A Guide to MPEG Fundamentals and Protocol Analysis

Updated with QoS/QoE Diagnostics and Troubleshooting

Primer

Tektronix®

APPLE 1050

Table of Contents

Section 1 – Introduction to MPEG 1
1.1 Convergence 1
1.2 Why Compression Is Needed 1
1.3 Principles of Compression 1
1.4 Compression in Television Applications 2
1.5 Introduction to Digital Video Compression 3
1.6 Introduction to Audio Compression 5
1.7 MPEG Streams 6
1.8 Need for Monitoring and Analysis 7
1.9 Pitfalls of Compression 7

Section 2 – Compression in Video 8
2.1 Spatial or Temporal Coding? 8
2.2 Spatial Coding 8
2.3 Weighting 10
2.4 Scanning 11
2.5 Entropy Coding 12
2.6 A Spatial Coder 12
2.7 Temporal Coding 13
2.8 Motion Compensation 14
2.9 Bidirectional Coding 16
2.10 I-, P- and B-pictures 16
2.11 An MPEG Compressor 18
2.12 Preprocessing 21
2.13 Wavelets 22

Section 3 – Audio Compression 23
3.1 The Hearing Mechanism 23
3.2 Subband Coding 24
3.3 MPEG Layer 1 25
3.4 MPEG Layer 2 26
3.5 Transform Coding 26
3.6 MPEG Layer 3 27
3.7 MPEG-2 Audio 27
3.8 MPEG-4 Audio 27
3.9 AC-3 28

Section 4 – The Standards 29
4.1 What is MPEG 29
4.2 Who are ITU 29
4.3 Evolution of the Standards 29
4.3.1 Video Performance 30
4.3.2 Licensing 31
4.4 MPEG-1 31
4.5 MPEG-2 31
4.5.1 Profiles and Levels in MPEG-2 32
4.6 MPEG-4 34
4.6.1 MPEG-4 Standards Documents 35
4.6.2 Object Coding 35
4.6.3 Video and Audio Coding 36
4.6.4 Scalability 36
4.6.5 Variable Macroblock Size 37
4.6.6 Profiles and Levels in MPEG-4 part 10 (AVC/H.264) 37
4.6.7 MP4 Container Format 38
4.7 SMPTE VC-1 38
4.7.1 Variable Block Size Transform 39
4.7.2 Motion Compensation 39
4.7.3 Deblocking and Smoothing 39
4.7.4 Fading Compensation 40
4.7.5 Profiles and Levels 40
4.8 MPEG-7 and MXF 40
4.9 MPEG-21 40

Section 5 – Elementary Streams 41
5.1 Video Elementary Stream Syntax 41
5.2 Audio Elementary Streams 42

Section 6 – Packetized Elementary Streams (PES) 43
6.1 PES Packets 43
6.2 Time Stamps 43
6.3 PTS/DTS 44

Section 7 – Program Streams 45
7.1 Recording vs. Transmission 45
7.2 Introduction to Program Streams 45

Section 8 – Transport Streams 46
8.1 The Job of a Transport Stream 46
8.2 Packets 47
8.3 Program Clock Reference (PCR) 47
8.4 Packet Identification (PID) 48
8.5 Program Specific Information (PSI) 48

Section 9 – Digital Modulation 50
9.1 Principles of Modulation 50
9.2 Analog Modulation 50
9.3 Quadrature Modulation 50
9.4 Simple Digital Modulation Systems 51
9.5 Phase Shift Keying 52
9.5.1 Turbo 8PSK 52
9.5.2 DVB S2 52
9.6 Quadrature Amplitude Modulation - QAM 53
9.7 Vestigial Sideband Modulation – VSB 54
9.8 Coded Orthogonal Frequency Division Multiplex – COFDM 55
9.9 China Digital Multimedia Broadcast-Terrestrial/Handheld - DMB-T/H 56
9.10 Integrated Services Data Broadcasting (ISDB) 56
9.10.1 ISDB-S Satellite System 56
9.10.2 ISDB-C Cable System 56
9.10.3 ISDB-T Terrestrial Modulation 56
9.10.4 ISDB-Tss Mobile Modulation 57
9.10.5 ISDB in Summary 57
9.11 Mobile Broadcast Systems 57
9.11.1 DVB-H 58
9.11.2 T-DMB 58
9.11.3 DAB IP 58
9.11.4 FLO 60
9.11.5 ISDB Single Segment 60

Section 10 – Introduction to DVB & ATSC 61
10.1 An Overall View 61
10.2 Remultiplexing 62
10.3 Service Information (SI) 62
10.4 Error Correction 63
10.4.1 Turbo Coding 64
10.5 Channel Coding; Energy Dispersion 65
10.6 Inner Coding 66
10.7 Transmitting Digits 67

Section 11 – Data Broadcast 68
11.1 Applications 68
11.1.1 Program Related Data 68
11.1.2 Opportunistic Data 68
11.1.3 Network Data 69
11.1.4 Enhanced TV 69
11.1.5 Interactive TV 69
11.2 Content Encapsulation 69
11.2.1 MPEG Data Encapsulation 69
11.2.2 DVB Data Encapsulation 73
11.2.3 ATSC A/90 Data Encapsulation 73
11.2.4 ARIB Data Encapsulation 73
11.3 Broadcasting Data Content 74
11.3.1 How Object Carousels Are Broadcast 74
11.3.2 MPEG-2 Data Synchronization 74
11.3.3 Signaling Data Contents 75
11.4 Content Presentation 75
11.4.1 Set Top Box Middleware 75
11.4.2 History of Open Middleware Standards 76
11.4.3 The DVB Multimedia Home Platform (MHP) 77
11.4.4 ATSC Advanced Common Application Platform (ACAP) 77

Section 12 – MPEG QoS Testing 78
12.1 Testing Requirements - QoS 78
12.2 Analyzing a Transport Stream 78
12.3 Hierarchic View 79
12.4 Standards Compliance Checking 81
12.4.1 DVB Measurement Guidelines, TR 101 290 81
12.4.2 ATSC Transport Stream Verification, A/78 82
12.5 Packet View 83
12.6 Syntax and CRC Analysis 83
12.7 Timing Analysis 84
12.8 Elementary Stream Testing 86
12.9 Elementary Stream Analysis 86
12.10 Audio Visual Synchronization 88
12.11 Creating a Transport Stream 88
12.12 PCR Inaccuracy Generation 89
12.13 Data Broadcast Carousel Analysis 90

Section 13 – MPEG QoE Testing 92
13.1 Testing Requirements - QoE 92
13.2 How are QoS and QoE interrelated? 94
13.3 Using QoS and QoE to identify “real” problems 95
13.4 Summary 99

Glossary 99

References 103

Section 1 – Introduction to MPEG

MPEG is one of the most popular audio/video compression techniques because it is not just a single standard. Instead, it is a range of standards suitable for different applications but based on similar principles. MPEG is an acronym for the Moving Picture Experts Group, part of the Joint Technical Committee, JTC1, established by the ISO (International Standards Organization) and IEC (International Electrotechnical Commission). JTC1 is responsible for Information Technology; within JTC1, Sub Group SG29 is responsible for "Coding of Audio, Picture, and Multimedia and Hypermedia Information." There are a number of working groups within SG29, including JPEG (Joint Photographic Experts Group), and Working Group 11 for compression of moving pictures. ISO/IEC JTC1/SG29/WG11 is MPEG.

MPEG can be described as the interaction of acronyms. As ETSI stated, "The CAT is a pointer to enable the IRD to find the EMMs associated with the CA system(s) that it uses." If you can understand that sentence you don't need this book!

1.1 Convergence

Digital techniques have made rapid progress in audio and video for a number of reasons. Digital information is more robust and can be coded to substantially eliminate error. This means that generation-losses in recording and losses in transmission may be eliminated. The compact disc (CD) was the first consumer product to demonstrate this.

While the CD has an improved sound quality with respect to its vinyl predecessor, comparison of quality alone misses the point. The real point is that digital recording and transmission techniques allow content manipulation to a degree that is impossible with analog. Once audio or video is digitized, the content is in the form of data. Such data can be handled in the same way as any other kind of data; therefore, digital video and audio become the province of computer technology.

The convergence of computers and audio/video is an inevitable consequence of the key inventions of computing and pulse code modulation (PCM). Digital media can store any type of information, so it is easy to use a computer storage device for digital video. The nonlinear workstation was the first example of an application of convergent technology that did not have an analog forerunner. Another example, multimedia, combines the storage of audio, video, graphics, text and data on the same medium. Multimedia has no equivalent in the analog domain.

1.2 Why Compression Is Needed

The initial success of digital video was in post-production applications, where the high cost of digital video was offset by its limitless layering and effects capability. However, production-standard digital video generates over 200 megabits per second of data, and this bit rate requires extensive capacity for storage and wide bandwidth for transmission. Digital video could only be used in wider applications if the storage and bandwidth requirements could be eased; this is the purpose of compression.

Compression is a way of expressing digital audio and video by using less data. Compression has the following advantages:

- A smaller amount of storage is needed for a given amount of source material.
- When working in real time, compression reduces the bandwidth needed.
  Additionally, compression allows faster-than-real-time transfer between media, for example, between tape and disk.
- A compressed recording format can use a lower recording density and this can make the recorder less sensitive to environmental factors and maintenance.

1.3 Principles of Compression

There are two fundamentally different techniques that may be used to reduce the quantity of data used to convey information content. In practical compression systems, these are usually combined, often in very complex ways.

The first technique is to improve coding efficiency. There are many ways of coding any given information, and most simple data representations of video and audio contain a substantial amount of redundancy. The concept of *entropy* is discussed below.

Many coding tricks can be used to reduce or eliminate redundancy; examples include *run-length coding* and variable-length coding systems such as Huffman codes. When properly used, these techniques are completely reversible so that after decompression the data is identical to that at the input of the system. This type of compression is known as *lossless*. Archiving computer programs such as PKZip employ lossless compression.

Obviously, lossless compression is ideal, but unfortunately it does not usually provide the degree of data reduction needed for video and audio applications. However, because it is lossless, it can be applied at any point in the system and is often used on the data output of lossy compressors.

If the elimination of redundancy does not reduce the data as much as needed, some information will have to be discarded. Lossy compression systems achieve data reduction by removing information that is irrelevant, or of lesser relevance. These are not general techniques that can be applied to any data stream; the assessment of relevance can only be made in the context of the application, understanding what the data represents and how it will be used. In the case of television, the application is the presentation of images and sound to the human visual and hearing systems, and the human factors must be well understood to design an effective compression system.

Some information in video signals cannot be perceived by the human visual system and is, therefore, truly irrelevant in this context. A compression system that discards only irrelevant image information is known as *visually lossless*.

1.4 Compression in Television Applications

Television signals, analog or digital, have always represented a great deal of information, and bandwidth reduction techniques have been used from a very early stage. Probably the earliest example is interlace. For a given number of lines, and a given rate of picture refresh, interlace offers a 2:1 reduction in bandwidth requirement. The process is lossy; interlace generates artifacts caused by interference between vertical and temporal information, and reduces the usable vertical resolution of the image. Nevertheless, most of what is given up is largely irrelevant, so interlace represented a simple and very valuable trade-off in its time. Unfortunately interlace and the artifacts it generates are very disruptive to more sophisticated digital compression schemes. Much of the complexity of MPEG-2 results from the need to handle interlaced signals, and there is still a significant loss in coding efficiency when compared to progressive signals.

The next major steps came with the advent of color. Color cameras produce GBR signals, so nominally there is three times the information of a monochrome signal – but there was a requirement to transmit color signals in the same channels used for monochrome.

The first part of the solution was to transform the signals from GBR to a brightness signal (normally designated Y) plus two color difference signals, *U* and *V*, or *I* and *Q*. Generation of a brightness signal went a long way towards solving the problem of compatibility with monochrome receivers, but the important step for bandwidth minimization came from the color difference signals.

It turns out that the human visual system uses sensors that are sensitive to brightness, and that can “see” a very high-resolution image. Other sensors capture color information, but at much lower resolution. The net result is that, within certain limits, a sharp monochrome image representing scene brightness overlaid with fuzzy (low-bandwidth) color information will appear as a sharp color picture. It is not possible to take advantage of this when dealing with GBR signals, as each signal contains both brightness and color information. However, in YUV space, most of the brightness information is carried in the Y signal, and very little in the color difference signals. So, it is possible to filter the color difference signals and drastically reduce the information to be transmitted.

This is an example of eliminating (mostly) irrelevant information. Under the design viewing conditions, the human visual system does not respond significantly to the high frequency information in the color difference signals, so it may be discarded. NTSC television transmissions carry only about 500 kHz in each color difference signal, but the pictures are adequately sharp for many applications.

The final step in the bandwidth reduction process of NTSC and PAL was to “hide” the color difference signals in unused parts of the spectrum of the monochrome signal. Although the process is not strictly lossless, this can be thought of as increasing the coding efficiency of the signal.

Some of the techniques in the digital world are quite different, but similar principles apply. For example, MPEG transforms signals into a different domain to permit the isolation of irrelevant information. The transform to color-difference space is still employed, but digital techniques permit filtering of the color difference signal to reduce vertical resolution for further savings.

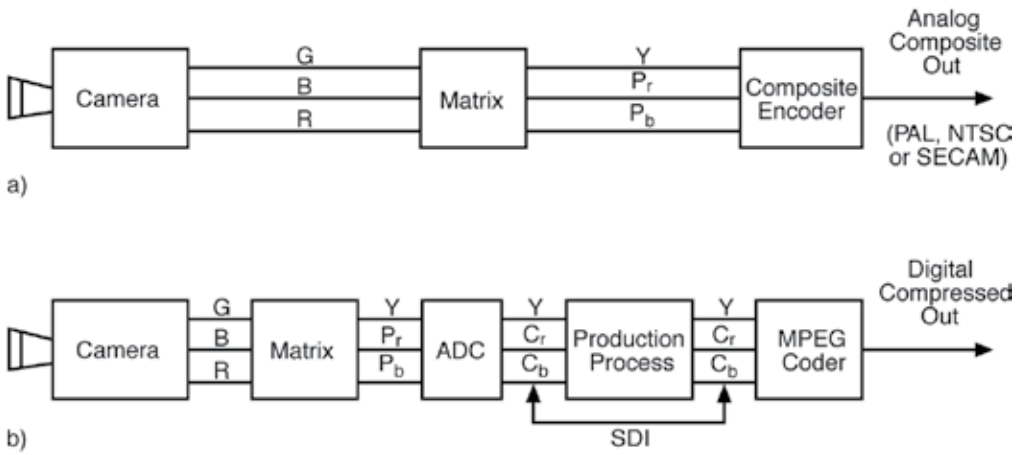


Figure 1-1. Component video translations.

Figure 1-1a shows that in traditional television systems, the GBR camera signal is converted to Y, P_b, P_r components for production and encoded into analog composite for transmission. Figure 1-1b shows the modern equivalent. The Y, P_b, P_r signals are digitized and carried as Y, C_b, C_r signals in SDI form through the production process prior to being encoded with MPEG for transmission. Clearly, MPEG can be considered by the broadcaster as a more efficient replacement for composite video. In addition, MPEG has greater flexibility because the bit rate required can be adjusted to suit the application. At lower bit rates and resolutions, MPEG can be used for video conferencing and video telephones.

Digital Video Broadcasting (DVB) and Advanced Television Systems Committee (ATSC) (the European- and American-originated digital-television broadcasting standards) would not be viable without compression because the bandwidth required would be too great. Compression extends the playing time of DVD (digital video/versatile disk) allowing full-length movies on a single disk. Compression also reduces the cost of ENG and other contributions to television production. DVB, ATSC and digital video disc (DVD) are all based on MPEG-2 compression.

In tape recording, mild compression eases tolerances and adds reliability in Digital Betacam and Digital-S, whereas in SX, DVC, DVCPRO and DVCAM, the goal is miniaturization. In disk-based video servers, compression lowers storage cost. Compression also lowers bandwidth, which allows more users to access a given server. This characteristic is also important for VOD (video on demand) applications.

1.5 Introduction to Digital Video Compression

In all real program material, there are two types of components of the signal: those that are novel and unpredictable and those that can be anticipated. The novel component is called entropy and is the true information in the signal. The remainder is called redundancy because it is not essential. Redundancy may be spatial, as it is in large plain areas of picture where adjacent pixels have almost the same value. Redundancy can also be temporal as it is where similarities between successive pictures are used. All compression systems work by separating entropy from redundancy in the encoder. Only the entropy is recorded or transmitted and the decoder computes the redundancy from the transmitted signal. Figure 1-2a (see next page) shows this concept.

An ideal encoder would extract all the entropy and only this will be transmitted to the decoder. An ideal decoder would then reproduce the original signal. In practice, this ideal cannot be reached. An ideal coder would be complex and cause a very long delay in order to use temporal redundancy. In certain applications, such as recording or broadcasting, some delay is acceptable, but in videoconferencing it is not. In some cases, a very complex coder would be too expensive. It follows that there is no one ideal compression system.

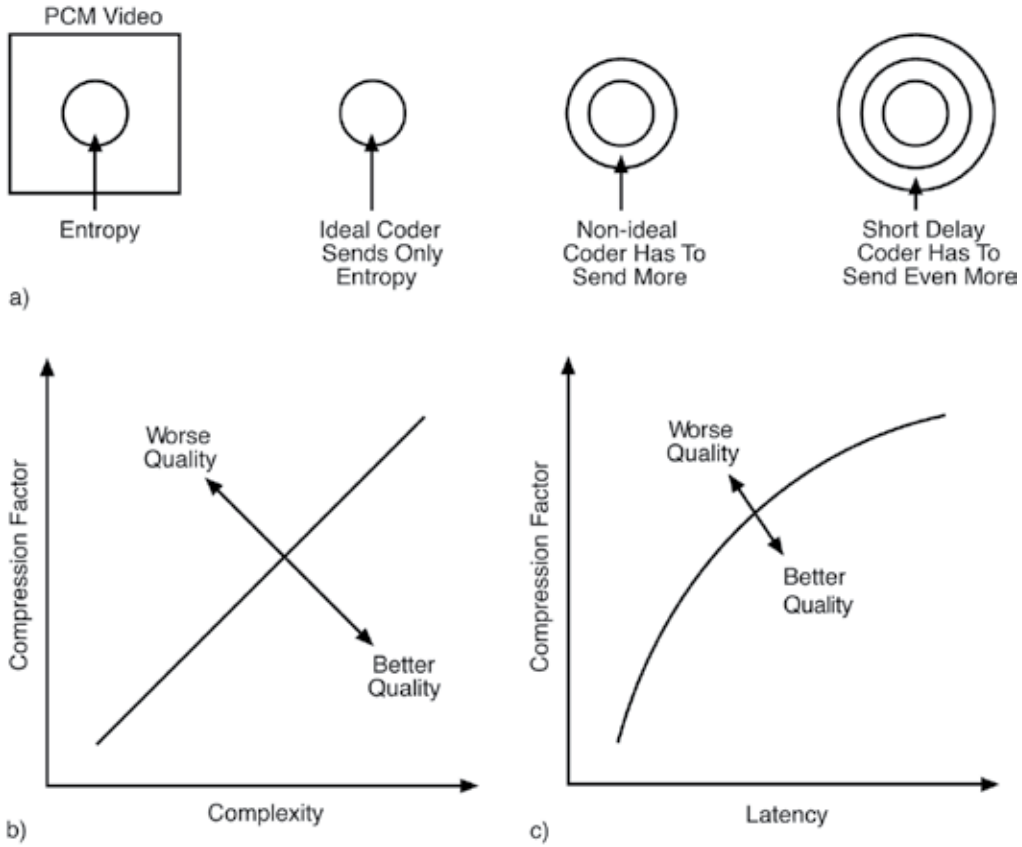


Figure 1-2. Trade-offs between quality, complexity, and latency.

In practice, a range of coders is needed which have a range of processing delays and complexities. The power of MPEG is that it is not a single compression format, but a range of standardized coding tools that can be combined flexibly to suit a range of applications. The way in which coding has been performed is included in the compressed data so that the decoder can automatically handle whatever the coder decided to do.

In MPEG-2 and MPEG-4 coding is divided into several profiles that have different complexity, and each profile can be implemented at a different level depending on the resolution of the input picture. Section 4 considers profiles and levels in detail.

There are many different digital video formats and each has a different bit rate. For example a high definition system might have six times the bit rate of a standard definition system. Consequently, just knowing the bit rate out of the coder is not very useful. What matters is the compression factor, which is the ratio of the input bit rate to the compressed bit rate, for example 2:1, 5:1 and so on.

Unfortunately, the number of variables involved makes it very difficult to determine a suitable compression factor. Figure 1-2a shows that for an ideal coder, if all of the entropy is sent, the quality is good. However, if the compression factor is increased in order to reduce the bit rate, not all of the entropy is sent and the quality falls. Note that in a compressed system when the quality loss occurs, it is steep (Figure 1-2b). If the available bit rate is inadequate, it is better to avoid this area by reducing the entropy of the input picture. This can be done by filtering. The loss of resolution caused by the filtering is subjectively more acceptable than the compression artifacts.

To identify the entropy perfectly, an ideal compressor would have to be extremely complex. A practical compressor may be less complex for economic reasons and must send more data to be sure of carrying all of the entropy. Figure 1-2b shows the relationship between coder complexity and performance. The higher the compression factor required, the more complex the encoder has to be.

The entropy in video signals varies. A recording of an announcer delivering the news has much redundancy and is easy to compress. In contrast, it is more difficult to compress a recording with leaves blowing in the wind or one of a football crowd that is constantly moving and therefore has less redundancy (more information or entropy). In either case, if all the entropy is not sent, there will be quality loss. Thus, we may choose between a constant bit-rate channel with variable quality or a constant quality channel with variable bit rate. Telecommunications network operators tend to prefer a constant bit rate for practical purposes, but a buffer memory can be used to average out entropy variations if the resulting increase in delay is acceptable. In recording, a variable bit rate may be easier to handle and DVD uses variable bit rate, using buffering so that the average bit rate remains within the capabilities of the disk system.

Intra-coding (intra = within) is a technique that exploits spatial redundancy, or redundancy within the picture; inter-coding (inter = between) is a technique that exploits temporal redundancy. Intra-coding may be used alone, as in the JPEG standard for still pictures, or combined with inter-coding as in MPEG.

Intra-coding relies on two characteristics of typical images. First, not all spatial frequencies are simultaneously present, and second, the higher the spatial frequency, the lower the amplitude is likely to be. Intra-coding requires analysis of the spatial frequencies in an image. This analysis is the purpose of transforms such as wavelets and DCT (discrete cosine transform). Transforms produce coefficients that describe the magnitude of each spatial frequency. Typically, many coefficients will be zero, or nearly zero, and these coefficients can be omitted, resulting in a reduction in bit rate.

Inter-coding relies on finding similarities between successive pictures. If a given picture is available at the decoder, the next picture can be created by sending only the picture differences. The picture differences will be increased when objects move, but this magnification can be offset by using motion compensation, since a moving object does not generally change its appearance very much from one picture to the next. If the motion can be measured, a closer approximation to the current picture can be created by shifting part of the previous picture to a new location. The shifting process is controlled by a pair of horizontal and vertical displacement values (collectively known as the motion vector) that is transmitted to the decoder. The motion vector transmission requires less data than sending the picture-difference data.

MPEG can handle both interlaced and non-interlaced images. An image at some point on the time axis is called a “picture,” whether it is a field or a frame. Interlace is not ideal as a source for digital compression because it is in itself a compression technique. Temporal coding is made more complex because pixels in one field are in a different position to those in the next.

Motion compensation minimizes but does not eliminate the differences between successive pictures. The picture difference is itself a spatial image and can be compressed using transform-based intra-coding as previously described. Motion compensation simply reduces the amount of data in the difference image.

The efficiency of a temporal coder rises with the time span over which it can act. Figure 1-2c shows that if a high compression factor is required, a longer time span in the input must be considered and thus a longer coding delay will be experienced. Clearly, temporally coded signals are difficult to edit because the content of a given output picture may be based on image data which was transmitted some time earlier. Production systems will have to limit the degree of temporal coding to allow editing and this limitation will in turn limit the available compression factor.

1.6 Introduction to Audio Compression

The bit rate of a PCM digital audio channel is only about 1.5 megabits per second, which is about 0.5% of 4:2:2 digital video. With mild video compression schemes, such as Digital Betacam, audio compression is unnecessary. But, as the video compression factor is raised, it becomes important to compress the audio as well.

Audio compression takes advantage of two facts. First, in typical audio signals, not all frequencies are simultaneously present. Second, because of the phenomenon of masking, human hearing cannot discern every detail of an audio signal. Audio compression splits the audio spectrum into bands by filtering or transforms, and includes less data when describing bands in which the level is low. Where masking prevents or reduces audibility of a particular band, even less data needs to be sent.

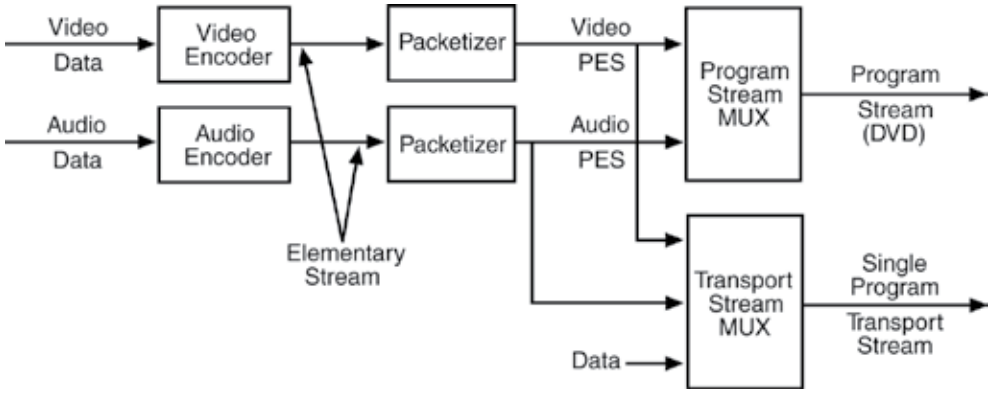


Figure 1-3. Encoding and encapsulation steps in Program and Transport Stream creation.

Audio compression is not as easy to achieve as video compression because of the acuity of hearing. Masking only works properly when the masking and the masked sounds coincide spatially. Spatial coincidence is always the case in mono recordings but not in stereo recordings, where low-level signals can still be heard if they are in a different part of the sound stage. Consequently, in stereo and surround sound systems, a lower compression factor is allowable for a given quality. Another factor complicating audio compression is that delayed resonances in poor loudspeakers actually mask compression artifacts. Testing a compressor with poor speakers gives a false result, and signals that are apparently satisfactory may be disappointing when heard on good equipment.

1.7 MPEG Streams

The output of a single MPEG audio or video coder is called an elementary stream. An elementary stream is an endless near real-time signal. For convenience, the elementary stream may be broken into data blocks of manageable size, forming a packetized elementary stream (PES). These data blocks need header information to identify the start of the packets and must include time stamps because packetizing disrupts the time axis.

Figure 1-3 shows that one video PES and a number of audio PES can be combined to form a program stream, provided that all of the coders are locked to a common clock. Time stamps in each PES can be used to ensure lip-sync between the video and audio. Program streams have variable-length packets with headers. They find use in data transfers to and from optical and hard disks, which are essentially error free, and in which files of arbitrary sizes are expected. DVD uses program streams.

For transmission and digital broadcasting, several programs and their associated PES can be multiplexed into a single transport stream. A transport stream differs from a program stream in that the PES packets are further subdivided into short fixed-size packets and in that multiple programs encoded with different clocks can be carried. This is possible because a transport stream has a program clock reference (PCR) mechanism that allows transmission of multiple clocks, one of which is selected and regenerated at the decoder. A single program transport stream (SPTS) is also possible and this may be found between a coder and a multiplexer. Since a transport stream can genlock the decoder clock to the encoder clock, the SPTS is more common than the Program Stream.

A transport stream is more than just a multiplex of audio and video PES. In addition to the compressed audio, video and data, a transport stream includes metadata describing the bit stream. This includes the program association table (PAT) that lists every program in the transport stream. Each entry in the PAT points to a program map table (PMT) that lists the elementary streams making up each program. Some programs will be open, but some programs may be subject to conditional access (encryption) and this information is also carried in the metadata.

The transport stream consists of fixed-size data packets, each containing 188 bytes. Each packet carries a program identifier code (PID). Packets in the same elementary stream all have the same PID, so that the decoder (or a demultiplexer) can select the elementary stream(s) it wants and reject the remainder. Packet continuity counts ensure that every packet that is needed to decode a stream is received. An effective synchronization system is needed so that decoders can correctly identify the beginning of each packet and deserialize the bit stream into words.

1.8 Need for Monitoring and Analysis

The MPEG transport stream is an extremely complex structure using interlinked tables and coded identifiers to separate the programs and the elementary streams within the programs. Within each elementary stream, there is a complex structure, allowing a decoder to distinguish between, for example, vectors, coefficients and quantization tables.

Failures can be divided into two broad categories. In the first category, the transport system correctly delivers information from an encoder/multiplexer to a decoder with no bit errors or added jitter, but the encoder/multiplexer or the decoder has a fault. In the second category, the encoder/multiplexer and decoder are fine, but the transport of data from one to the other is defective. It is very important to know whether the fault lies in the encoder/ multiplexer, the transport or the decoder if a prompt solution is to be found.

Synchronizing problems, such as loss or corruption of sync patterns, may prevent reception of the entire transport stream. Transport stream protocol defects may prevent the decoder from finding all of the data for a program, perhaps delivering picture but not sound. Correct delivery of the data but with excessive jitter can cause decoder timing problems.

If a system using an MPEG transport stream fails, the fault could be in the encoder, the multiplexer or in the decoder. How can this fault be isolated? First, verify that a transport stream is compliant with the MPEG-coding standards. If the stream is not compliant, a decoder can hardly be blamed for having difficulty. If the stream is compliant, the decoder may need attention.

Traditional video testing tools, the signal generator, the waveform monitor and vectorscope, are not appropriate in analyzing MPEG systems, except to ensure that the video signals entering and leaving an MPEG system are of suitable quality. Instead, a reliable source of valid MPEG test signals is essential for testing receiving equipment and decoders. With a suitable analyzer, the performance of encoders, transmission systems, multiplexers and remultiplexers can be assessed with a high degree of confidence. As a long standing supplier of high grade test equipment to the video industry, Tektronix continues to provide test and measurement solutions as the technology evolves, giving the MPEG user the confidence that complex compressed systems are correctly functioning and allowing rapid diagnosis when they are not.

1.9 Pitfalls of Compression

MPEG compression is lossy in that what is decoded is not identical to the original. The entropy of the source varies, and when entropy is high, the compression system may leave visible artifacts when decoded. In temporal compression, redundancy between successive pictures is assumed. When this is not the case, the system may fail. An example is video from a press conference where flashguns are firing. Individual pictures containing the flash are totally different from their neighbors, and coding artifacts may become obvious.

Irregular motion or several independently moving objects on screen require a lot of vector bandwidth and this requirement may only be met by reducing the bandwidth available for picture-data. Again, visible artifacts may occur whose level varies and depends on the motion. This problem often occurs in sports-coverage video.

Coarse quantizing results in luminance contouring and posterized color. These can be seen as blotchy shadows and blocking on large areas of plain color. Subjectively, compression artifacts are more annoying than the relatively constant impairments resulting from analog television transmission systems.

The only solution to these problems is to reduce the compression factor. Consequently, the compression user has to make a value judgment between the economy of a high compression factor and the level of artifacts.

In addition to extending the encoding and decoding delay, temporal coding also causes difficulty in editing. In fact, an MPEG bit stream cannot be arbitrarily edited. This restriction occurs because, in temporal coding, the decoding of one picture may require the contents of an earlier picture and the contents may not be available following an edit. The fact that pictures may be sent out of sequence also complicates editing.

If suitable coding has been used, edits can take place, but only at splice points that are relatively widely spaced. If arbitrary editing is required, the MPEG stream must undergo a decode-modify-recode process, which will result in generation loss.

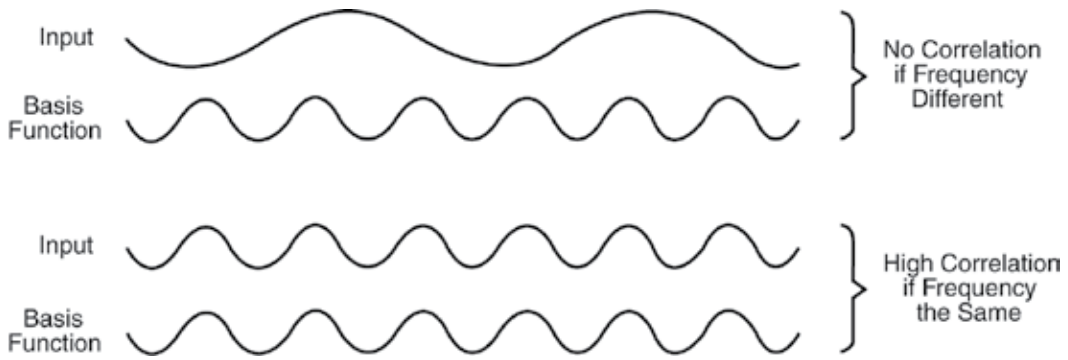


Figure 2-1. Related and unrelated waveforms.

Section 2 – Compression in Video

This section shows how video compression is based on the perception of the eye. Important enabling techniques, such as transforms and motion compensation, are considered as an introduction to the structure of an MPEG coder.

2.1 Spatial or Temporal Coding?

As was seen in Section 1, video compression can take advantage of both spatial and temporal redundancy. In MPEG, temporal redundancy is reduced first by using similarities between successive pictures. As much as possible of the current picture is created or "predicted" by using information from pictures already sent. When this technique is used, it is only necessary to send a difference picture, which eliminates the differences between the actual picture and the prediction. The difference picture is then subject to spatial compression. As a practical matter it is easier to explain spatial compression prior to explaining temporal compression.

Spatial compression relies on similarities between adjacent pixels in plain areas of picture and on dominant spatial frequencies in areas of patterning. The JPEG system uses spatial compression only, since it is designed to transmit individual still pictures. However, JPEG may be used to code a succession of individual pictures for video. In the so-called "Motion JPEG" application, the compression factor will not be as good as if temporal coding was used, but the bit stream will be freely editable on a picture-by-picture basis.

2.2 Spatial Coding

The first step in spatial coding is to perform an analysis of spatial frequencies using a transform. A transform is simply a way of expressing a waveform in a different domain, in this case, the frequency domain. The output of a transform is a set of coefficients that describe how much of a given frequency is present. An inverse transform reproduces the original waveform. If the coefficients are handled with sufficient accuracy, the output of the inverse transform is identical to the original waveform.

The most well known transform is the Fourier transform. This transform finds each frequency in the input signal. It finds each frequency by multiplying the input waveform by a sample of a target frequency, called a basis function, and integrating the product. Figure 2-1 shows that when the input waveform does not contain the target frequency, the integral will be zero, but when it does, the integral will be a coefficient describing the amplitude of that component frequency.

The results will be as described if the frequency component is in phase with the basis function. However if the frequency component is in quadrature with the basis function, the integral will still be zero. Therefore, it is necessary to perform two searches for each frequency, with the basis functions in quadrature with one another so that every phase of the input will be detected.

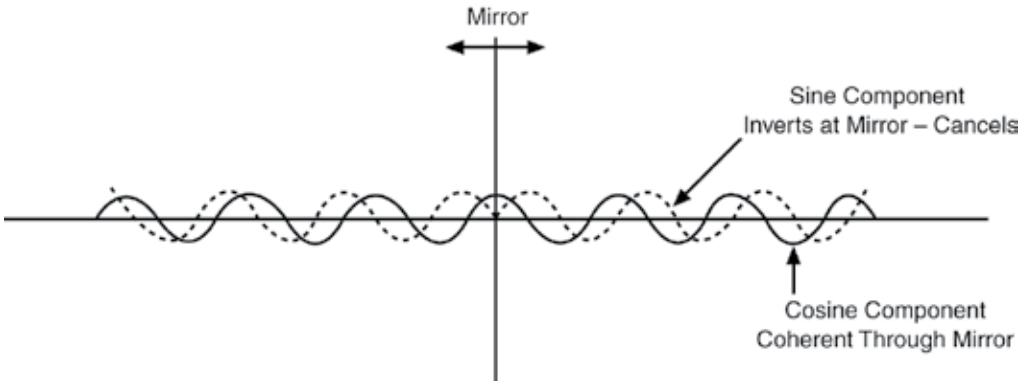


Figure 2-2. Relationship of sine and cosine components.

The Fourier transform has the disadvantage of requiring coefficients for both sine and cosine components of each frequency. In the cosine transform, the input waveform is time-mirrored with itself prior to multiplication by the basis functions. Figure 2-2 shows that this mirroring cancels out all sine components and doubles all of the cosine components. The sine basis function is unnecessary and only one coefficient is needed for each frequency.

The discrete cosine transform (DCT) is the sampled version of the cosine transform and is used extensively in two-dimensional form in MPEG. A block of 8x8 pixels is transformed to become a block of 8x8 coefficients. Since the transform requires multiplication by fractions, there is wordlength extension, resulting in coefficients that have longer wordlength than the pixel values. Typically an 8-bit pixel block results in an 11-bit coefficient block. Thus, a DCT does not result in any compression; in fact it results in the opposite. However, the DCT converts the source pixels into a form where compression is easier.

Figure 2-3 shows the results of an inverse transform of each of the individual coefficients of an 8x8 DCT. In the case of the luminance signal, the top-left coefficient is the average brightness or DC component of the whole block. Moving across the top row, horizontal spatial frequency increases. Moving down the left column, vertical spatial frequency increases. In real pictures, different vertical and horizontal spatial frequencies may occur simultaneously and a coefficient at some point within the block will represent all possible horizontal and vertical combinations.

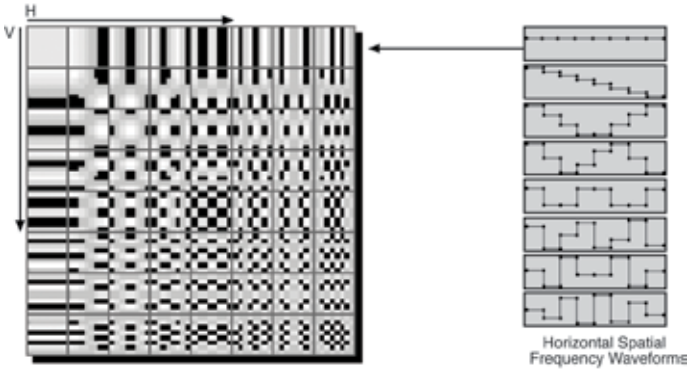


Figure 2-3. 64 unique coefficients of a DCT.

Figure 2-3 also shows 8 coefficients as one-dimensional horizontal waveforms. Combining these waveforms with various amplitudes and either polarity can reproduce any combination of 8 pixels. Thus combining the 64 coefficients of the 2-D DCT will result in the original 8x8 pixel block. Clearly for color pictures, the color difference samples will also need to be handled. Y, Cb and Cr data are assembled into separate 8x8 arrays and are transformed individually.

In much real program material, many of the coefficients will have zero or near-zero values and, therefore, will not be transmitted. This fact results in significant compression that is virtually lossless. If a higher compression factor is needed, then the wordlength of the non-zero coefficients must be reduced. This reduction will reduce accuracy of these coefficients and will introduce losses into the process. With care, the losses can be introduced in a way that is least visible to the viewer.

2.3 Weighting

Figure 2-4 shows that the human perception of noise in pictures is not uniform but is a function of the spatial frequency. More noise can be tolerated at high spatial frequencies. Also, video noise is effectively masked by fine detail in the picture, whereas in plain areas it is highly visible. The reader will be aware that traditional noise measurements are frequently weighted so that technical measurements relate more closely to the subjective result.

Compression reduces the accuracy of coefficients and has a similar effect to using shorter wordlength samples in PCM; that is, the noise level rises. In PCM, the result of shortening the word-length is that the noise level rises equally at all frequencies. As the DCT splits the signal into different frequencies, it becomes possible to control the spectrum of the noise. Effectively, low-frequency coefficients are rendered more accurately than high-frequency coefficients by a process of weighting.

Figure 2-5 shows that, in the weighting process, the coefficients from the DCT are divided by constants that are a function of two-dimensional frequency. Low-frequency coefficients will be divided by small numbers, and high-frequency coefficients will be divided by large numbers. Following the division, the result is truncated to the nearest integer. This truncation is a form of requantizing. In the absence of weighting, this requantizing would have the effect of uniformly increasing the size of the quantizing step, but with weighting, it increases the step size according to the division factor.

