] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Lutra, “Overview of H.264/AVC Video Coding Standard”, IEEE Transaction on Circuits and Systems for Video technology, Vol. 13, pp. 560-576, July 2003.

[14] D. Marpe, G Blättermann and T Wiegand, “Adaptive Codes for H.26L”, ITU-T SG16/6 document

[15] T. Wiegand, H. Schwarz, et al., “Rate-Constrained Coder Control and Comparison of Video Coding Standards”, IEEE Transaction on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 688-703, July 2003.

[16] D. Wood, “Everything You Wanted to Know about Video Codecs – but Were too Afraid to Ask”, EBU Technical Review, July 2003.

[17] K. McCann, “HDTV in Europe: Theory and Practice”, EBU Journal, No 18, June 2006, pp. 7.

[18] K. McCann, “HDTV in Europe: Theory and Practice”, DVB-Scene, June 2006.

[19] EBU Broadcasting Technology Management Committee, “Maximising the Quality of SDTV in the Flat-Panel Environment”, EBU Technical Review, April, 2004.

[20] T. Wiegand, “Scalable Video Model 3.0”, Joint Video Team (JVT), Hong Kong, China, Doc. JVT-N015, January 2005

[21] J. Ohm, “Three-Dimensional Subband Coding with Motion Compensation”, IEEE Trans. on Image Processing, vol. IP-3, No. 5, pp. 559-571, September 1994.

[22] R. Atta and M. Ghanbari, “Spatio-temporal Scalability-based Motion Compensated 3-D subband/DCT Video Coding”, IEEE Trans. on Circuits and Systems for Video Technology, Vol 16 No 1, January 2006, pp. 43-55.

[23] M. Ghanbari, “Two-layer Coding of Video Signals for VBR Networks”, IEEE Journal on Selected Areas in Communications, Special Issue on Packet Speech and Video, **7:5** (June 1989) pp. 771–781.

[24] H. Radha, M. van der Schaar, and Y. Chen, “The MPEG-4 Fine-grained Scalable Video Coding Method for Multimedia Streaming over IP”, IEEE Trans. in Multimedia, 3(1):53-68, 2001.

[25] H. Schwarz, D. Marpe, and T. Wiegand, “MCTF and Scalability Extension of H.264/AVC”, in the Picture Coding Symposium, December, 2004.

[26] J-R. Ohm, “Standardization in JVT: Scalable Video Coding”, Workshop on Video and Image Coding and Applications (VICA), July 2005.

[27] M. Barnsley, “Fractals Everywhere”, 2nd edition, publ. Morgan Kaufman, 1993.

[28] P.C. Cosman, R. M. Gray, and M. Vetterli, “Vector Quantization of Image Subbands: A Survey”, IEEE Trans. on Image Processing, 5(2):202-255,1996.

[29] E.A.B. da Silva, D.G. Sampson, and M. Ghanbari, “A Successive Approximation Vector Quantizer for Wavelet Transform Image Coding“, IEEE Trans. on Image Processing, 5(2):299-310,1996.

[30] L. Corte-Real, and A. Pimenta, “A Very Low Bit Rate Video Coder Based on Vector Quantization”, IEEE Trans. on Image Processing, 5(2):263-273,1996.

[31] D.E.Pearson, “Developments in Model-based Coding”, Proc. of IEEE, 83:892-906, 1995.

[32] R. Koenen, F. Pereira, and L. Chiariglione, “MPEG-4: Context and Objectives”, Signal Processing Image Communication Journal, 9(4):295-304, 1997.

[33] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, “H.264/AVC Baseline Profile Decoder Complexity Analysis”, IEEE Trans. on Circuits and Systems for Video Technology, 13(17):704-716, 2003.

[34] T. Agerwala and S. Chatterjee, “Computer Architecture: Challenges and Opportunities for the Next Decade”, IEEE Micro, 25(3): 58-69, 2005.

[35] D. Marpe, T. Wiegand, G. J. Sullivan, “The H.264/MPEG4 Advanced Video Coding Standard and its Applications”, IEEE Communications, 44(8):134-143, Aug. 2006.

[36] K.-H. Lam, C-H. Tsui, “Reducing Power Consumption of Block Matching Motion Estimation Using Adaptive Algorithm Selection”, in Asia Pacific Conf. on Multimedia Technology & Applications, 2000.

[37] S. Balam, and D. Sconfeld, “Associative Processors for Video Coding Applications”, IEEE Trans. on Circuits and Systems for Video Technology, 16(2):241-250, 2006.

[38] S. Pekovsky and K. Maalej, “DVB-H Architecture for Mobile Communications Systems”, available from http://www.rfdesign.com, April 2005.

[39] G. Faria, J. A. Henriksson, E. Stare and P. Talmola, “DVB-H: Digital Broadcast Services to Handheld Devices”, IEEE Proc., 94(1): 194-209, Jun. 2006.

[40] F. Bellifermine, A. Chimienti, R. Picco, “Evolution and Trends of HDTV”, 5th Annual European Computer Conference, pp. 155-163, May 1991.

[41] Q. Wang and M. Ghanbari, “Scalable Coding of Very High Resolution Video using the Virtual Zero-tree”, IEEE Trans. on Circuits and Systems for Video Technology, 7(5):719-729, Oct. 1997.

[42] D. Taubman, M. Marcellin, “JPEG2000: Image Compression Fundamentals, Compression, and Practice”, publ. Kluwer, 2001.

[43] A. Fawcett, “iTunes for TV? IPTV in the UK – a Viable Fourth Digital TV platform?”, EBU Technical Review, July, 2005.

[44] P. Christian, “Let a Thousand Channels Bloom”, IET Engineering and Technology, pp. 28-31, Oct. 2006.

[45] T. Oelbaum, V. Baroncini, T. K. Tan, and C. Fenimore, “Subjective Quality Assessment of the Emerging AVC/H.264 Video Coding Standard”, International Broadcasting Conference, September, 2004.

[46] T.-C. Chen, C-J. Lian and L.-G. Chen, “Hardware Architecture Design of an H.264/AVC Video Codec”, Asian South Pacific Design Automation Conf., pp. 750-757, 2006.

[47] D. I. Crawford, “Spectrum for Mobile Multimedia Services”, IEEE Symposium on Consumer Electronics, ISCE-2006, St Petersburg, June 2006.

[48] H. Hoffman, T. Itagaki, D. Wood, “1080P/50: The 3rd Generation System – is it a Feasible Option for Production and Emission?”, IBC Conference Publication, IBC-2006, Amsterdam: 51-64, September 2006.

[49] T. Borer and T. Davies, “Dirac – Video compression using Open Technology”, EBU Technical Review, pp. 1-9, July, 2005.

[50] EBU Technical Information I35-2003, “Further Considerations on the Impact of Flat Panel Home Displays on the Broadcasting Chain”, pp. 1-14, 2003.

[51] EBU Broadcast Technology Management Group, “Maximising the Quality of SDTV in the Flat Panel Environment”, EBU Technical Review, April 2004.

[52] R. Salmon, “The Changing World of TV Displays”, EBU Technical Review, April 2004.

[53] EBU Technical Information I34-2002, “The Potential Impact of Flat Panel Displays on Broadcast Delivery of Television”, 2002.

[54] M. Armstrong, “Video Quality and Digital TV Broadcasting”, BBC Research and Development White Paper WHP 131, May 2006.

Appendix A: Hardware issues

The bit-rate performance of future codecs is based on an ideal premise that future hardware will be able to match computational demands. Complexity analysis based on machine instruction counts is likely to be too optimistic by a factor of 3.6 according to a study of a basic H.264 decoder [1] (References are at the end of the Appendix). Moore's law can also no longer be relied upon to predict future computational speed based on a simple-minded regard to clock speed. This is an important point as to some extent Moore's law has allowed complexity to be increased by algorithm developers without overly worrying about available hardware. Moore's law is an empirical rule that predicts a halving of feature size on microchips every 18 months, from which it follows that clock speed doubles at that rate. Unfortunately, Intel had been forced to cancel proposed further extensions to their Pentium general-purpose processor (GPP) because of the problem of power dissipation. In [2], power issues are the limiting factor in server system designs, with DRAM memory, processor and cooling fans contributing respectively by 30%, 28%, and 23% to power consumption. While there is likely to be a compound annual growth rate (cagr) of 80% [2] in computational demand, for example in digital movie processing, clock speed increase over the next five years is estimated to be 15-20% cagr. Therefore, future single-chip GPPs will host multiple processors but the rate of increase in computational speed cannot be relied upon to continue to be exponential over the next five to ten years.

Nonetheless, in the short term it is expected that system-on-chip (SoC) designs, with heterogeneous processors and assorted hardware on a chip will match the immediate needs of H.264/AVC codec HD (1024-interlaced and 720-progressive) video. In fact, the first SoC chips for the H.264 codec have emerged in the last six months such as the A199 from Ambarbella Inc, aimed at high-end embedded applications such as digital camcorders. The main and high profiles of H.264 are implemented (both encoder and decoder) with a low-power ARM-9 embedded processor and runs as 216 MHz to encode HD video at 15 Mbit/s dissipating 1 W (manufacturer's figures). As a general comment, Microsoft's VC-1 codec was *not* implemented on the A199 because of its weaker compression performance and because H.264 decoders are already available from several third-party suppliers. MPEG-2 HD video SoC were developed from 1998 onwards. Broadcom Corp. (via Sand Video Inc.) have also developed an early H.264/AVC high profile SoC decoder, the BCM7411, which also incorporates digital TV features as would be needed for set-top-boxes, as well as being able to decode compressed HD video stored on DVD. Similarly Conextant (via Amphion Ltd.), STMicroelectronics, WISchip, and Texas Instruments are producing HD decoders.

Note that in the case of Ambarbella's A199, 15 Mbit/s represents the high-end of the data rates for the HDTV high profile, which are expected to range from 2 Mbit/s to 16 Mbit/s. Therefore, the Ambarbella solution is not optimal in terms of video storage on a camcorder. The power dissipation of this state-of-the-art device for HD video is also beyond that normally expected for a mobile device. Indeed, only by reducing spatial resolution to Standard Definition (SD) video is the power usage reduced to 450 mW (100 mW is ideal). However, apart from camcorders it is unlikely that encoder power consumption for HD video will be an issue.