As a result, coefficients representing low spatial frequencies are requantized with relatively small steps and suffer little increased noise. Coefficients representing higher spatial frequencies are requantized with large steps and suffer more noise. However, fewer steps means that fewer bits are needed to identify the step and compression is obtained.

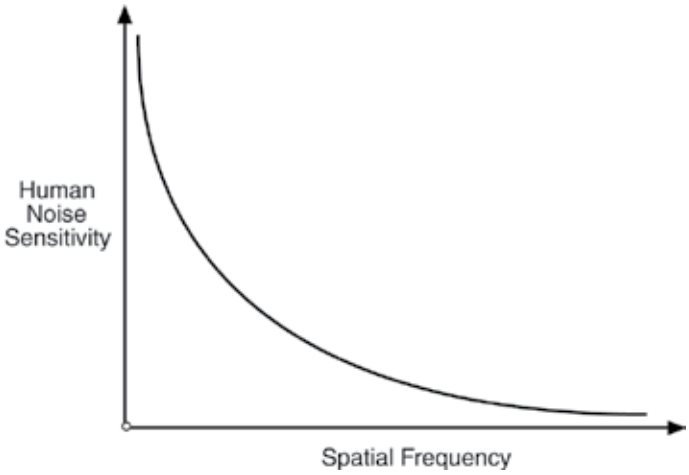
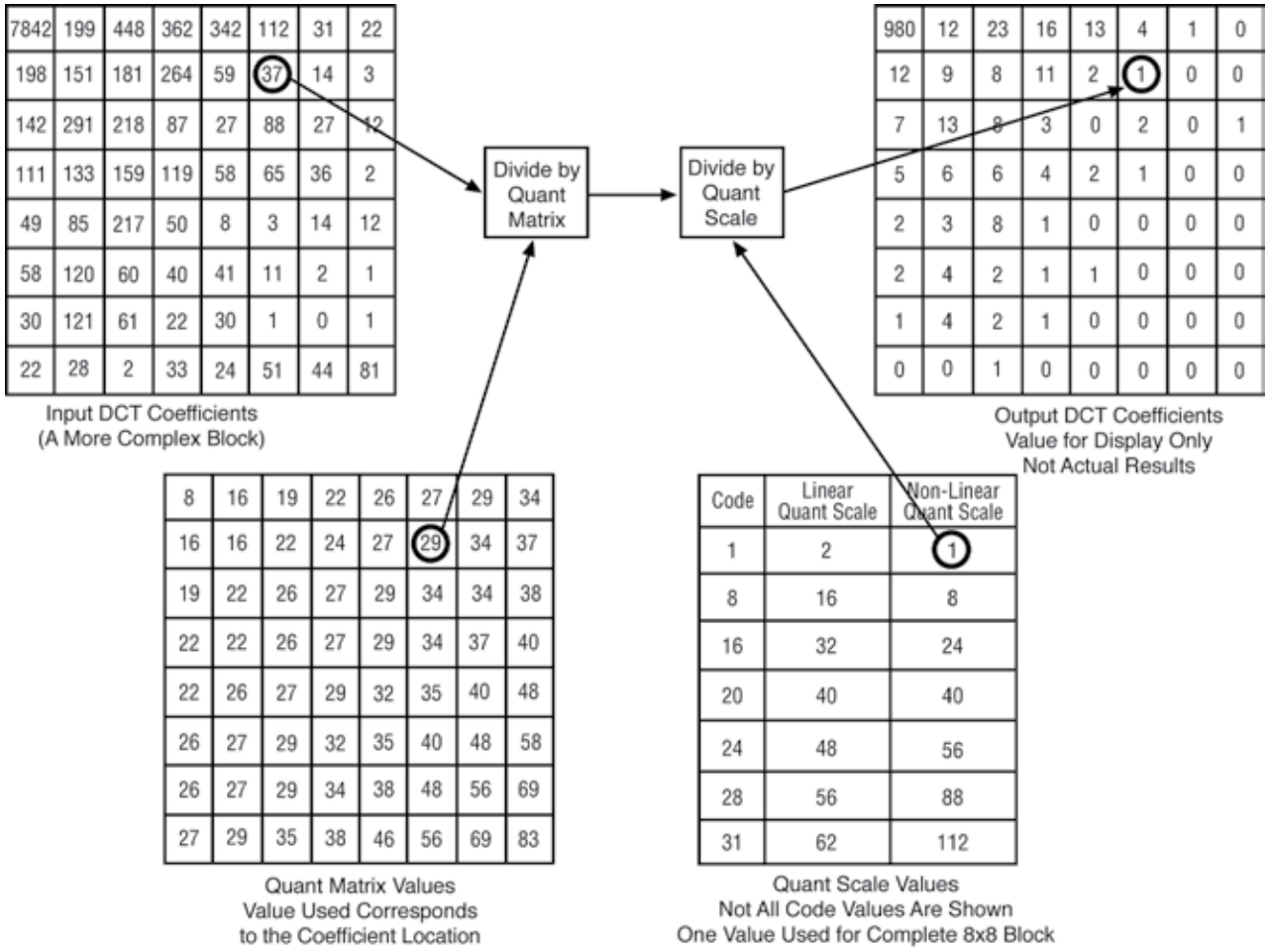


Figure 2-4. Human perception of noise to spatial frequency.

In the decoder, low-order zeros will be added to return the weighted coefficients to their correct magnitude. They will then be multiplied by inverse weighting factors. Clearly, at high frequencies the multiplication factors will be larger, so the requantizing noise will be greater. Following inverse weighting, the coefficients will have their original DCT output values, plus requantizing error, which will be greater at high frequency than at low frequency.

As an alternative to truncation, weighted coefficients may be nonlinearly requantized so that the quantizing step size increases with the magnitude of the coefficient. This technique allows higher compression factors but worse levels of artifacts.

Clearly, the degree of compression obtained and, in turn, the output bit rate obtained, is a function of the severity of the requantizing process. Different bit rates will require different weighting tables. In MPEG, it is possible to use various different weighting tables and the table in use can be transmitted to the decoder, so that correct decoding is ensured.



7842	199	448	362	342	112	31	22
198	151	181	264	59	37	14	3
142	291	218	87	27	88	27	12
111	133	159	119	58	65	36	2
49	85	217	50	8	3	14	12
58	120	60	40	41	11	2	1
30	121	61	22	30	1	0	1
22	28	2	33	24	51	44	81

980	12	23	16	13	4	1	0
12	9	8	11	2	1	0	0
7	13	8	3	0	2	0	1
5	6	6	4	2	1	0	0
2	3	8	1	0	0	0	0
2	4	2	1	1	0	0	0
1	4	2	1	0	0	0	0
0	0	1	0	0	0	0	0

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	48	56	69
27	29	35	38	46	56	69	83

Code	Linear Quant Scale	Non-Linear Quant Scale
1	2	1
8	16	8
16	32	24
20	40	40
24	48	56
28	56	88
31	62	112

Figure 2-5. DCT coefficients are weighted by Quant Matrix and Quant Scale.

2.4 Scanning

In typical program material, the most significant DCT coefficients are generally found in or near the top-left corner of the matrix. After weighting, low-value coefficients might be truncated to zero. More efficient transmission can be obtained if all of the non-zero coefficients are sent first, followed by a code indicating that the remainder are all zero. Scanning is a technique that increases the probability of achieving this result, because it sends coefficients in descending order of magnitude probability. Figure 2-6a (see next page) shows that in a non-interlaced system, the probability of a coefficient having a high value is highest in the top-left corner and lowest in the bottom-right corner. A 45 degree diagonal zigzag scan is the best sequence to use here.

In Figure 2-6b, an alternative scan pattern is shown that may be used for interlaced sources. In an interlaced picture, an 8x8 DCT block from one field extends over twice the vertical screen area, so that for a given picture detail, vertical frequencies will appear to be twice as great as horizontal frequencies. Thus, the ideal scan for an interlaced picture will be on a diagonal that is twice as steep. Figure 2-6b shows that a given vertical spatial frequency is scanned before scanning the same horizontal spatial frequency.

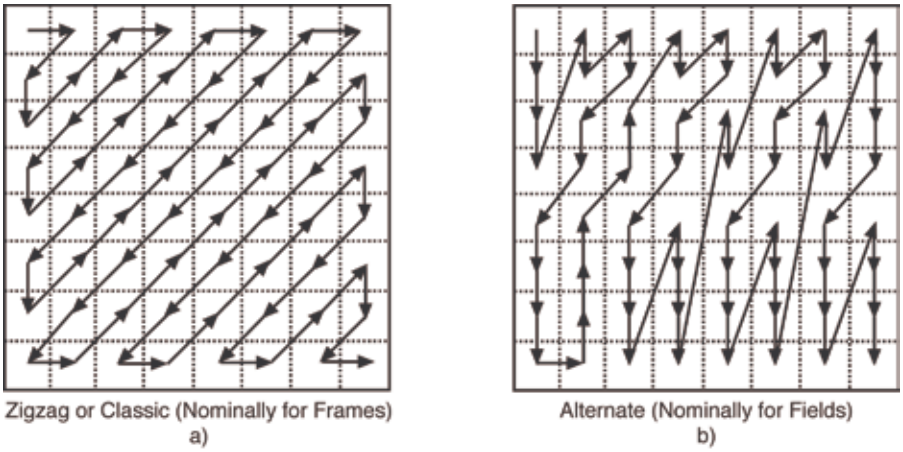


Figure 2-6. Two methods for serializing the DCT table. The alternate method is often used with interlaced video.

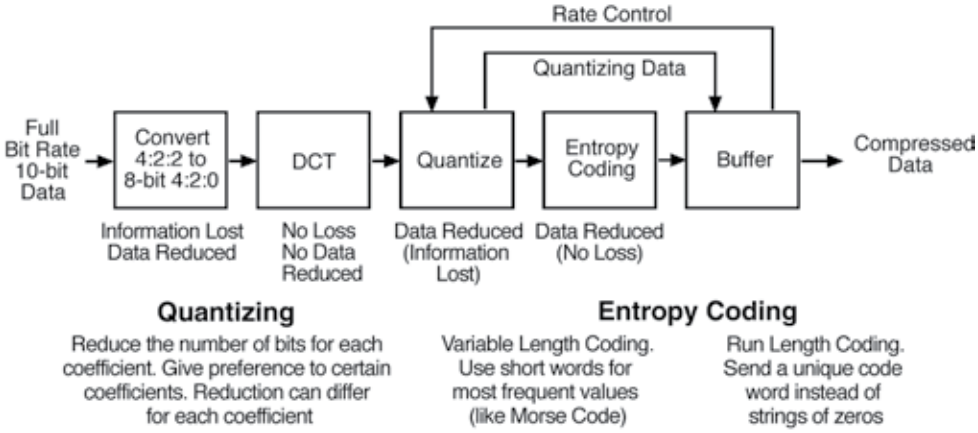


Figure 2-7. Rate Control and Quantization determine the output rate and quality.

2.5 Entropy Coding

In real video, not all spatial frequencies are present simultaneously; therefore, the DCT coefficient matrix will have zero terms in it. Requantization will increase the number of zeros by eliminating small values. Despite the use of scanning, zero coefficients will still appear between the significant values. Run length coding (RLC) allows these coefficients to be handled more efficiently. Where repeating values, such as a string of zeros, are present, RLC simply transmits the number of zeros rather than each individual bit.

The probability of occurrence of particular coefficient values in real video can be studied. In practice, some values occur very often; others occur less often. This statistical information can be used to achieve further compression using variable length coding (VLC). Frequently occurring values are converted to short code words, and infrequent values are converted to long code words. To aid decoding, no code word can be the prefix of another.

2.6 A Spatial Coder

Figure 2-7 ties together all of the preceding spatial coding concepts. The input signal is assumed to be 4:2:2 SDI (Serial Digital Interface), which may have 8- or 10-bit wordlength. MPEG uses only 8-bit resolution; therefore, a rounding stage will be needed when the SDI signal contains 10-bit words. Most MPEG profiles operate with 4:2:0 sampling; therefore, a vertical low-

pass filter/interpolation stage will be needed. Rounding and color subsampling introduces a small irreversible loss of information and a proportional reduction in bit rate. The raster scanned input format will need to be stored so that it can be converted to 8x8 pixel blocks.

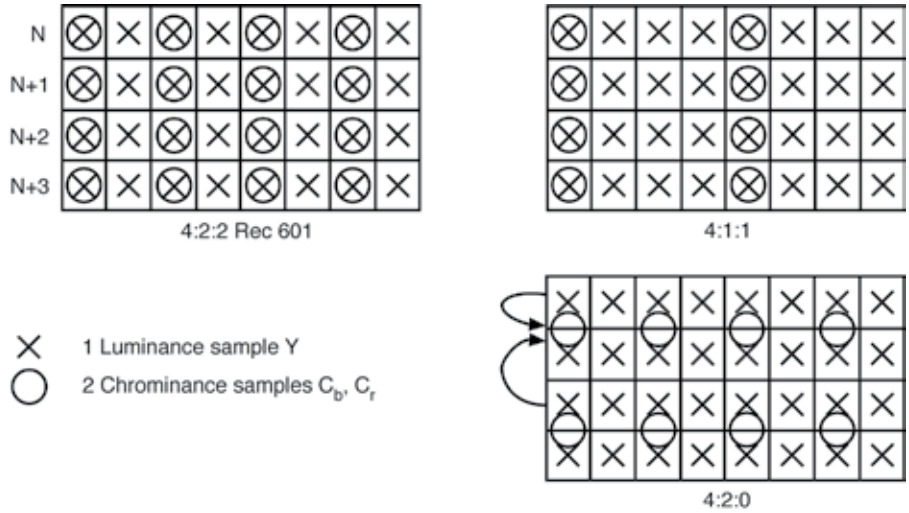


Figure 2-8. Luma/Chroma sample tables.

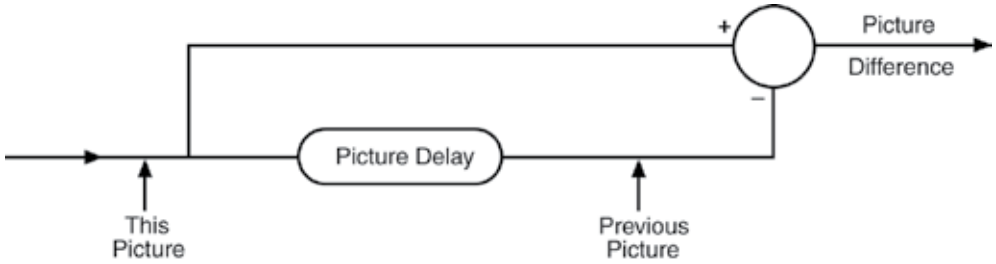


Figure 2-9. Picture differencing.

The DCT stage transforms the picture information to the frequency domain. The DCT itself does not achieve any compression. Following DCT, the coefficients are weighted and truncated, providing the first significant compression. The coefficients are then zigzag scanned to increase the probability that the significant coefficients occur early in the scan. After the last non-zero coefficient, an EOB (end of block) code is generated.

Coefficient data are further compressed by run-length and variable-length coding. In a variable bit-rate system, the quantizing may be fixed, but in a fixed bit-rate system, a buffer memory is used to absorb variations in coding difficulty. Highly detailed pictures will tend to fill the buffer, whereas plain pictures will allow it to empty. If the buffer is in danger of overflowing, the requantizing steps will have to be made larger, so that the compression factor is raised.

In the decoder, the bit stream is deserialized and the entropy coding is reversed to reproduce the weighted coefficients. The coefficients are placed in the matrix according to the zigzag scan, and inverse weighting is applied to recreate the block of DCT coefficients. Following an inverse transform, the 8x8 pixel block is recreated. To obtain a raster-scanned output, the blocks are stored in RAM, which is read a line at a time. To obtain a 4:2:2 output from 4:2:0 data, a vertical interpolation process will be needed as shown in Figure 2-8.

The chroma samples in 4:2:0 are positioned half way between luminance samples in the vertical axis so that they are evenly spaced when an interlaced source is used.

2.7 Temporal Coding

Temporal redundancy can be exploited by inter-coding or transmitting only the differences between pictures. Figure 2-9 shows that a one-picture delay combined with a subtracter can compute the picture differences. The picture difference is an image in its own right and can be further compressed by the spatial coder as was previously described. The decoder reverses the spatial coding and adds the difference picture to the previous picture to obtain the next picture.

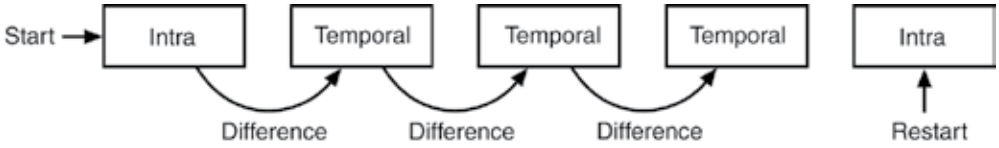


Figure 2-10. Bandwidth reduction achieved by sending Difference-frames between Intra-frames.

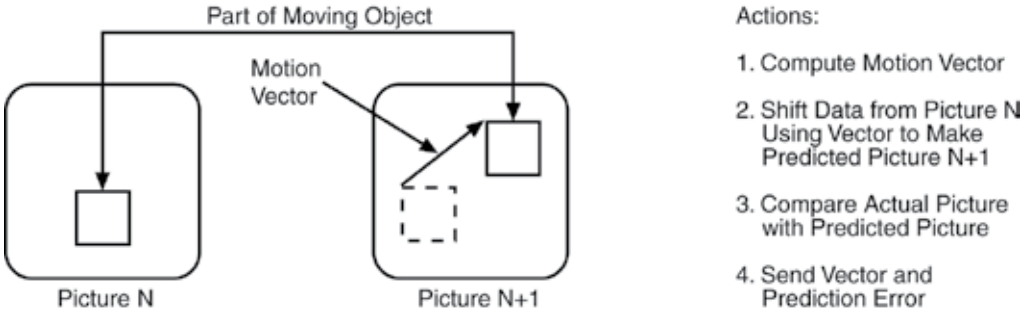


Figure 2-11. Picture N+1 created by sending Motion Vectors rather than DCT values.

There are some disadvantages to this simple system. First, as only differences are sent, it is impossible to begin decoding after the start of the transmission. This limitation makes it difficult for a decoder to provide pictures following a switch from one bit stream to another (as occurs when the viewer changes channels). Second, if any part of the difference data is incorrect, the error in the picture will propagate indefinitely.

The solution to these problems is to use a system that is not completely differential. Figure 2-10 shows that periodically complete pictures are sent. These are called Intra-coded pictures (or I-pictures), and they are obtained by spatial compression only. If an error or a channel switch occurs, it will be possible to resume correct decoding at the next I-picture.

2.8 Motion Compensation

Motion reduces the similarities between pictures and increases the data needed to create the difference picture. Motion compensation is used to increase the similarity. Figure 2-11 shows the principle. When an object moves across the TV screen, it may appear in a different place in each picture, but it does not change in appearance very much. The picture difference can be reduced by measuring the motion at the encoder. This is sent to the decoder as a vector. The decoder uses the vector to shift part of the previous picture to a more appropriate place in the new picture.

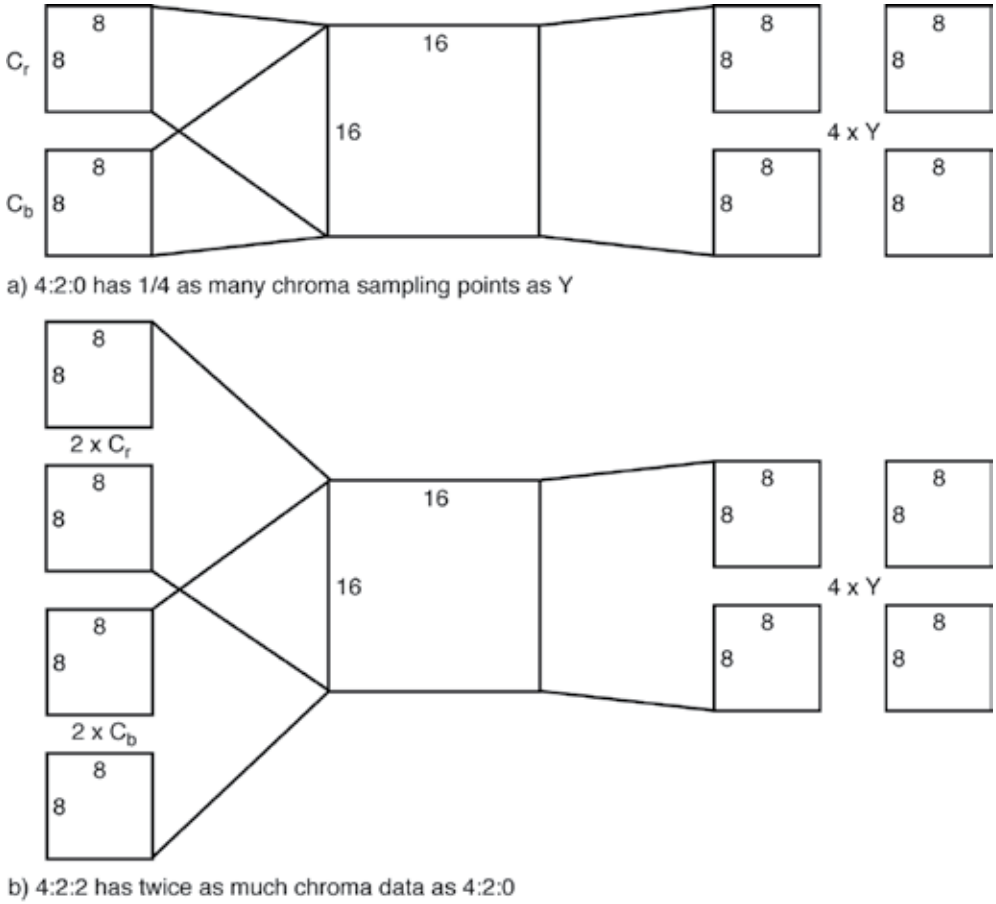


Figure 2-12. 4:2:0 sample size vs. 4:2:2 sample size.

One vector controls the shifting of an entire area of the picture that is known as a macroblock. The size of the macroblock is determined by the DCT coding and the color subsampling structure. Figure 2-12a shows that, with a 4:2:0 system, the vertical and horizontal spacing of color samples is exactly twice the spacing of luminance. A single 8x8 DCT block of color samples extends over the same area as four 8x8 luminance blocks; therefore this is the minimum picture area that can be shifted by a vector. One 4:2:0 macroblock contains four luminance blocks: one C_b block and one C_r block.

In the 4:2:2 profile, color is only subsampled in the horizontal axis. Figure 2-12b shows that in 4:2:2, a single 8x8 DCT block of color samples extends over two luminance blocks. A 4:2:2 macroblock contains four luminance blocks: two C_b blocks and two C_r blocks.

The motion estimator works by comparing the luminance data from two successive pictures. A macroblock in the first picture is used as a reference. The correlation between the reference and the next picture is measured at all possible displacements with a resolution of half a pixel over the entire search range. When the greatest correlation is found, this correlation is assumed to represent the correct motion.

The motion vector has a vertical and horizontal component. In typical program material, a moving object may extend over a number of macroblocks. A greater compression factor is obtained if the vectors are transmitted differentially. When a large object moves, adjacent macroblocks have the same vectors and the vector differential becomes zero.

Motion vectors are associated with macroblocks, not with real objects in the image and there will be occasions where part of the macroblock moves and part of it does not. In this case, it is impossible to compensate properly. If the motion of the moving part is compensated by transmitting a vector, the stationary part will be incorrectly shifted, and it will need difference data to be corrected. If no vector is sent, the stationary part will be correct, but difference data will be needed to correct the moving part. A practical compressor might attempt both strategies and select the one that required the least data.

2.9 Bidirectional Coding

When an object moves, it conceals the background at its leading edge and reveals the background at its trailing edge. The revealed background requires new data to be transmitted because the area of background was previously concealed and no information can be obtained from a previous picture. A similar problem occurs if the camera pans; new areas come into view and nothing is known about them. MPEG helps to minimize this problem by using bidirectional coding, which allows information to be taken from pictures before and after the current picture. If a background is being revealed, it will be present in a later picture, and the information can be moved backwards in time to create part of an earlier picture.

Figure 2-13 shows the concept of bidirectional coding. On an individual macroblock basis, a bidirectionally-coded picture can obtain motion-compensated data from an earlier or later picture, or even use an average of earlier and later data. Bidirectional coding significantly reduces the amount of difference data needed by improving the degree of prediction possible. MPEG does not specify how an encoder should be built, only what constitutes a compliant bit stream. However, an intelligent compressor could try all three coding strategies and select the one that results in the least data to be transmitted.

2.10 I-, P- and B-pictures

In MPEG, three different types of pictures are needed to support differential and bidirectional coding while minimizing error propagation:

I-pictures are intra-coded pictures that need no additional information for decoding. They require a lot of data compared to other picture types, and therefore they are not transmitted any more frequently than necessary. They consist primarily of transform coefficients and have no vectors. I-pictures are decoded without reference to any other pictures, so they allow the viewer to switch channels, and they arrest error propagation.

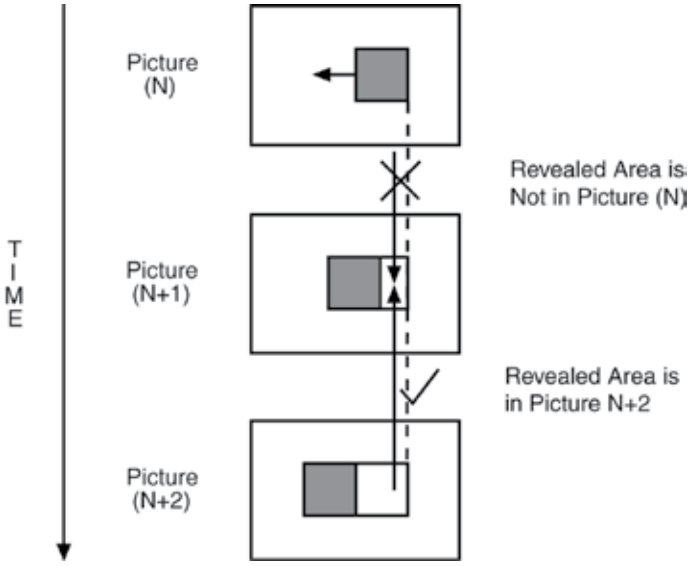


Figure 2-13. Area revealed as a result of moving a block.

P-pictures are forward predicted from an earlier picture, which could be an I-picture or a P-picture. P-picture data consists of vectors describing where, in the previous picture, each macroblock should be taken from, and transform coefficients that describe the correction or difference data that must be added to that macroblock. Where no suitable match for a macroblock could be found by the motion compensation search, intra data is sent to code that macroblock. P-pictures require roughly half the data of an I-picture.

B-pictures are bidirectionally predicted from earlier and/or later I- or P-pictures. B-picture data consists of vectors describing where in earlier or later pictures data should be taken from. It also contains the intracoded data that provide necessary corrections. Again, when no suitable match for a macroblock is found by the motion compensation search, intra data is sent to code that macroblock. Bidirectional prediction is quite effective, so most macroblocks in a B-picture will be coded largely by motion vectors. Also, a B-picture is never used as a reference for coding other pictures, so there is no possibility of error propagation. This permits encoders to use more aggressive requantization for correction data. A typical B-picture requires about one quarter the data of an I-picture.

Note that a B-picture does not have to use both directions of prediction; in some circumstances only one direction is employed. This option may be used when constructing closed groups of pictures (GOP).

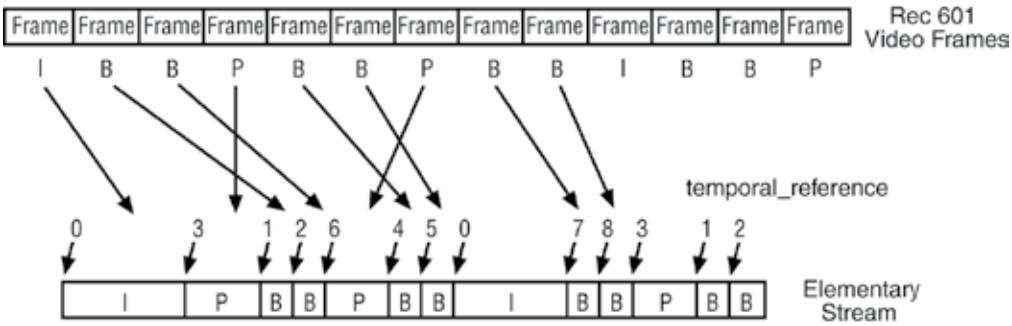


Figure 2-14. Input frame sequence sent out of order to accommodate revealed areas.

Figure 2-14 introduces the concept of the GOP. The GOP represents the structure of I-, P-, and B-pictures in the sequence. Generally the GOP structure repeats through the sequence, but the GOP length and structure may be changed at any time. There are no formal limits on the length of a GOP, but for transmission purposes a typical length is 12 or 15 pictures.

The nature of MPEG's temporal compression means that the transmission order of pictures is not the same as the display order. A P-picture naturally follows the I- or P-picture from which it is predicted, so there are no special requirements. A bidirectionally-coded B-picture, however, cannot be decoded until both of its reference pictures have been received and decoded. Figure 2-14 shows the pictures of a GOP in display order at the top, and in transmission order below. Note that, in transmission order, the B-pictures always follow the two reference pictures from which they are predicted.

There are two types of GOP, open and closed. A closed GOP requires no reference outside itself. In display order, it may begin with an I-picture and end with a P-picture. In transmission order there will usually be B-pictures following the last P-picture, but these are pictures that will be displayed before that last P-picture.

It is possible to start and or end a *closed GOP* with B-pictures (in display order), but in this case the starting and ending B-pictures must be coded using only a single direction of prediction. B-pictures at the start of a closed GOP must use backward prediction only. B-pictures at the end of a closed GOP may use forward prediction only – similar to a P-picture, but B-picture rules would be used for requantization etc.

An *open GOP* does not have these restrictions on prediction vectors. For example, B-pictures at the end of the GOP can use forward prediction from the last P-picture and backward prediction from the first I-picture of the next GOP. This structure is slightly more efficient, but predictions can cross any picture boundary. It is much more difficult to splice bit streams, and events such as channel changes are more likely to cause picture errors.

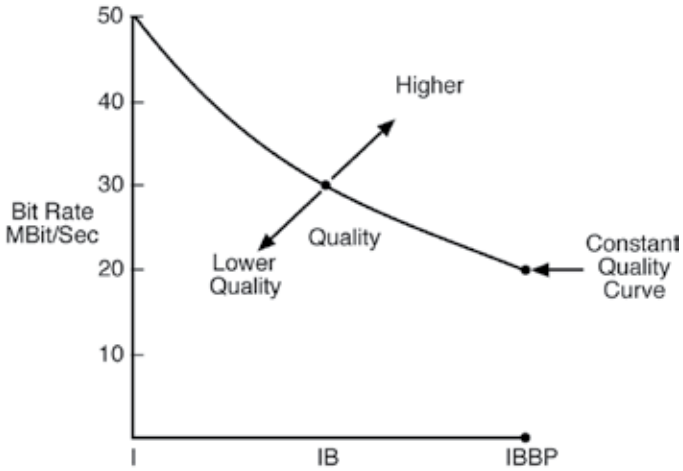


Figure 2-15. Bandwidth and quality decreases as GOP increases.

The GOP structure may be altered by the encoder when there are scene changes. Predictions across a scene change will usually fail, since there will be large entropy between the two pictures either side of the scene change. An encoder may choose to detect the scene change, use a closed GOP leading up to the scene change, and start a new GOP (open or closed) with an I-picture representing the first picture of the new scene.

Sending picture data out of sequence requires additional memory at the encoder and decoder and also causes delay. The number of bidirectionally-coded pictures between intra- or forward-predicted pictures must be restricted to reduce cost, and to minimize delay if this is an issue.

Figure 2-15 shows the trade-off that must be made between compression factor and coding delay. For a given quality, sending only I-pictures requires more than twice the bit rate of an IBBP sequence.

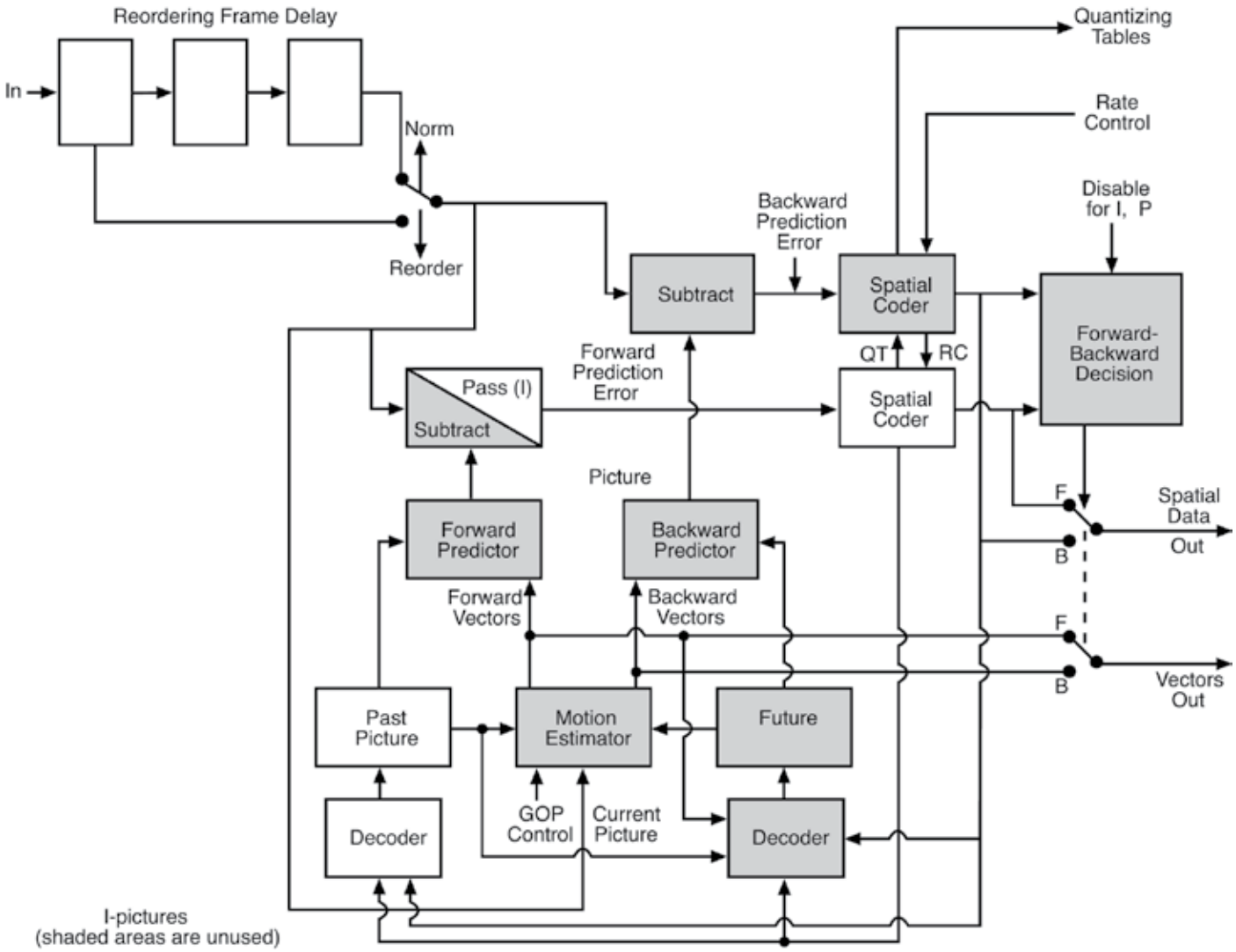


Figure 2-16a. MPEG-2 video encoder flowchart (I-frames highlighted).

2.11 An MPEG Compressor

Figures 2-16a, b, and c show a typical bidirectional motion compensator structure. Pre-processed input video enters a series of frame stores that can be bypassed to change the picture order. The data then enter the subtracter and the motion estimator. To create an I-picture, the end of the input delay is selected and the subtracter is turned off so that the data pass straight through to be spatially coded (see Figure 2-16a). Subtracter output data also pass to a frame store that can hold several pictures. The I-picture is held in the store.

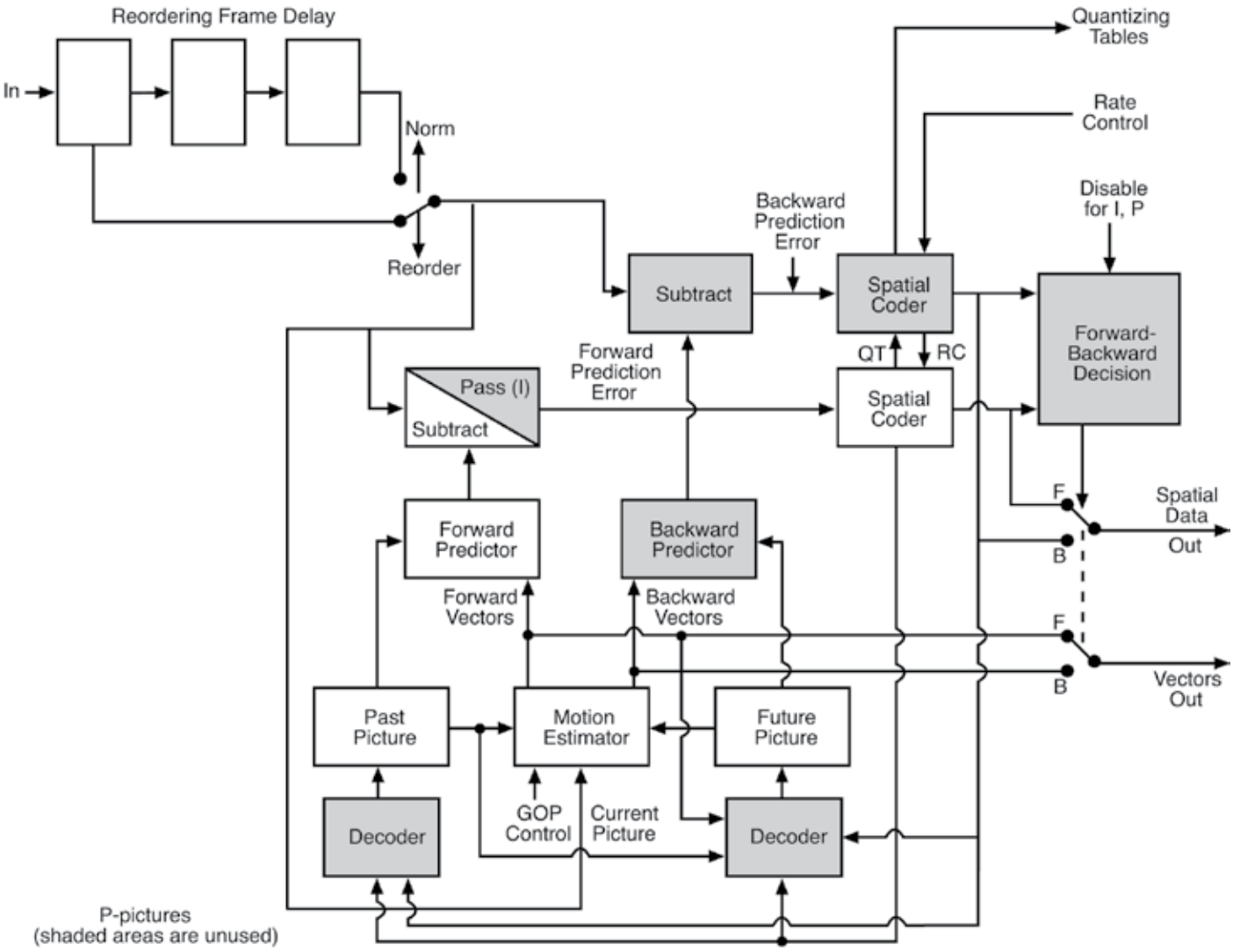


Figure 2-16b. MPEG-2 video encoder flowchart (P-frames highlighted).

To encode a P-picture, the B-pictures in the input buffer are bypassed, so that the future picture is selected (see Figure 2-16b). The motion estimator compares the I-picture in the output store with the P-picture in the input store to create forward motion vectors. The I-picture macroblocks are shifted by these vectors to make a predicted P-picture. The predicted P-picture is subtracted from the actual P-picture to produce the prediction error, which is spatially coded and sent along with the vectors. The prediction error is also added to the predicted P-picture to create a locally decoded P-picture that also enters the output store.