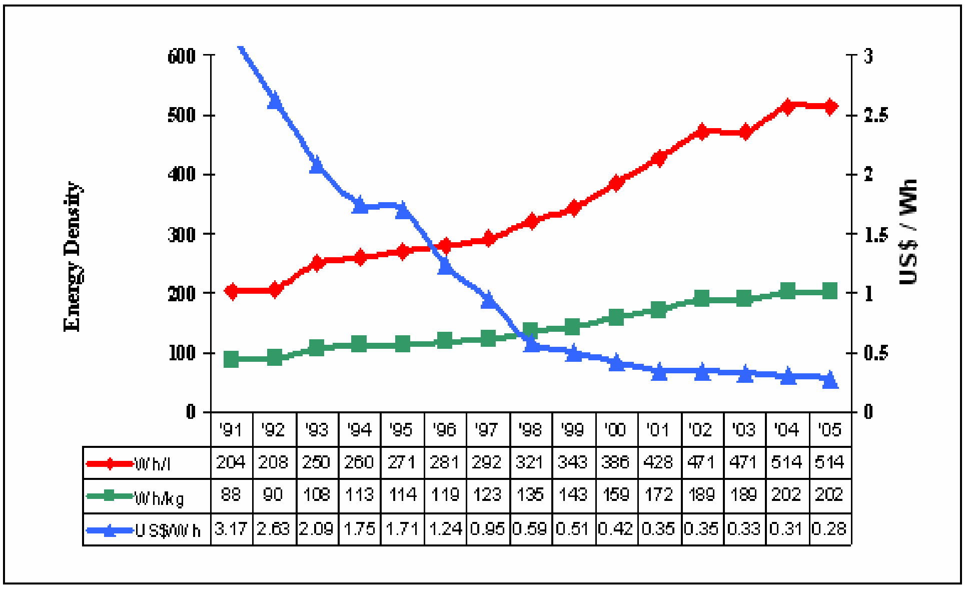
It is still interesting to see that as resolution reduces from progressive HD to progressive SD by a factor of 2.7, power dissipation only reduces by a factor of 2.2. Apart from computational complexity, which involves active per-pixel power usage, passive power consumption also exists in maintaining memory banks, though the A199 has apparently an optimised architecture in that respect.

The real bottleneck for data intensive multimedia applications lies in memory bandwidth and not in computational speed. This is certainly true for GPPs, which is why multimedia streaming architectures (see later analysis) have been developed in the last five years or so, though these have not generally passed into the consumer marketplace. GPPs represent a programmable solution that allows a variety of multimedia tasks to be performed, thereby leading to a more cost-effective solution. For SoC on Applications Specific Integrated Circuits (ASICs), power is less of an issue but form factor (the spatial size) is also considered, as well as cost, which in turn depends on the potential market --- usually a run of over 100,000 devices to make fabrication worthwhile. However, there already is a market for H.264 decoders because the next generation of digital satellite set-top boxes are being specialised to H.264 for (say) Echostar satellites in the U.S.

Complexity analysis [1] shows that a H.264 decoder adds a time factor of 2-3 over its H.263 predecessor across a range of resolutions and encoding rates when applied to the baseline profile (without CABAC) on a GPP, specifically Intel's Pentium III (P3). Memory usage is also up due to H.264's multiple reference frames, being 355 kB for CIF (352 × 288 pixel) and 92 kB for QCIF (176 × 144 pixel) resolution. Both these resolutions are likely on mobile devices and further memory is required for receive buffers, when transmitting over a network. Playout buffers dominate H.264 memory usage. Going from QCIF to CIF, an increase by a factor of four in resolution increases decoder memory by a factor of 3.9, *i.e.* similarly. In [1], it is also reported that for equivalent quality (judged by the quantization parameter) H.264 achieves a 35%-50% lower bit rate. At CIF resolution on a 600 MHz P3 (i.e. at the low end of today's Pentiums) the decode speed was 1.5 to 2 times faster than real-time. Even allowing for file access and display the decode rates are significantly faster than video rate. Therefore, H.264 is easily accomplishable by software implementations on PCs for reduced resolution.

A.1 Battery future

Battery capacity and power density are critical for mobile devices such as cellular `phones and PDAs. Lithium-ion (LI-ion) batteries dominate the market, with their high nominal voltage, 3.6V, being suitable for 3G cell `phones. The form factor and/or volume of the battery are also relevant for portable devices. In terms of Watt hours/litre (Wh/l), equivalent volume has steadily reduced by 90% over the decade starting in 1995 to about about 500 Wh/l, while over the same period the density has increased by 77% to about 200 Wh/kg (refer to Figure 20). Capacity is typically 1200 mAh (milli-Amp-hours). Other factors include the number of charge/discharge cycles before replacement is advisable (300-400 times for typical cell `phone batteries) and the charge time (3 hrs before the charge current drops below 3% ---- faster charge times, even down to a few minutes may be a result of not fully charging the battery --- there is no memory effect for Li-ion batteries). Notice that Li-ion battery life is not just dependent on charge/discharge cycles, as their capacity drops by 20% per year irrespective of re-charging.



	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
◆ Wh/l	204	208	250	260	271	281	292	321	343	386	428	471	471	514	514
■ Wh/kg	88	90	108	113	114	119	123	135	143	159	172	189	189	202	202
▲ US$/Wh	3.17	2.63	2.09	1.75	1.71	1.24	0.95	0.59	0.51	0.42	0.35	0.35	0.33	0.31	0.28

Figure 20: Trends in Li-ion battery charge-density/volume and cost over the last 15 years, after [3].

Unlike Moore's law for feature size, there is no continuous increase in capacity, as Lithium-ion (Li-ion) based batteries have resulted in a stepped increase in capacity, contributing to an average of 8-10% increase year-by-year[14]. Therefore, while Moore's law *has* predicted an increase in processing speed of around double every 2 years, batteries double their capacity every ten years. Notice that military super-high capacity Li-ion batteries may well have high capacities but are not deemed safe in consumer products. Li-ion polymer batteries were thought to offer an alternative safer form of battery (and flexible form factor) but, unfortunately, they are unable to supply the current bursts needed (crucially) for wireless communications devices (or for powering laptop hard-drives).

Battery cost (Figure 20) has reduced from 3$ (1991) to 0.25 US$/Wh but not unnaturally the rate of reduction had bottomed out by 2005. In terms of ability to supply the anticipated high-demand (rising in value in the US from about 2000 million US$ in 2005 to 3800 million US$ in 2012 (according to a market survey by Freedonia Inc. quoted in [3]), Japan has a falling shared of the market, down to 60% in 2005, market capacity being replaced by China and S. Korea.

Recent advances in Li-ion batteries have come from improving the material of the cathode by means of nanotechnology. Increases in capacity of three to four times are claimed in patents by Altair Nanotechnology Inc. The main gains are from increasing the area of the electrode, and, as for Altair, Toshiba and others have a fast, under 6 mins., charging battery. The number of charge/discharge cycles may also increase to 1000 (from around 3000). Finally, the life of the battery will not be limited by the higher temperatures existing in laptops. Nanotechnology is also involved in MIT's use of viruses to build nano-wires, resulting in a three times increase in energy density. However, it is apparent from all these battery developments that *no* order of magnitude increase in capacity and charge/density is on the horizon for portable devices.

[14] I. Buchmann, "High-power lithium-ion, a new area in portable power", March 2006 from `http://www.buchmann.ca/Article27-Page1.asp`

Toshiba's small-sized Methanol fuel cell, weighing 0.3 g, does have a capacity of 3 to 4 times that of the Li-ion battery. Quoted figures are 20 hrs at 100 mW for 2 cc charge of concentrated methanol. The cell can then by re-charged by the user. Though Toshiba have obviously reduced its size, the power density is still only 12 mW/g, compared to a 160 mW/g Li-ion `cell phone battery and to a 500 mW/g for military Li-sulphate batteries. A principle problem with fuel cells is that they are not suited to supplying bursts of power [4] as required for (say) DVB-H time-slicing at the receiver. Therefore, a solution that may migrate from its present development in the military sphere is a hybrid technology. The fuel cell is responsible for steady-state power and re-charge of a battery, whereas the battery provides burst energy for communication. Another development that may ease the gap between steady-state and burst energy is super-capacitors made from carbon-aerogel, which has an extremely high relative surface area. Their energy density is about 70% of Li-ion batteries.

A.2 Low-power video processing

Software and hardware approaches to power reduction and battery management are becoming essential, leading to low-power codecs. In some cases, the low-power techniques are standalone. In others, the techniques are tied to the underlying hardware and power consumption reductions may be relative to that hardware. Power usage goes as $P=CV^2f$, where C is capacitance, V is voltage and f is switching frequency. Low-power software techniques reduce computational complexity, so that the encoder/decoder can operate at a reduced clock but still process at video rates. This is most effective at the encoder because of the computational dominance of motion estimation (m.e.). In applications such as videophones and camcorders, the encoder is required alongside the decoder.

Battery management techniques are a new development, not yet directly applied to codecs but likely to become prominent. Battery life is not dependent on average power but on peak power consumption and the power discharge profile. Detailed models of a battery [5] establish task-by-task behaviour. By careful scheduling of tasks executed in parallel [6], the peak power can be reduced and sudden drops in power usage can be avoided. These techniques may prolong battery life by about 10%.

Compared to a GPP such as the Pentium 4 with a power budget of about 20 W, then a Palm-top processor such as an ARM-9 requires a 200-fold decrease in consumption to about 100 mW. A DVB-H mobile device is projected to consume 1.2 W and DAB for video may only require 400 mW, respectively a 17-fold and 200-fold decrease on a typical GPP [7]. Within the lower power budget, the functionality of a complex codec must be supported, aside from communications, display, and various other applications.