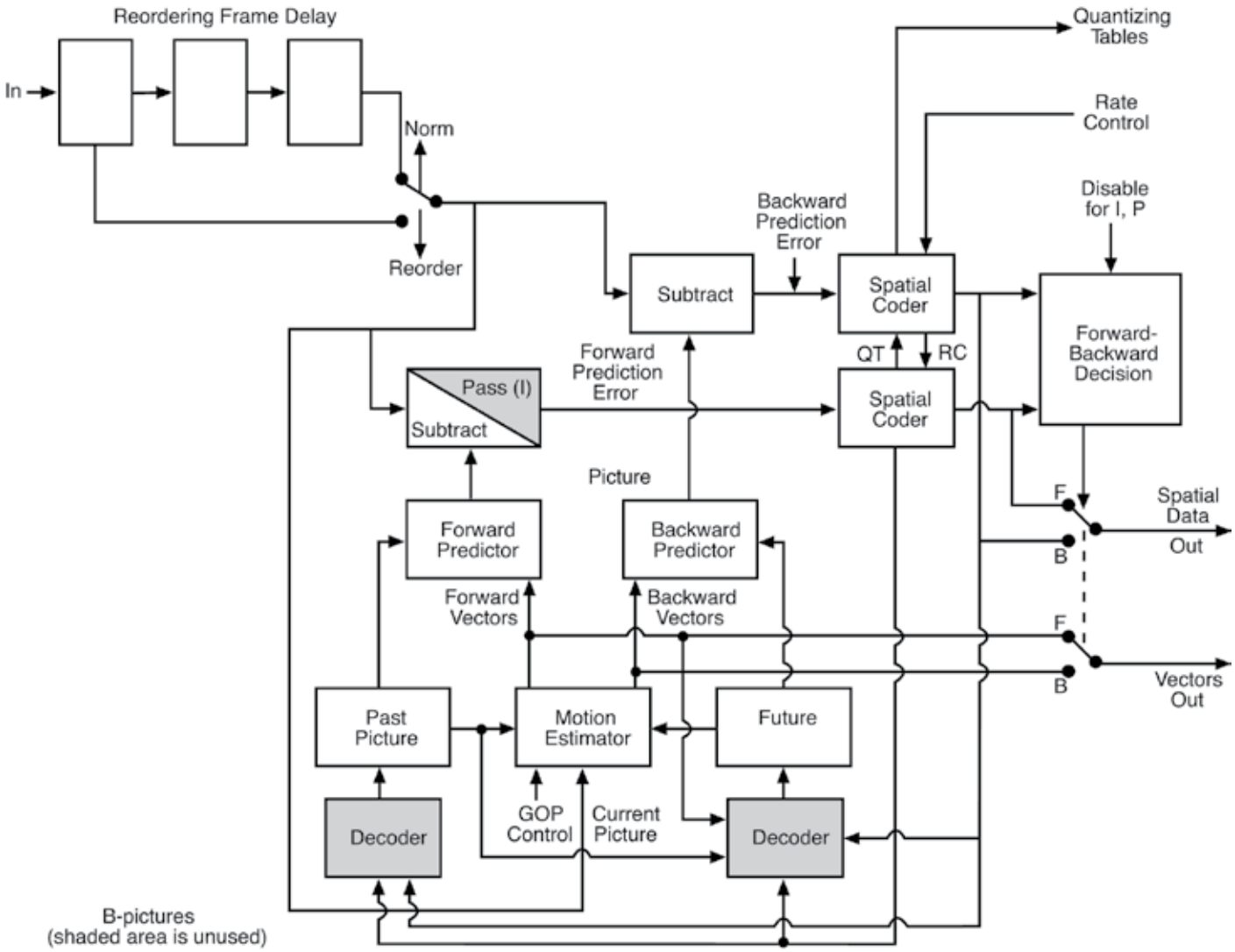


Figure 2-16c. MPEG-2 video encoder flowchart (B-frames highlighted).

The output store then contains an I-picture and a P-picture. A B-picture from the input buffer can now be selected. The motion compensator will compare the B-picture with the I-picture that precedes it and the P-picture that follows it to obtain bidirectional vectors (see Figure 2-16c). Forward and backward motion compensation is performed to produce two predicted B-pictures. These are subtracted from the current B-picture. On a macroblock- by-macroblock basis, the forward or backward data are selected according to which represent the smallest differences. The picture differences are then spatially coded and sent with the vectors.

When all of the intermediate B-pictures are coded, the input memory will once more be bypassed to create a new P-picture based on the previous P-picture.

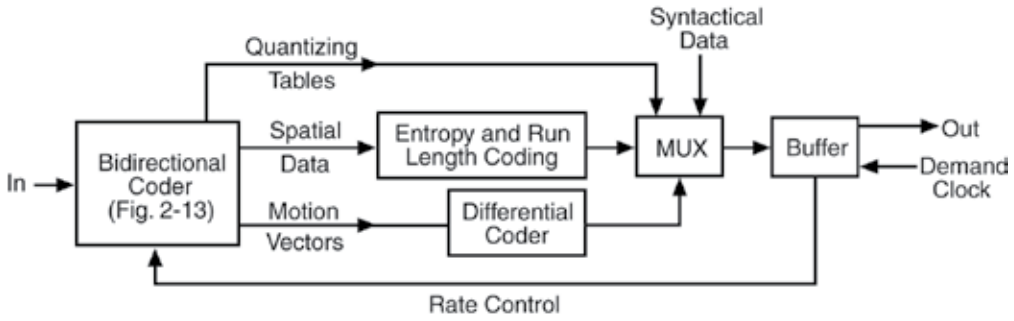


Figure 2-17. Rate controller diagram.

Figure 2-17 shows an MPEG coder. The motion compensator output is spatially coded and the vectors are added in a multiplexer. Syntactical data is also added which identifies the type of picture (I, P, or B) and provides other information to help a decoder (see Section 5 – Elementary Streams). The output data are buffered to allow temporary variations in bit rate. If the mean bit rate is too high, the buffer will tend to fill up. To prevent overflow, quantization will have to be made more severe. Equally, should the buffer show signs of underflow, the quantization will be relaxed to maintain the average bit rate.

2.12 Preprocessing

A compressor attempts to eliminate redundancy within the picture and between pictures. Anything that reduces that apparent redundancy, that is not picture content, is undesirable. Noise and film grain are particularly problematic because they generally occur over the entire picture. After the DCT process, noise results in more non-zero coefficients, and the coder cannot distinguish this information from genuine picture data. Heavier quantizing will be required to encode all of the coefficients, reducing picture quality. Noise also reduces similarities between successive pictures, increasing the difference data needed.

Residual subcarrier in video decoded from composite video is a serious problem because it results in high, spatial frequencies that are normally at a low level in component programs. Subcarrier also alternates in phase from picture to picture causing an increase in difference data. Naturally, any composite decoding artifact that is visible in the input to the MPEG coder is likely to be reproduced at the decoder.

Any practice that causes unwanted motion is to be avoided. Unstable camera mountings, in addition to giving a shaky picture, increase picture differences and vector transmission requirements. This will also happen with telecine material if sprocket hole damage results in film weave or hop. In general, video that is to be compressed must be of the highest quality possible. If high quality cannot be achieved, then noise reduction and other stabilization techniques will be desirable.

If a high compression factor is required, the level of artifacts can increase, especially if input quality is poor. In this case, it may be better to reduce the entropy presented to the coder by using pre-filtering. The video signal is subject to two-dimensional, low-pass filtering, which reduces the number of coefficients needed and reduces the level of artifacts. The picture will be less sharp, but less sharpness is preferable to a high level of artifacts.

In most MPEG-2 applications, 4:2:0 sampling is used, which requires a chroma downsampling process if the source is 4:2:2. In MPEG-1,the luminance and chroma are further downsampled to produce an input picture or CIF (common image format) that is only 352-pixels wide. This technique reduces the entropy by a further factor. For very high compression, the QCIF (quarter common image format) picture, which is 176-pixels wide, is used. Downsampling is a process that combines a spatial low-pass filter with an interpolator. Downsampling interlaced signals is problematic because vertical detail is spread over two fields that may de-correlate due to motion.

When the source material is telecine, the video signal has different characteristics than normal video. In 50-Hz video, pairs of fields represent the same film frame, and there is no motion between them. Thus, the motion between fields alternates between zero and the motion between frames. In 60-Hz video, 3:2 pulldown is used to obtain 60 Hz from 24 Hz film. One frame is made into two fields; the next is made into three fields, and so on.

Consequently, one field in five is completely redundant. MPEG handles film material best by discarding the third field in 3:2 systems. A 24-Hz code in the transmission alerts the decoder to recreate the 3:2 sequence by re-reading a field store. In 50- and 60-Hz telecine, pairs of fields are deinterlaced to create frames, and then motion is measured between frames. The decoder can recreate interlace by reading alternate lines in the frame store.

A cut is a difficult event for a compressor to handle because it often results in an almost complete prediction failure, requiring a large amount of correction data. If a coding delay can be tolerated, a coder may detect cuts in advance and modify the GOP structure dynamically, so that an I-picture is inserted to coincide with the cut. In this case, the cut is handled with very little extra data. The last B-pictures before the I frame will almost certainly need to use forward prediction. In some applications that are not real-time, such as DVD mastering, a coder could take two passes at the input video: one pass to identify the difficult or high entropy areas and create a coding strategy, and a second pass to actually compress the input video.

2.13 Wavelets

All transforms suffer from uncertainty because the more accurately the frequency domain is known, the less accurately the time domain is known (and vice versa). In most transforms such as discreet Fourier transport (DFT) and discreet cosine transform (DCT), the block length is fixed, so the time and frequency resolution is fixed. The frequency coefficients represent evenly spaced values on a linear scale. Unfortunately, because human senses are logarithmic, the even scale of the DFT and DCT gives inadequate frequency resolution at one end and excess resolution at the other.

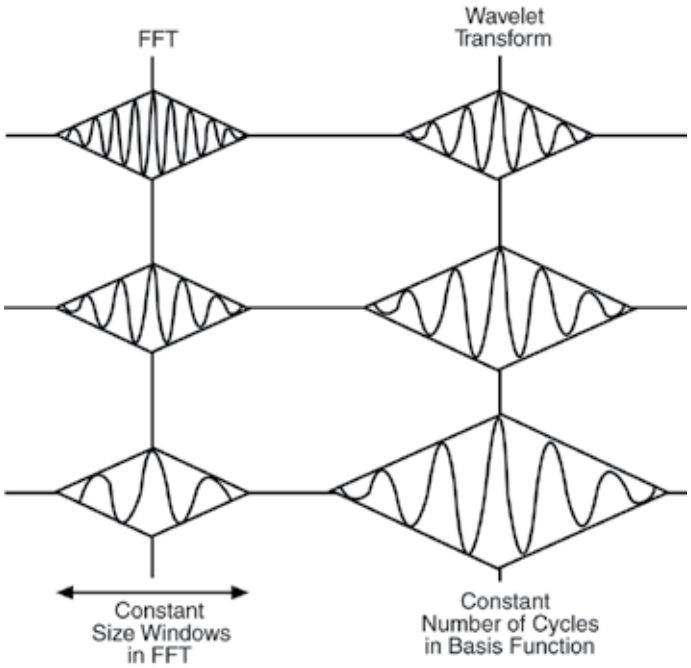


Figure 2-18. MPEG audio use fixed-size FFT while wavelet uses variable size.

The wavelet transform is not affected by this problem because its frequency resolution is a fixed fraction of an octave and therefore has a logarithmic characteristic. This is done by changing the block length as a function of frequency. As frequency goes down, the block becomes longer. Thus, a characteristic of the wavelet transform is that the basis functions all contain the same number of cycles, and these cycles are simply scaled along the time axis to search for different frequencies. Figure 2-18 contrasts the fixed block size of the DFT/DCT with the variable size of the wavelet.

Wavelets are especially useful for audio coding because they automatically adapt to the conflicting requirements of the accurate location of transients in time and the accurate assessment of pitch in steady tones.

For video coding, wavelets have the advantage of producing resolution-scalable signals with almost no extra effort. In moving video, the advantages of wavelets are offset by the difficulty of assigning motion vectors to a variable size block, but in still-picture or I-picture coding this difficulty is not an issue. Wavelet coding has shown particular benefits for very-low bit rate applications. The artifacts generated by excessive quantization of wavelet coefficients generally appear as “smearing,” and these are much less objectionable than the “blockiness” that results from excessive quantization of DCT coefficients.

Section 3 – Audio Compression

Lossy audio compression is based entirely on the characteristics of human hearing, which must be considered before any description of compression is possible. Surprisingly, human hearing, particularly in stereo, is actually more critically discriminating than human vision, and consequently audio compression should be undertaken with care. As with video compression, audio compression requires a number of different levels of complexity according to the required compression factor.

3.1 The Hearing Mechanism

Hearing comprises physical processes in the ear and nervous/mental processes that combine to give us an impression of sound. The impression we receive is not identical to the actual acoustic waveform present in the ear canal because some entropy is lost. Audio compression systems that lose only that part of the entropy that will be lost in the hearing mechanism will produce good results.

The physical hearing mechanism consists of the outer, middle, and inner ears. The outer ear comprises the ear canal and the eardrum. The eardrum converts the incident sound into a vibration, in much the same way as a microphone diaphragm. The inner ear works by sensing vibrations transmitted through a fluid. The impedance of fluid is much higher than that of air and the middle ear acts as an impedance-matching transformer that improves power transfer.

Figure 3-1 shows that vibrations are transferred to the inner ear by the stirrup bone, which acts on the oval window. Vibrations in the fluid in the ear travel up the cochlea, a spiral cavity in the skull (shown unrolled in Figure 3-1 for clarity). The basilar membrane is stretched across the cochlea. This membrane varies in mass and stiffness along its length. At the end near the oval window, the membrane is stiff and light, so its resonant frequency is high. At the distant end, the membrane is heavy and soft and resonates at low frequency. The range of resonant frequencies available determines the frequency range of human hearing, which in most people is from 20 Hz to about 15 kHz.

Different frequencies in the input sound cause different areas of the membrane to vibrate. Each area has different nerve endings to allow pitch discrimination. The basilar membrane also has tiny muscles controlled by the nerves that together act as a kind of positive-feedback system that improves the Q-factor of the resonance.

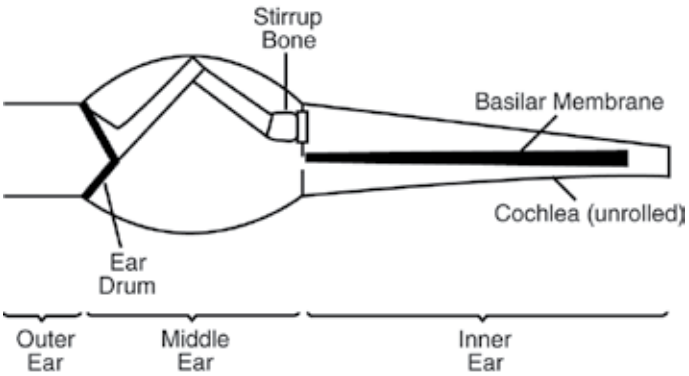


Figure 3-1. Components of the human ear.

The resonant behavior of the basilar membrane is an exact parallel with the behavior of a transform analyzer. According to the uncertainty theory of transforms, the more accurately the frequency domain of a signal is known, the less accurately the time domain is known. Consequently, the more able a transform is to discriminate between two frequencies, the less able it is to discriminate between the time of two events. Human hearing has evolved with a certain compromise that balances time-uncertainty discrimination and frequency discrimination; in the balance, neither ability is perfect.

The imperfect frequency discrimination results in the inability to separate closely spaced frequencies. This inability is known as auditory masking, defined as the reduced sensitivity to sound in the presence of another.

Figure 3-2a (see next page) shows that the threshold of hearing is a function of frequency. The greatest sensitivity is, not surprisingly, in the speech range. In the presence of a single tone, the threshold is modified as in Figure 3-2b. Note that the threshold is raised for tones at higher frequency and to some extent at lower frequency. In the presence of a complex input spectrum, such as music, the threshold is raised at nearly all frequencies. One consequence of this behavior is that the hiss from an analog audio cassette is only audible during quiet passages in music. Companding makes use of this principle by amplifying low-level audio signals prior to recording or transmission and returning them to their correct level afterwards.

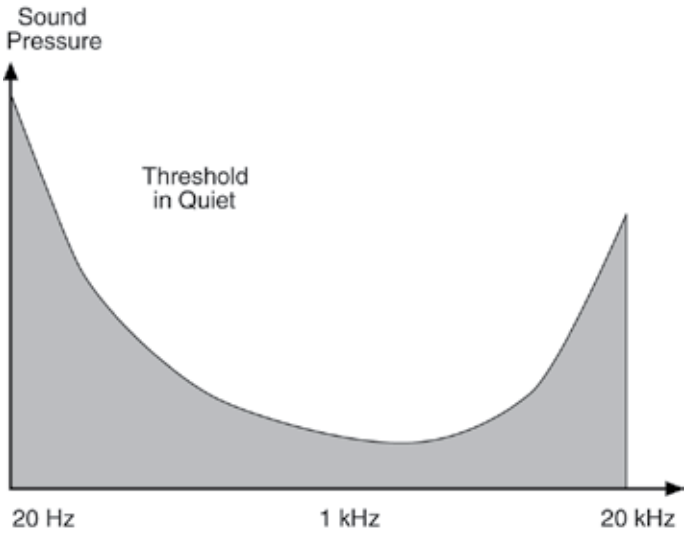


Figure 3-2a. Frequency range of the human ear.

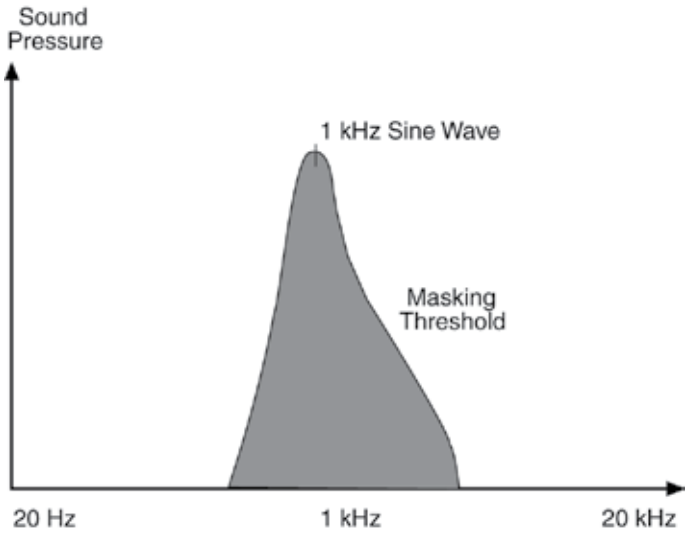


Figure 3-2b. MPEG masking threshold, based upon the ear being more sensitive to the human voice range.

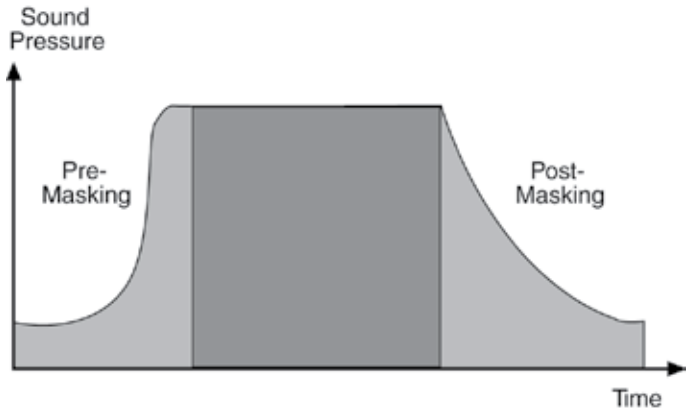


Figure 3-3. Temporal masking applied due to human ear recovery time as amplitude changes.

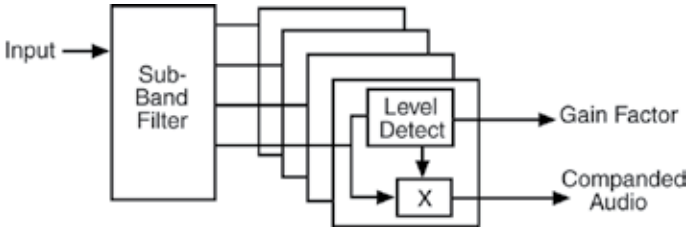


Figure 3-4. Audio compression uses different gain levels on different frequency bands.

The imperfect time discrimination of the ear is due to its resonant response. The Q-factor is such that a given sound has to be present for at least about 1 millisecond before it becomes audible. Because of this slow response, masking can still take place even when the two signals involved are not simultaneous. Forward and backward masking occur when the masking sound continues to mask sounds at lower levels before and after the masking sound's actual duration. Figure 3-3 shows this concept.

Masking raises the threshold of hearing, and compressors take advantage of this effect by raising the noise floor, which allows the audio waveform to be expressed with fewer bits. The noise floor can only be raised at frequencies at which there is effective masking. To maximize effective masking, it is necessary to split the audio spectrum into different frequency bands to allow introduction of different amounts of companding and noise in each band.

3.2 Subband Coding

Figure 3-4 shows a band-splitting compander. The band-splitting filter is a set of narrow-band, linear-phase filters that overlap and all have the same bandwidth. The output in each band consists of samples representing a waveform. In each frequency band, the audio input is amplified up to maximum level prior to transmission. Afterwards, each level is returned to its correct value. Noise picked up in the transmission is reduced in each band. If the noise reduction is compared with the threshold of hearing, it can be seen that greater noise can be tolerated in some bands because of masking. Consequently, in each band, after companding it is possible to reduce the wordlength of samples. This technique achieves compression because the quantization noise introduced by the loss of resolution is masked.

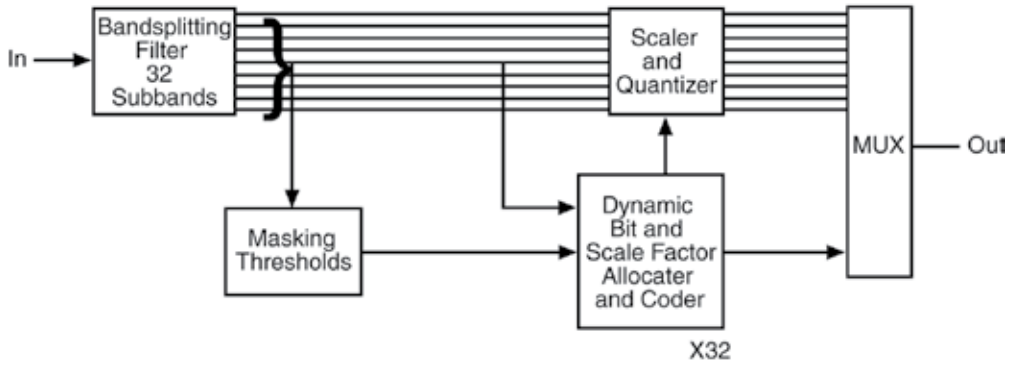


Figure 3-5. MPEG-1 using 32-subband encoder.

Figure 3-5 shows a simple band-splitting coder as is used in MPEG layer 1. The digital audio input is fed to a band-splitting filter that divides the spectrum of the signal into a number of bands. In MPEG, this number is 32. The time axis is divided into blocks of equal length. In MPEG layer 1, there are 384 input samples, so there are 12 samples in each of 32 bands in the output of the filter. Within each band, the level is amplified by multiplication to bring the level up to maximum. The gain required is constant for the duration of a block, and a single scale factor is transmitted with each block for each band in order to allow the process to be reversed at the decoder.

The filter bank output for MPEG layer 1 is analyzed using a 512-point FFT to determine the spectrum of the input signal. This analysis drives a masking model that determines the degree of masking that can be expected in each band. The more masking available, the less accurate the samples in each band need to be. The sample accuracy is reduced by requantizing to reduce wordlength. This reduction is also constant for every word in a band, but different bands can use different wordlengths. The wordlength needs to be transmitted as a bit allocation code for each band to allow the decoder to deserialize the bit stream properly.

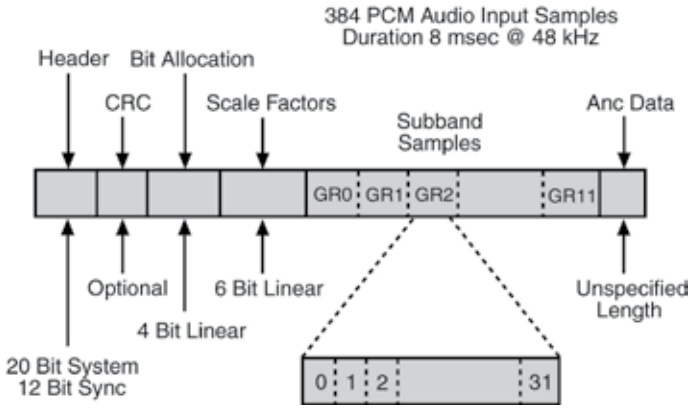


Figure 3-6. MPEG-1 frame headers and payload.

3.3 MPEG Layer 1

Figure 3-6 shows an MPEG layer 1 audio bit stream. Following the synchronizing pattern and the header, there are 32-bit allocation codes of four bits each. These codes describe the wordlength of samples in each subband. Next come the 32 scale factors used in the companding of each band. These scale factors determine the gain needed in the decoder to return the audio to the correct level. The scale factors are followed, in turn, by the audio data in each band.

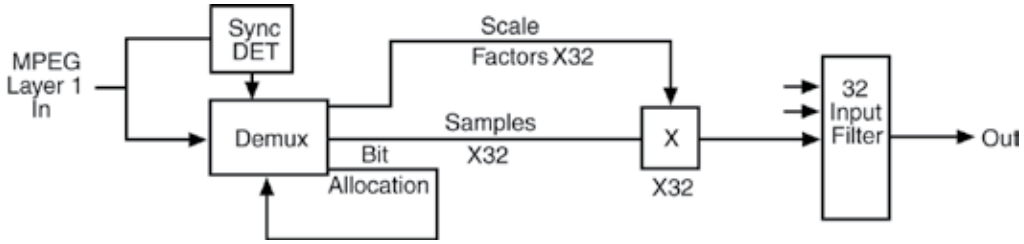


Figure 3-7. MPEG 32-subband decoder.

Figure 3-7 shows the layer 1 decoder. The synchronization pattern is detected by the timing generator, which deserializes the bit allocation and scale factor data. The bit allocation data then allows deserialization of the variable length samples. The requantizing is reversed and the compression is reversed by the scale factor data to put each band back to the correct level. These 32 separate bands are then combined in a combiner filter that produces the audio output.

3.4 MPEG Layer 2

Figure 3-8 shows that when the band-splitting filter is used to drive the masking model, the spectral analysis is not very accurate, since there are only 32 bands and the energy could be anywhere in the band. The noise floor cannot be raised very much because, in the worst case shown, the masking may not operate. A more accurate spectral analysis would allow a higher compression factor. In MPEG layer 2, the spectral analysis is performed by a separate process. A 1024-point FFT (Fast Fourier Transform) working directly from the input is used to drive the masking model instead. To resolve frequencies more accurately, the time span of the transform has to be increased, which is done by raising the block size to 1152 samples.

While the block-companding scheme is the same as in layer 1, not all of the scale factors are transmitted, since they contain a degree of redundancy on real program material. The scale factor of successive blocks in the same band differs by 2 dB less than 10 percent of the time, and advantage is taken of this characteristic by analyzing sets of three successive scale factors. When the sub- band content is approximately constant (known as stationary or quasi-stationary program), only one scale factor out of three is sent. As transient content increases in a given subband, two or three scale factors will be sent. A scale factor select code is also sent to allow the decoder to determine what has been sent in each subband. This technique effectively halves the scale factor bit rate.

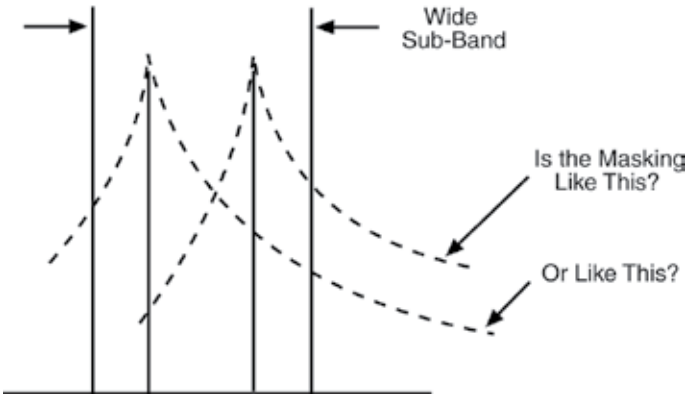


Figure 3-8. 32-subband filtering creates abiguity and lower quality compared to the 1024-subband filtering in MPEG-2.

3.5 Transform Coding

Layers 1 and 2 are based on band-splitting filters in which the signal is still represented as a waveform. However, layer 3 adopts transform coding similar to that used in video coding. As was mentioned above, the ear performs a kind of frequency transform on the incident sound and, because of the Q-factor of the basilar membrane, the response cannot increase or reduce rapidly. Consequently, if an audio waveform is transformed into the frequency domain, the coefficients do not need to be sent very often. This principle is the basis of transform coding. For higher compression factors, the coefficients can be requantized, making them less accurate. This process produces noise that will be placed at frequencies where the masking is the greatest. A by-product of a transform coder is that the input spectrum is accurately known, so a precise masking model can be created.

3.6 MPEG Layer 3

This complex level of coding is really only required when the highest compression factor is needed. It has a degree of commonality with layer 2. A discrete cosine transform is used having 384 output coefficients per block. This output can be obtained by direct processing of the input samples, but in a multi-level coder, it is possible to use a hybrid transform incorporating the 32-band filtering of layers 1 and 2 as a basis. If this is done, the 32 subbands from the QMF (quadrature mirror filter) are each further processed by a 12-band MDCT (modified discreet cosine transform) to obtain 384 output coefficients.

Two window sizes are used to avoid pre-echo on transients. The window switching is performed by the psycho-acoustic model. It has been found that pre-echo is associated with the entropy in the audio rising above the average value. To obtain the highest compression factor, non-uniform quantizing of the coefficients is used along with Huffman coding. This technique allocates the shortest wordlengths to the most common code values.

3.7 MPEG-2 Audio

Although originally designated MPEG audio levels 1, 2 and 3, the systems are now more accurately known as MPEG-1 Level 1, etc. MPEG-2 defined extensions to MPEG-1 audio, and a new advanced coding system.

MPEG-2 permits the use of sampling at lower rates than MPEG-1. This is not strictly backward compatible, but requires only additional tables in an MPEG-1 decoder for interoperability.

MPEG-2 BC (backward compatible) audio provides for 5.1 channels (five full-bandwidth channels plus a low-bandwidth low frequency effects channel). MPEG-2 BC has an MPEG-1 (2 channel) bit stream at its core and adds the multi-channel extensions in a form that will be ignored by an MPEG-1 decoder.

MPEG-2 AAC (advanced audio coding) is a more sophisticated system with higher resolution filter banks and additional coding tools. It offers significantly higher coding efficiency, but is not backward compatible.

3.8 MPEG-4 Audio

MPEG-4 coding is based on objects. (See Section 4.4.2.) MPEG-4 audio objects can represent natural or synthetic sounds. For natural audio coding, the MPEG-4 toolkit includes MPEG-2 AAC as well as a variety of other tools, including parametric encoding for very low bit rates and a technique known as code excited linear predictive (CELP) coding for speech coding in the mid range of bit rates. Various forms of scalability are supported, including bitstream scalability that can be applied at points in the transmission system.

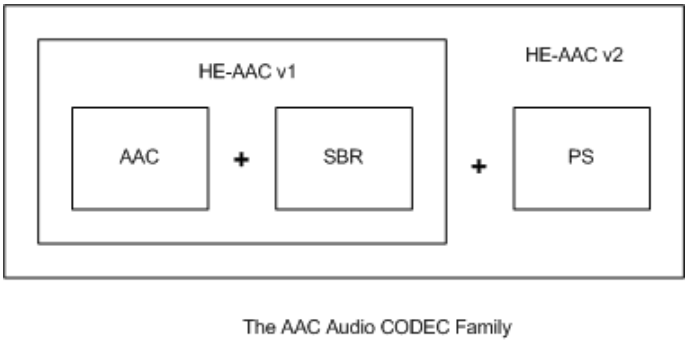


Figure 3-9. MPEG-4 audio with v2 as a superset of v1. v2 adds Parametric Stereo.

The use of object coding permits choices to be made at the decoding point. For example, a concerto may be transmitted as two objects, orchestra and solo. Normal decoding would present the complete work, but an instrumentalist could decode only the orchestra object and perform the solo part "live." Similar approaches could permit coding of programs so that listeners could select a "mix minus" mode to eliminate commentary from, say, a sporting event.

In 2001, a High Efficiency variant of MPEG-4 AAC was introduced. This is intended to provide superior performance at lower bit rates, typically around 128 Kbit/s. HE-AAC, also known as aacPlus, is really a superset of AAC that adds an enhancement technique, known as Spectral Band Replication (SBR). SBR is a bandwidth extension technique that enables audio CODECs to deliver the same listening experience at approximately half the bit rate that the core CODEC would require, if operated on its own.

More recently, HE-AAC v2 was introduced, as illustrated by Figure 3-9. Again, this is a superset of v1 and adds a Parametric Stereo (PS) enhancement. Parametric Stereo increases the coding efficiency by exploiting a parametric representation of the stereo image of a given input signal.

Given this superset architecture, HE-AAC v2 decoders are also capable of decoding plain AAC bitstreams, as well as bitstreams incorporating AAC and SBR data components. This provides a high degree of flexibility for broadcasters, as it contains all the technical components necessary for audio compression over a wide range of bit rates.

3.9 AC-3

The AC-3 audio coding technique, invented by Dolby Laboratories, is used with the ATSC system instead of one of the MPEG audio coding schemes. It is documented as ATSC Standard a/52. Subsequently, AC-3 was adopted as an optional component of DVB, and by the Motorola Digicypher II system. AC-3 is a transform-based system that obtains coding gain by requantizing frequency coefficients.

The PCM input to an AC-3 coder is divided into overlapping windowed blocks as shown in Figure 3-9-1. These blocks contain 512 samples each, but because of the complete overlap, there is 100 percent redundancy. After the transform, there are 512 coefficients in each block, but because of the redundancy, these coefficients can be decimated to 256 coefficients using a technique called time domain aliasing cancellation (TDAC).

The input waveform is analyzed, and if there is a significant transient in the second half of the block, the waveform will be split into two to prevent pre-echo. In this case, the number of coefficients remains the same, but the frequency resolution will be halved and the temporal resolution will be doubled. A flag is set in the bit stream to indicate to the decoder that this has been done.

The coefficients are output in floating-point notation as a mantissa and an exponent. The representation is the binary equivalent of scientific notation. Exponents are effectively scale factors. The set of exponents in a block produce a spectral analysis of the input to a finite accuracy on a logarithmic scale called the spectral envelope. This spectral analysis is the input to the masking model that determines the degree to which noise can be raised at each frequency.

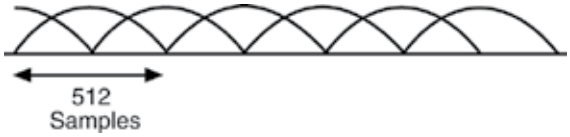


Figure 3-9-1. Overlapping blocks in AC-3 provide 100% redundancy.

The masking model drives the requantizing process, which reduces the accuracy of each coefficient by rounding the mantissa. A significant proportion of the transmitted data consist of mantissa values.

The exponents are also transmitted, but not directly as there is further redundancy within them that can be exploited. Within a block, only the first (lowest frequency) exponent is transmitted in absolute form. The remaining exponents are transmitted differentially and the decoder adds the difference to the previous value. Where the input audio has a smooth spectrum, the exponents in several frequency bands may be the same. Exponents can be grouped into sets of two or four with flags that describe what has been done.

Sets of six blocks are assembled into an AC-3 sync frame. The first block of the frame always has full exponent data, but in cases of stationary signals, later blocks in the frame can use the same exponents.

Section 4 – The Standards

Sections 2 and 3 introduced the technologies of video and audio compression, and many of the tools used in the standards. Over the years, there have been two main standards bodies undertaking parallel development of video and audio compression algorithms and standards, namely MPEG and ITU. This section examines the history, structure and evolution of the broadcasting related standards.

4.1 What is MPEG

MPEG is the Moving Pictures Experts Group, a committee that comes under the joint control of the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC). IEC handles international standardization for electrical and electronic technologies; ISO handles virtually everything else. At the start of the information technology age, ISO and IEC formed a joint technical committee (JTC1) to address IT issues. JTC1 has a number of working groups, including JPEG (Joint Photographic Experts Group) and WG11, which is MPEG.

The committee was formed in 1988 under the leadership of MPEG convener Dr. Leonardo Chiariglione of Italy. It established an enviable track record of generating standards that achieved widespread adoption: MPEG-1, MPEG-2, and the MP3 audio compression standard (MPEG-1 Audio, layer 3). This reputation was somewhat tarnished by MPEG-4, not because of deficiencies in the standard, but as a result of the long delay in publishing licensing terms, and the strong adverse reaction to the first terms that were eventually published in early 2002.

4.2 Who are ITU

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organization of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing recommendations on them with a view to standardizing telecommunications on a world-wide basis. In some areas of information technology that fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

A steering group known as, ITU-T SG 16 Q.6, is responsible for developing the actual standards. This group is better known as VCEG (Video Coding Experts Group).

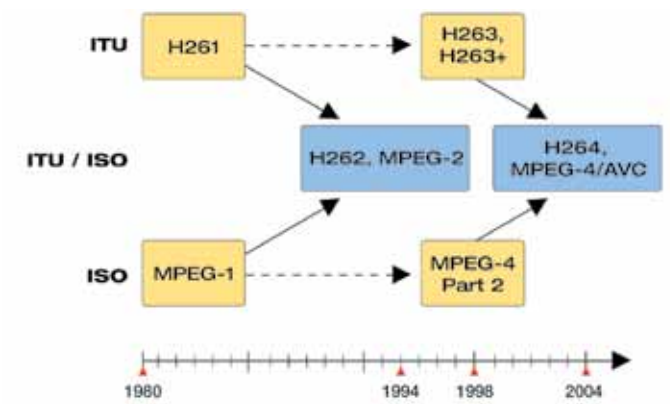


Figure 4-3. Evolution of MPEG.

4.3 Evolution of the Standards

Figure 4.3 illustrates the evolution of the video standards. It can be seen that on occasions the ITU-T VCEG group and MPEG WG 11 groups have collaborated to develop new video coding standards. The resultant CODECs have then been issued by both bodies under separate names. The MPEG-2 video CODEC standard that arose from the partnership is known in the ITU-T community as H.262.

The MPEG-4 part 2 video standard is a large standard and still has a number of issues for commercial implementation. It is over-complex, containing many profiles, the vast majority of which are unused by commercial applications. It includes a number of technical compromises, such that the syntax is not as well structured and clear as it could be, making implementation and interoperability more difficult. Some elements are not entirely clear and are open to interpretation. There are some errors, such as the standard, the normative reference software and normative bitstreams are sometimes at odds with each other. In the extended period it took to develop and standardize MPEG-4, technology moved onto H.264/AVC, which undoubtedly gives better compression.

H.264 is a name related to the ITU-T line of H.26x video standards, while AVC relates to the ISO/IEC MPEG side of the partnership project that completed the work on the standard, after earlier development done in the ITU-T as a project called H.26L. It is usual to call the standard H.264/AVC (or AVC/H.264 or H.264/ MPEG-4 AVC or MPEG-4 Part 10) to emphasize the common heritage. The name H.26L, harkening back to its ITU-T history, is far less common, but still used. Occasionally, it has also been referred to as “the JVT CODEC”, in reference to the JVT organization that developed it.

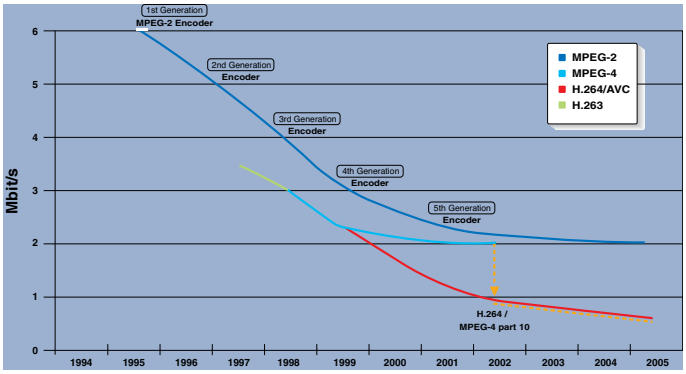


Figure 4-3-1. Bandwidth requirements drop as codecs improve over time.