Examples of purely software techniques are: adaptive selection of m.e. algorithm, with 13%-60% reduction in power for only 0.3 dB PSNR reduction [8]; and, as 72% of complexity in H.264 is in multi-reference frame m.e., stored motion vectors reduce complexity by factor of four or more [9]. Attention in the H.264/AVC has been given to reducing transform and quantization complexity [10], by ensuring the new integer DCT can be accomplished in 16-bit arithmetic and avoiding division at the encoder when applying the quantization matrix values (through shifts), at a cost of only 0.02 dB. (Normalization may also occur through shifts.) However, there remain many impediments to a hardware implementation of H.264/AVC [11], which include a high memory bandwidth, especially for m.e.; a long macro-block processing pipeline, with many sequentially dependent operations; data dependencies between neighbouring macro-blocks, causing delay and storage problems. Nonetheless, in [11] a basic profile SDTV and single reference frame HDTV encoder is implemented on a single chip with an extended four-stage macro-block pipeline and the use of many on-chip memory stores. Turning to the reduction of wireless transmission power, it is also possible to include power consumption into optimisation algorithms that match data rate with distortion. The approach [12] requires trading source coding against channel coding for error control. Notice that in cellular phone systems such schemes should account for the overall reduction of power amongst the cells occupants and the need to protect key video frames [13].

Examples of hardware techniques include monitoring of cache usage to reduce memory access, specifically for the ARM processor [14], a processor already optimised for portable multimedia loads [15]. Cache power consumption is the largest or second largest on recent embedded processors. As video processing applications are data dominated it is also important to optimise memory accesses themselves. Given that the access patterns are known in advance, it is possible to match array dimension, size, and access stride to the characteristics of the memory type, as occurs for SDRAM in [16]. Reductions in data-width by ignoring least-significant bits are also an effective way of reducing power consumption, given the dominance of m.e. In [17], a 70% reduction in power is reported for a 4-bit reduction in bit-width. However, this technique's performance is not only video sequence dependent but it also requires custom hardware, such as the systolic array of [17]. Related to this is reduction of frame memory access power [18] by adaptively reducing the bit-width of the access, depending on the motion present in the video sequence. Frame memory access is estimated to take up 35% of the power budget on portable devices. Again through a systolic array architecture, in which data are clock-synchronously pumped through a processor array in parallel, the extent of an m.e. search can be modified according motion activity in the video sequence. The estimate is based on the maximum motion search displacement in past macroblocks for the current and previous frames, known as a window-follower algorithm. A reduction in power consumption of between 70-90% for a 30 Hz CIF sequence was reported in [19] through `window-following'.

A.3 Low-power hardware architectures

Hardware techniques for power reduction are widely practiced and choice of architecture is also important [20]. GPPs employ a super-scalar architecture in which instructions are dynamically scheduled onto available sub-units. This avoids compiler dependency but the hardware complexity results in significant power consumption.

Voltage scaling [21] exploits the V^2 dependency of power consumption. There is a limit to how far this approach can be pursued, as reducing the voltage increases latency and, hence, throughput. Apart from a global reduction in voltage it is also possible to have multiple voltage domains, or to adaptively alter the voltage. Another circuit-level technique is to reduce switching loss on data and address buses, *e.g.* by means of Gray code addressing or by sending only the difference between successive addresses. By judicious arrangement of memory banks, it is possible to reduce passive power, as only relevant memory banks need be accessed.

Video processing and multimedia benefits from heterogeneous processor types, which in System-on-Chip designs can be found in co-existence. The following paragraphs explore the multimedia architectural `menagerie'.

Very Long Instruction Word (VLIW**)** processors are a version of superscalar GPP but without a dynamic hardware instruction scheduler. Given that the application is static, reliance on a specialized compiler is not an advantage, and in recent times processors such as the Philips TriMedia [22] have become prominent. Relatively low clock speeds (less than 200 MHz) also reduce power consumption. As always, reductions in clock speeds come from increased parallelism in execution. However, though it is true to say that there has been a recent advance in understanding how to exploit VLIW processors [23], the spotlight appears to have moved on to streaming architectures. Though massive exploitable parallelism has been identified from video processing execution traces [24], the reality is that most practical VLIWs have less than ten parallel units (10-way).

As has previously been mentioned Single Instruction Multiple Data (SIMD) parallel processing element arrays, and more specifically, systolic processing are well suited to m.e. and regular algorithms in general. In SIMD, a controller sequences the array through a set of instructions, with designs differing in the amount of flexibility available at decision points. SIMDs are not suited to sequential algorithms or those with frequent branches, and, hence, will often occur on a SoC in association with a RISC processor. The ability to reduce data-widths has its attractions for power reduction purposes. In another example in [25], a 44% reduction in power consumption was achieved for the MPEG-4 shape adaptive DCT, with only a 0.3 dB drop in video quality.

Processor computation speed greatly outstrips memory access technology. Moreover, cache-based memory hierarchies, as exist on all GPPs, are not well suited to video or multimedia processing, as there is limited temporal locality. Therefore, the major US universities both have projects to introduce in-order instruction execution (unlike GPPs) without caches, in a manner similar to vector supercomputers. Berkeley's Vector IRAM [26] includes large banks of on-chip DRAM (there has been a recent resolution to the electro-magnetic incompatibilities from including DRAM and reconfigurable logic on-chip), which buffer data entering and leaving the vector processor core. The vector IRAM is a factor of six times faster than 5 or 8-way VLIWs [27]. Stanford's Imagine stream processor [28] appears to be a more complex design, achieving 18.3 Giga Operations/s (GOPS) in an MPEG-2 encoder for a rate of 105 fps, with 720 × 480 pixel frames, consuming 2.2 W (still too much for portable devices but already a magnitude less than GPPs). Streaming processors, such as the RSVP [29], are now actively under development for mobile imaging applications, achieving a two times speed-up over an ARM (the current market leader) and 28 times speed-ups for other auxiliary operations. However, power consumption is unreported.

The various processor types are likely to appear as embedded cores on an SoC, thereby reducing the area occupied on mobile devices. For example, TI's DM310 [30] includes on a single chip: an ARM RISC controller; a memory management unit (MMU); a digital signal processor (DSP) (C54x) at 72 MHz; a video encoder unit, and various co-processors running at 144 MHz: 8-way SIMD unit (for m.e), a Variable Length Coding (VLC) unit, and a Quantization unit (the latter two are needed for data dependent operations). The DM310 consumes 400 mW in processing VGA (640 × 480 pixel frames) at 30 fps.

Though GPPs have all introduced instruction-level SIMD instructions (restricted to 8-way parallelism) and streaming sub-systems, these are not ideal because of the high overhead involved in data format conversion [31]. Merging GPP with custom instruction-level SIMD is capable of a 6% to 35% speedup improvement on image and video codecs with the CSI architecture [32] over conventional SIMD enhancements. These developments along with the streaming processors already mentioned, which may be destined for the GPP/PC market, are likely to make Internet TV more attractive.

Associative processors (APs) are claimed as being ideally suited to video codecs [33], though presently there are apparently no commercial developments of this architecture. As in SIMD, processing elements (PEs) work in lock-step but clusters of PEs work on different instruction streams [34], allowing data dependent operation. The PE design is simplified by means of content addressable memory (CAM), with the results of operations found from the CAM. Compared to a RISC, an AP halves the data traffic for m.e. and DCT algorithms, with proportionate reduction in power consumption. As with all PE arrays, SIMD and systolic included, the weakness is limited clock frequency, because of the difficulty of synchronizing operations, and this may result in unacceptable delay for HDTV applications. An interesting variant on SIMD [35] is one in which individual PEs are actually VLIW processors, with the number of processors limited to four or sixteen, reducing the problem of synchronization. This experimental processor echoes the mainstream industries move towards multi-core, hyper-threading, and multi-threading.

Though more related to transmission that to codecs, it is worth mentioning that in DVB-H and other mobile TV standards (DMB and DAB for video), as Orthogonal Frequency Division Multiplexing (OFDM) is used, specialist processing is required at the demodulator. For example, in the Hi-Par DSP [36], VLIW core exists alongside a RISC controller of an SIMD array, which is capable of performing a 4K FFT in 153.6 ms at 100 MHz system clock speed.

It is worth considering what scope exists in terms of parallelism [37] in video processing applications. As clock speed on portable devices is severely limited in order to preserve power, and as clock speed has reached a plateau on GPPs, parallelism represents the best way to improve throughput. Lack of exploitable parallelism may well turn out to be the decisive impediment to adoption of MPEG-4's objects and sprites (objects subject to perspective projection), with the diversity of algorithms requiring some form of functional partitioning [38]. In contrast, there is a huge investment in macro-block processing and exploitable parallelism is only limited by the number of processors feasible for a single chip solution. Within each macroblock typically a ×4 speed-up is possible through data parallelism, though a ×18 speed-up is possible. At the task level the degree of parallelism is probably limited to about ten-way [41]. At the instruction level, massive unexploited parallelism is reported for video processing (with perfect scheduling between 32 and 1000 in [41]) but as we have seen, in VLIW processors to date this has hardly been exploited. It should be noted, that Amdahl's law predicts a decisive restriction on achievable parallelism, whatever the parallelism within individual algorithms, through residual sequential bottlenecks. Entropic coding represents such a bottleneck, though sequential processing of CABAC is heavily optimized [42].

Hardware Issues References

[1] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, "H.264/AVC Baseline Profile Decoder Complexity Analysis, IEEE Trans. on Circuits and Systems for Video Technology, 13(17):704-716, 2003.

[2] T. Agerwala and S. Chatterjee, "Computer Architecture: Challenges and Opportunities for the Next Decade", IEEE Micro, 25(3): 58-69, 2005.

[3] I. Buchmann, "Batteries in a Portable World - A Handbook on Rechargeable Batteries for Non-Engineers", publ. Cadex Electronics, 2001.

[4] P. Patel-Pred, "Traveling Light", IEEE Spectrum, July, p. 12, 2006.

[5] K. Lahiri, A. Raghunathan, and S. Dey, "Efficient Power Profiling for Battery-driven Embedded System Design", IEEE Trans. on Computer-aided Design for Integrated Circuits and Systems, 23(6):919-931, 2004.