H.264/AVC concentrates solely on the video, and as a result has been easier and quicker to standardize than MPEG-4 part 2. In addition, it does not have to interoperate with many other parts of the same standard. The H.264/AVC standards body ensured that systems elements were taken care of elsewhere – for example, with a minor modification to the MPEG-2 Transport Stream it can carry H.264/AVC video as well as MPEG-2 video. As a result, H.264/AVC has become highly attractive to the whole broadcast industry, which uses MPEG-2 Transport Stream ('TS'). The approach of embedding H.264/AVC video in an MPEG-2 TS means that it is backwards compatible with the majority of the existing highly expensive infrastructure – satellite transmission, TV/cable transmission, etc. To take advantage of the improved video quality and HD possibilities that H.264/AVC offers, only the video elements at the source and destination endpoints need updating.

Although complex in itself (see 'Complexity of H.264/AVC'), the standard has been well designed and well written. Relatively good reference software is also available. The result is that for engineers implementing H.264/AVC, although still a very large task, it is generally clearer to implement than MPEG-4 part 2. Many companies recognize the maturity of the standard, so there is now real commercial impetus behind it.

In recent years Microsoft has entered the broadcast market with an alternative video CODEC to AVC/H.264. Microsoft originally developed the Windows Media Video (WMV) format as a proprietary CODEC for Internet streaming purposes. With WMV version 9 it added a new Advanced Profile for broadcast applications. In 2003, WMV-9 was submitted to SMPTE for standardization. The standard was officially approved in March 2006 as SMPTE 421M, which is also known as VC-1.

VC-1 provides similar performance to H.264/AVC, and potentially offers reduced processing and memory requirements. With SMPTE standardization, it should find widespread use.

It should be noted that competing formats for the next generation DVD standard, HD DVD and Blu-ray, both mandate support for AVC/H.264 and VC-1.

4.3.1 Video Performance

CODEC performance is simply the measure of how many bits it takes to achieve a particular visual quality in a video sequence. However, it is much more complex than it appears, as the performance and "visual quality" is a subjective combination of many things, e.g.:

- Sharpness of the video.
- Visual artifacts such as "blockiness" or "ringing".
- Color fidelity.
- Video/audio synchronization.
- Smoothness of motion.

These visual features derive from the different types of encoding used throughout the sequence, the differing frame types and the allocation of bit usage variation during the sequence. Whilst the subjective elements are hard to measure, it is possible to quantify objective elements, such as the encoding and bit usage. There is some correlation between the visual quality and the quantifiable parameters, for example, such as PSNR (Peak Signal-to-Noise Ratio) measurement.

4.3.2 Licensing

The MPEG committee itself has no role in licensing. Under the auspice of ISO and IEC, it requires that technology included in its standards be licensable under "reasonable and non-discriminatory terms", but there is no accepted definition of "reasonable." Licensing is the responsibility of the holders of the relevant patents, and typically this means many organizations throughout the world that have contributed research and development, and wish to see some recompense.

For MPEG-2, the patent holders grouped together and formed MPEG-LA (MPEG licensing authority). All the essential patents are certified by this group, and are licensed as a block to any organization wishing to implement the standards. This worked well for MPEG-2, but as noted above, greater difficulties were experienced with MPEG-4 part 2, and certainly impacted its commercial success.

H.264/AVC (aka MPEG-4 part 10) also suffered from protracted licensing issues, which is one reason why VC-1 was introduced. Eventually two separate groups of patent holders were formed and are represented by MPEG-LA and Via Licensing. A license is required from both bodies for any organization wishing to implement the H.264/AVC standard.

4.4 MPEG-1

The MPEG-1 system, ISO/IEC 11172, is the first international compression standard for motion imagery and was developed between 1988 and 1992. It uses DCT transforms, coefficient quantization, and variable length coding in a similar manner to JPEG, but also includes motion compensation for temporal compression.

It is in three parts:

- System ISO/IEC 11172-1, the multiplex structure.
- ISO/IEC 11172-2, video coding.
- ISO/IEC 11172-3, audio coding.

MPEG-1 represented a remarkable technical achievement. It was designed to compress image streams with SIF picture size, 352x288 (25-Hz PAL) or 352x240 (30 Hz NTSC), and associated audio, to approximately 1.5 Mbits/s total compressed data rate. This rate is suitable for transport over T1 data circuits and for replay from CD-ROM, and corresponds approximately to the resolution of a consumer video recorder. A measure of this achievement may be seen by comparing the numbers for an audio CD. A normal audio CD, carrying two-channel audio, at 16-bit resolution with a sampling rate of 44.1 kHz, has a data transfer rate of up to 1.5 Mbit/s. MPEG-1 succeeds in compressing video and audio so that both may be transmitted within the same data rate!

The CIF format is a compromise between European and American SIF (source input format) formats: spatial resolution for 625 SIF (352x288) and Temporal Resolution 525 SIF (29.97 Hz). This is the basis for video conferencing.

MPEG-1 can was designed for CIF images, and has no tools to handle interlaced images, so it had little obvious impact in the world of broadcast television.

Before leaving MPEG-1, it is important to note what is actually included in the standard, and how interoperability is achieved. The standard defines a tool set, the syntax of the bit stream, and the operation of the decoder. It does not define the operation of the encoder – any device that produces a syntactically valid bit stream that can be decoded by a compliant decoder is a valid MPEG encoder. Also, it does not define the quality of the picture, nor encoding quality. This allows for the evolution of encoding technology without change to the standard, and without rendering existing decoders obsolete. This model is used throughout the MPEG standards. The success of this strategy is obvious; although MPEG-2 is used for video, MPEG-1, layer 2 audio is still in use as the principal audio compression system in the DVB transmission systems today.

4.5 MPEG-2

MPEG-1 was frozen (i.e., subsequent changes were allowed to be editorial only) in 1991. In the same year the MPEG-2 process was started, and MPEG-2 eventually became a standard in 1994. The initial goals were simple; there was a need for a standard that would accommodate broadcast quality video width. This required the coding of "full size" standard definition images (704x480 at 29.97 Hz, and 704x576 at 25 Hz), and the ability to code interlaced video efficiently.

In many ways, MPEG-2 represents the "coming of age" of MPEG. The greater flexibility of MPEG-2, combined with the increased availability of large-scale integrated circuits, meant that MPEG-2 could be used in a vast number of applications. The success of MPEG-2 is best highlighted by the demise of MPEG-3, intended for high-definition television. MPEG-3 was soon abandoned when it became clear that MPEG-2 could accommodate this application with ease. MPEG-2 is, of course, the basis for both the ATSC and DVB broadcast standards, and the compression system used by DVD.

HIGH		4:2:0 1920x1152 80 Mbps I, P, B	4:2:2 1920x1080 300 Mbps I, P, B			4:2:0, 4:2:2 1920x1152 100 Mbps I, P, B
HIGH – 1440		4:2:0 1440x1152 60 Mbps I, P, B			4:2:0 1440x1152 60 Mbps I, P, B	4:2:0, 4:2:2 1440x1152 80 Mbps I, P, B
MAIN	4:2:0 720x576 15 Mbps I, P	4:2:0 720x576 15 Mbps I, P, B	4:2:2 720x608 50 Mbps I, P, B	4:2:0 720x576 15 Mbps I, P, B		4:2:0, 4:2:2 720x576 20 Mbps I, P, B
LOW		4:2:0 352x288 4 Mbps I, P, B		4:2:0 352x288 4 Mbps I, P, B		
LEVEL / PROFILE	SIMPLE	MAIN	4:2:2 PROFILE	SNR	SPATIAL	HIGH

Figure 4-5. MPEG-2 levels and profiles.

MPEG-2 was also permitted to be a moving target. By the use of profiles and levels, discussed below, it was possible to complete the standard for one application, but then to move on to accommodate more demanding applications in an evolutionary manner. Work on extending MPEG-2 continues into 2002.

MPEG-2 is documented as ISO/IEC 13818, currently in 10 parts. The most important parts of this standard are:

- ISO/IEC 13818-1 Systems (transport and programs streams), PES, T-STD buffer model and the basic PSI tables: CAT, PAT, PMT and NIT.
- ISO/IEC 13818-2 video coding.
- ISO/IEC 13818-3 audio coding.
- ISO/IEC 13818-4 MPEG test and conformance.
- ISO/IEC 13818-6 data broadcast and DSMCC.

One of the major achievements of MPEG-2 defined in 13818-1, the transport stream, is described in Section 8. The flexibility and robustness of this design have permitted it to be used for many applications, including transport of MPEG-4 and MPEG-7 data.

Note: DVB and ATSC transport streams carry video and audio PES within “program” groupings, which are entirely different than “program streams” (these are used on DVD & CD).

MPEG Transport Streams are normally constant bit rate but program streams are normally variable bit rate.

4.5.1 Profiles and Levels in MPEG-2

With certain minor exceptions, MPEG-1 was designed for one task; the coding of fixed size pictures and associated audio to a known bit rate of 1.5 Mbits/sec. The MPEG-1 tools and syntax can and have been used for other purposes, but such use is outside the standard, and requires proprietary encoders and decoders. There is only one type of decoder compliant to the MPEG-1 standard.

At the outset, there was a similar goal for MPEG-2. It was intended for coding of broadcast pictures and sound, nominally the 525/60 and 625/50 interlaced television systems. However, as the design work progressed, it was apparent that the tools being developed were capable of handling many picture sizes and a wide range of bit rates. In addition, more complex tools were developed for scalable coding systems. This meant that in practice there could not be a single MPEG-2 decoder. If a compliant decoder had to be capable of handling high-speed bit streams encoded using all possible tools, it would no longer be an economical decoder for mainstream applications. As a simple example, a device capable of decoding high-definition signals at, say, 20 Mbits/sec would be substantially more expensive than one limited to standard-definition signals at around 5 Mbits/sec. It would be a poor standard that required the use of an expensive device for the simple application.

MPEG devised a two-dimensional structure of profiles and levels for classifying bit streams and decoders. Profiles define the tools that may be used. For example, bidirectional encoding (B-frames) may be used in the main profile, but not in simple profile. Levels relate just to scale. A high level decoder must be capable of receiving a faster bit stream, and must have more decoder buffer and larger frame stores than a main level decoder. However, main profile at high level (MP@HL) and main profile at main level (MP@ML) use exactly the same encoding/decoding tools and syntax elements.

Figure 4-5 shows the pairings of profile and level that are defined by MPEG-2 (Profiles on the horizontal axis, Levels on the vertical axis). Note that not all combinations are valid; only the completed pairings are defined in the standard. It is a requirement of conformity to the standard that a decoder at any profile/level must be able to decode lower profiles and levels. For example, an MP@ML decoder must be able to decode main profile at low level (MP@LL) and simple profile at main level (SP@ML) bit streams.

The simple profile does not support bidirectional coding, and so only I-and P-pictures will be output. This reduces the coding and decoding delay and allows simpler hardware. The simple profile has only been defined at main level.

The Main Profile is designed for a large proportion of uses. The low level uses a low-resolution input having only 352 pixels per line. The majority of broadcast applications will require the MP@ML subset of MPEG, which supports SDTV (standard definition TV).

The high-1440 level is a high definition scheme that doubles the definition compared to the main level. The high level not only doubles the resolution but maintains that resolution with 16:9 format by increasing the number of horizontal samples from 1440 to 1920.

In compression systems using spatial transforms and requantizing, it is possible to produce scalable signals. A scalable process is one in which the input results in a main signal and a “helper” signal. The main signal can be decoded alone to give a picture of a certain quality, but if the information from the helper signal is added, some aspect of the quality can be improved.

For example, a conventional MPEG coder, by heavily requantizing coefficients, encodes a picture with moderate signal-to-noise ratio results. If, however, that picture is locally decoded and subtracted pixel-by-pixel from the original, a quantizing noise picture results. This picture can be compressed and transmitted as the helper signal. A simple decoder only decodes the main, noisy bit stream, but a more complex decoder can decode both bit streams and combine them to produce a low-noise picture. This is the principle of SNR (signal-to-noise ratio) scalability.

As an alternative, coding only the lower spatial frequencies in a HDTV picture can produce a main bit stream that an SDTV receiver can decode. If the lower definition picture is locally decoded and subtracted from the original picture, a definition-enhancing picture would result. This picture can be coded into a helper signal. A suitable decoder could combine the main and helper signals to recreate the HDTV picture. This is the principle of spatial scalability.

The high profile supports both SNR and spatial scalability as well as allowing the option of 4:2:2 sampling.

The 4:2:2 profile has been developed for improved compatibility with digital production equipment. This profile allows 4:2:2 operation without requiring the additional complexity of using the high profile. For example, an HP@ML decoder must support SNR scalability, which is not a requirement for production. The 4:2:2 profile has the same freedom of GOP structure as other profiles, but in practice it is commonly used with short GOPs making editing easier. 4:2:2 operation requires a higher bit rate than 4:2:0, and the use of short GOPs requires an even higher bit rate for a given quality.

The concept of profiles and levels is another development of MPEG-2 that has proved to be robust and extensible; MPEG-4 uses a much more complex array of profiles and levels, to be discussed later.

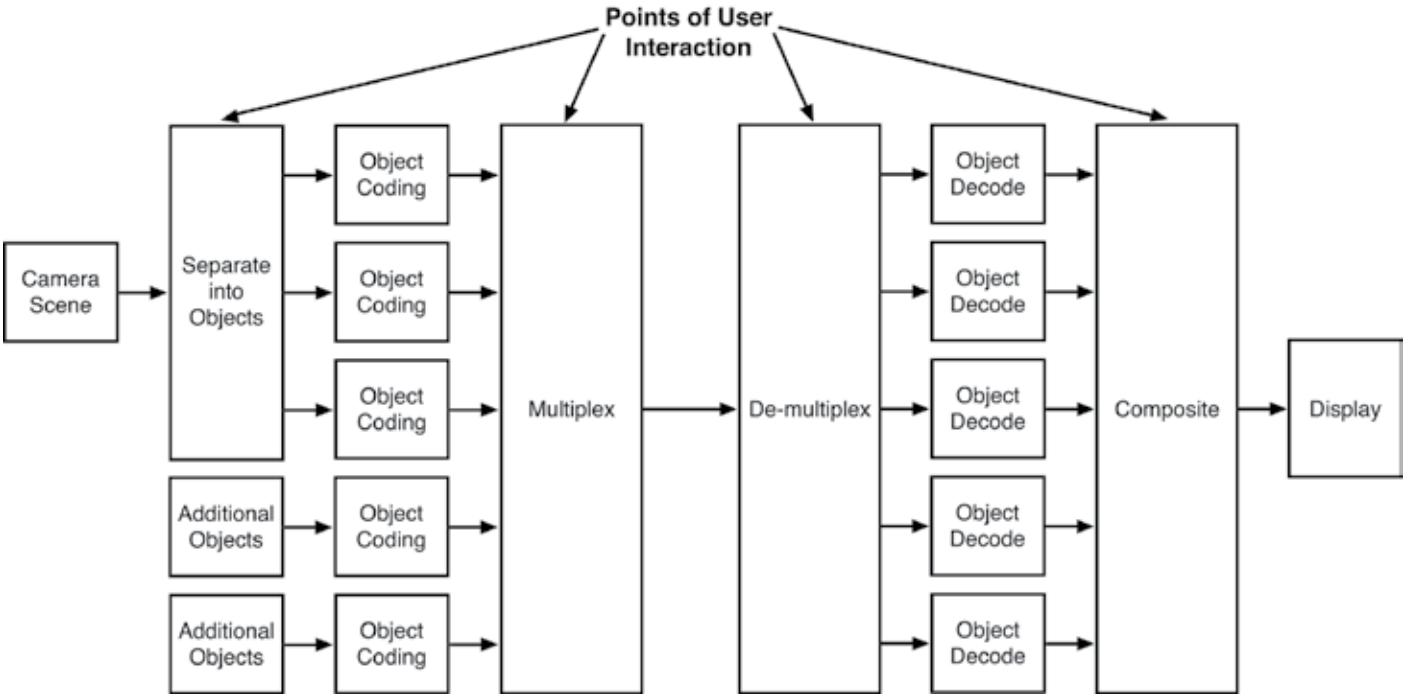


Figure 4-6. MPEG-4, Part 2 codec diagram.

4.6 MPEG-4

International standardization is a slow process, and technological advances often occur that could be incorporated into a developing standard. Often this is desirable, but continual improvement can mean that the standard never becomes final and usable. To ensure that a standard is eventually issued there are strict rules that prohibit substantive change after a certain point in the standardization process. So, by the time a standard is officially adopted there is often a backlog of desired enhancements and extensions. Such was the case with MPEG-2. As discussed above, MPEG-3 had been started and abandoned, so the next project became MPEG-4.

At first the main focus of MPEG-4 was the encoding of video and audio at very low rates. In fact, the standard was explicitly optimized for three bit rate ranges:

- Below 64 Kbits/s.
- 64 to 384 Kbits/s.
- 384 Kbit/s to 4 Mbits/s.

Performance at low bit rates remained a major objective and some very creative ideas contributed to this end. Great attention was also paid to error resilience, making MPEG-4 very suitable for use in the error-prone environments, such as transmission to personal handheld devices. However, other profiles and levels use bit rates up to 38.4 Mbits/s, and work is still proceeding on studio- quality profiles and levels using data rates up to 1.2 Gbits/s.

More importantly, MPEG-4 became vastly more than just another compression system – it evolved into a totally new concept of multimedia encoding with powerful tools for interactivity and a vast range of applications. Even the official “overview” of this standard spans 67 pages, so only a brief introduction to the system is possible here.

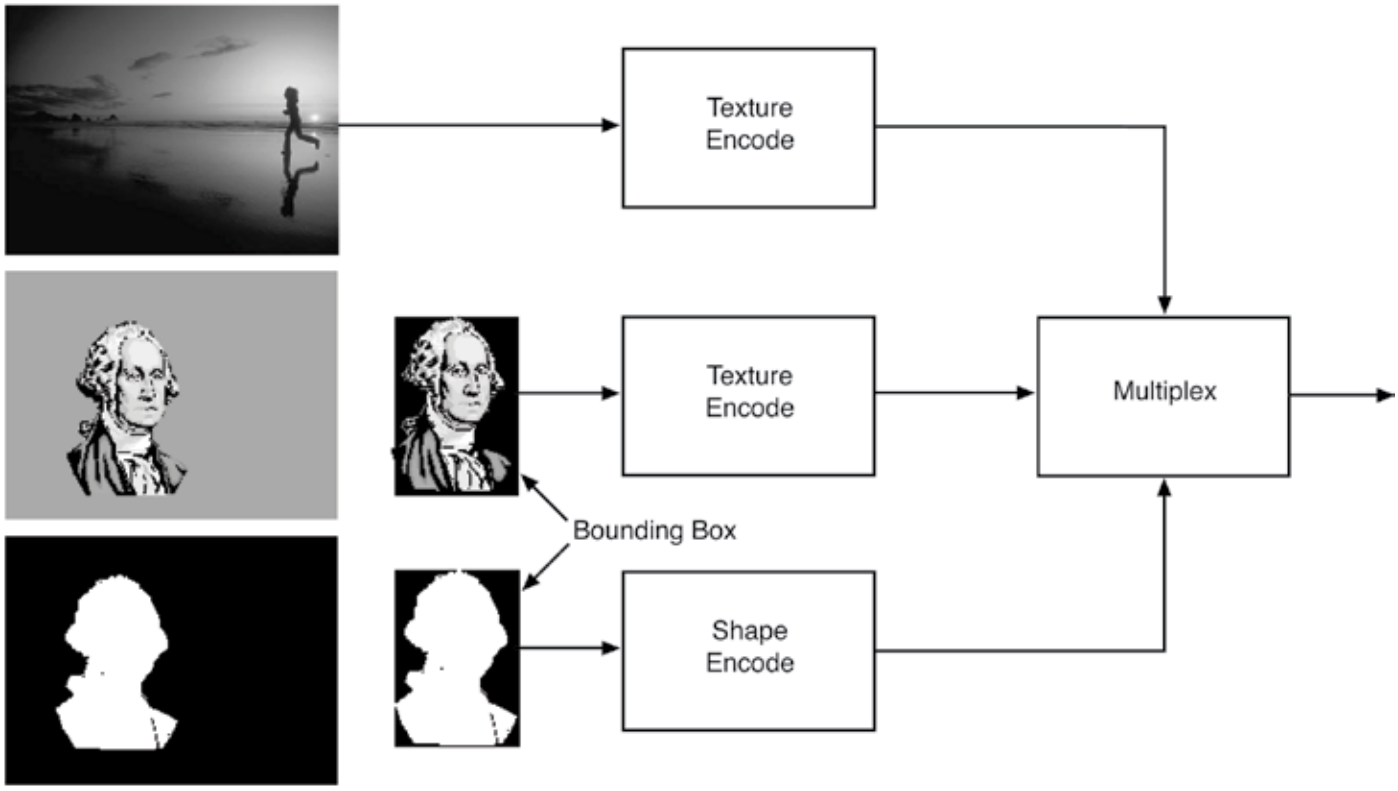


Figure 4-6-3. MPEG-4, Part 2 feature extract process.

4.6.1 MPEG-4 Standards Documents

The principal parts of the MPEG-4 standards are:

- ISO/IEC 14496-1 Systems.
- ISO/IEC 14496-2 Visual.
- ISO/IEC 14496-3 Audio.
- ISO/IEC 14496-4 Conformance Testing.
- ISO/IEC 14496-6 Delivery Multimedia Integration Framework (DMIF).
- ISO/IEC 14496-10 Advanced Video Coding.
- ISO/IEC 14496-12 ISO Base Media File Format.
- ISO/IEC 14496-14 MP4 File Format.
- ISO/IEC 14496-15 Advanced Video Coding (AVC) file format.

4.6.2 Object Coding

The most significant departure from conventional transmission systems is the concept of objects. Different parts of the final scene can be coded and transmitted separately as video objects and audio objects to be brought together, or composited, by the decoder. Different object types may each be coded with the tools most appropriate to the job. The objects may be generated independently, or a scene may be analyzed to separate, for example, foreground and background objects. In one interesting demonstration, video coverage of a soccer game was processed to separate the ball from the rest of the scene. The background (the scene without the ball) was transmitted as a “teaser” to attract a pay-per-view audience. Anyone could see the players and the field, but only those who paid could see the ball!

The object-oriented approach leads to three key characteristics of MPEG-4 streams:

- Multiple objects may be encoded using different techniques, and composited at the decoder.
- Objects may be of natural origin, such as scenes from a camera, or synthetic, such as text.
- Instructions in the bit stream, and/or user choice, may enable several different presentations from the same bit stream.

The generalized system for object coding in MPEG-4 is shown in Figure 4-6. This diagram also emphasizes the opportunities for user interaction within MPEG-4 systems – a powerful feature, particularly for video game designers.

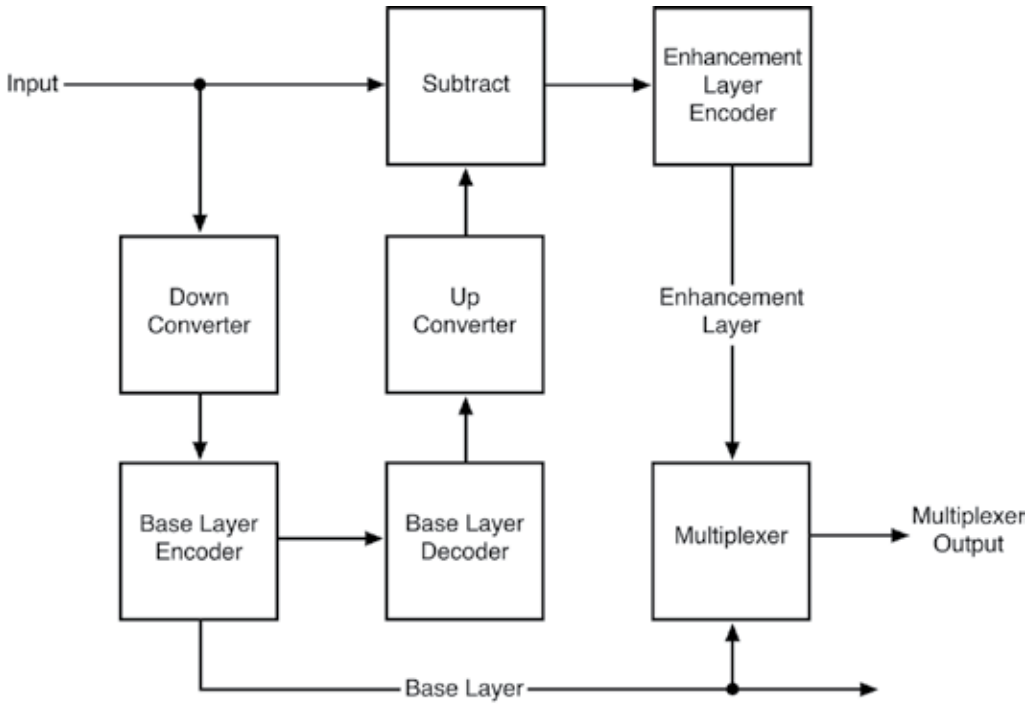


Figure 4-6-4. MPEG-4, Part 2 encoder diagram.

These capabilities do not have to be used – MPEG-4 provides traditional coding of video and audio, and improves on MPEG-2 by offering improved efficiency and resilience to errors. However, the true power of MPEG-4 comes from the architecture described above. The coding of objects independently offers a number of advantages. Each object may be coded on the most efficient manner, and different spatial or temporal scaling (see 4.6.3) may be used as appropriate.

4.6.3 Video and Audio Coding

Many of the video coding tools in MPEG-4 are similar to those of MPEG-2, but enhanced by better use of predictive coding and more efficient entropy coding. However, the application of the tools may differ significantly from earlier standards.

MPEG-4 codes video objects. In the simplest model a video is coded in much the same way as in MPEG-2, but it is described as a single video object with a rectangular shape. The representation of the image is known as texture coding. Where there is more than one video object, some may have irregular shapes, and generally all will be smaller than a full-screenbackground object. This means that only the active area of the object need be coded, but the shape and position must also be represented. The standard includes tools for shape coding of rectangular and irregular objects, in either binary or gray-scale representations (similar to an alpha channel). The concept is shown in Figure 4-6-3.

Similarly, MPEG-4 uses tools similar to those of MPEG-1 and MPEG-2 for coding live audio, and AAC offers greater efficiency. Multiple audio “objects” may be encoded separately and composited at the decoder. As with video, audio objects may be natural or synthetic.

4.6.4 Scalability

In the context of media compression, scalability means the ability to distribute content at more than one quality level within the same bitstream. MPEG-2 and MPEG-4 part 2 both provide scalable profiles using a conventional model; the encoder generates a base-layer and one or more enhancement layers, as shown in Figure 4-6-4. The enhancement layer(s) may be discarded for transmission or decoding if insufficient resources are available. This approach works, but all decisions about quality levels have to be made at the time of encoding, and in practice the number of enhancement layers is severely limited (usually to one).

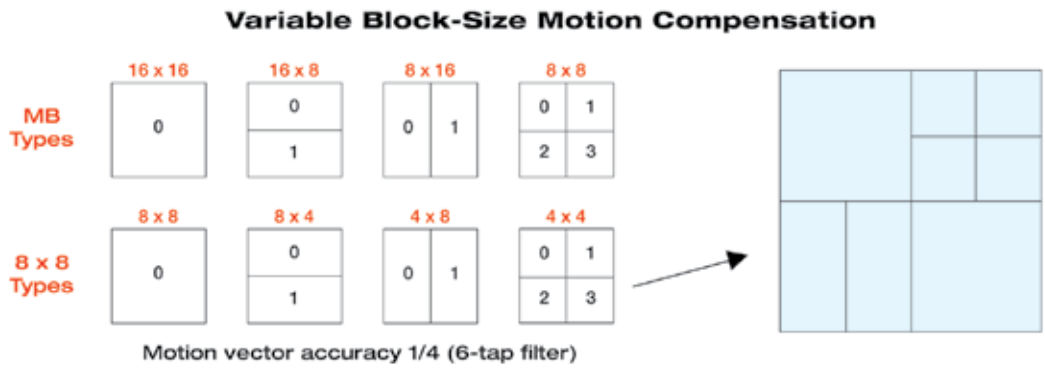


Figure 4-6-5. MPEG-4, Part 10 still uses 16x16 macroblock sizes, but adds many new block sizes and combinations for better efficiency.

4.6.5 Variable Macroblock Size

One of the advantages of AVC/H.264 over MPEG-4 part 2 is variable size macroblocks, as illustrated by Figure 4-6-5. H.264/ AVC allows the 16x16 macroblocks to be sub-divided down to blocks as small as 4x4, to enhance the compression efficiency. In MPEG-2 and MPEG-4 part 2, there is just one type of 'Intra' macroblock, containing a compressed image, which does not refer to any other macroblock.

To incorporate this efficiency, however, there are 26 types of AVC/ H.264 Intra macroblock. There are many other complex elements in H.264/AVC, such as the 'CABAC' Entropy coding, where bits are shared such that fractions of 1 bit may effectively represent a specific syntax element.

On the converse side, substantial attempts have been made to understand and reduce the computing power required. For example, instead of a DCT, a simple well defined integer 4x4 transform could be utilized that only uses fast-to-execute shift, subtract and add operations.

4.6.6 Profiles and Levels in MPEG-4 part 10 (AVC/H.264)

In a similar manner to MPEG-2, a series of profiles and levels have been specified for classifying bitstreams and decoders. The following seven profiles, or capabilities, are intended to target different types of application:

- **Baseline Profile:** Mainly targeted at applications with limited computing resources, this profile is widely used in videoconferencing and mobile devices.
- **Main Profile:** This profile has largely been superseded by High profile as the mainstream profile for consumer broadcast and storage devices.
- **Extended Profile:** A profile focussed on video streaming, this has relatively high compression capabilities together with good resilience to data losses.
- **High Profile:** The primary profile for broadcast and disk storage devices, particularly for HDTV applications. It includes extra coding efficiencies, for example, with the introduction of an 8x8 block transform. This is the profile used by HD DVD and Blu-ray disk formats.
- **High 10 Profile:** This profile extends the High profile used predominately for consumer devices; in particular, support for up to 10 bits per sample decoded picture precision.
- **High 4:2:2 Profile:** Primarily targeting professional devices that use interlaced video, this profile extends the High 10 Profile; in particular, support for the 4:2:2 chroma sampling format.
- **High 4:4:4 Profile:** This profile extends the High 4:2:2 profile. It supports up to 4:4:4 chroma sampling, up to 12 bits per sample, efficient loss less region coding and RGB video coding transform that reduces color-space transformation error. Note that this profile has been removed from the standard in favor of developing a new improved 4:4:4 profile.

H.264/AVC Profiles

Profile	Chroma Format	Max Sample Depth	Max Bit Rate Level 3	Max Bit Rate Level 5
Baseline	4:2:0	8 bits	10 Mb/s	135 Mb/s
Main	4:2:0	8 bits	10 Mb/s	135 Mb/s
Extended	4:2:0	8 bits	10 Mb/s	135 Mb/s
High	4:2:0	8 bits	12.5 Mb/s	168.75 Mb/s
High 10	4:2:0	10 bits	30 Mb/s	405 Mb/s
High 4:2:2	"4:2:0, 4:2:2"	10 bits	40 Mb/s	540 Mb/s
High 4:4:4	"4:2:0, 4:2:2, 4:4:4"	12 bits	40 Mb/s	540 Mb/s

Figure 4-6-6. H.264 profiles.

There are five main levels and a number of sub-levels associated with each profile. Levels essentially relate to scale. Each level specifies a maximum video bit rate and number of macroblocks per frame for example, as well as display resolution. Display resolution may vary from 128x96 in level 1 to 4096x2304 in level 5.1. This covers the QCIF formats used on mobile phones through to HDTV displays. Note that 1920x1080 and 4096x2048 are referred to as 2K and 4K modes respectively in HDTV parlance.

4.6.7 MP4 Container Format

MP4 is a multimedia container format standard. It is most-commonly used to store MPEG-4 audio and video bitstreams, but also can be used to store other data such as subtitles and still images. The MP4 format is based upon Apple's QuickTime container format, but has been improved upon in many ways. Like most modern container formats, MP4 allows streaming over the Internet. The official file name extension for MPEG-4 Part 14 files is .mp4, thus the container format is often referred to simply as MP4. Devices that play .mp4 files are referred to as MP4 players.

MPEG-4 part 15 specifies the format for carriage of H.264/AVC elementary streams within an MP4 file.

4.7 SMPTE VC-1

The basic mode of operation of VC-1 is the same as most other video compression standards since H.261 and MPEG-1 and involves block-based motion compensation and spatial transforms. A number of enhancements have been made to give superior rate-distortion performance compared with older CODECs, which allows it to compete with H.264/AVC.

Together with innovations to improve quality, it has also been designed to have a relatively low level of complexity, which helps it to target a wide range of applications and bit rates. Some of the technologies used in VC-1 are described on the next page.

VC-1 Profiles

Profile	Level	Chroma Format	Max Sample Depth	Max Bit Rate (b/s)	Interlace Support
Simple	Low	4:2:0	8 bits	96 k	
	Medium	4:2:0	8 bits	384 k	
Main	Low	4:2:0	8 bits	2 M	
	Medium	4:2:0	8 bits	10 M	
	High	4:2:0	8 bits	20 M	
Advanced	L0	4:2:0	8 bits	2 M	
	L1	4:2:0	8 bits	10 M	X
	L2	4:2:0	8 bits	20 M	X
	L3	4:2:0	8 bits	45 M	X
	L4	4:2:0	8 bits	135 M	X

Figure 4-7. SMPTE VC-1 profiles.

4.7.1 Variable Block Size Transform

VC-1 shares the ability of AVC/H.264 to use smaller macroblocks than more traditional CODECs in order to increase coding efficiency and reduce the appearance of artifacts such as ‘ringing’ at block boundaries. The standard 8x8 transform can still be used, but this can be subdivided into multiple 8x4, 4x8 or 4x4 transforms to produce the best picture quality for each macroblock. The transform size can be specified per block or per macroblock, but can also be signaled at the frame level which can be useful in low bit rate situations to reduce the overhead.

4.7.2 Motion Compensation

The model for motion compensation in VC-1 is similar to that used in MPEG-1 and many other CODECs in which a prediction for a block is made based upon a reference frame and a displacement vector.

There are four modes of motion compensation available in VC-1 which are tailored to different decoder scenarios. For lower bit rates and complexity 16x16 blocks together with 1/2 pixel resolution can be used. For higher bit rates, motion vectors to 1/4 pixel resolution may be coded and the block size reduced to 8x8.

Unlike H.264/AVC, motion vectors for 4x4 blocks can't be sent; this means that the complexity is lower, but at the cost of reduced flexibility.

4.7.3 Deblocking and Smoothing

Various methods are employed to reduce the ‘blocking’ artifacts that are often visible in block-based CODECs, especially at lower bit rates. VC-1 uses an in-loop deblocking filter to reduce these artifacts, which can affect multiple frames if used as a predicted block in future frames.

A technique of overlap smoothing is also used to reduce unintended block boundary discontinuities. This uses transforms which slightly overlap the block boundaries to mitigate the edge artifacts.

4.7.4 Fading Compensation

Motion compensation tends to perform poorly when attempting to compensate for fading effects in a scene. Fade compensation allows for a linear transform of pixel values to be made to account for broad changes in luminance. This can in turn produce better results from the motion compensation step.

4.7.5 Profiles and Levels

Figure 4-7 shows that VC-1 has three profiles; Simple, Main and Advanced. The Simple and Main profiles need to be encapsulated within an ASF wrapper, while Advanced profile streams can exist as independent elementary streams.

The Advanced profile addresses many of the issues associated with the broadcast industry in that it supports interlaced as well as progressive encoding. It is also easily encapsulated in container and transport formats.

The three profiles have between two and five levels which broadly correspond to the scale of the video both in terms of maximum bit rate and resolution. This gives a range suitable to applications from mobile streaming to High Definition video.

4.8 MPEG-7 and MXF

MPEG-7 is not about compression; it is about metadata, also known as the “bits about the bits.” Metadata is digital information that describes the content of other digital data. In modern parlance, the program material or content, the actual image, video, audio or data objects that convey the information are known as data essence. The metadata tells the world all it needs to know about what is in the essence.

Anyone who has been involved with the storage of information, be it videotapes, books, music, etc., knows the importance and the difficulty of accurate cataloging and indexing. Stored information is useful only if its existence is known, and if it can be retrieved in a timely manner when needed.

This problem has always been with us, and is addressed in the analog domain by a combination of labels, catalogs, card indexes, etc. More recently, the computer industry has given us efficient, cost-effective, relational databases that permit powerful search engines to access stored information, provided the information is present in a searchable form.

MPEG-7 defines a structure of descriptors and description schemes that can characterize almost anything. In theory at least, primitive elements such as color histograms and shapes can be combined to represent complex entities such as individual faces. It may be possible to index material automatically such that the database can be searched for scenes that show, for example, President Clinton and U.S. Federal Reserve Chairman Greenspan together.

However, MPEG-7 has not been widely adopted and instead so-called ‘wrapper’ or ‘container’ formats such MXF are increasing in popularity, particularly in Production environments. The key features of MXF are:

- Inter-operability.
- Extensibility.
- Platform independence.

MXF supports a number of different essence tracks, encoded with a variety of CODECs together with accompanying metadata. It uses the KLV [Key-Length-Value] metadata system, which may refer to a local dictionary or external public registries. An MXF metadata dictionary is standardized by SMPTE.

4.9 MPEG-21

The MPEG-21 standard is intended to define an open framework for multimedia delivery from content creator to content consumer. It is intended to enable transparent and augmented use of multimedia resources across a wide range of networks and devices. A major aspiration of the standard is that it will put an end to illicit file sharing.

At its heart is a “Rights Expression Language” standard, as a means of sharing digital rights, permissions and restrictions. It is an XML-based standard designed to communicate machine-readable license information and do so in an "ubiquitous, unambiguous and secure" manner.

MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction, termed the Digital Item, and the concept of users interacting with them. MPEG-21 provides a framework in which one user interacts with another user, and the object of that interaction is a Digital Item. MPEG-21 essentially defines the technology needed to support users who wish to exchange, access, consume, trade or manipulate Digital Items in an efficient and transparent way. That is, it is a collection of technologies that designers of, say DRM systems, could use.

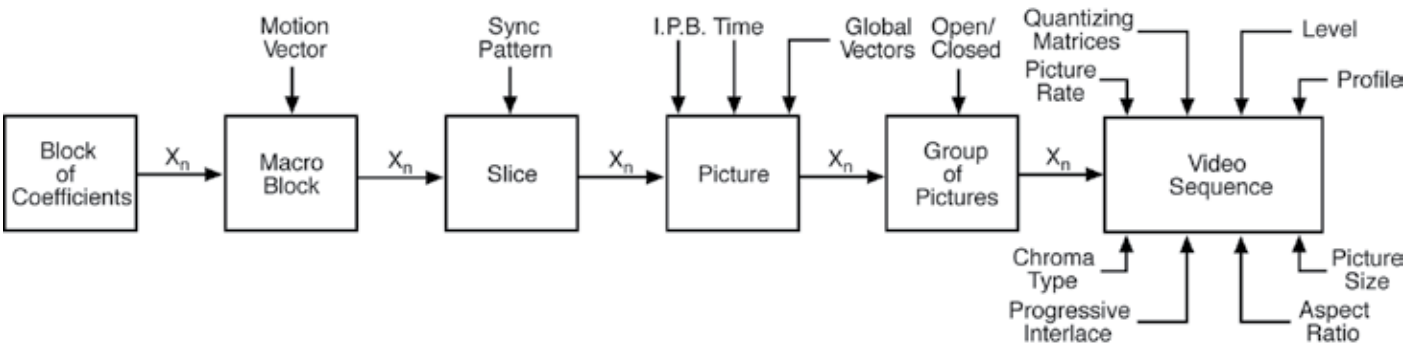


Figure 5-1. Building a video sequence starting with a block of DCT coefficients.

Section 5 – Elementary Streams

An elementary stream is basically the raw output of an encoder and contains no more than is necessary for a decoder to approximate the original picture or audio. The syntax of the compressed signal is rigidly defined in MPEG so that decoders can be guaranteed to operate on it. The encoder is not defined except that it must somehow produce the right syntax.