[6] A. Lahiri, A. Basu, M. Choudhary, and S. Mitra, "Battery-Aware Code Partitioning for a Text to Speech System", in Design Automation and Test in Europe Conf., pp. 672-677, 2006.

[7] A. Sieber and C. Weck, "What's the Difference between DVB-H and DAB in the Mobile Environment?", European Broadcasting Union (EBU) Techical Review, pp. 1-9, July, 2004.

[8] K.-H. Lam, C-H. Tsui, "Reducing Power Consumption of Block Matching Motion Estimation Using Adaptive Algorithm Selection", in Asia Pacific Conf. on Multimedia Technology & Applications, 2000.

[9] Y.-W. Huang, and B.-Y. Hsuieh, and S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Analysis and Complexity Reduction of Multiple Reference Frames Motion Estimation in H.264/AVC", IEEE Trans. on Circuits and Systems for Video Technology, 16(4): 507-522, 2006.

[10] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Korofsky, "Low-Complexity Transform and Quantization in H.264/AVC ", IEEE Trans. on Circuits and Systems for Video Technology, 13(7): 598-603, 2003.

[11] T.-C. Chen at al., "Analysis and Architecture Design of an HDTV720p 20 Frames/s H.264/AVC Encoder", IEEE Trans. on Circuits and Systems for Video Technology, 16(6);673-688, 2006.

[12] Q. Zhang, Z. Li, W. Zhu, Y.-Q. Zhang, "Power-Minimized Bit Allocation for Video Communication over Wireless Channels", IEEE Trans. on Circuits and Systems for Video Technology, 12(6): 398-410, 2002.

[13] I.-M. Kim, H.-M. Kim, "Transmit Power Optimization for Video Transmission Over Slowly-Varying Rayleigh-Fading Channels in CDMA Systems", IEEE Trans. On Wireless Communications, 3(5):1411-1415, 2004.

[14] C.-L. Yang, H.-W. Tseng, C.-C. Ho, and J.-L. Wu, "Software-Controlled Cache Architecture

for Energy Efficiency", IEEE Trans. on Circuits and Systems for Video Technology, 15(5): 634-644, 2005.

[15] J. Kin, M. Gupta, and W.H. Mangione-Smith, "The Filter Cache: An Energy Efficient Memory Structure", Int. Symposium on Microarchitecture, pp. 184-193, 1997.

[16] H. Kim and I.-C. Park, "High-Performance and Low-Power Memory-Interface Architecture for Video Processing Applications, IEEE Trans. on Circuits and Systems for Video Technology, 11(11): 1160-1170, 2001.

[17] Z.-L. He, K.-K. Chan, C.-Y. Tsui, and M.L. Liou, "Low Power Motion Estimation Design Using Adaptive Pixel Truncation", IEEE Trans. on Circuits and Systems for Video Technology, 10(5): 669-678, 2000.

[18] V. G. Moshnyaga, "Reducing Energy Dissipation of Frame Memory by Adaptive Bi-width Compression", IEEE Trans. on Circuits and Systems for Video Technology, 12(8): 713-718, 2002.

[19] S. Saponara, and L. Fanucci, "Data-Adaptive Motion Estimation Algorithm and VLSI Architecture Design for Low-Power Video Systems", IEE Proc. in Computer Digital Techniques, 155(1):51-59, 2004.

[20] V. Muresan, N. O'Connor, N. Murphy, S. Marlow, and S. McGrath, "Low Power Techniques for Video Compression", In Irish Signals and Systems Conf., 2002.

[21] K. Usami *et al.*, "Design Methodology of Ultra Low-Power MPEG-4 Codec Core Exploiting Voltage Scaling Techniques", in 35th IEEE Design Automation Conf., pp. 438-488, 1998.

[22] S. Rathmam and G. Slavenburg, "Processing the New World of Interactive Media: The Trimedia VLIW CPU Architecture", IEEE Signal Processing, pp. 108-117, March, 1998.

[23] P. Farabuschi, G. Desoli, and J. A. Fisher, "The Latest Word in Digital and Media Processing", IEEE Signal Processing, pp. 59-85, March, 1998.

[24] Z. Wu and W. Wolf, "Trace-driven Studies of VLIW Video Signal Processors", in ACM Symposium on Parallel Algorithms & Architectures, pp. 289-297, June, 1998.

[25] K.-H. Chen, J.-I. Guo, J.-S. Wang, C.-W. Chei, and J.-W. Chen, "An Energy-Aware IP Core Design for the Variable Length DCT/IDCT Targeting at MPEG4 Shape-Adaptive Transforms", IEEE Trans. on Circuits and Systems for Video Technology, 15(5): 704-715, 2005.

[26] D. Patterson *et al.*, "A Case for Intelligent RAM: IRAM", IEEE Micro, 17(2):35-44, 1997.

[27] C. Kozyrakis and D Patterson, "Vector Vs. Superscalar and VLIW Architectures for Embedded Multimedia Architectures", in IEEE Int. Symposium on Microarchitecture, pp. 283-293, 2002.

[28] B. Khailany *et al*., "IMAGINE: Media Processing with Streams", IEEE Micro, 21(2):35-46, 2001.