The advantage of this approach is that it suits the real world in which there are likely to be many more decoders than encoders. By standardizing the decoder, they can be made at low cost. In contrast, the encoder can be more complex and more expensive without a great cost penalty, but with the potential for better picture quality as complexity increases. When the encoder and the decoder are different in complexity, the coding system is said to be asymmetrical.

The MPEG approach also allows for the possibility that quality will improve as coding algorithms are refined while still producing bit streams that can be understood by earlier decoders. The approach also allows the use of proprietary coding algorithms, which need not enter the public domain.

5.1 Video Elementary Stream Syntax

Figure 5-1 shows the construction of the elementary video stream. The fundamental unit of picture information is the DCT (discrete cosine transform) block, which represents an 8x8 array of pixels that can be Y, C_b or C_r. The DC coefficient is sent first and is represented more accurately than the other coefficients. Following the remaining coefficients, an end of block (EOB) code is sent.

Blocks are assembled into macroblocks, which are the fundamental units of a picture and which can be motion compensated. Each macroblock has a two-dimensional motion vector in the header. In B-pictures, the vectors can be backward as well as forward. The motion compensation can be field or frame based and this is indicated. The scale used for coefficient requantizing is also indicated. Using the vectors, the decoder obtains information from earlier and later pictures to produce a predicted picture. The blocks are inverse-transformed to produce a correction picture that is added to the predicted picture to produce the decoded output. In 4:2:0 coding, each macroblock will have four Y blocks and two color-difference blocks. To make it possible to identify which block describes which component, the blocks are sent in a specified order.

Macroblocks are assembled into slices that must always represent horizontal strips of picture from left to right. In MPEG, slices can start anywhere and be of arbitrary size, but in ATSC they must start at the left-hand edge of the picture. Several slices can exist across the screen width. The slice is the fundamental unit of synchronization for variable length and differential coding. The first vectors in a slice are sent absolutely, whereas the remaining vectors are transmitted differentially. In I-pictures, the first DC coefficients in the slice are sent absolutely and the remaining DC coefficients are transmitted differentially. In difference pictures, correlation of these coefficients is not to be expected, and this technique is not appropriate.

In the case of a bit error in the elementary stream, either the deserialization of the variable length symbols will break down, or subsequent differentially coded coefficients or vectors will be incorrect. The slice structure allows recovery by providing a resynchronizing point in the bit stream.

A number of slices are combined to make a picture that is the active part of a field or a frame. The picture header defines whether the picture was I, P or B coded and includes a temporal reference so that the picture can be presented at the correct time. In the case of pans and tilts, the vectors in every macroblock will be the same. A global vector can be sent for the whole picture, and the individual vectors then become differences from this global value.

Pictures may be combined to produce a GOP that must begin (in transmission order) with an I-picture. The GOP is the fundamental unit of temporal coding. In the MPEG standard, the use of a GOP is optional, but it is a practical necessity. Between I-pictures, a variable number of P- and/or B-pictures may be placed as was described in Section 2. A GOP may be open or closed. In a closed GOP, the last B-pictures do not require the I-picture in the next GOP for decoding and the bit stream could be cut at the end of the GOP.

If GOPs are used, several GOPs may be combined to produce a video sequence. The sequence begins with a sequence start code, followed by a sequence header and ends with a sequence end code. Additional sequence headers can be placed throughout the sequence. This approach allows decoding to begin part way through the sequence, as might happen in playback of digital video disks and tape cassettes. The sequence header specifies the vertical and horizontal size of the picture, the aspect ratio, the chroma subsampling format, the picture rate, the use of progressive scan or interlace, the profile, level, and bit rate, and the quantizing matrices used in intra and inter-coded pictures.

Without the sequence header data, a decoder cannot understand the bit stream, and therefore sequence headers become entry points at which decoders can begin correct operation. The spacing of entry points influences the delay in correct decoding that may occur when the viewer switches from one television channel to another.

5.2 Audio Elementary Streams

Various types of audio can be embedded in an MPEG-2 multiplex. These types include audio coded according to MPEG layers 1, 2, 3, or AC-3. The type of audio encoding used must be included in a descriptor that a decoder will read in order to invoke the appropriate type of decoding.

The audio compression process is quite different from the video process. There is no equivalent to the different I, P, and B frame types, and audio frames always contain the same amount of audio data. There is no equivalent of bidirectional coding and audio frames are not transmitted out of sequence.

In MPEG-2 audio, the descriptor in the sequence header contains the layer that has been used to compress the audio and the type of compression used (for example, joint stereo), along with the original sampling rate. The audio sequence is assembled from a number of access units (AUs) that will be coded audio frames.

If AC-3 coding is used, as in ATSC, this usage will be reflected in the sequence header. The audio access unit (AU) is an AC-3 sync frame as described in Section 3.7. The AC-3 sync frame represents a time span equivalent of 1536 audio samples and will be 32 ms for 48-kHz sampling and 48 ms for 32 kHz.

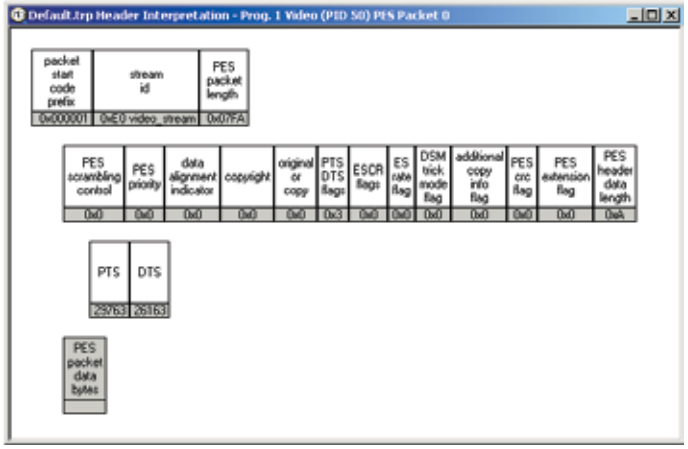


Figure 6-1. Header parameters within a Packetized Elementary Stream.

Section 6 – Packetized Elementary Streams (PES)

For practical purposes, the continuous elementary streams carrying audio or video from compressors need to be broken into packets. These packets are identified by headers that contain time stamps for synchronizing. PES packets can be used to create Program Streams or Transport Streams.

6.1 PES Packets

In the PES, an endless elementary stream is divided into packets of a convenient size for the application. This size might be a few hundred kilobytes, although this would vary with the application.

Each packet is preceded by a PES packet header. Figure 6-1 shows the contents of a header. The packet begins with a start-code prefix of 24 bits and a stream ID that identifies the contents of the packet as video or audio and further specifies the type of audio coding. These two parameters (start code prefix and stream ID) comprise the packet start code that identifies the beginning of a packet. It is important not to confuse the packet in a PES with the much smaller packet used in transport streams that, unfortunately, shares the same name.

Because MPEG only defines the transport stream, not the encoder, a designer might choose to build a multiplexer that converts from elementary streams to a transport stream in one step. In this case, the PES packets may never exist in an identifiable form, but instead, they are logically present in the Transport Stream payload.

6.2 Time Stamps

After compression, pictures are sent out of sequence because of bidirectional coding. They require a variable amount of data and are subject to variable delay due to multiplexing and transmission. In order to keep the audio and video locked together, time stamps are periodically incorporated in each picture.

A time stamp is a 33 bit number that is a sample of a counter driven by a 90 kHz clock. This clock is obtained by dividing the 27 MHz program clock by 300. Since presentation times are evenly spaced, it is not essential to include a time stamp in every presentation unit. Instead, time stamps can be interpolated by the decoder, but they must not be more than 700 ms apart in either program streams or transport streams.

Time stamps indicate where a particular access unit belongs in time. Lip sync is obtained by incorporating time stamps into the headers in both video and audio PES packets. When a decoder receives a selected PES packet, it decodes each access unit and buffers it into RAM. When the time-line count reaches the value of the time stamp, the RAM is read out. This operation has two desirable results. First, effective timebase correction is obtained in each elementary stream. Second, the video and audio elementary streams can be synchronized together to make a program.

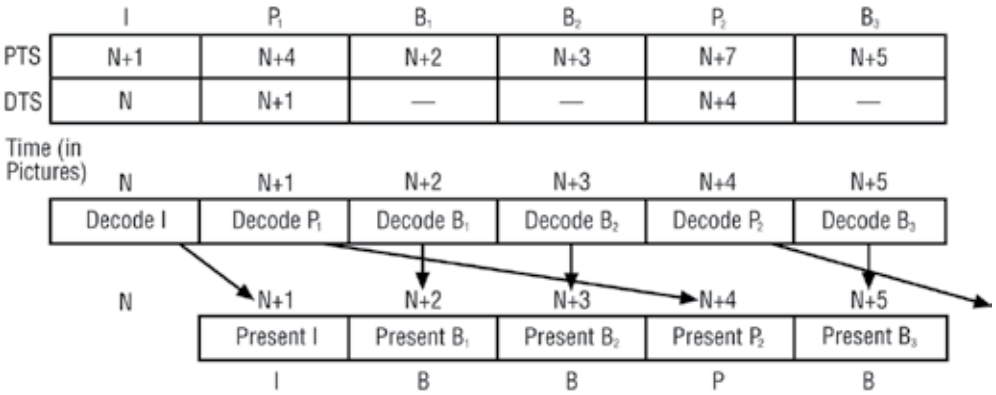


Figure 6-2. PTS and DTS values (based upon PCR values) tell the decoder when to decode a picture and then when to present the picture.

6.3 PTS/DTS

When bidirectional coding is used, a picture may have to be decoded some time before it is presented so that it can act as the source of data for a B-picture. Although, for example, pictures can be presented in the order IBBP, they will be transmitted in the order IPBB. Consequently, two types of time stamp exist. The decode time stamp (DTS) indicates the time when a picture must be decoded, whereas a presentation time stamp (PTS) indicates when it must be presented to the decoder output.

B-pictures are decoded and presented simultaneously so that they only contain PTS. When an IPBB sequence is received, both I- and P- must be decoded before the first B-picture. A decoder can only decode one picture at a time; therefore the I-picture is decoded first and stored. While the P-picture is being decoded, the decoded I-picture is output so that it can be followed by the B-pictures.

Figure 6-2 shows that when an access unit containing an I-picture is received, it will have both DTS and PTS in the header and these time stamps will be separated by one picture period. If bidirectional coding is being used, a P-picture must follow and this picture also has a DTS and a PTS time stamp, but the separation between the two stamp times is three picture periods to allow for the intervening B-pictures. Thus, if IPBB is received, I is delayed one picture period, P is delayed three picture periods, the two Bs are not delayed at all, and the presentation sequence becomes IBBP. Clearly, if the GOP structure is changed such that there are more B-pictures between I and P, the difference between DTS and PTS in the P-pictures will be greater.

The PTS/DTS flags in the packet header are set to indicate the presence of PTS alone or both PTS and DTS time stamp. Audio packets may contain several access units and the packet header contains a PTS. Because audio packets are never transmitted out of sequence, there is no DTS in an audio packet.

Section 7 – Program Streams

Program streams are one way of combining several PES packet streams and are advantageous for recording applications such as DVD.

7.1 Recording vs. Transmission

For a given picture quality, the data rate of compressed video will vary with picture content. A variable bit rate channel will give the best results. In transmission, most practical channels are fixed and the overall bit rate is kept constant by the use of stuffing (meaningless data).

In a DVD, the use of stuffing is a waste of storage capacity. However, a storage medium can be slowed down or speeded up, either physically or, in the case of a disk drive, by changing the rate of data transfer requests. This approach allows a variable-rate channel to be obtained without capacity penalty. When a medium is replayed, the speed can be adjusted to keep a data buffer approximately half full, irrespective of the actual bit rate, which can change dynamically. If the decoder reads from the buffer at an increased rate, it will tend to empty the buffer, and the drive system will simply increase the access rate to restore balance. This technique only works if the audio and video were encoded from the same clock; otherwise, they will slip over the length of the recording.

To satisfy these conflicting requirements, program streams and transport streams have been devised as alternatives. A program stream works well on a single program with variable bit rate in a recording environment; a transport stream works well on multiple programs in a fixed bit rate transmission environment.

The problem of genlocking to the source does not occur in a DVD player. The player determines the time base of the video with a local synchronizing pulse generator (internal or external) and simply obtains data from the disk in order to supply pictures on that time base. In transmission, the decoder has to recreate the time base at the encoder or it will suffer overflow or underflow. Thus, a transport stream uses program clock reference (PCR), whereas a program stream has no need for the program clock.

7.2 Introduction to Program Streams

A program stream is a PES packet multiplex that carries several elementary streams that were encoded using the same master clock or system time clock (STC). This stream might be a video stream and its associated audio streams, or a multichannel audio-only program. The elementary video stream is divided into access units (AUs), each of which contains compressed data describing one picture. These pictures are identified as I, P, or B and each carries an AU number that indicates the correct display sequence. One video AU becomes one program-stream packet. In video, these packets vary in size. For example, an I-picture packet will be much larger than a B-picture packet. Digital audio access units are generally of the same size and several are assembled into one program-stream packet. These packets should not be confused with transport-stream packets that are smaller and of fixed size. Video and audio AU boundaries rarely coincide on the time axis, but this lack of coincidence is not a problem because each boundary has its own time-stamp structure.

Section 8 – Transport Streams

A transport stream is more than a multiplex of many PES packets. In program streams, time stamps are sufficient to recreate the time axis because the audio and video are locked to a common clock. For transmission down a data network over distance, there is an additional requirement to recreate the clock for each program at the decoder. This requires an additional layer of syntax to provide PCR signals.

8.1 The Job of a Transport Stream

The Transport Stream carries many different programs and each may use a different compression factor and a bit rate that can change dynamically even though the overall bit rate stays constant. This behavior is called statistical multiplexing and it allows a program that is handling difficult material to borrow bandwidth from a program handling easy material. Each video PES can have a different number of audio and data PESs associated with it. Despite this flexibility, a decoder must be able to change from one program to the next and correctly select the appropriate audio and data channels. Some of the programs can be protected so that they can only be viewed by those who have paid a subscription or fee. The Transport Stream must contain CA information to administer this protection. The Transport Stream contains PSI to handle these tasks.

The transport layer converts the PES data into small packets of constant size (adding stuffing bits if necessary) that are self-contained. When these packets arrive at the decoder, there may be jitter in the timing. The use of time division multiplexing also causes delay, but this factor is not fixed because the proportion of the bitstream allocated to each program need not be fixed. Time stamps are part of the solution, but they only work if a stable clock is available. The Transport Stream must contain further data allowing the re-creation of a stable clock.

The operation of digital video production equipment is heavily dependent on the distribution of a stable system clock for synchronization. In video production, genlocking is used, but over long distances, the distribution of a separate clock is not practical. In a Transport Stream, the different programs may have originated in different places that are not necessarily synchronized. As a result, the Transport Stream has to provide a separate means of synchronizing for each program.

This additional synchronization method is called a PCR and it recreates a stable reference clock that can be divided down to create a time line at the decoder, so that the time stamps for the elementary streams in each program become useful. Consequently, one definition of a program is a set of elementary streams sharing the same timing reference.

In a Single Program Transport Stream (SPTS), there will be one PCR channel that recreates one program clock for both audio and video. The SPTS is often used as the communication between an audio/video coder and a multiplexer. It is also used in VOD applications. However, most broadcast transport streams include multiple programs (MPTS). This larger aggregate contains control information, better known as signaling (SI) required to coordinate the entire transmission, such as an electronic program guide (EPG), as well as program specifics.

```
{} Transport Packet
  {} Transport_Packet
    {} transport_packet_header
      sync_byte = 0x47
      transport_error_indicator = 0
      payload_unit_start_indicator = 0
      transport_priority = 0
      PID = 670 [0x29E]
      transport_scrambling_control = 0 (Not scrambled)
      adaptation_field_control = 3 (Adaptation_field followed by payload)
      continuity_counter = 9 [0x9]
    {} adaptation_field
      adaptation_field_length = 7 [0x7]
      discontinuity_indicator = 0
      random_access_indicator = 0
      elementary_stream_priority_indicator = 0
      PCR_flag = 1
      OPCR_flag = 0
      splicing_point_flag = 0
      transport_private_data_flag = 0
      adaptation_field_extension_flag = 0
      {} PCR
        program_clock_reference_base = 1407451743 [0x53E4025F]
        reserved = 0x3F
        program_clock_reference_extension = 180 [0xB4]
    {} payload
      raw_bytes = ... [length = 176 bytes]
```

Figure 6-1. ISO/IEC 13818-1 Transport Stream header parameters.

8.2 Packets

Figure 8-1 shows the structure of a transport stream packet. The size is a constant 188 bytes and it is always divided into a header and a payload. Figure 8-1a shows the minimum header length is 4 bytes and contains no Adaptation field. In this header, the most important information is:

- The sync byte. This byte is recognized by the decoder so that the header and the payload can be deserialized.
- The transport error indicator. This indicator is set if the error correction layer above the transport layer is experiencing a raw-bit error rate (BER) that is too high to be correctable. It indicates that the packet may contain errors. See Section 10 – Introduction to DVB & ATSC for details of the error correction layer.
- The packet identification (PID). This thirteen-bit code is used to distinguish between different types of packets. More will be said about PID later.
- The continuity counter. This four-bit value is incremented by the multiplexer as each new packet having the same PID is sent. It is used to determine if any packets are lost, repeated, or out of sequence.

In some cases, more header information is needed, and if this is the case, the adaptation field control bits are set to indicate that the header is larger than normal. Figure 8-1 shows that when this happens the extra header length is described by the adaptation field length code. Where the header is extended, the payload becomes smaller to maintain constant packet length.

8.3 Program Clock Reference (PCR)

The encoder used for a particular program will have a 27-MHz program clock. In the case of an SDI (serial digital interface) input, the bit clock can be divided by 10 to produce the encoder program clock. Where several programs originate in the same production facility, it is possible that they will all have the same clock. In case of an analog video input, the H-sync period will need to be multiplied by a constant in a phase-locked loop to produce 27 MHz.

The adaptation field in the packet header is used periodically to include the PCR code that allows generation of a locked clock at the decoder. If the encoder or a remultiplexer has to switch sources, the PCR may have a discontinuity. The continuity count can also be disturbed. This event is handled by the discontinuity indicator, which tells the decoder to expect a disturbance. Otherwise, a discontinuity is an error condition.

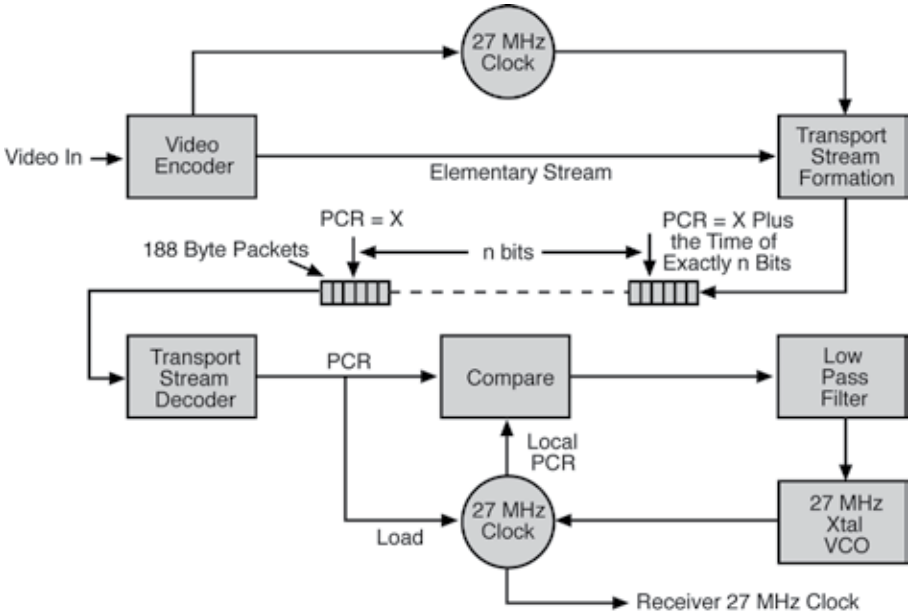


Figure 8-2. Adding a PCR time stamp into a transport stream and then extraction and synchronizing with a decoder.

Figure 8-2 shows how the PCR is used by the decoder to recreate a remote version of the 27-MHz clock for each program. The encoder clocks drive a constantly running binary counter, and the value of these counters are sampled periodically and placed in the header adaptation fields as the PCR. The PCR is a 42-bit number that is represented by a 33-bit PCR base, plus a 9-bit PCR extension to provide higher resolution. (The PCR base, like the PTS, is a 33-bit number that is a sample of a counter driven by a 90-kHz clock). The packets generated by each encoder are given a different PID. The decoder recognizes the packets with the correct PID for the selected program and ignores others. At the decoder, a voltage controlled oscillator (VCO) generates a nominal 27 MHz clock and this drives a local PCR counter. The local PCR is compared with the PCR from the packet header and the difference is the PCR phase error. This error is filtered to control the VCO that eventually will bring the local PCR count into step with the header PCRs. Heavy VCO filtering ensures that jitter in PCR transmission does not modulate the clock. The discontinuity indicator will reset the local PCR count and, optionally, may be used to reduce the filtering to help the system quickly lock to the new timing.

MPEG requires that PCRs be sent at a rate of at least 10 PCRs per second, whereas DVB specifies a minimum of 25 PCRs per second.

8.4 Packet Identification (PID)

A 13-bit field in the transport packet header contains the Packet Identification code (PID). The PID is used by the demultiplexer to distinguish between packets containing different types of information. The transport-stream bit rate must be constant, even though the sum of the rates of all of the different streams it contains can vary. This requirement is handled by the use of null packets. If the real payload rate falls, more null packets are inserted. Null packets always have the same PID, which is 8191 (thirteen ones in the binary representation).

In a given transport stream, all packets belonging to a given elementary stream will have the same PID. The demultiplexer can easily select all data for a given elementary stream simply by accepting only packets with the right PID. Data for an entire program can be selected using the PIDs for video, audio and data streams such as subtitles or teletext. The demultiplexer can correctly select packets only if it can correctly associate them with the elementary stream to which they belong. The demultiplexer can do this task only if it knows what the right PIDs are. This is the function of the PSI.

8.5 Program Specific Information (PSI)

PSI is carried in packets having unique PIDs, some of which are standardized and some of which are specified by the program association table (PAT), conditional access table (CAT) and the transport stream description table (TSDT).

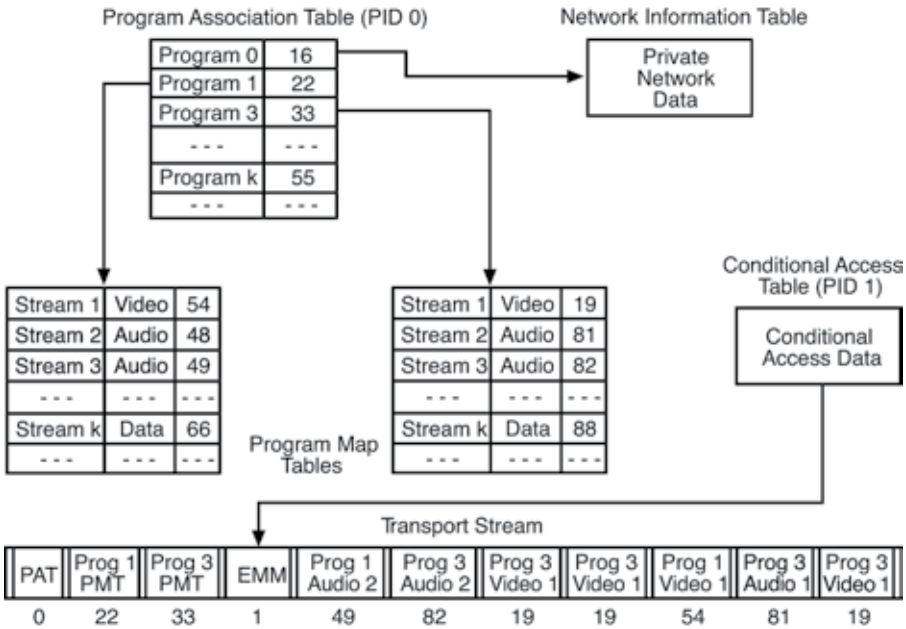


Figure 8-3. Table hierarchy in interleaved transport stream.

These packets must be included periodically in every transport stream. The PAT always has a PID of 0, the CAT always has a PID of 1, and the TSDT always has a PID of 2. These values and the null-packet PID of 8191 are the only PIDs fixed by the MPEG standard. The demultiplexer must determine all of the remaining PIDs by accessing the appropriate tables. However, there are some constraints in the use of PIDs in ATSC and DVB. In this respect (and in some others), MPEG and DVB/ATSC are not fully interchangeable. All DVB and ATSC transport stream must be MPEG-2 compliant (ISO/IEC 13818-1), but not all MPEG-2 transport streams will be compliant with the ATSC (A65) or DVB (EN 300 468) standards.

The programs that exist in the transport stream are listed in the program association table (PAT) packets (PID = 0) that carries the PID of each PMT packet. The first entry in the PAT, program 0, is reserved for network data and contains the PID of network information table (NIT) packets. Usage of the NIT is optional in MPEG-2, but is mandatory in DVB.

The PIDs for entitlement control messages (ECM) and entitlement management messages (EMM) are listed in the conditional access table (CAT) packets (PID = 1).

As Figure 8-3 shows, the PIDs of the video, audio, and data elementary streams that belong in the same program are listed in the Program Map Table (PMT) packets. Each PMT packet normally has its own PID, but MPEG-2 does not mandate this. The program number within each PMT will uniquely define each PMT.

A given network information table (NIT) contains details of more than just the transport stream carrying it. Also included are details of other transport streams that may be available to the same decoder, for example, by tuning to a different RF channel or steering a dish to a different satellite. The NIT may list a number of other transport streams and each one must have a descriptor that specifies the radio frequency, orbital position, and so on. In DVB, additional metadata, known as DVB-SI, is included, and the NIT is considered to be part of DVB-SI. This operation is discussed in Section 10 – Introduction to DVB & ATSC. When discussing the subject in general, the term PSI/SI is used.

Upon first receiving a transport stream, the demultiplexer must look for PIDs 0 and 1 in the packet headers. All PID 0 packets contain the PAT. All PID 1 packets contain CAT data.

By reading the PAT, the demultiplexer can find the PIDs of the NIT and of each program map table (PMT). By finding the PMTs, the demultiplexer can find the PIDs of each elementary stream.

Consequently, if the decoding of a particular program is required, reference to the PAT and then the PMT is all that is needed to find the PIDs of all of the elementary streams in the program. If the program is encrypted, access to the CAT will also be necessary. As demultiplexing is impossible without a PAT, the lockup speed is a function of how often the PAT packets are sent. MPEG specifies a maximum interval of 0.5 seconds for the PAT packets and the PMT packets that are referred to in those PAT packets. In DVB and ATSC, the NIT may reside in packets that have a specific PID.

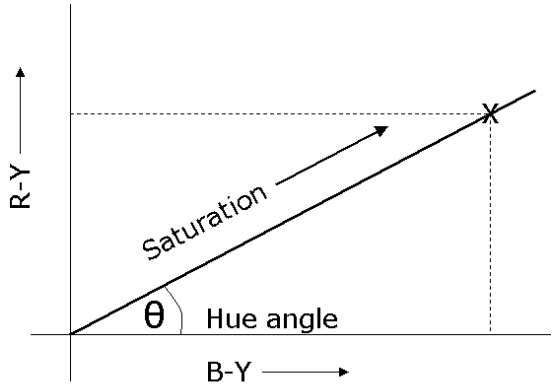


Figure 9-1. Polar axis graph for analog video. Similar to vectorscope display.

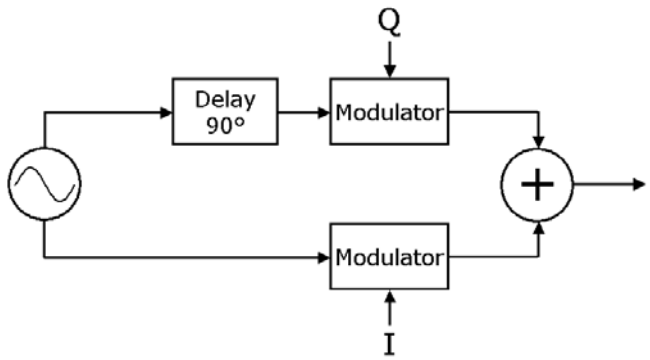


Figure 9-2. Quadrature Modulation.

Section 9 – Digital Modulation

MPEG systems encode and package video, audio, and other data. For storage, the resulting data stream may be recorded on a hard drive or perhaps DVD. For other applications, the MPEG data, usually in the form of a transport stream, has to be sent from one place to another, or to many places. Television systems use cable, satellite and terrestrial transmission in various ways for contribution, distribution and broadcast. All of these transport mechanisms require that the data be modulated on some carrier. This section provides a brief introduction to the digital modulation schemes that are used to achieve this.

This section discusses just the modulation techniques. Other processes are necessary to make a data stream suitable for transmission, and will depend on both the modulation chosen and the channel characteristics. These techniques will be discussed in Section 10.

9.1 Principles of Modulation

A continuous carrier conveys no information unless it is modified in some way by the information to be transmitted. A carrier can be modified in three ways, by changing its amplitude, frequency or phase. Frequency and phase are, of course, closely related. Although generally treated as separate modulation types, the distinction can become very blurred; some “frequency modulation” systems are implemented by “phase modulators” after performing integration on the modulating signal. Generally these types of modulators are called “angle modulators” and are useful when constant envelope power in the modulated signal is desired.

9.2 Analog Modulation

In the analog world, the amplitude or frequency of a carrier is changed (modulated), according to the amplitude of an audio or video signal, usually according to a linear law. An example where two quadrature-phase subcarriers are amplitude modulated (as described later in Figure 9-3) can be found in the NTSC and PAL color modulation systems. When these two linearly, amplitude modulated sub-carriers are combined, the resulting signal has both simultaneous amplitude and phase modulation.

This familiar operation provides a useful example. We can think of the color information in two ways. It can be expressed in polar form as a phase angle (related to hue) plus an amplitude (related to saturation and color intensity), or in rectangular, “Cartesian” xy form as values of two color difference signals, B-Y and R-Y. If the color difference components are treated as “x” and “y” values on a graph, the two representations are seen to be closely related, as illustrated in Figure 9-1. This figure is similar to the well-known vectorscope display which is essentially a B-Y and R-Y color component constellation diagram.

9.3 Quadrature Modulation

Figure 9.2 shows how we can modulate a carrier with two different signals, using the technique known as “quadrature modulation.” A single carrier is split into two paths, and one path is delayed by a time equal to one-quarter of the cycle time of the carrier or otherwise shifted in phase by 90 degrees to create two carriers in quadrature-phase. The two carriers are each amplitude modulated by an appropriate signal, and the two modulated carriers are then added together. This generates a single signal with amplitude and phase determined by the amplitudes of the two modulating signals.

Demodulation is achieved by an almost identical process; the received signal is sent to two demodulators. In each demodulator the signal is multiplied by a local oscillator signal, synchronized to the received signal. The two local oscillator signals are 90 degrees apart, and each demodulator recovers one axis of the quadrature modulation.

9.4 Simple Digital Modulation Systems

Most digital modulation systems use some form of quadrature modulation; the two axes are usually designated I (In phase) and Q (Quadrature phase). Sometimes only one modulation axis is used. In all cases, digital modulation is really just analog modulation where particular modulating signal levels are used to uniquely define a particular combination of binary bits. These pre-defined analog signal levels are called symbols. All digital modulation schemes represent a compromise of bandwidth efficiency, robustness and complexity. Symbol rate is the principal factor in determining the bandwidth of the transmitted signal. The symbol rate is the rate at which the modulation is changed. With ideal filtering the positive-frequency bandwidth can be as little as half the symbol rate but this requires precise symbol time-sampling and filtering in the receiver. Therefore, typically the symbols are filtered with a little extra bandwidth of 5 to 20% to ease the receiver filtering and sampling. So, with the typical double-sideband signals of quadrature modulation, the bandwidth of the modulated signal is about 5 to 20% larger than the symbol rate. Since the receiver needs to be band-limited to optimize signal-to-noise ratio, the symbol filtering is typically split between the transmitter (limiting the occupied channel bandwidth) and the receiver (limiting the noise bandwidth). The most common way to split the symbol filtering is to take the square root of the desired symbol response and place identical square root filters shapes in both the transmitter output and receiver input. This way the combined response squares the square root to recover the desired symbol response.

Some simple digital modulation systems carry only one bit of information per symbol. In other words, each symbol may represent one of two possible states, representing a binary zero or a binary one via the symbol polarity. In this case the bit rate of the system is the same as the symbol rate. However, other systems have many possible states for each symbol, so they can convey more than one bit of information per symbol. Generally the number of states is made to be a power of two, so the bit rate of the system is some integer multiple of the symbol rate.

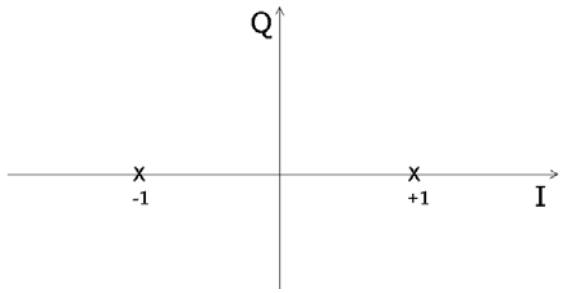


Figure 9-3. BPSK modulation plot (only two possible values).

Digital modulation systems are frequently labeled by the modulation type, preceded by a number representing the number of states for each symbol. For example, 4QAM describes quadrature amplitude modulation with four possible states for each symbol. Four states can convey two bits of information (00, 01, 10, 11), so the bit rate of a 4QAM system is twice the symbol rate.

The simplest digital modulation systems convey one bit of information per symbol. Each symbol has two possible states, representing binary zero and binary one. The states may be created by amplitude, frequency or phase modulation, but frequency and phase modulation are the most common.

Binary frequency shift keying (BFSK or 2FSK) uses one carrier frequency to represent a binary zero, and a different frequency to represent a binary one. Sometimes the frequency difference is very small, and is achieved by a phase modulator after integrating the symbols.

Binary phase shift keying (BPSK or 2PSK) uses one phase of the (constant amplitude) carrier to represent binary zero, and the inverse (180 degrees phase shift) to represent binary one. The different possible states of a symbol are usually shown in a constellation diagram showing the various combinations resulting from permitted values of the I and Q modulating signals. The constellation diagram for BPSK, shown in Figure 9-3, is very simple; only one axis is used, and there are only two permitted values.

These systems can be very robust; the receiver needs only enough signal (or signal-to-noise ratio) to determine which of two possible states has been transmitted for each symbol. However, they do not use spectrum efficiently; the bandwidth is nominally the same as the bit-rate so the bandwidth may be too large at the required bit rate. These systems are used on very difficult transmission paths, such as deep-space telemetry where signal-to-noise, rather than bandwidth, is the limiting factor.

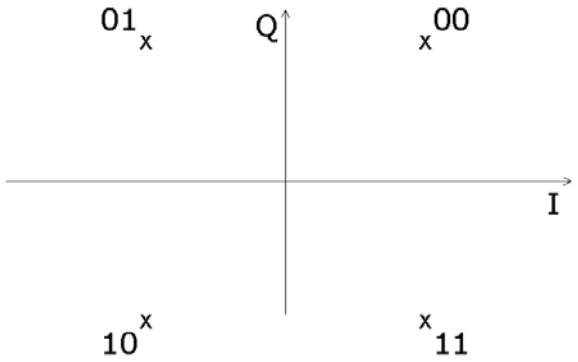


Figure 9-4. QPSK modulation plot (four possible values).

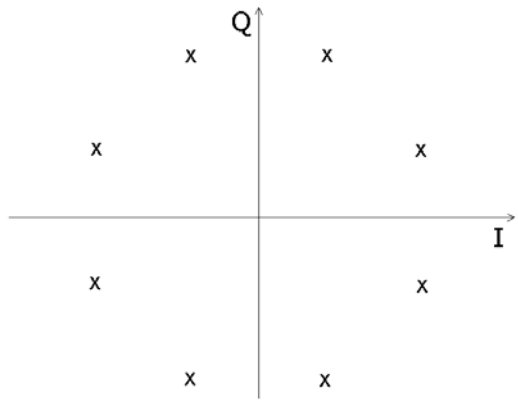


Figure 9-5. 8PSK modulation plot (eight possible values).

9.5 Phase Shift Keying

BPSK or 2PSK was described in the previous section. Other forms of PSK modulation use both the I and Q axes. Quaternary phase shift keying (QPSK, also known as quadrature phase shift keying) is the most common, and uses two values on each axis. The constellation diagram is shown in Figure 9-4. QPSK has four possible states per symbol, so each symbol carries two bits of information; one possible mapping of states to binary values is shown in the Figure. If the I and Q symbols are properly timed, the QPSK signal will have a constant envelope power. Therefore, QPSK is used extensively in satellite communications where saturated power amplifiers with fixed output envelope power are used.

8PSK is less common, but is used in professional satellite transmission systems, particularly in Japan and the USA. The constellation diagram is shown in Figure 9-5. 8PSK carries three bits of information in each symbol, so the bit rate is three times the symbol rate and without special coding requires more carrier power or a larger receive dish, since the signal-to-noise must be higher than that for QPSK.

9.5.1 Turbo 8PSK

Turbo 8PSK uses the advanced techniques of Turbo Coding forward error correction to increase the throughput of satellite broadcast links by using 8PSK without increasing antenna size and carrier power. Turbo coding FEC utilizes statistical techniques for probabilistic data recovery and is discussed in more detail in section 10 on advanced FEC techniques.

9.5.2 DVB S2

DVB-S2 Digital Video Broadcast – Satellite – Second generation is an improved and updated specification designed to supplement the DVB-S specification. It offers higher bandwidth than the original DVB-S specification at the cost of additional complexity and has been driven commercially by the need for high definition satellite broadcast services. DVB-S2 can closely approach the Shannon limit for the maximum data rate of transmission achievable in a given channel bandwidth with a defined carrier to noise ratio (C/N) and offers up to 30% greater capacity than DVB-S, but still provides backward compatibility.

DVB-S2 has two innovations over DVB-S, changing channel encoding parameters in real time using VCM (Variable Coding and Modulation) and ACM (Adaptive Coding and Modulation)

Four modulation modes are available:

- QPSK as in DVB-S.
- 8 PSK.
- 16APSK.
- 32APSK.

DVB-S2 FEC uses a combination of binary BCH coding combined with a very large, low-density parity check code (LDPC) first developed by Gallager in 1958. LDPC coding has renewed interest, eclipsing Turbo coding, due to the large computing power available today. LDPC decoders can provide receiver error thresholds within 0.7 dB of the Shannon limit. In fact, in some cases, the signal-to-noise ratio can be negative. Also, combined BCH and LDPC FEC are used in the China DMB-T/H standard described later.

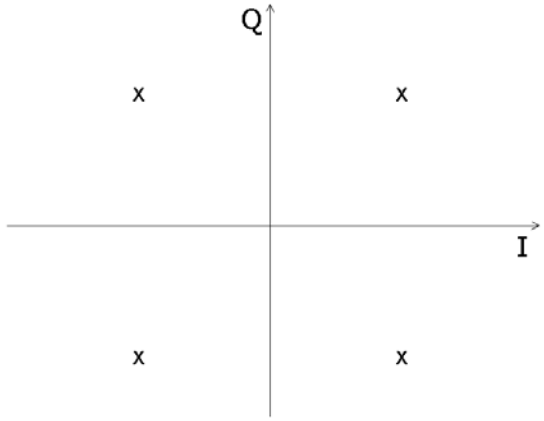


Figure 9-6. 4QAM modulation plot (four possible values).

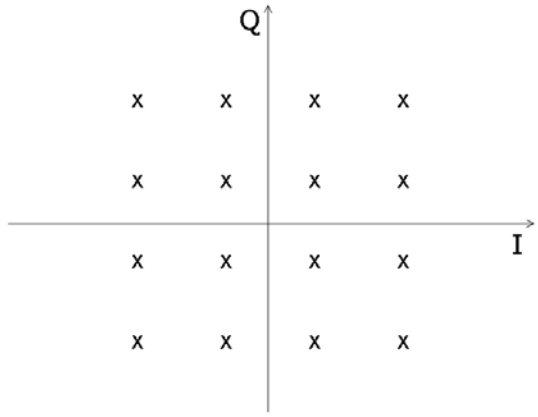


Figure 9-7. 16QAM modulation plot (16 possible values).