[29] S. M. Chai *et al*., "Streaming Processors for Next Generation Mobile Imaging Applications", IEEE Communications, 43(12):81-89, 2005.

[30] D. Talla *et al.*, "Anatomy of a Portable Digital Mediaprocessor", IEEE Micro, 24(2):32-39, 2004.

[31] D. Talla, L. K. John, D. Burgerm "Bottlenecks in Multimedia Processing with SIMD Style Extensions and Architectural Enhancements", IEEE Trans. on Computers, 52(8):1015-1031, 2003.

[32] D. Cheresiz, B. Juurlink, S. Vassiliadis, and H. A. G. Wijshoff, "The CSI Multimedia Architecture", IEEE Trans. on Very Large Scale Integration (VLSI) Systems, 13(1):1-13, 2005.

[33] S. Balam, and D. Sconfeld, "Associative Processors for Video Coding Applications", IEEE Trans. on Circuits and Systems for Video Technology, 16(2):241-250, 2006.

[34] W. Gehrke and K. Gaedke, "Associative Controlling for Monolithic Parallel Processing Architectures", IEEE Trans. on Circuits and Systems for Video Technology, 5(5):453-464, 1995.

[35] W. Hinrichs *et al.*, "A 1.3 –GOPS Parallel DSP for High-Performance Image-Processing Applications", IEEE J. of Solid-State Circuits, 35(5):946-952, 2000.

[36] H. Kloos, L. Friebe, J.P. Wittenburg, W. Hinrichs, H. Lieske, and P. Pirsch, "HiPar-DSP 16, a new DSP for Onboard Real-Time SAR Systems", in 15th Aerospace Conf. on Photonic and Quantum Tech., 2001.

[37] P. Pirsch, A. Freimann, C. Klar, and J. P. Wittenburg, "Processor Architectures for Multimedia Applications", in Workshop on System Architecture, Modelling and Simulation (SAMOS), pp. 188-206, 2002 (LNCS 2268).

[38] J. Kneip, S. Bauer, J. Volmer, B. Schmale, P. Kuhn, M. Reiymann, "The MPEG-4 Video Coding Standard – a VLSI Point of View", IEEE Int. Workshop on Signal Processing Systems, 1998.

[40] P. Pirsch, J. Kneip, K. Rönner,"Parallelization Resources of Image Processing Algorithms and their Mapping on a Programmable Parallel Videosignal Processor", in Int. Symposium on Circuits and Systems. I-562-565, 1995.

[41] H. Liao and A. Wolfe, "Available Parallelism in Video Applications", IEEE/ACM Symposium on Microarchitecture, pp. 321-329, December, 1997.

[42] D. Marpe, H. Schwarz, and T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", IEEE Trans. on Circuits and Systems for Video Technology, 13(7):620-636, 2003.

Glossary of Acronyms

ATSC	Advanced Television Systems Committee
AVC	Advanced Video Codec
ASIC	Applications Specific Integrated Circuits
Aps	Associative Processors
CIF	Common Intermediate Format
CAM	Content Addressable Memory
CABAC	Context-Adaptive Binary Arithmetic Coding
CAVLC	Context-Adaptive Variable Length Coding
dB	Decibels
DPCM	Differential Pulse Code Modulation
DSP	Digital Signal Processor
DVB	Digital Video Broadcasting
DCT	Discrete Cosine Transform
DTT	Digital Terrestial Television
DWT	Discrete Wavelet Transform
EDTV	Enhanced Definition Television
ES	Elementary Stream
FGS	Fine-Grained Scalability
FEC	Forward Error Control
GPP	General-Purpose Processor
GIPS	Giga Instructions Per Second
GOF	Group of Frames
GOP	Group of Pictures
HDTV	High Definition TV
IDCT	Integer DCT
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
LZ	Lempel-Ziv
LI-ion	Lithium-ion
LUT	Look-Up-Table
MB	Macroblock
Mb/s	Mega-bit per second
MMU	Memory Management Unit
MCDCT-TF	Motion Compensated DCT Temporal Filtering
MCTF	Motion Compensated Temporal Filtering
MVs	Motion Vectors
MPEG	Moving Picture Experts Group
MBMS	Multimedia Broadcast Multicast Service
MIMO	Multiple Input Multiple Output
MPE	Multi-Protocol Encapsulation
OFDM	Orthogonal Frequency Division Multiplexing
OBMC	Overlapped Block Motion Compensation
PACC	Partitioning,Aggregation and Conditional Coding
PSNR	Peak-Signal-to-Noise Ratio
QoS	Quality-of-Service
QCIF	Quarter-CIF
SVC	Scalable Video Coding
SDI	Serial Digital Interface
SNR	Signal-to-Noise Ratio
SFNs	Single Frequency Networks
SIMD	Single Instruction Multiple Data
SDTV	Standard Definition Television

SoC	System-on-Chip
TS	Transport Stream
UEP	Unequal Error Protection
VLC	Variable Length Coding
VQ	Vector Quantisation
VLIW	Very Long Instruction Word
VoD	Video-on-demand
Wh/l	Watt hours/litre