9.6 Quadrature Amplitude Modulation - QAM

Quadrature amplitude modulation (QAM) is the basis of many transmission systems. Both the I and Q axes are used for modulation and, depending on the particular variant, two or more amplitude levels are permitted for each axis.

The simplest variant is 4QAM where just two values are used for each axis, providing four possible states for each symbol. The constellation diagram is shown in Figure 9-6, where it will be seen that 4QAM is identical to QPSK, and carries two bits per symbol.

16QAM uses four values on each axis, providing 16 possible states. 16QAM systems carry four bits per symbol. If six values are permitted for each modulation axis, there are a total of 36 possible states. Five bits may be carried using only 32 states, so four of the possible combinations are not used in 32QAM. The constellation diagram for 16QAM is shown in Figure 9-7 and for 32QAM is shown in Figure 9-8. In 32QAM the four “corner” states are not used; these are the states that would represent the highest amplitude and, therefore, the highest transmitter power.

Figures 9-7 and 9-8 also help to show the trade off between bit rate and robustness. In the presence of noise or jitter, the closer spacing of the states in 32QAM (at equal transmitted power) will make decoding errors more likely. Put another way, the more possible states per symbol the better the signal to-noise ratio required for a given error rate.

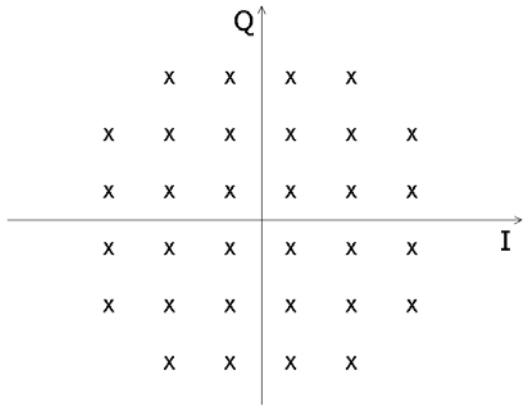


Figure 9-8. 32QAM modulation plot (32 possible values).

When good signal-to-noise can be guaranteed, even greater constellation densities may be used. 64QAM uses eight values on each axis, and carries six bits per symbol. 64QAM is the most extensively used modulation scheme in cable systems worldwide, as it provides a good trade-off between robustness and compatibility with legacy cable infrastructures. 256QAM, used in some of the latest cable television systems, has 16 permissible values for each modulation axis, and carries eight bits per symbol.

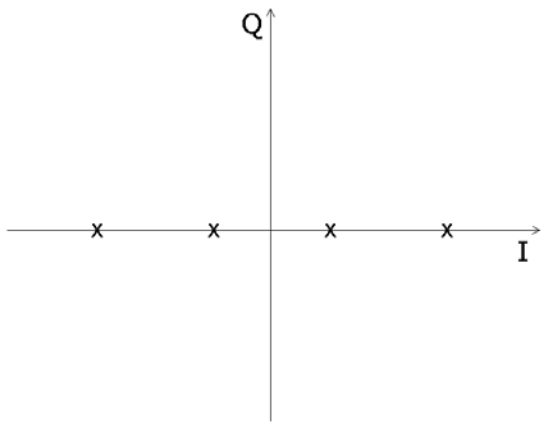


Figure 9-9. 4VSB modulation plot (four possible values).

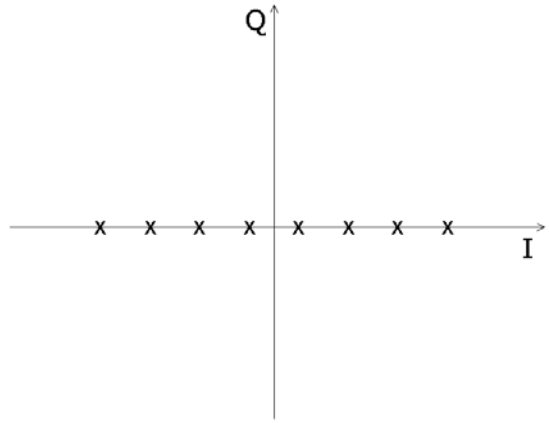


Figure 9-10. 8VSB modulation plot (eight possible values).

9.7 Vestigial Sideband Modulation – VSB

When a carrier is modulated, sidebands are generated above and below the carrier frequency. For example, a QAM system with a symbol rate of 6 megasymbols per second will have upper and lower sidebands each just over 3 MHz, requiring a nominal channel bandwidth of just over 6 MHz. However, using quadrature carriers, at least two bits per symbol are available if for the simplest constellation.

To recover both amplitude and phase information (or both axes of the quadrature modulation), both sidebands must be recovered at the receiver.

Vestigial sideband systems eliminate most of one sideband prior to transmission, so only one modulation axis can be recovered since the quadrature carrier component is not cancelled due the missing lower sideband. (An alternative interpretation is to say that the other modulation axis is used so as to suppress the unwanted sideband.) 2VSB has the same constellation as BPSK. However, since only one sideband is transmitted, the quadrature carrier cannot be used so QPSK is not possible. Recall that QPSK (or 4-QAM) is 2-bits per symbol thereby needing only half the symbol-rate for the same bit rate as 2VSB. Therefore, since QPSK needs a bandwidth about equal to the symbol-rate, 2VSB with only one sideband would occupy the same bandwidth as QPSK for the same bit rate. Figure 9-9 shows the constellation diagrams for 4VSB and Figure 9-10 shows 8VSB, carrying respectively two and three bits per symbol.

8VSB modulation is used in the United States by the ATSC digital television standard. 4VSB was proposed originally, providing two bits per symbol. However, it was found that 8VSB, in conjunction with and extra parity bit using Trellis coding (see Section 10.6), and convolutional inner error correction provides the same date rate with improved signal-to-noise performance.

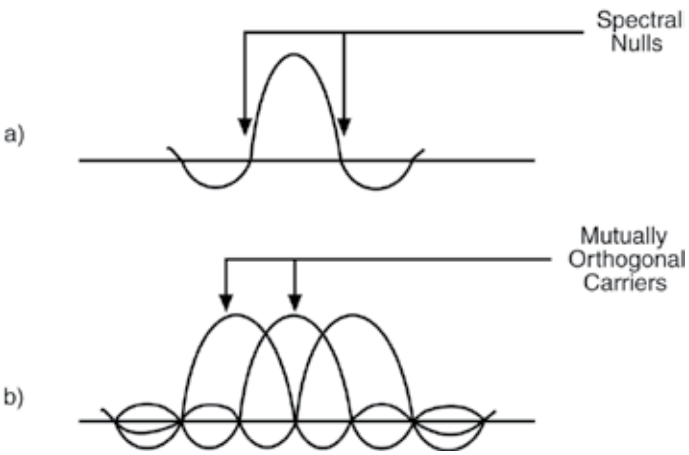


Figure 9-11. Single carrier vs. multiple carrier.

9.8 Coded Orthogonal Frequency Division Multiplex – COFDM

In the above systems, a baseband signal is supplied to modulators that operate on a single carrier to produce the transmitted sideband(s). An alternative to a wideband system is one that produces many narrowband carriers at carefully regulated spacing. Figure 9-11a shows that a digitally modulated carrier has a spectral null at each side. Another carrier of slightly higher or lower frequency can be placed here without interference as long as it has an integer number of cycles more or less per symbol. This creates an orthogonal carrier pair that can be independently modulated. Therefore, the two are mutually orthogonal as Figure 9-11b shows. This is the principle of OFDM (orthogonal frequency division multiplexing). In practice, a variant known as coded orthogonal frequency division multiplexing (COFDM) improves the performance dramatically in non-ideal channel conditions by the use of FEC such as concatenated Viterbi convolutional and Reed Solomon coding, described in the next section. COFDM is used in the digital video broadcasting (DVB) terrestrial system, DVB-T, the Japanese ISDB-T terrestrial system and in all mobile broadcast systems DVB-H, T-DMB, DAB-IP, ISDB-Tss and FLO.

Each carrier in an OFDM system may be modulated by any of the techniques described in this section. In practice QAM is generally used, 16QAM and 64QAM being most common. It is not necessary to use all of the carriers. For example, if one part of the channel is known to be subjected to a high degree of interference, the affected carriers may be omitted.

The number of carriers in an OFDM system can be very large. DVB-T has options for 1705 or 6817 carriers (known as 2k and 8k systems). Because the bandwidth allotted to each carrier is small, the symbol rate is correspondingly reduced, and the length of time to transmit each symbol is increased. This is the key to OFDM's tolerance to multipath interference.

In a single-carrier system, such as 8VSB, a high data rate means that the symbol time is very short. In the case of the ATSC digital television system, some 11 million symbols are transmitted each second, giving a symbol duration of less than 100 ns. This means that even a very short multipath delay will create inter-symbol interference because the delayed signal representing one symbol will arrive during reception of a subsequent symbol.

In contrast, an OFDM system with thousands of carriers will have a symbol time in the order of hundreds of microseconds (depending on the data rate, the number of carriers, and the modulation used). Inter-symbol interference may be virtually eliminated by adding a "guard band" to each symbol,deliberately making the symbol longer than needed. This reduces the symbol rate, but only to a relatively small degree. For example, if the nominal symbol duration is 200 µs, a guard band of 50 µs will reduce the symbol rate by only 20% – and the elimination of inter-symbol interference may allow a higher order constellation to be used, perhaps more than compensating for this loss. This technique is not practical on a wide-band single-carrier system. As an example, the same 50 µs guard band with a 100 ns symbol time would reduce the data rate to a fraction of one percent!

However, OFDM is not fundamentally immune to multipath due to self-symbol interference causing carrier nulls. Therefore, any pilot carriers that are needed at the receiver must be scattered among the various data carriers. In fact, COFDM receivers typically use adaptive equalization to mitigate multipath as required with single carrier, high symbol-rate systems. This tolerance to multipath interference also makes COFDM systems well suited to single frequency networks where two or more synchronized transmitters emit the same signal. A receiver may, depending on its location and antenna system, receive signals from more than one transmitter at different times. If the path lengths are radically different, the main signal will probably be very much stronger than the secondary signal, and interference will be minimal. If path lengths and signal strengths are similar, the guard band will prevent inter-symbol interference.

9.9 China Digital Multimedia Broadcast-Terrestrial/Handheld - DMB-T/H

China has announced its own indigenous digital TV standard designed for both fixed and mobile coverage. Called Digital Multimedia Broadcast-Terrestrial/Handheld DMB-T/H (GB 20600-2006), it is derived from two other standards: ADTB-T (similar to ATSC-T) and a variant of T-DMB.

DMB-T/H is derived from work at Tsinghua University in Beijing and Jiaotong University in Shanghai and combines technology developed by both Universities.

DMB-T/H provides SDTV and HDTV services, Digital Audio Broadcast, Multimedia Broadcasting and data services to fixed receivers. For mobile receivers it provides SDTV, Digital Audio Broadcasting, Multimedia Broadcast and data services.

The standard uses the following technologies for transmission in a 8 MHz channel the UHF and VHF bands and has a number of modes of operation defined for specific multimedia delivery types:

- Support for Multi Frequency Networks (MFNs) and Single Frequency Networks (SFNs).
- Bit randomization.
- FEC (binary BCH and LDPC coding).
- QAM Modulation within OFDM – 64QAM, 32QAM, 16QAM, 4QAM, 4QAM-NR.
- Very deep time domain interleaving.
- 3 code rates, 3 frame sizes and 2 interleaving depths allowing a wide range of bit rates.
- 3780 carrier mode (2 KHz spacing) with Frequency domain interleaving.
- Dual pilot, single carrier mode.
- Spread Spectrum techniques for SI transmission.

9.10 Integrated Services Data Broadcasting (ISDB)

Integrated services data broadcasting (ISDB) is a development that uses many modulation schemes and has been developed for digital television services in Japan. It is designed to support hierarchical systems of many levels. It could be used, for example, to provide simultaneously low data rate reception under exceptionally difficult mobile conditions, intermediate data rate (standard definition) for fringe-area static reception, and high data rate (perhaps for HDTV) for good reception conditions.

There are four ISDB modulation systems currently in use in Japan:

9.10.1 ISDB-S Satellite System

Launched in December 2000, ISDB-S enabled two broadcasters to share a satellite transponder. It is also referred to as BS-digital or CS-digital when the space segment is a broadcast satellite or a communication satellite, respectively.

Up to eight transport streams can be used in all in a manner to be agreed between the broadcasters sharing the transponder. The aggregate bit rate will depend on the transponder bandwidth and the modulation mode used. For example, for a transponder of 34.5 MHz, the maximum rate including forward error correction is 56.610 Mbits/s.

Hierarchical modulation allows the modulation mode to be varied on a packet-by-packet basis within a 48-packet frame. Each packet is assigned a modulation slot. Four modulation modes are supported BSPK(1/2), QPSK (to 7/8) and TC8PSK. The number of slots vary according to the mode used.

9.10.2 ISDB-C Cable System

The main feature of this system is that it transmits multiple transport streams on a single 64QAM carrier. The system was developed in order to be able to retransmit efficiently the information carried on ISDB-S signals. A maximum of 52.17 Mbits/s of information are transmitted typically on a BS-digital carrier. The information rate of a 64QAM/6 MHz signal is 29.162 Mbits/s. Hence, at least two cable television channels must be used to retransmit information of a single BS carrier. The full BS digital service consists of four broadcasters and occupies approximately 174 MHz including guard bands. Using ISDB-C 8 cable channels would be required to carry this information, whereas as many as 29 channels would be required using conventional cable transmission of one transport stream per carrier.

There are 52 modulation slots plus 1 slot for the synchronizing header, TSMF (transport stream multiplexing frame).

9.10.3 ISDB-T Terrestrial Modulation

The ISDB-T channel is divided into 13 segments (typically 400-500 kHz wide), and a separate COFDM transmission is used for each segment. All of the parameters affecting robustness (number of carriers, guard band length, modulation type, convolution coding) may be chosen separately for each layer of the hierarchy. For example, the most robust segment might use a long guard band, QPSK modulation, and 1/2 convolution coding. The highest level could use a shorter guard band, 64QAM, and 7/8 convolution coding – providing many times the data rate of the robust segment.

The center segment may be used for partial reception, designed to allow a narrow band receiver to receive this segment only.

Figure 9-11-2. Mobile phone including mobile TV functionality.

In normal OFDM the complete channel bandwidth represents a single layer. The carriers used are spaced out across the bandwidth at set multiples of a certain frequency. In ISDB-T the channel bandwidth of 5.6 MHz is divided up into 13 segments each having a bandwidth of 429 kHz. Hierarchical transmission of ISDB-T is achieved by transmitting OFDM segment groups having different transmission parameters. These groups of layers constitute the layers. In non-hierarchical modulation the same modulation scheme is used for all 13 segments.

9.10.4 ISDB-Tss Mobile Modulation

Single Segment ISDB-T (ISDB-Tss) uses the central segment of the 13 segment group to transmit services to mobile terminals and is covered in more detail in section 9.11.5.

9.10.5 ISDB in Summary

ISDB-S provides a means of applying different modulation modes to multiple transport streams and transmitting them in a 34.5 MHz channel on a single carrier. ISDB-C provides a means of transmitting multiple transport streams in a single 6 MHz channel on a single carrier but with a common modulation mode. ISDB-T is concerned with up to three transmissions of the same transport stream in a single 6 MHz channel.

MPEG compression is already being used in broadcasting and will become increasingly important in the future. This section discusses the additional requirements for digital television broadcasting, as implemented by the two principal DTV Standards.

9.11 Mobile Broadcast Systems

Mobile broadcast systems are increasingly important as cellular phones and highly efficient digital video compression techniques merge to enable TV and multimedia reception on the move. COFDM channel coding techniques dominate this area due to their robustness and ability to deal with multipath interference and large Doppler shifts of carriers due to the requirement for high speed reception in cars and trains.

The majority of mobile broadcast networks are designed to be Single Frequency Networks (SFNs), i.e. the same transmissions are made from all broadcast points in the network at exactly the same frequency and time aligned, using a GPS derived clock. This has the advantage of spectral efficiency and relies on the ability of OFDM signals to reject self interference from multipath signals using the inter-symbol guard band.

Several standards for mobile TV have been authored by industry bodies and commercial companies and all are deploying, with some predominating in their parent geographic region. Several competing systems are deploying in parallel, leading to a requirement for multi-standard handsets able to roam between networks using different transmission standards. All mobile standards have focused on providing low power, and hence long battery life with highly robust RF transmissions. In a mobile environment handsets will have a small antenna at a height of 1.5-2 m from the ground. This has lead to a series of innovations in advanced channel prediction and FEC techniques. All the transmission standards discussed are true broadcast, point to multipoint systems which only require a forward link to the receiving terminal, most use a cellular back channel to provide interactive services, for example links to web pages.

Most mobile TV transmission systems use a number of common methods to improve the transmission links immunity to the challenging RF environment. These include:

- Robust sync acquisition mode to improve sync time and provide channel mode signaling.
- Deep time interleaving to reduce the impact of impulsive interference.
- Time slicing to reduce power consumption by enabling the handset to only receive on a low duty cycle.
- Variable Fast Fourier Transform (FFT) depth to provide trade offs between SFN size.
- SFN operation for frequency efficiency.

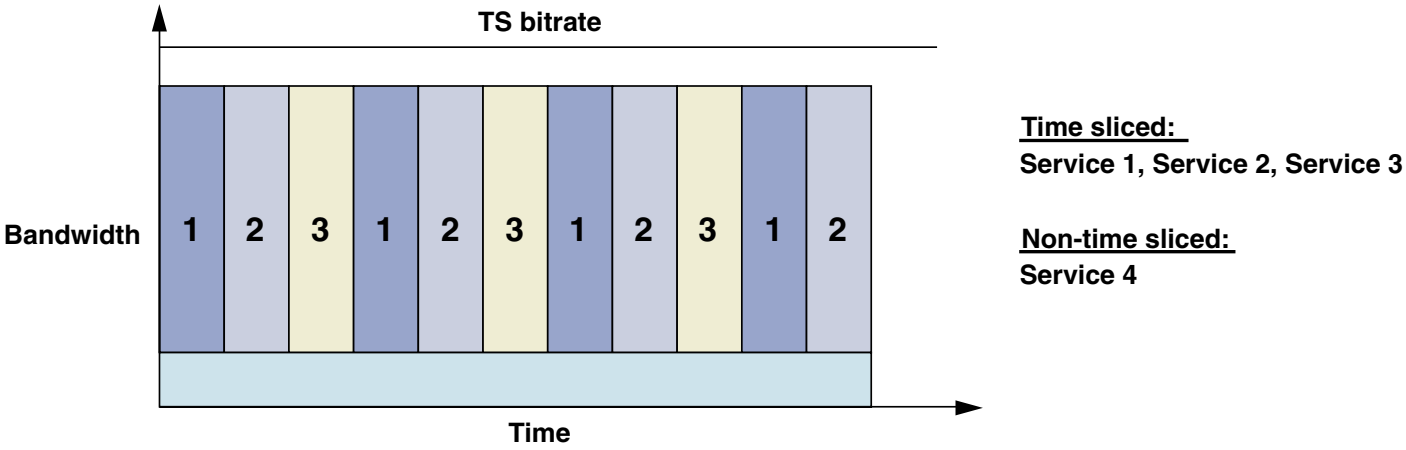


Figure 9-11-3. Services divided into time domain slots so mobile devices do not need to receive and parse the entire stream.

9.11.1 DVB-H

Digital Video Broadcast – Handheld DVB-H has been defined by the DVB as its standard for transmission to mobile terminals, and allows bandwidths of up to ~14 Mbits/s to be received by mobile battery powered terminals. Robust reception at mobile speeds of up to several 100 KM/h is achievable.

The standard has deployed a series of techniques to reduce power and increase robustness of reception, including time slicing of transmissions and an IP based transmission link with a Multiprotocol Encapsulation forward Error Correction (MPE-FEC). Its technical characteristics are:

- Time slicing.
- MPE IP transmission stream.
- MPE FEC (Optional).
- Deep interleaving of data.
- 2K, 4K & 8K FFT Mode.
- Transmission Parameter Signaling (TPS) bits.

The video coding standard used for DVB-H transmission has been left for the network operator to select from DVB approved CODECs. In practice, SMPTE VC-1 and MPEG-4 Part 10 (H.264/AVC) have predominated in operational systems.

DVB-H has been specified to operate in both the existing TV VHF and UHF bands and the L Band with channel bandwidths of 5,6,7 and 8 MHz. While the UHF band is broadly preferred spectrum, commercial issues have seen networks deploying in the L Band successfully.

ETSI ratified standards for DVB-H include amongst others EN 302 304 Transmission System for Handheld Terminals (DVB-H) and TR 102 377 DVB-H Implementation Guidelines standards.

9.11.2 T-DMB

Digital Multimedia Broadcasting-Terrestrial (T-DMB) transmission is based on the Eureka 147 Digital Audio Broadcasting (DAB) standard. Like DAB, T-DMB is designed for transmissions on radio frequency band III (VHF) and L Band. The stream is Reed Solomon FEC encoded (an additional FEC layer to audio DAB) with a parity word of 16 bytes length; convolutional interleaving is then applied.

In order to diminish the channel effects such as fading and shadowing, the DMB modem uses OFDM-4DPSK modulation. T-DMB is an ETSI standard (TS 102 427 and TS 102 428).

T-DMB uses MPEG-4 Part 10 (H.264/AVC) for the video and MPEG-4 Part 3 BSAC or HE-AAC V2 for the audio. The audio and video is encapsulated in MPEG-2 Transport Streams.

9.11.3 DAB IP

DAB IP is a transmission system again derived from the Eureka 147 DAB audio standard, but using the ETSI DAB standard EN 300 401 v1.4.1 for transmission using Internet Protocol (IP) as the transport mechanism.

DAB IP has the advantage of having spectrum in Europe available and the ability to use the existing DAB audio infrastructure for transmission hence saving network rollout cost. The downside is the limited RF channel bandwidth common to all DAB based technologies of 1.7 MHz which limits the bit rate and therefore number of TV channels that can be carried. In addition to TV services DAB radio is also provided by most handsets.

IP transport has been chosen for its flexibility and the ability to integrate with existing transmission networks and offer future migration to higher bandwidth radio transmission systems.

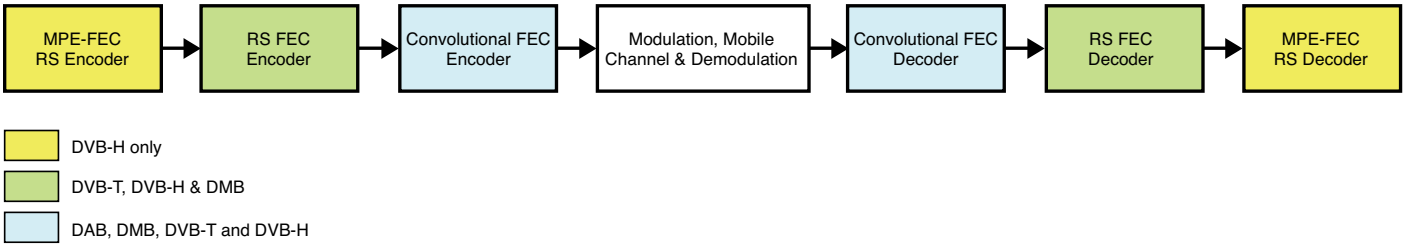


Figure 9-11-4. Forward error correction and redundancy diagram for RF transmission.

Parameter	T-DMB	DVB-H
Channel Bandwidths (MHz)	1.712	5, 6, 7, 8
Error Protection	Convolutional code + RS FEC	Convolutional code + RS FEC+ MPE-FEC
MPE-FEC code rate	No MPE-FEC	Free selection (most likely 1⁄2 to 7⁄8)
Time interleaving	384 ms	Practically up to 1000 ms depending on MPE-FEC selection, typically 200-500 ms
Time Slicing	Micro time slicing	Time slicing
Protocol stack	Raw MPEG-4 (i.e. no IP layer)	IP Layer
Practical data rate	1.06 (@ 1.712 MHz channel, 1⁄4 GI 1⁄2 CR + RS-FEC)	3.32 –13.8 (@ 8 MHz channel, 1⁄4 GI QPSK 1⁄2 CR MPE-FEC 2⁄3 – 1⁄8 GI 16QAM 3⁄4 CR MPE-FEC 5⁄6)
FFT Sizes	2 k, 1 k, 0.5 k, 0.25 k	8 k, 4 k, 2 k
Guard Intervals (us)	246, 123, 62, 31	224, 112, 56, 28, 14, 7
Inner Modulations	Differential-QPSK	QPSK, 16QAM, 64QAM
Convolutional code rates	1⁄4, 3⁄8, 1⁄2, 3⁄4	4⁄9, 4⁄7, 2⁄3, 4⁄5, 1⁄2, 2⁄3, 3⁄4, 5⁄6, 7⁄8

Figure 9-11-5. T-DMB and DVB-H comparison chart.

Mode	Mode 1	Mode 2	Mode 3
Number of Segments		1 or 3	
Segment Bandwidth		6000/14 = 428.57 kHz	
Bandwidth	432.5 kHz (1 segment) 1.289 MHz (3 segments)	430.5 kHz (1 segment) 1.287 MHz (3 segments)	429.5 kHz (1 segment) 1.286 MHz (3 segments)
Carrier Spacing	250⁄63 = 3.968 kHz	125⁄63 = 1.984 kHz	125⁄126 = 0.992 kHz
Number of Carriers	109	217	433
Carrier Modulation	QPSK, 16QAM, 64QAM, DQPSK		
Effective Symbol Duration	252 µs	504 µs	1008 µs
Guard Interval	63 µs (1⁄4) 31.5 µs (1⁄8) 15.75 µs (1⁄16) 7.875 µs (1⁄32)	126 µs (1⁄4)63 µs (1⁄8) 31.5 µs (1⁄16) 15.75 µs (1⁄32)	252 µs (1⁄4) 126 µs (1⁄8) 63 µs (1⁄16) 31.5 µs (1⁄32)
Inner Code	Convolutional Code (1⁄2, 2⁄3, 3⁄4, 5⁄6, 7⁄8)		
Outer Code	Reed-Solomon (204, 188)		
Interleaving	Time and Frequency		
Information Bit Rates (TS Rates)	280.85 kbps (DQPSK, Inner Code = 1⁄2, GI = 1⁄32) -1.787 Mbps (16QAM, Inner Code = 7⁄8, GI = 1⁄4) (1 segment) 0.842 Mbps (DQPSK, Inner Code = 1⁄2, GI = 1⁄32) -5.361 Mbps (16QAM, Inner Code = 7⁄8, GI = 1⁄4) (3 segments)		

Figure 9-11-4. Three modes used by ISDB-T modulation (often, all three at the same time).

9.11.4 FLO

Forward Link Only – FLO is Qualcomm's proprietary mobile TV standard. The system is being submitted to standards bodies for ratification and is expanding from its initial deployment in the USA. FLO is again based on OFDM modulation using a 4 K FFT, but uses hierarchical modulation to provide service coverage to the edges of the transmission area with a reduced bit rate.

FLO also uses time interleaving to reduce susceptibility to impulsive interference and turbo coding for forward error correction (see section 10 on FEC). Its technical characteristics are:

- OFDM 4 K FFT mode based.
- Hierarchical modulation.
- Service to the edge of the cell.
- Time Interleaving.
- Turbo coding.

9.11.5 ISDB Single Segment

ISDB Single Segment transmission is the Japanese standard for mobile TV transmission and unlike the other mobile standards was designed into the Japanese digital terrestrial transmission standard from inception, as part of its hierarchical mode.

ISDB single segment utilizes the fact that ISDB-T splits the transmission channel into 13 segments each of which is an OFDM block. A transport capacity of 280.85 Kbps to 1.17873 Mbps is available per block depending on the chosen coding rate.

The transmission parameters for ISDB are specified on a block by block basis, so that for mobile transmission a robust channel can be coded. This feature is used with the central single segment (ISDB-Tss) to provide portable reception.

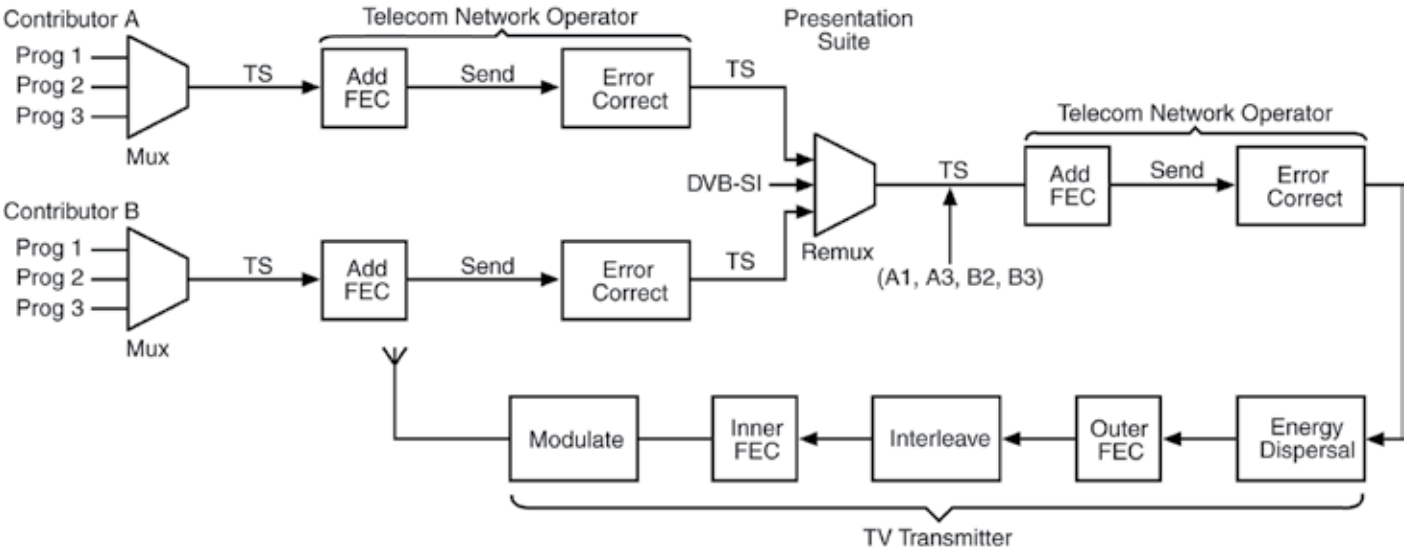


Figure 10-1. Multiplexing, modulation, and transmission.

Section 10 – Introduction to DVB & ATSC

MPEG compression is already being used in broadcasting and will become increasingly important in the future. This section discusses the additional requirements for digital television broadcasting, as implemented by the two principal DTV Standards.

10.1 An Overall View

ATSC (Advanced Television Systems Committee) is a U.S. organization that defines standards for terrestrial digital broadcasting. DVB refers to the Digital Video Broadcasting Project and to the standards and practices established by the DVB Project. This project was originally a European project, but produces standards and guides accepted in many areas of the world. These standards and guides encompass all transmission media, including satellite, cable and terrestrial broadcasting.

Digital broadcasting has different distribution and transmission requirements, as is shown in Figure 10-1. Broadcasters will produce transport streams that contain several television programs. Transport streams have no protection against errors, and in compressed data, the effect of errors is serious. Transport streams need to be delivered error-free to transmitters, satellite uplinks and cable head ends. In this context, error free means a bit error rate (BER) of 1 in 10-11 or better. This task is normally entrusted to telecommunications network operators, who will use an additional layer of error correction as needed (error correction strategies are selected depending on the transmission channel). This layer should be transparent to the destination.

A particular transmitter or cable operator may not want all of the programs in a transport stream. Several transport streams may be received and a selection of channels may be made and encoded into a single output transport stream using a remultiplexer. The configuration may change dynamically.

Broadcasting in the digital domain consists of conveying the entire transport stream to the viewer. Whether the channel is cable, satellite or terrestrial, the problems are much the same. Metadata describing the transmission must be encoded into the transport stream in a standardized way. In DVB, this metadata is called service information (DVB-SI) and includes services such as teletext as well as details of programs carried both within itself and within other multiplexes.

In broadcasting, there is much less control of the signal quality and noise or interference is a possibility. This requires some form of forward error correction (FEC) layer. Unlike the FEC used by the telecommunications network operators, which can be proprietary, (or standardized as per European Telecommunications Standard Institute (ETSI), which defines DVB transmission over SDH and PDH networks), the forward error connection (FEC) used in broadcasting must be standardized so that receivers will be able to handle it.

The addition of error correction obviously increases the bit rate as far as the transmitter or cable is concerned. Unfortunately, reliable, economical radio and cable-transmission of data requires more than serializing the data. Practical systems require channel coding.

10.2 Remultiplexing

This is a complex task because a remultiplexer has to output a compliant bit stream that is assembled from parts of others. The required data from a given input transport stream can be selected with reference to the program association table and the program map tables that will disclose the PIDs of the programs required. It is possible that the same PIDs have been used in two input transport streams; therefore, the PIDs of one or more elementary streams may have to be changed. The packet headers must pass on the program clock reference (PCR) that will allow the final decoder to recreate a 27 MHz clock. As the position of packets containing PCR may be different in the new multiplex, the remultiplexer may need to edit the PCR values to reflect their new position on the time axis.

The program map tables and program association tables will need to be edited to reflect the new transport stream structure, as will the conditional access tables (CAT).

If the sum of the selected program stream bit rates is less than the output bit rate, the remultiplexer will create stuffing packets with suitable PIDs. However, if the transport streams have come from statistical multiplexers, it is possible that the instantaneous bit rate of the new transport stream will exceed the channel capacity. This condition might occur if several selected programs in different transport streams simultaneously contain high entropy. In this case, the only solution is to recompress and create new, shorter coefficients in one or more bit streams to reduce the bit rate.

10.3 Service Information (SI)

In the future, digital delivery will mean that there will be a large number of programs, teletext and services available to the viewer and these may be spread across a number of different transport streams. Both the viewer and the integrated receiver decoder (IRD) will need help to display what is available and to output the selected service. This capability requires metadata beyond the capabilities of MPEG-PSI (program specific information) and is referred to as DVB-SI (service information). DVB-SI is considered to include the NIT, which is optional in MPEG transport streams.

DVB-SI is embedded in the transport stream as additional transport packets with unique PIDs and carries technical information for IRDs. DVB-SI also contains electronic program guide (EPG) information, such as the nature of a program, the timing and the channel on which it can be located, and the countries in which it is available. Programs can also be rated so that parental judgment can be exercised.

DVB-SI must include the following tables over and above MPEG-PSI:

- Network Information Table (NIT). Information in one transport stream that describes many transport streams. The NIT conveys information relating to the physical organization of the multiplex, transport streams carried via a given network and the characteristics of the network itself. Transport streams are identified by the combination of an original network ID and a Transport Stream ID in the NIT.
- Service Description Table (SDT). Each service in a DVB transport stream can have a service descriptor and these descriptors are assembled into the service description table. A service may be television, radio or teletext. The service descriptor includes the name of the service provider.
- Event Information Table (EIT). EIT is a table for DVB that contains program names, start times, durations and so on.
- Time and Date Table (TDT). The TDT is a table that embeds a UTC time and date stamp in the transport stream.

DVB-SI also defines other optional tables including: bouquet association table (BAT), running status table (RST), time offset table (TOT) and the stuffing table (ST).

The ATSC, like DVB, used the MPEG-2 private section table to define several new tables. This set of new mandatory tables defined by ATSC in A/65A is part of the program and system information protocol (PSIP). ATSC PSIP must include the following tables over and above the MPEG-PSI:

- Terrestrial Virtual Channel Table (TVCT) defining, at a minimum, MPEG-2 programs embedded in the transport stream in which the TVCT is carried.
- Master Guide Table (MGT) defining the type, packet identifiers and versions for all the other PSIP tables in the transport stream, except for the system time table (STT).
- Rating Region Table (RRT) defining the TV parental guideline system referenced by any content advisory descriptor carried within the transport stream.
- System Time Table (STT) defining the current date and time of day.
- Event Information Table (EIT-n) defining the first four Event Information Tables (EIT-0, EIT-1, EIT-2 and EIT-3) describing 12 hours of events (TV programs), each with a coverage of 3 hours, and including all of the virtual channels listed in the TVCT.

10.4 Error Correction

Error correction is necessary because conditions on long transmission paths cannot be controlled. In some systems, error detection is sufficient because it can be used to request a retransmission. Clearly, this approach will not work with real-time signals such as television. Instead, FEC is used in which sufficient extra bits, known as redundancy, are added to the data to allow the decoder to perform corrections in real time.

Briefly, R-S codes (a special case of the general BCH codes) add redundancy by adding extra parity data bytes (N bytes for example) to a block of data bytes so that the R-S decoder in the receiver can correct any or all of the bits in each (N/2) byte(s) in the block of bytes. The R-S decoder also can indicate erred bytes that it cannot correct as long as the bytes in error in each block are not excessive. If none of the bits of any of the bytes within the block (including the extra parity bytes) are in error a zero byte-error syndrome is returned. This zero condition is obtained irrespective of the data and makes checking easy. In Transport Streams, the packets are always 188 bytes long prior to the addition of error-correction data. The addition of 16 bytes of R-S redundancy (parity) produces a packet length of 204 bytes. In practice, transport streams may use 204 byte packets even when FEC is not present. The use of 16 stuffing bytes avoids reclocking the stream when FEC is added or deleted. Note that the R-S parity bytes do not always have to be 8 bits to match the packet data byte size. For example, QAM-J83B uses a 7 bit R-S byte size on blocks of 8 bit data bytes so there is not a 1:1 correspondence to the packet data bytes.

In the event that the syndrome is non-zero, solving the simultaneous equations will result in two values needed for error correction: the erred byte and the bits within that byte that are in error. However, if the number of erred bytes exceeds half the number of parity bytes, the bits in error cannot be corrected. Unfortunately, in typical transmission channels, the signal quality is statistical. This means that while single bits may be in error due to noise, on occasion a large number of bits, known as a burst, can be corrupted together. This corruption might be due to lightning or interference from electrical equipment and will overwhelm the block of R-S coded data and easily exceed the (N/2) number of correctable bytes.

It is not economic to protect every code word against such bursts, because they do not occur often enough. The solution is to use a technique known as interleaving. Figure 10-2 shows that when interleaving is used, the source data are FEC coded, but prior to transmission, they are fed into a RAM buffer. Figure 10-3 shows one possible technique in which

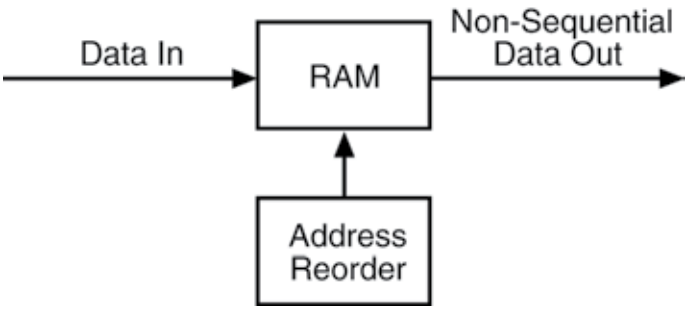


Figure 10-2. Large RAM buffer for byte re-ordering.

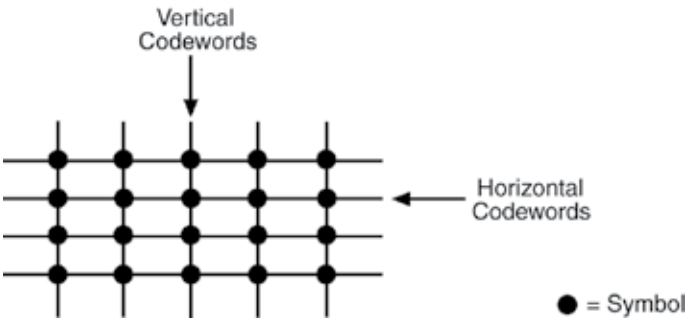


Figure 10-3. Sequentially filling the array (left to right, top down).

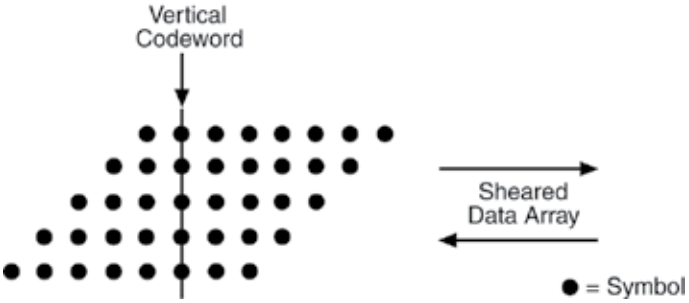


Figure 10-4. Sending out data using vertical columns. With FEC, a burst of lost symbols will not be impacted.

data enters the RAM in rows and is then read out in columns. The reordered data are now transmitted. Upon reception, the data are put back to their original order, or de-interleaved, by using a second RAM. The result of the interleaving process is that a burst of errors in the channel after de-interleaving becomes a large number of single-byte errors in each R-S data block spanning a large number of R-S data blocks, which are more readily correctable.

When a burst error reaches the maximum correctable size, the system is vulnerable to random bit errors that make code words uncorrectable. The use of an inner code applied after interleave and corrected before de-interleave can prevent random errors from entering the de-interleave memory.

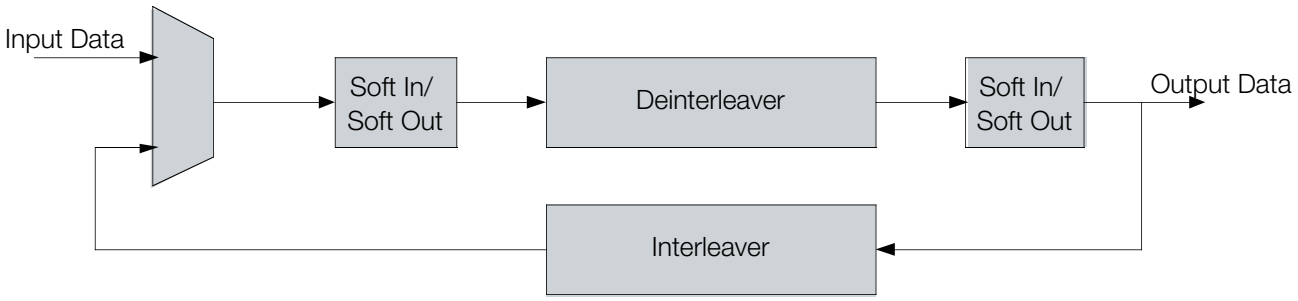


Figure 10-5. Interleaver diagram.

As Figure 10-3 shows, when this approach is used with a block interleave structure, the result is a product code. Figure 10-4 shows that interleave can also be convolutional, in which the data array is sheared by applying a different delay to each row. Convolutional (Forney), or cross interleave, has the advantage that less memory is needed to interleave and de-interleave.

10.4.1 Turbo Coding

Turbo coding is a class of recently developed (1992) error correction codes designed to achieve maximal information transfer over a limited bandwidth channel which is subject to noise.

The key difference from established error coding techniques is that the receiver is designed to provide not only a received data bit stream, but also a likelihood measure for each bit being correct.

At the transmit encoder, data is split into three sub blocks of bits, as shown in Figure 10-5. The first sub block of m bits is the data payload. The second sub block is n/2 parity bits for the payload data computed using a Recursive Systematic Convolution (RSC) code. The third block is n/2 parity bits calculated for a known permutation of the payload data. The transmitted data is therefore m+n bits and includes two redundant but different sub blocks of parity data.

At the decoder the front end receiver produces two streams of integers from the two n/2 parity streams for each bit of the data stream which corresponds to the likelihood that the bit is a 0 or 1. This is referred to as the soft bit. This integer data is in the range -127 to +127 with -127 meaning certainly 0, -100 means very likely 0, 0 means could be 1 or 0 and similar values of + probability for 1 bits.

Key to turbo code is the use of the two likelihood streams generated by the receiver to resolve differences between the two decoded parity bit streams. Each of the two parity decoders generates a hypothesis for the payload sub block pattern of m bits. The hypothesis bit patterns are compared and if they differ the two decoders exchange the likelihoods they have for each bit in the hypothesized m bit payload. Each decoder then incorporates the derived likelihood estimates for the bits from the other decoder to generate a new hypothesis for m. They then compare the new hypothesis until agreement is achieved.

This iterative process continues until the decoders reach a converged solution for m on which they agree to be correct. A characteristic of some Turbo decoders is that the corrected BER can have a floor where the BER does not improve with increases in S/N but this BER is typically designed to be much better than the required BER.

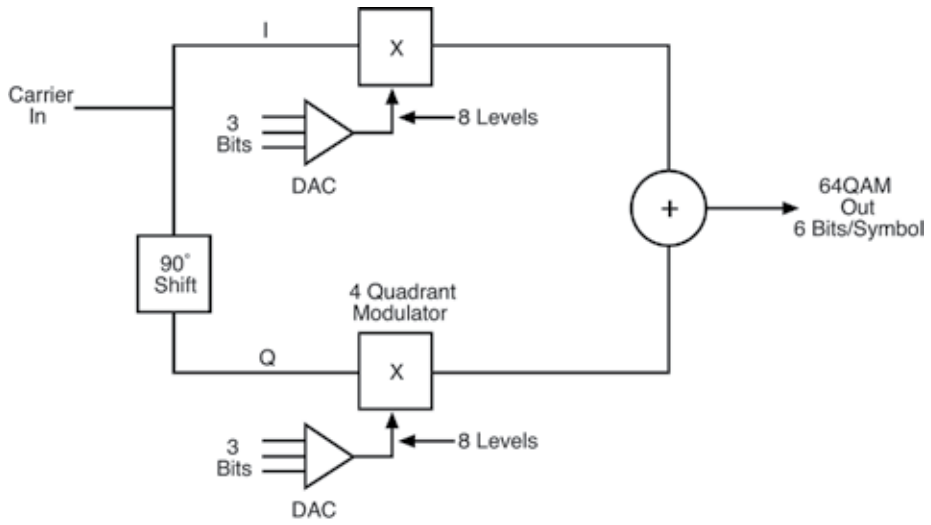


Figure 10-6. QAM modulator.

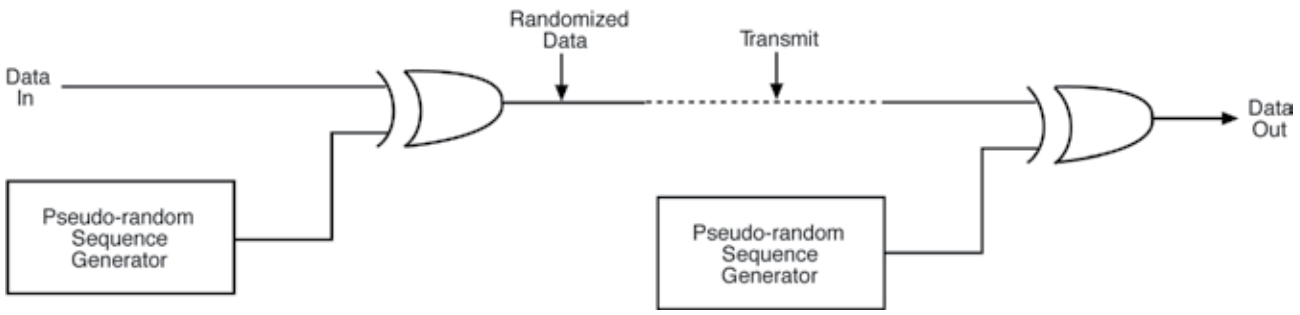


Figure 10-7. PRBS used with input stream to disperse energy.

10.5 Channel Coding; Energy Dispersion

Raw serial binary data is unsuitable for transmission for several reasons. Runs of identical bits cause DC offsets and lack a bit clock. There is no control of the spectrum and the bandwidth required is too great. In practical radio and cable systems, a modulation scheme called a channel code is necessary. Digital modulations schemes are discussed in Section 9. Figure 10-6 shows the application of these principles to a 64QAM modulator.

In the schemes described above, the transmitted signal spectrum is signal dependent. Some parts of the spectrum may contain high energy and cause interference to other services, whereas other parts of the spectrum may contain little energy and be susceptible to interference. In practice, randomizing is necessary to de-correlate the transmitted spectrum from the data content. Figure 10-7 shows that when randomizing or energy dispersal is used, a pseudo-random sequence is added to the serial data before it is input to the modulator. The result is that the transmitted spectrum is noise-like with relatively stationary statistics. Clearly, an identical and synchronous sequence must be subtracted at the receiver as shown. Randomizing is typically (DVB and ATSC) not applied to sync patterns like the TS sync bytes, to simply detection but this is not always true. For example the new China standard does randomize the TS sync bytes (payload agnostic) and determines frame timing from the symbol synchronization.

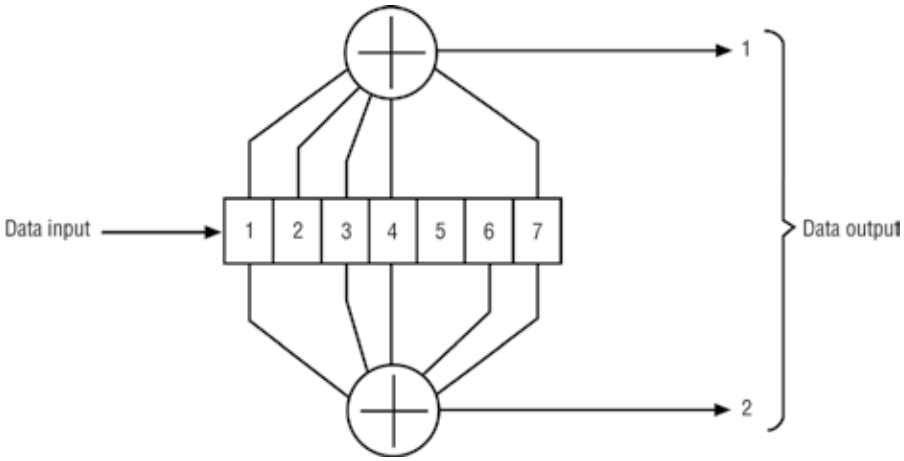


Figure 10-8. Convolutional coding.

10.6 Inner Coding

The inner code of a FEC system is designed to prevent random errors from reducing the power of the interleave scheme. A suitable inner code can prevent such errors by giving an apparent increase to the SNR of the transmission. Typical inner codes are convolution codes where, for example, an extra bit is added based on the previous bit sequence. The ratio of data bits to total bits is called the convolution code rate. In Trellis coding (developed by G. Ungerboeck in 1982), which can be used with multi-level signaling, the convolution code is combined with the symbol constellation mapping where several multi-level symbols are associated into a group. The waveform that results from a particular group of symbols is called a Trellis. If each symbol can have eight levels, then in three symbols there can be 512 possible trellises. Prior to Ungerboeck the inner convolution coding was independent of the symbol mapping or constellation, as in Figure 10-8. By combining the processes, a substantial improvement in effective coding gain or S/N improvement was achieved.

In Trellis coding, the data are coded such that only certain Trellis waveforms represent valid data. If only 64 of the Trellises represent error-free data, then two data bits per symbol can be sent instead of three. The remaining bit is a form of redundancy because Trellises other than the correct 64 must be due to errors. If a Trellis is received in which the level of one of the symbols is ambiguous due to noise, the ambiguity can be resolved because the correct level must be the one which gives a valid trellis. This technique is known as maximum-likelihood decoding made famous by A. Viterbi.

The 64 valid Trellises should be made as different as possible to make the system continue to work with a poorer signal-to-noise ratio. If the Trellis coder makes an error, the outer code will correct it.

In DVB, standard Viterbi inner convolutional coding may be used. Figure 10-8 shows that following interleave, the data are fed to a shift register. The contents of the shift register produce two outputs that represent different parity checks on the input data so that bit errors can be corrected. Clearly, there will be two output bits for every input bit; therefore the coder shown is described as a 1/2 rate coder. Any rate between 1/1 and 1/2 would still allow the original data to be transmitted, but the amount of redundancy would vary. Failing to transmit the entire 1/2 output is called puncturing and it allows any required balance to be obtained between bit rate and correcting power.

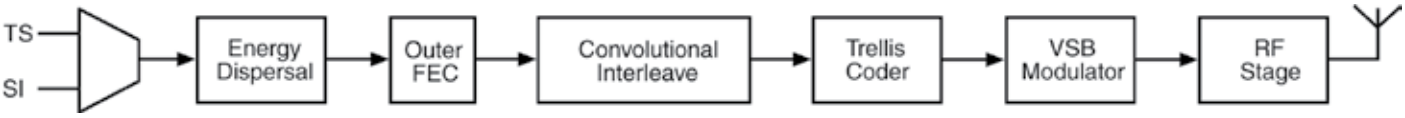


Figure 10-9. ATSC 8VSB RF from TS to modulation and upconversion.

Figure 10-10. DVB COFOM RF from TS to modulation and upconversion.

10.7 Transmitting Digits

Figure 10-9 shows the elements of an ATSC digital transmitter. Service Information describing the transmission is added to the Transport Stream. This stream (except for the packet 0x47 sync byte) is then randomized prior to routing to an outer R-S error correction coder that adds redundant bytes to the data of each TS packet. A convolutional interleave process then reorders the data so that adjacent data in the Transport Stream are not adjacent in the transmission. An inner trellis coder and interleave is then used to produce a multi-level signal for the vestigial sideband (VSB) modulator. In ATSC the 0x47 sync byte is actually never coded and sent by the transmitter. In place of that byte a special, 2-level, sync pulse is sent and the receiver always inserts a 0x47 into that position so it does not have to decode the sync byte from the randomization, R-S coding and Trellis coding.

Figure 10-10 shows a DVB-T transmitter. Service information is added as before, followed by the randomizing stage for energy dispersal (except for the packet 0x47 sync byte). Outer R-S check symbols are added prior to interleaving. After the interleaver, the inner coding process takes place, and the coded data is fed to a COFDM modulator. The modulator output is then upconverted to produce the RF output.

At the receiver, the bit clock is extracted and used to control the timing of the whole system. The channel coding is reversed to obtain the raw data plus the transmission errors. The inner code corrects random errors and may identify larger errors to help the outer coder after de-interleaving. The randomizing is removed and the result is the original transport stream. The receiver must identify the PAT, the Service Information (SI) and PMT that the PAT points to so the viewer can be told what is available in the Transport Stream and the selected program can be located in the multiplex.

Section 11 – Data Broadcast

The previous sections have looked at the basics of an MPEG-2 transport stream and their main application of carrying compressed video and audio streams, similar to conventional analog broadcasts. However one of the major advantages of an MPEG-2 transport stream is that it can carry data as well as video and audio. Although analog television systems can and do carry data, their maximum data bandwidth is severely limited compared with the bandwidths possible on a transport stream.

This section will provide an overview of the different methods provided by MPEG-2 and the regional variants (DVB, ATSC and ARIB (Association of Radio Industries and Businesses)) to encapsulate data within a transport stream. The next section will then deal with how the data is presented to the viewer by the set top box (which does not strictly fall within the scope of MPEG-2).

11.1 Applications

There are many different types of applications for data broadcast over a transport stream, and each application type may require different types of data with different timing requirements. For example, the type of data involved in sending Internet traffic is very different from that needed to provide a firmware update for a set top box. A non-real time update of pricing information has very different timing requirements from a quiz application where answers and questions must be sent in close synchronization with video/audio.

MPEG-2 provides a large variety of different techniques to send data. The choice of technique is a trade-off between optimizing bandwidth (and hence the cost of providing the service) while meeting the timing requirements of the application.

The types of applications can be loosely grouped by their real-time requirements and level of interactivity as described in the next sections.

11.1.1 Program Related Data

The base MPEG-2 specification does not have any provision for an EPG that can give information about the TV channels or individual programs being carried on the transport stream. The regional variants ATSC, DVB and ARIB have used the MPEG-2 private table syntax such as the EIT to provide additional data about programs. These tables are required to be broadcast at regular intervals and give start times, synopsis and other information about services and programs.

However even this level of information may not be enough and so there are a number of proprietary EPGs that extend the amount of information available and also provide facilities like enhanced searching, favorite channels and other augmented services. These enhanced EPGs are usually permanently resident on the set top box and use a combination of the standard table information and proprietary data formats. Usually this sort of program information is not time critical and so cheap low bit rate techniques are used. In many cases the data is designed to be cached by the set top box so it appears to be immediately available to the user, even if the actual total transmission time is relatively long.

11.1.2 Opportunistic Data

It is rarely possible to utilize the complete bandwidth of a transport stream with video and audio streams, not least because of the need to handle the complex constraints of remultiplexing or table insertion. Opportunistic data systems make use of this spare capacity by replacing some null packets with useful data. However the bandwidth of this data cannot be guaranteed and may be very small. Hence it can only be used for applications with no real-time constraints.

Some applications using this sort of data could be file transfers such as price lists or stock level data distribution via satellite to all company locations. The only constraint is that the transfer must take place overnight and so the low bandwidth and hence cheap opportunistic data services can be used.

11.1.3 Network Data

A very common application is to simply use the broadcast transport stream as a carrier of network data. In the simplest case, a satellite link would provide a very high bandwidth interconnect between two geographically separated companies. This network link could be used to carry virtually any type of network protocol.

However the most common current use is for high speed Internet downloads using just the IP protocol. In this scenario a low-speed Internet connection is used to send commands and page requests to an Internet server. However the Internet server will then route any requested data via the satellite link at speeds much higher than that possible by an ISDN or average broadband connection.

A variant on this application is used when only a slow back channel is available. In this case, for example, data related to the current TV program can be delivered on demand over the transmitted signal without viewers being aware that they are accessing an Internet site.

11.1.4 Enhanced TV

In an enhanced TV application there is no back channel, and so all data required must be sent over the transmitted transport stream. Although this imposes obvious constraints on the design and range of an application, it is still possible to produce a very wide range of games, quizzes and infomercials where the viewer may be completely unaware that all interaction is with the TV only. This is especially the case with pre-recorded material, where it is relatively straightforward to synchronize data pop-ups, such as quiz questions and answers, with the audio/video.

11.1.5 Interactive TV

The key difference between enhanced and interactive TV is that interactive TV has a back channel to send or receive highly personalized data. This greatly increases the range of applications that can be supported, including the ability to provide real time interaction with other viewers, such as multi-player gaming or voting, as well as truly personalized feedback.

A further extension of this is when high-speed Internet connection is combined with a true back channel. This offers seamless integration of personalized data with broadcast television, while also permitting true VOD delivery of programming or other material.

11.2 Content Encapsulation

The first stage in a data broadcast is to encapsulate the data into a form suitable for transmission on a transport stream. There are a wide variety of different data types and so the MPEG-2 standards provide a number of different encapsulation methods. The various country specific standards such as the DVB and ATSC have further enhanced the basic MPEG-2 options to meet regional requirements, but these all build upon the core MPEG-2 standards.

11.2.1 MPEG Data Encapsulation

11.2.1.1 Data Piping

Data piping is used for simple asynchronous delivery of data to a target set top box on the network. Its main application is to send proprietary data in closed systems where the target set top box has been pre-programmed to look for specific data on certain PIDs. Data is carried directly in the payload of MPEG-2 TS packets without any timing information. A typical application might be a nightly update of product stock levels to all sales outlets throughout a region.

11.2.1.2 Data Streaming

Data streaming is used for the end-to-end delivery of data in asynchronous, synchronous or synchronized fashion. Data is carried as PES packets in a similar method to video and audio services. In fact, video and audio are really specific examples of a synchronous data streaming service. As PES packets are signaled in the service information (SI) tables, and can carry timing information, this approach is more flexible than data piping but in practice is used in a similar fashion in proprietary closed systems.

11.2.1.3 DSMCC – Digital Storage Medium Command and Control

The MPEG-2 DSM-CC specification (ISO/IEC 13818-6) provides further ways of broadcasting data in the sections of a standard MPEG-2 private table. It was originally devised as a way of supporting VOD delivery of program material across a network on a Transport Stream. The protocol has been extended to be able to cope with both on-demand delivery (using the MPE paradigm) as well as periodic delivery (using the carousel paradigm) of data across multiple network providers

11.2.1.4 MPE – Multi-protocol Encapsulation

Multi-protocol encapsulation (MPE) allows a datagram of any communication protocol to be transmitted in the section of a DSM-CC table via a Transport Stream. A datagram is a logical structure that contains all defining information about the data, i.e., its size and contents, where it should be going and how it should get there.

The most common application is Internet traffic where the TCP/IP datagram carries information about the logical (IP) addresses of the source and destination (target) as well as the Media Access Control (MAC) address (a unique network address) of the target. However MPE supports nearly any type of network protocol and is certainly not restricted to only TCP/IP data. Indeed, The DVB-H standard makes use of MPE to stream audio and video elementary streams to mobile phones. DVB-H typically uses the UDP/IP protocols instead of TCP/IP, as the data is being broadcast. Many implementations also use the real-time protocol (RTP) to facilitate audio and video synchronisation (lip sync). An example DVB-H protocol stack is illustrated in Figure 11-1.

11.2.1.5 Carousels

Carousels are intended for the periodic transmission of information over a transport stream. Although the content of a carousel can be changed in response to a request from a target user it is more usual for the carousel to be regularly repeated regardless of whether any target is listening or needs that data at that moment. A target that needs a specific data item is expected to simply wait until it is retransmitted.

There are two different types of carousels, object carousels and data carousels, and the main differences between them are that:

- Data carousels contain only modules of data of unspecified content; it is up to the target to know what to do with the data it receives.
- Object carousels contain identifiable data objects such as pictures, text files, or executable application files and contains a directory listing of all objects in the carousel.

Data carousels are often used for downloading new system software to a set top box whereas an object carousel is used for shopping services, EPGs and to send applications and games.

In both data and object carousels, items are repeated at periodic intervals. However, object carousels make it easy to vary the repetition rate of individual objects. For example, the EPG for the next hours viewing may repeat far more often than that for next month. The repetition rates for objects may be a commercial decision made by the service provider to maximize bandwidth utilization.

Both object and data carousels are based upon the DSM-CC extensions to the MPEG-2 specification ISO13818-6, with specific extensions for the DVB, ARIB and ATSC systems.

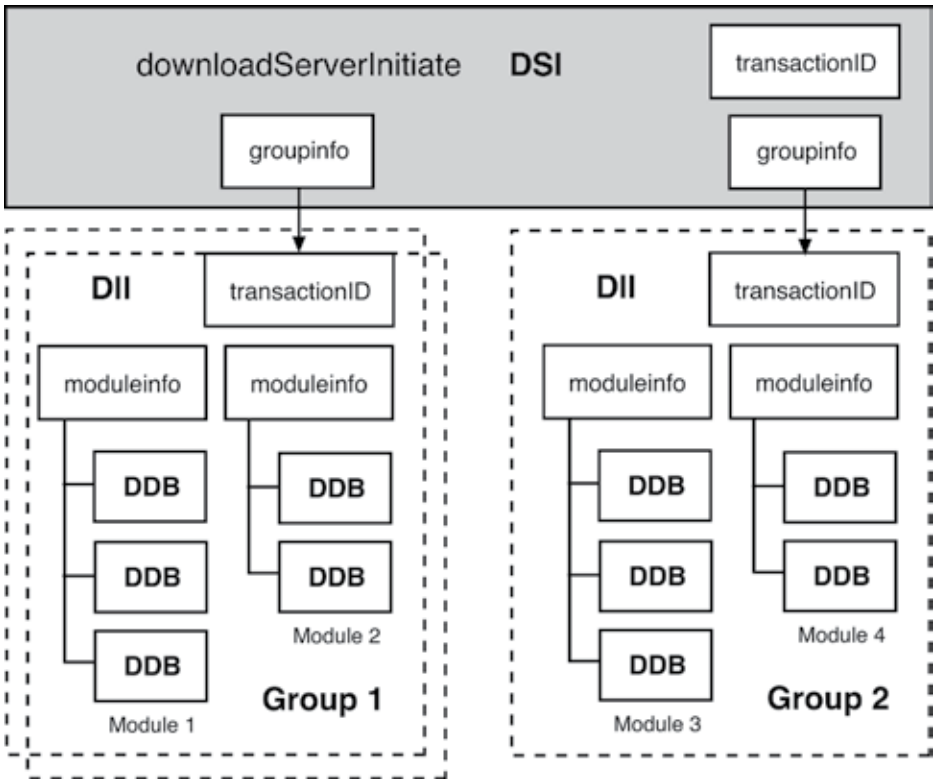


Figure 11-1. MPEG Data Carousel structure.

11.2.1.6 Data Carousels

A data carousel does not contain any individual data items or directory structure but a single monolithic chunk of data. It is up to the target user to know what the data is and what to do with it.

The structure is shown in Figure 11-1. A complete single item of data is defined as a "module." Transmission modules are split up into one or more blocks. Each block is sent as a section in the payload of a DownloadDataBlock (DDB) message, which follows the MPEG-defined private table syntax. DDB messages can be sent in any order or at any periodicity; hence a mechanism is needed to identify which DDBs belong to what modules.

A DownloadInfoIndication (DII) message is used to link the DDBs for a module together. The information for more than one module can be in a single DII message; this forms a Group. Usually a group will contain logically related data modules.

If there are more related modules than can be grouped together into a single DII message then a Supergroup can be formed from multiple DII messages. These are referenced by a DownloadServerInitiate (DSI) message.

A one-layer data carousel contains a small number of modules referenced in a single DII.

A two-layer data carousel contains DSI messages referencing one or more DII messages. It may be smaller or larger than a single carousel.

A typical use for a 2-layer carousel would be for multi-

language support. One group might convey the executable program in one module along with English text in a second module. The second group could then just carry a single module with just French text, saving the overhead of repeating the application module.

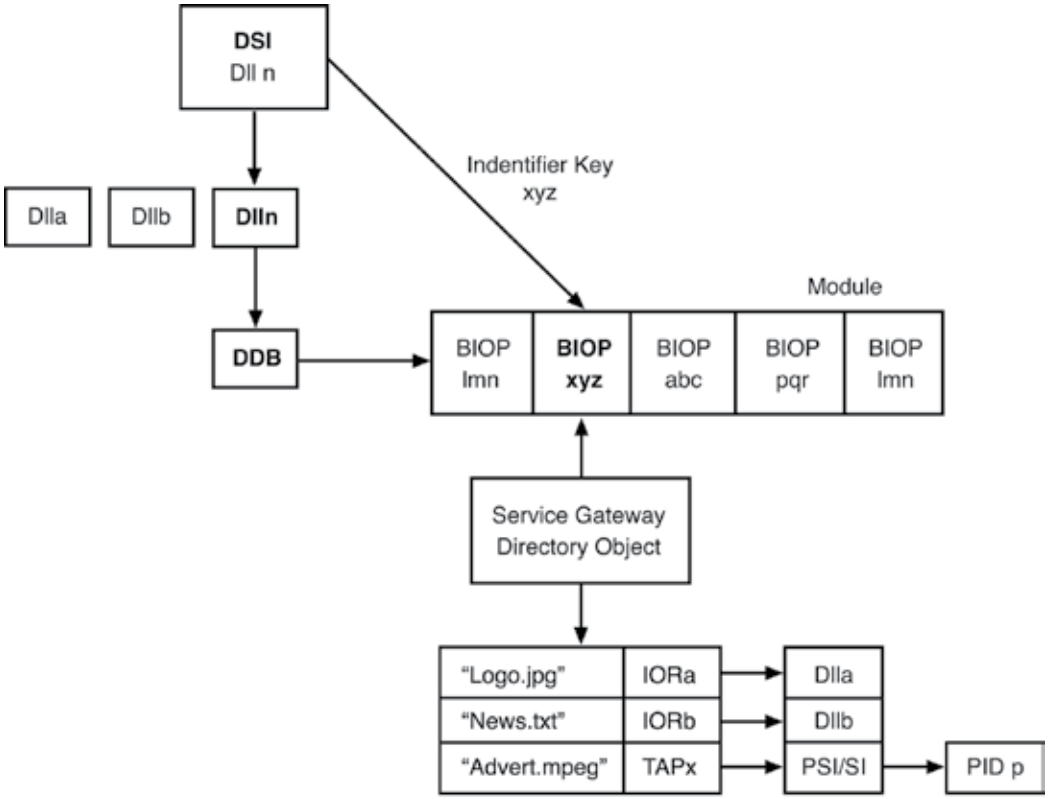


Figure 11-2. MPEG Object Carousel structure.

11.2.1.7 Object Carousels

Object carousels are used to broadcast individually identifiable items of identified data from a server to a receiver. These items are called objects and may be pictures, text files, programs, a pointer to a video PID, a directory listing or service gateway of what is available in the carousel. Related objects grouped and sent together as a single carousel form a service domain. Objects can be sent as often as required and different objects may have very different repetition rates.

A key feature of object carousels is that all objects are sent using the BIOP (broadcast inter-ORB protocol) paradigm. Conventional software developers have been using ORB (object request brokerage) for many years. BIOP extends the basic system to support identifying and using objects in a broadcast environment across different networks from different service providers.

In essence, a BIOP is a method to exchange information about an object being broadcast in the carousel. The BIOP may contain the object or may simply provide a pointer to the object. The BIOP may also indicate how to use the object, including providing a link to where to download the application software needed to use the object.

Object carousels are similar to data carousels in that groups of objects are combined together to form modules. The basic data carousel methodology is then used to transmit that data using blocks, modules and DIIs. The key difference is that the DSI is used to point directly to the Service Gateway directory object, which can then be used to find all other objects in the carousel. This arrangement is shown in Figure 11-2.

11.2.2 DVB Data Encapsulation

DVB has adopted the standard MPEG-2 encapsulation methods with only very minor variations, mainly to remove possible ambiguities that emerged from the original specifications (EN 301 192). These include specifying a slightly different MPE format and imposing restrictions on the maximum data PES packet size. DVB has also taken responsibility for ensuring that there can be no ambiguity between data services that use IORs by providing unique allocation of network IDs and server IDs.

DVB has also defined two specific data streaming PES formats for DVB subtitling and for DVB teletext and have a standardized format for using data carousels to transmit set top box (STB) firmware updates. The DVB have also defined a number of data descriptors and additional tables that are discussed in the signaling and announcement sections below. Most useful of these additions is the ability to use a descriptor to provide a simple mask for the MAC addresses in an MPE datagram. This enables simple support for narrow, multi or broadcast MPE services.

11.2.3 ATSC A/90 Data Encapsulation

The ATSC A/90 specification includes some significant differences from both the DVB and the MPEG-2 standards. First, A/90 does not use object carousels and the CORBA/IOR system at all but instead uses a separate table, the Network Resources Table (NRT), to provide the binding information that links a referenced datum with its real location. Apart from arguably simplifying the system operation it also enables direct support for Internet URLs that is not supported by DVB or MPEG.

Secondly, the A/90 uses only data carousels that can effectively contain only a single object. The data carousel format has been extended to optionally include a DSM-CC adaptation header that may contain a PTS to enable synchronous data delivery. The MPEG concept of NPT within a program is not supported. A/90 also defines additional data types including its own version of DSM-CC MPE as well allowing IP data to be streamed using the PES format.

Mainly due to its lack of support for object carousels, there has been little adoption of A/90.

11.2.4 ARIB Data Encapsulation

The Japanese ARIB B24 standard is arguably the simplest of all systems. It does not support data piping or object carousels. Instead it uses the data carousels format to send one or more entities by imposing a specific directory structure upon the data carousel content. ARIB does not allow references to any entities that are not directly referenced in the PMT and so has no need for either CORBA (common object request broker architecture) or NRT type linkages.

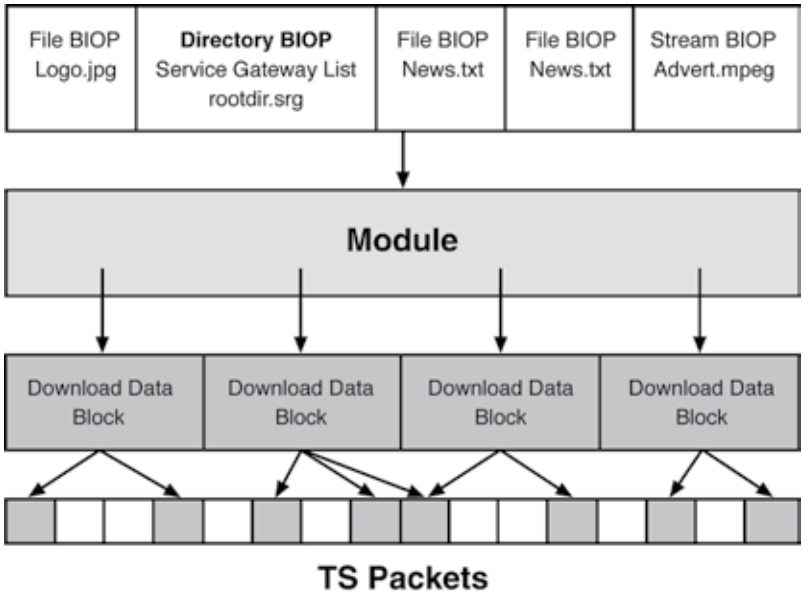


Figure 11-3. Time division multiplexing DSMCC sections in TS packets.

11.3 Broadcasting Data Content

11.3.1 How Object Carousels Are Broadcast

A full explanation is beyond the scope of this document; the following description is a brief and much-simplified overview. (Also see Figure 11-3.)

Directory, file and stream objects are sent in the same method as data carousels i.e., in modules split into blocks are sent as sections in the payload of a DownloadDataBlock (DDB).

A DownloadServerInitiate (DSI) message contains the location of a special directory object called the service gateway. DSI messages are referenced in the SI and so form the starting point to work out what is in a specific object carousel. The DSI references the DownloadInfoIndication (DII) that references the DDB that contain the module in which the service gateway object is sent.

Objects are referenced in a directory object using IORs (inter-operable object references). This contains all the information needed to access an object in the same service domain or on another object carousel (including those broadcast on other Transport Streams).

The name given to the structure in the IOR that describes the location of an object is called a profile body that comes in two flavors:

BIOP profile body – used only for objects within this service domain.

Lite Options Profile Body – used for objects on other servers or transport streams.

An IOR can contain more than one profile body if the object is available on more than one carousel and the set top box can choose the easiest/quickest one to access.

Taps are used to solve the problem that the actual PIDs used to broadcast DIIs, DDBs and video/audio streams are not known until immediately before transmission. Within the carousel therefore all references to PIDs are only made in terms of a tap; the association between a tap and a real PID is made in the SI. This vastly simplifies re-multiplexing streams across different networks.

11.3.2 MPEG-2 Data Synchronization

There is a need for data broadcasts to be synchronized in some way with programs being broadcast. It is not really practical to use the real-time delivery of a datum as the synchronization method, except in very non-critical real time applications such as updating sports scores where a few seconds or more error is of no practical significance. However even a second or two could have a very big impact on, for example, quiz shows where revealing the answer too early could have serious consequences.

MPEG-2 provides different timing mechanisms for the different types of data encapsulation. Data piping and MPE do not support any form of timing mechanism apart from near real-time delivery of the data.

Data streaming PES packets can contain presentation time stamp (PTS) and possibly decoding time stamp (DTS) timestamps for synchronization with the system clock derived from the PCR values in the stream. The mechanism is exactly the same as for video or audio synchronization and decode.

MPEG-2 data carousels have no timing information. However, object carousels can contain a special object called a "stream event" which contains timing information relative to the normal play time (NPT) of an individual television program. The NPT is not the same as the PCR as the NPT clock can be paused during, for example, a commercial break. In other words the NPT of a program can remain in full synchronization with the program timeline, regardless of when it is transmitted or how it is split into commercial breaks.

11.3.3 Signaling Data Contents

Data services are first encapsulated according to the relevant national or international standard and then transmitted via a transport stream to an STB.

However in order to make use of a data service the STB must first know that a data service exists and when it will be available ("announcement"). Secondly it must then be possible to find and identify the various components of the data service ("signaling") when it is actually being broadcast. MPEG-2 only defines the PAT and PMT so in effect all announcement and signaling can only be made by inspection of the elementary stream type and some very basic descriptors.

Although usable there are a large number of limitations to just using only the PAT and PMT, not least that scheduling of data services or events is not possible. Both the DVB and ATSC have therefore defined additional tables that significantly increase the amount of information provided about any data services present. Note that in all cases the data-carrying PIDs must be signaled as present with an entry in at least one PMT.

11.3.3.1 DVB Announcement

The DVB consider data services to be either associated with an individual event (for example a single quiz program) or to be part of a service such as a sports channel. In effect it conforms to the MPEG-2 paradigm where a data service is logically indistinguishable from a conventional audio or video stream.

It therefore has simply defined a number of additional descriptors that are placed into either the EIT or the SDT table which announce that a data service or event will be available. The actual location of data services and components is then signaled using the techniques described above.

The only new table is the application information table, which is defined for use with MHP services only. This contains a list of all applications within a carousel, a pointer to their boot class and optionally allows applications to be auto-started or killed upon a channel change.

11.3.3.2 ATSC Announcement

In contrast to the DVB the A/90 specification treats data events and services totally separately from video and audio services and has added a number of tables to support this paradigm. These are the data services table (DST), which list the data services being sent and the data event table (DET). The DET is exactly analogous to the EIT for conventional program scheduling information and uses an identical syntax. A third table, the long term services table, is similar to the DET but provides long term schedule information beyond the maximum 16 days provided by the DET.

11.4 Content Presentation

The final stage of data broadcasting is running the application on the STB and presenting information to the viewer. It should be noted that a Transport Stream is entirely agnostic to the actual content it is carrying, provided of course that it is encapsulated in the correct protocol for video, audio or data. In other words the following section is almost completely divorced from MPEG-2, DVB or ATSC data broadcasting standards as applications and data are just delivered by the Transport Stream to be used by the STB.

11.4.1 Set Top Box Middleware

In some cases, such as Internet data, the environment in which the data is to be used is self-defined or relatively unimportant. However in most cases, such as games and applications it is critical that the data application is constructed to run properly on the target STB. In most cases the STB will have a software application program interface (API) to run that connects the high level function calls from the application to the software drivers that control the real hardware in the box and link to the data in the Transport Stream. This software layer, which sits between the application and the drivers, is known as the "middleware."

There are however a very large variety of STB in use in the world. Their performance and feature sets run from very basic boxes that can only just decode and display a basic standard definition picture through to the latest units with powerful processing capabilities. Some modern units have high performance graphics chip sets and run full operating systems such as Linux or Windows. They are equipped with hard disk drives for recording programs, have DVD ROMs for DVD replay and can run conventional programs such as Microsoft or Star Office packages.

A simple graphic designed for display on a new high performance box might not even be displayable on an older unit. Application developers would have to write programs completely differently using different programming languages for each type of set top box, middleware and operating system, clearly inefficient and increasing the development cost. There have been a number of attempts to standardize boxes or at least their minimum functionality, where the functionality of the box is clearly defined and controlled by the broadcasters themselves in a very closed fashion.

Another approach has been taken by companies that provide a complete solution by offering a middleware along with the application development tools to develop compliant programs. These systems are more open in that the middleware can be licensed to run on a variety of different set top boxes with different performance or features as long as they meet the basic required functionality.

A proprietary data transmission scheme and a “walled garden” interactive service have been widely used as an interactive system, providing true interactivity via a back channel. Other company offerings are also widely used throughout Europe and applications can be delivered using standard object carousels. There are many other proprietary systems.

A more open standard is an application environment defined by the Multimedia Hypermedia Experts Group (MHEG). A particular variant of this, MHEG-5, is optimized for low performance low memory applications and set top boxes in particular. MHEG-5 has had some success, notably in the UK where it is used for terrestrial broadcasts to STBs.

The current trend is therefore towards standardizing upon an open middleware with a clearly defined programming API. This has the benefit that the consumer can choose from a wide variety of set top boxes whilst enabling application developers to reuse code.

11.4.2 History of Open Middleware Standards

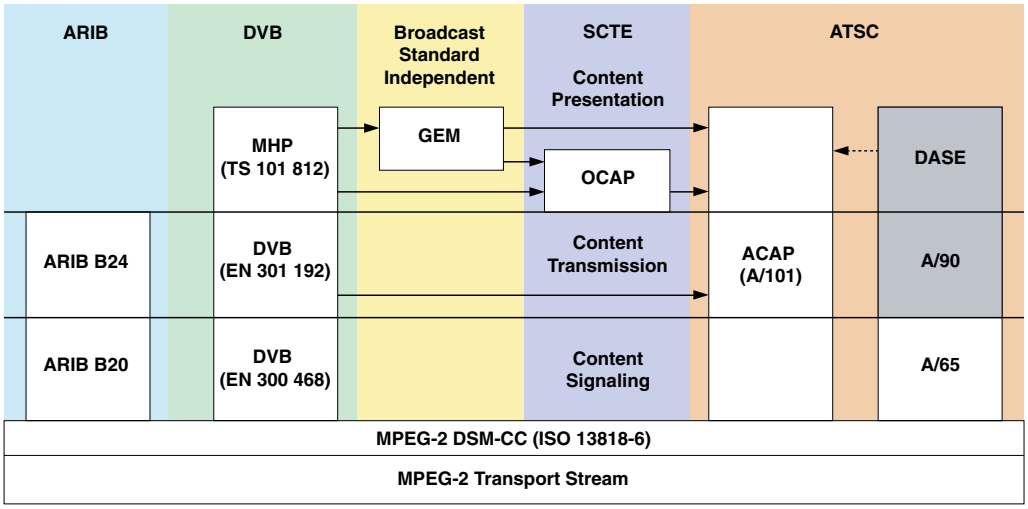
A couple of front-runners for a global Middleware standard quickly emerged, the Multimedia Home Platform (MHP) from the DVB and the Digital TV Application Software Environment (DASE) from ATSC. Both standards build upon their respective Data Encapsulation standards, EN 301 192 for the DVB and A/90 for ATSC, as illustrated in Fig. 11-4. The lack of interoperability and object carousel support has meant that DASE has not been widely adopted. CableLabs in the USA decided to develop a standard that addressed these deficiencies for the US cable TV market. They collaborated with DVB in its work on open middleware standards, leading to the MHP specification being used as the basis for the OpenCable™ Applications Platform (OCAP) in January 2002.

With several organizations around the world striving for the same goal in creating open middleware systems, it obviously made sense to seek some form of harmonization. For an industry that is as global as the television industry, this is even more important. The GEM (Globally Executable MHP) work item in DVB came about after a request from CableLabs to consider the unification of MHP with the original DASE standard from ATSC.

The GEM specification was first published in February 2003 and is a subset of MHP. It was designed to address interoperability issues between the various open standard middleware specifications. These issues included:

- Interoperability problems that beset previous middleware standards, such as OCAP and DASE.
- Elements of the transmission system; in particular, modulation systems, delivery mechanisms and CA systems.
- Specific network operator market requirements.

ATSC then developed the ACAP (Advanced Common Application Platform) standard. ACAP is primarily based on GEM and DASE, with additional functionality from OCAP. Indeed, equipment that is ACAP compliant, will also comply with the GEM and OCAP standards. ACAP was intended to provide access to the various advantages of these standards to the ATSC markets, thus replacing the outdated DASE.



Relationships between Open Middleware Standards

Figure 11-4. Each standard makes use of the 13818-6 standard.

11.4.3 The DVB Multimedia Home Platform (MHP)

The multimedia home platform (MHP) defines a very comprehensive API with several different profiles and levels to allow for different performance and cost STB and services. Enhanced TV and Interactive TV (with back channels) are supported, and it also provides support for Internet connections, which is otherwise not directly supported by the other DVB standards.

MHP applications are delivered using standard DVB object carousels and announced using the AIT table described earlier. This allows applications to be auto-started, paused or stopped and for resources to be explicitly cached in order to speed application boot times. The MHP standard is based on Java technology and encompasses a set of APIs derived from existing standards such as JavaTV, HAVI (home audio video interoperability) and DAVIC (Digital Audio Visual Council).

Applications can be either Java- or HTML-based and there are a number of MHP software development kits (SDK) available from several different suppliers. Applications will have to undergo a compliance test and be electronically signed to ensure both the interoperability and security of the system.

The MHP logo is also protected and STBs must undergo a compliance process before they can be sold with the logo. Regular interoperability tests have taken place over the last two years to ensure that all MHP applications can run on all MHP set top boxes from a variety of leading manufacturers.

MHP is the de facto standard in Europe, and services officially started in Finland in August 2001. However, MHP is not just a European standard as it has been adopted throughout the world. For example it is mandated in South Korea for satellite data broadcasting and in Australia for terrestrial broadcasting.

11.4.4 ATSC Advanced Common Application Platform (ACAP)

The ACAP middleware standard is intended to provide content and application authors the assurance that their programs and data will be received and run uniformly on all brands and models of receivers. It is intended to apply to broadcast systems and receivers for terrestrial broadcast and cable TV systems. In addition, the same architecture and facilities may be applied to other transport systems, such as satellite.

At the heart of ACAP is the GEM framework. GEM is aimed at allowing varied organizations to create harmony in technical specifications, such as the selection of a single execution engine and (where possible) a common set of APIs. The goal is such that applications and content will be interoperable across all GEM-based platforms.

Even though other standards will not be completely compatible with the full MHP specification, GEM ensures that compatibility will be maintained where it is feasible to do so. The set of technologies where functional equivalents are allowed is negotiated as part of the technical dialogue between the DVB and each of the organizations wishing to use GEM. Additionally, the GEM specification contains a list of those other specifications with which it can be used.

In keeping with the goal of interoperability, signaling of ACAP services is restricted purely to MPEG; that is, from PMTs only. This is in contrast with MHP, which includes DVB specific signaling.

ACAP supports two usage profiles, one is Java only and the other is a combination of Java and XHTML (extended HTML).

Section 12 – MPEG QoS Testing

The ability to analyze existing transport streams for compliance is essential, but this ability must be complemented by an ability to create transport streams that are known to be compliant.

12.1 Testing Requirements - QoS

Although the technology of MPEG differs dramatically from the technology that preceded it, the testing requirements are basically the same. On an operational basis, the user wants to have a simple, regular confidence check that ensures all is well. In the event of a failure, the location of the fault needs to be established rapidly. For the purpose of equipment design, the nature of problems needs to be explored in some detail. As with all signal testing, the approach is to combine the generation of known valid signals for insertion into a system with the ability to measure signals at various points.

One of the characteristics of MPEG that distances it most from traditional broadcast video equipment is the existence of multiple information layers, in which each layer is hoped to be transparent to the one below. It is very important to be able to establish in which layer any fault resides to avoid a fruitless search.

For example, if the picture monitor on an MPEG decoder is showing visible defects, these defects could be due to a number of possibilities. Perhaps the encoder is faulty, and the transport stream is faithfully delivering the faulty information. On the other hand, the encoder might be fine, but the transport layer is corrupting the data. In DVB, there are even more layers such as energy dispersal, error correction, and interleaving. Such complexity requires a structured approach to fault finding, using the right tools. The discussion of protocol analysis of the compressed data in this primer may help the user derive such an approach. Reading the discussion of another important aspect of testing for compressed television, picture-quality assessment, may also be helpful. This later discussion is found in the Tektronix publication, “A Guide to Video Measurements for Compressed Television Systems.”

12.2 Analyzing a Transport Stream

An MPEG Transport Stream has an extremely complex structure, but an analyzer such as the Tektronix MTS400 Series can break down this structure into a logical presentation of the contained data. This allows the user to determine Transport Stream correctness. Much of this analysis can take place in real time on a live Transport Stream. This includes the programs hierarchy and what proportion of the stream bit rate is allocated to each PID.

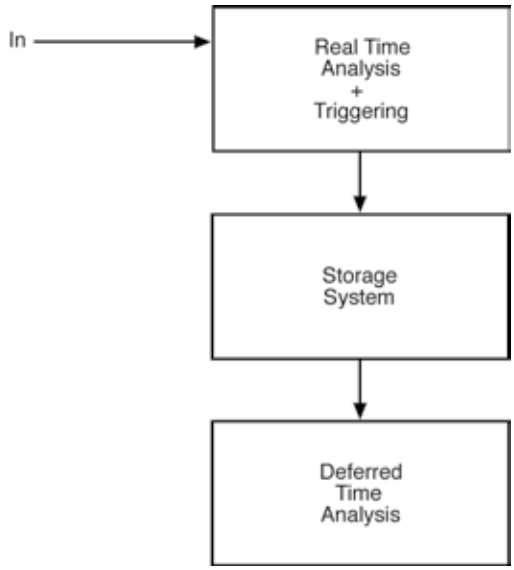


Figure 12-1. Real-time/Deferred-time diagram.

If more detailed analysis is required, this can be achieved by simply recording that part of the Transport Stream for further analysis. This technique is known as deferred-time testing and can be used, for example, to examine the contents of a time stamp.

When used for deferred-time testing, an MPEG Transport Stream analyzer is acting like a logic analyzer that provides data-interpretation tools specific to MPEG. As with all logic analyzers, a real-time triggering mechanism is required to determine the time or conditions under which a capture will take place. Figure 12-1 shows a typical block diagram of an analyzer illustrating the real-time section, storage section, and the deferred section. In real-time analysis, only the real-time section operates, and a signal source needs to be connected. For capture, the real-time section is used to determine when to trigger the capture. Most analyzers include tools known as filters that allow selective analysis to be applied before or after capture.

Once the capture is completed, the deferred section can operate on the captured data and the input signal is no longer necessary. There is also a good parallel in the storage oscilloscope which can display the real-time input directly or save it for later study.

The MTS400 Series Analyzers combine both the real-time and deferred analysis in a single tool. They also incorporate a sophisticated triggering mechanism, called CaptureVu, which causes the bitstream to be stored in volatile memory (RAM) for instant deferred analysis. This enables much faster debug iterations around the real-time analysis, capture and deferred analysis loop.

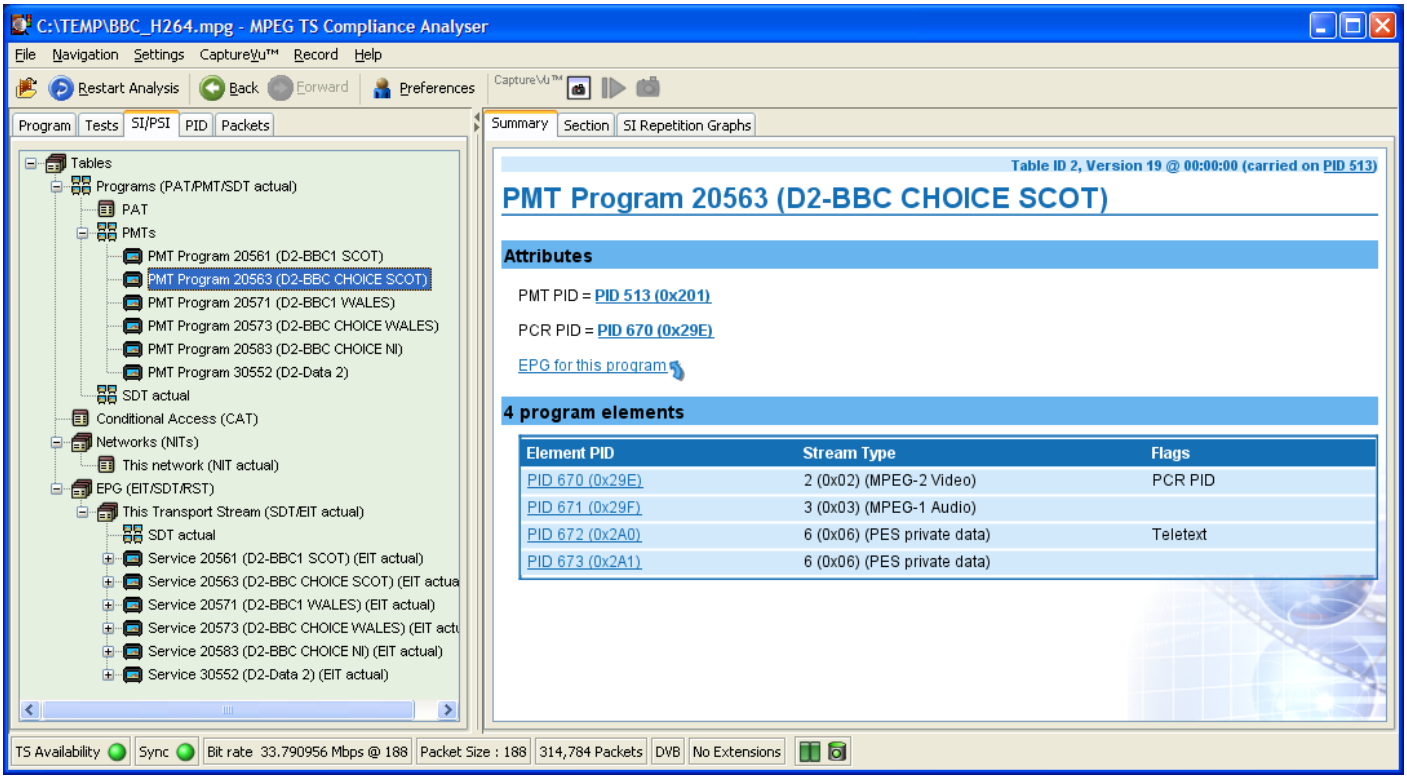


Figure 12-2. Transport stream hierarchy from PAT to PMT, to program elements and PIDs.

12.3 Hierarchic View

When analyzing an unfamiliar Transport Stream, the hierarchic view is an excellent starting point because it provides a graphic view of every table, or SI, component in the bitstream. Figure 12-2 shows an example of a hierarchic display such as that provided by the MTS400. The tree display shows the Transport Stream tables present together with contained stream components. The user can very easily see how many programs are present and the video and audio content of each.

An analyzer creates the Hierarchic view by using the PAT and PMT in the PSI data in the Transport Stream together with any standards specific tables such as the DVB-SI EIT and SDT. The PIDs from these tables are displayed in the Summary pane associated with each table. PAT and PMT data are fundamental to the operation of any demultiplexer or decoder; indeed, if missing or incorrect, it is unlikely that equipment further downstream will be able to interpret the Transport Stream at all. This situation is shown an MTS400 Series as PAT/PMT error.

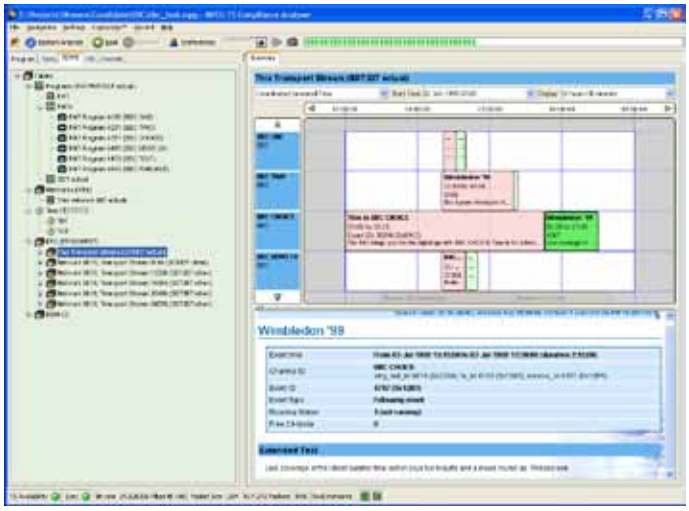


Figure 12-3. Electronic program guide.

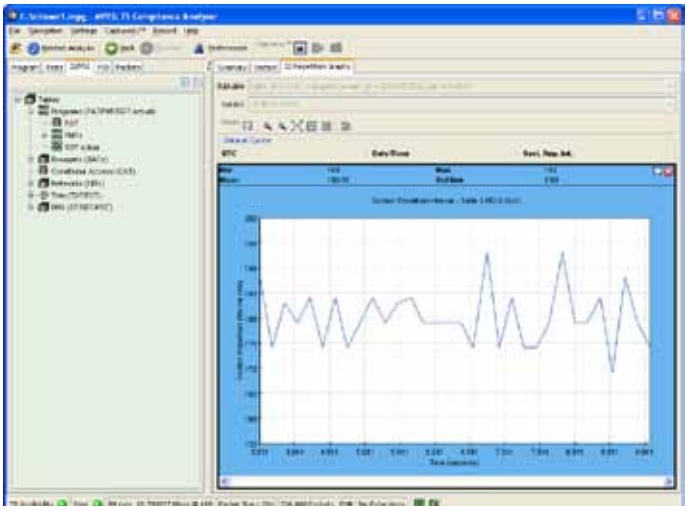


Figure 12-4. PSI/SI/PSIP table interval timing.

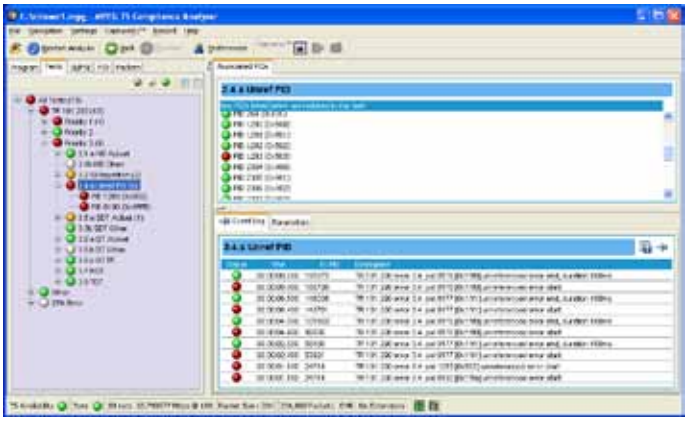


Figure 12-5. Isolated TR 101 290 tests, PIDs, and event logs.

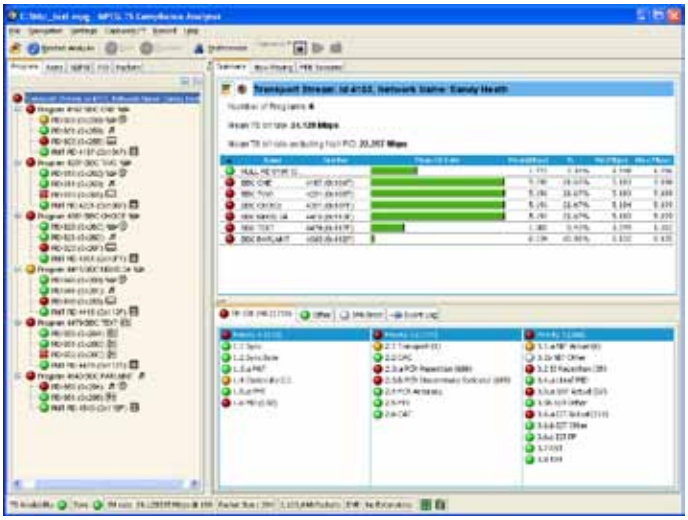


Figure 12-6. Bandwidth chart of each program.

Whilst the Hierarchic view is useful for displaying PSI and SI contents, it is not particularly useful for navigating between tables. The Summary pane in the MTS400 Series contains a series of hyperlinks for quickly navigating the logical connections between tables such as the DVB SDT and PMT as well as the PAT and PMT. In addition, the MTS400 includes an EPG view that comprises the EIT and SDT contents. The user may navigate directly to an individual EIT section of interest from the EPG view (see Figure 12-3). The EPG view also supports the ATSC and ARIB standards.

The ability of a demux or decoder to lock to a Transport Stream depends on the frequency with which the PSI data are sent. The PSI/SI rate option shown in Figure 12-4 displays the frequency of insertion of system information. PSI/SI information should also be consistent with the actual content in the bitstream. For example, if a given PID is referenced in a PMT, it should be possible to find PIDs of this value in the bitstream. The consistency-check function makes such a comparison. Figure 12-5 shows a consistency-error from a bitstream including two unreferenced packets.

A Program Summary view graphically displays the proportions of the transport stream allocated to each PID or program. Figure 12-6 shows an example display. The hierarchical view and the Program Summary show the number of elements in the transport stream and the proportion of bandwidth allocated.

12.4 Standards Compliance Checking

12.4.1 DVB Measurement Guidelines, TR 101 290

Founded in September 1993, the Digital Video Broadcasting Project (DVB) is an industry- led consortium of over 260 broadcasters, manufacturers, network operators, software developers, and regulatory bodies in over 35 countries committed to designing global standards for the delivery of digital television and data services. Although DVB Services are focused on Europe, its standards are available on every continent with more than 120 million DVB receivers deployed.

One of the most important DVB standards is TR 101 290. This provides guidelines for measurement in DVB satellite, cable and terrestrial digital television systems. It defines a number of measurement techniques, such that the results obtained are comparable when the measurement is carried out in compliance with the appropriate definition. The general aim of the tests is to provide a "health check" of the most important elements of the TS; that is, the list of tests is not exhaustive.

The tests are split into three priority groups according to their importance for monitoring purposes:

1. The first priority group lists a basic set of parameters which are considered necessary to ensure that the TS can be decoded. These are mainly Transport Stream synchronisation, dropped packets plus the PSI integrity (PAT and PMT tables).
2. The second group lists additional parameters which are recommended for continuous monitoring. These include the PCR and PTS timestamps and table CRCs.
3. The third group lists additional parameters which could be of interest for certain applications. These are mainly concerned with the integrity of the other SI tables, such as SDT and EIT.

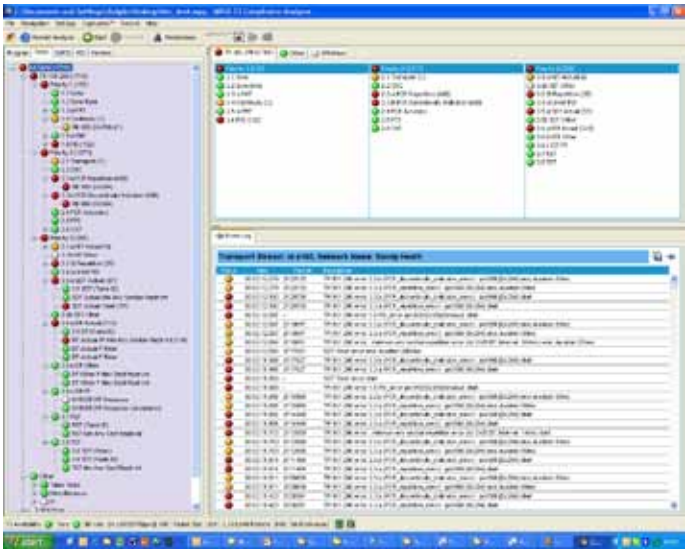


Figure 12-7. Summary of TR 101 290 tests and event log.

Figure 12-7 shows the DVB Tests view from the MTS400 Series. A colored LED shows the current state of each test, including amber for a test that was previously in error. The view allows the user to drill down to individual sub-tests, adjust test parameters and display the filtered log entries. Where applicable, it also displays a list of PIDs that are subject to each test.

Error Description	TOA	POA	CM	QOS	TNC
PAT repetition interval error (100 ms < cycle time <= 200 ms)					X
PAT repetition interval error (200 ms < cycle time <= 500 ms)				X	X
PAT not found for 500 ms (cycle time > 500 ms)	X	X	X	X	X

Table 12-2. ATSC A/78 softens the one hard timer limit defined by TR 101 290. ATSC adds multiple timer limits (i.e., 1x, 2x, or 5x).

12.4.2 ATSC Transport Stream Verification, A/78

The Advanced Television Systems Committee (ATSC) was formed in 1982, as an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast and associated industries. Specifically, ATSC develops television standards among different communications media focusing on digital television, interactive systems, and broadband multimedia communications. The largest users of the ATSC standards are the North American cable and terrestrial markets.

The A/78 is a Recommended Practice that provides a common methodology for describing TS conformance criteria. Specifically, the SI elements and parameters within an ATSC TS are verified for correct transmission. The RP uses the premise that the severity of an error depends upon the magnitude of deviation. It specifies 5 groups of priority tests associated with ranges of deviation from the standard for certain parameters. The 5 priorities, or severity levels, are:

1. Transport Stream Off Air (TOA) The station is technically off-air, as the TS errors are so severe. Receivers will not be able to tune and decode anything within this broadcast.
2. Program Off Air (POA) A major service is in error to the point where decoders will not be able to decode the program. This is likely to be caused by incorrect signaling.
3. Component Missing (CM) Typically, one of the audio or video elementary streams cannot be found.
4. Quality Of Service (QOS) Parameters are out of specification by such a margin that a significant fraction of the receivers can be expected to produce flawed outputs.
5. Technically Non-Conformant (TNC) Violates the letter of the standard, but in practice will have little effect on the viewing experience.

As might be expected, there is much commonality between the ATSC and DVB parameters monitored. The difference is that ATSC specifies the error threshold for escalating categorization. This can generally be described as:

- The official metric to twice the metric.
- Twice the official metric to 5 times the metric.
- Over five times the metric.

An example is the PAT repetition error, as shown in Table 12-2.

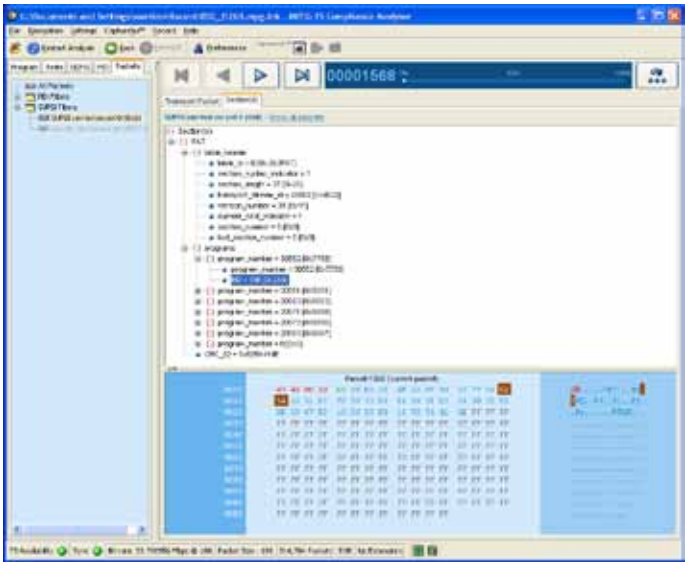


Figure 12-8. Packet view supporting TS packets, SI packets, and hex view.

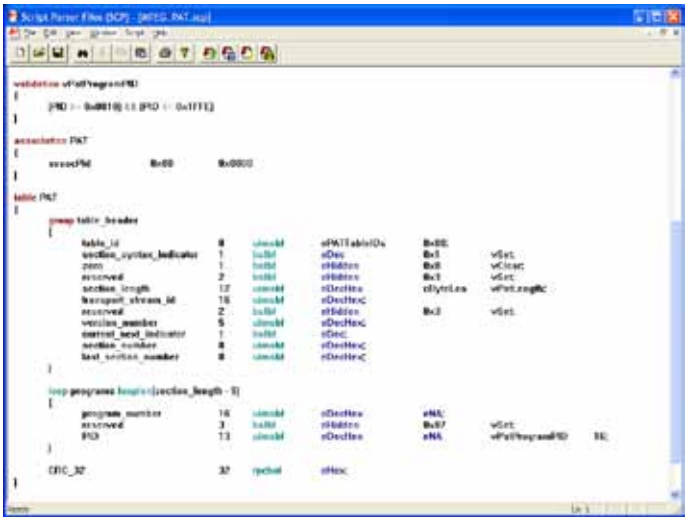


Figure 12-9. Script files define structures of all present and new sections and tables.

12.5 Packet View

As an alternative to checking for specific data in unspecified places, it is possible to analyze unspecified data in specific places, including individual Transport Stream packets, tables, or PES packets. This analysis on the MTS400 Series is known as the Packet view, because the analyzer automatically parses and decodes the data and then displays its meaning. Figure 12-8 shows an example of an MPEG transport packet in hex view as well as Packet view. A selected field in the Packet view automatically selects the corresponding bytes in the hex view for easier correlation.

Figure 12-8 shows an example of a PAT in the Packet view. The field structure in the Packet view is defined by script files. Script files are simply text files, such as that shown in figure 12-9 that make use of pre-defined keywords. This allows for quick updates if the standards change and also the interpretation of proprietary SI, as no changes to the software are required.

Figure 12-10. TR 101 290 flags CRC error based upon PSI and SI sections and script files.

12.6 Syntax and CRC Analysis

To ship program material, the transport stream relies completely on the accurate use of syntax by encoders. Without correct settings of fixed flag bits, sync patterns, packet-start codes, and packet counts, a decoder may misinterpret the bit stream. The syntax check function considers all bits that are not program material and displays any discrepancies. Spurious discrepancies could be due to transmission errors; consistent discrepancies point to a faulty encoder or multiplexer. Figure 12-10 shows a syntax-error as well as a missing cyclic redundancy check (CRC).

Many MPEG tables have checksums or CRCs attached for error detection. The analyzer can recalculate the checksums and compare them with the actual checksum. Again, spurious CRC mismatches could be due to stream-bit errors, but consistent CRC errors point to a hardware fault.

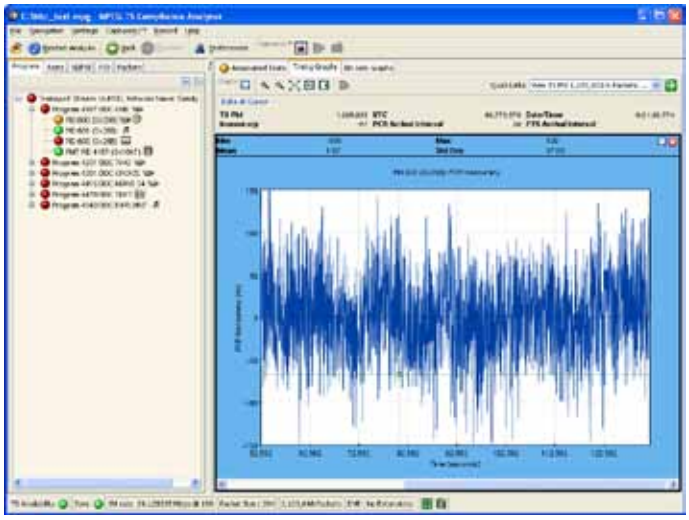


Figure 12-11. PCR plot over time.

12.7 Timing Analysis

The tests described above check for the presence of the correct elements and syntax in the transport stream. However, to display real-time audio and video correctly, the transport stream must also deliver accurate timing to the decoders. This task can be confirmed by analyzing the PCR and time-stamp data.

The correct transfer of program-clock data is vital because this data controls the entire timing of the decoding process. PCR analysis can show that, in each program, PCR data is sent at a sufficient rate and with sufficient accuracy to be compliant.

The PCR data from a multiplexer may be precise, but remultiplexing may put the packets of a given program at a different place on the time axis, requiring that the PCR data be edited by the remultiplexer. Consequently, it is important to test for PCR inaccuracies after the data is remultiplexed.

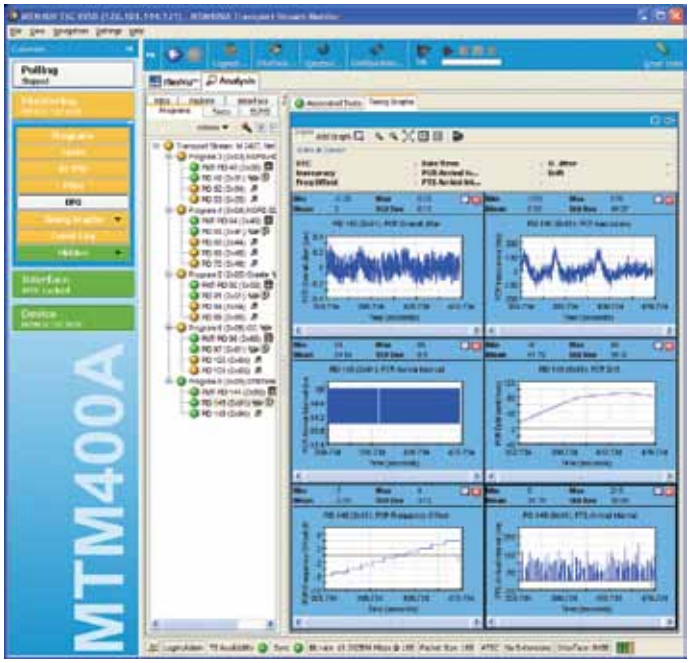


Figure 12-12. Real-time program clock plots.

Figure 12-11 shows a PCR display that indicates the positions at which PCRs were received with respect to an average clock. The Data at Cursor pane displays the data associated with any selected PCR. PCR inaccuracy measurement is based upon the PLL clock recovery in a receiver. The analyzer calculates the constant bitrate from the first 10 PCR values, which is then used as a basis for interpolating further values. These may then be compared with actual values. A high pass filter is used to ensure that the plotted inaccuracies are consistent with PLL jitter.

An alternate approach shown in Figure 12-12 provides a graphical display of PCR interval, inaccuracy, jitter, frequency offset and drift, which is updated in real-time.

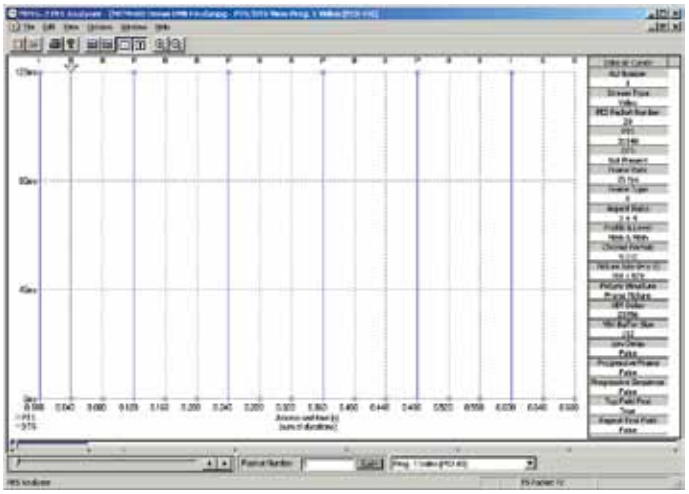


Figure 12-13. Packetized Elementary Stream Access Unit intervals and headers.

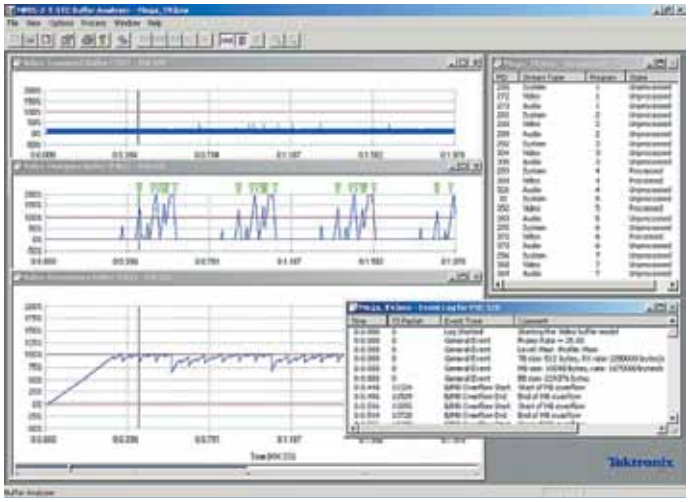


Figure 12-14. Transport, multiplex, and element buffer levels.

Figure 12-13 shows a time-stamp display for a selected elementary stream. The access unit, the presentation time, and, where appropriate, the decode times are all shown.

In MPEG, the reordering and use of different picture types causes delay and requires buffering at both encoder and decoder. A given elementary stream must be encoded within the constraints of the availability of buffering at the decoder. MPEG defines a model decoder called the T-STD (transport stream system target decoder); an encoder or multiplexer must not distort the data flow beyond the buffering ability of the T-STD. The transport stream contains parameters called VBV (video buffer verify) specifying the amount of buffering needed by a given elementary stream.

The T-STD analysis displays the buffer occupancy graphically so that overflows or underflows can be easily seen. Figure 12-14 shows a buffering display. The output of a normal compressor/multiplexer is of limited use because it is not deterministic. If a decoder defect is seen, there is no guarantee that the same defect will be seen on a repeat of the test because the same video signal will not result in the same transport stream. In this case, an absolutely repeatable transport stream is essential so that the defect can be made to occur at will for study or rectification.

Transport stream jitter should be within certain limits, but a well-designed decoder should be able to recover programs beyond this limit in order to guarantee reliable operation. There is no way to test for this capability using existing transport streams because, if they are compliant, the decoder is not being tested. If there is a failure, it will not be reproducible and it may not be clear whether the failure was due to jitter or some other noncompliance. The solution is to generate a transport stream that is compliant in every respect and then add a controlled amount of inaccuracy to it so that the inaccuracy is then known to be the only source of noncompliance. The editor feature of the MTS400 is designed to create such signals.

12.8 Elementary Stream Testing

Due to the flexible nature of an MPEG bitstream, the number of possibilities and combinations it can contain is almost incalculable. As the encoder is not defined, encoder manufacturers are not compelled to use every possibility; indeed, for economic reasons, this is unlikely. This fact makes testing quite difficult because the fact that a decoder works with a particular encoder does not prove compliance. That particular encoder may simply not be using the modes that cause the decoder to fail.

A further complication occurs because encoders typically will not produce the same bitstream if the video or audio input is repeated. There is little chance that the same alignment will exist between the resultant I-, P- and B-pictures. If a decoder fails a given test, it may not fail the next time the test is run, making fault-finding difficult. A failure with a given encoder does not determine whether the fault lies with the encoder or the decoder. The coding difficulty depends heavily on the nature of the program material, and any given program material will not necessarily exercise every parameter over the whole coding range.

To make tests that have meaningful results, two tools are required:

- A known source of test signals that deliberately explore the whole coding range. This range includes bitstreams compliant to the video standard concerned, but also streams containing deliberate bitstream errors. These signals must be deter ministic so that a decoder failure will give repeatable symptoms. The Tektronix Vclips product is a collection of bitstreams designed to perform this task.
- An elementary stream analyzer that allows the entire syntax from an encoder to be checked for compliance, such as the Tektronix MTS4EA.

12.9 Elementary Stream Analysis

An elementary stream is a payload that the transport stream must deliver transparently. The transport stream will do so whether or not the elementary stream is compliant. In other words, testing a transport stream for compliance simply means checking that it is delivering elementary streams unchanged. It does not mean that the elementary streams were properly assembled in the first place.The elementary stream structure or syntax is the responsibility of the encoder. Thus, an elementary stream test is essentially testing the encoder. It should be noted that an encoder can produce compliant syntax, and yet still have poor audio or video quality. However, if the syntax is incorrect, a decoder may not be able to interpret the elementary stream. An elementary stream may be intermittently noncompliant if some less common mode of operation is not properly implemented.

As transport streams often contain several programs that come from different encoders, elementary stream problems tend to be restricted to one program, whereas transport stream problems tend to affect all programs. If problems are noted with the output of a particular decoder, then the Vclips test streams should be run on that decoder. If these are satisfactory, the fault may lie in the input signal. If the Transport Stream syntax has been tested, or if other programs are working without fault, then the particular elementary stream in error should be analyzed.

The main component of elementary stream analysis is verifying compliance against the relevant standards. The MTS4EA produces an output log of all violations and non-conformances. The user may select any log entry and synchronise the video display and other views to that point in the bitstream for fast pinpointing of problems. A useful feature is overlays; there are many different types including average quantisation and bits per macroblock.

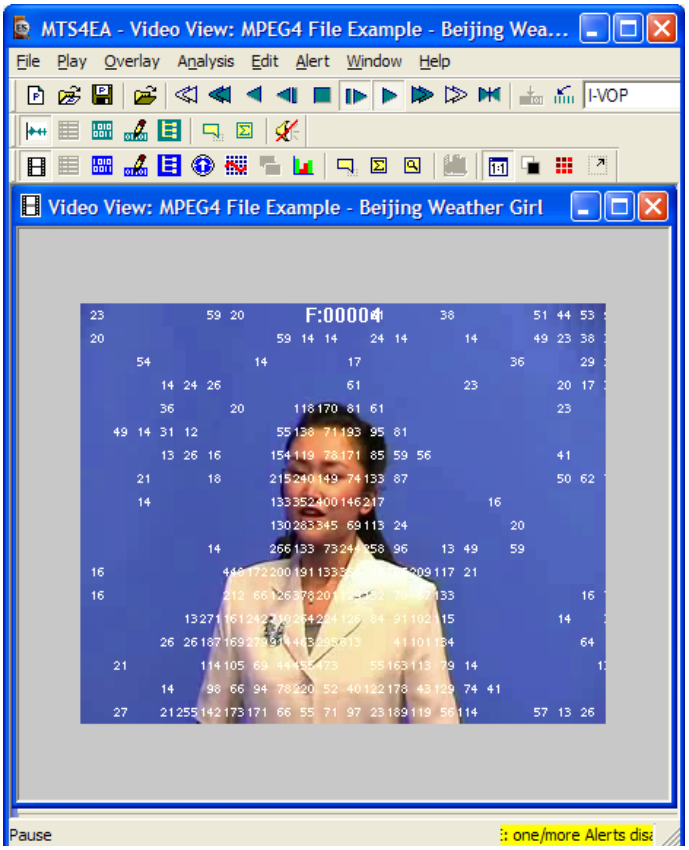


Figure 12-15. Advanced elementary stream analyzer showing bits per block.

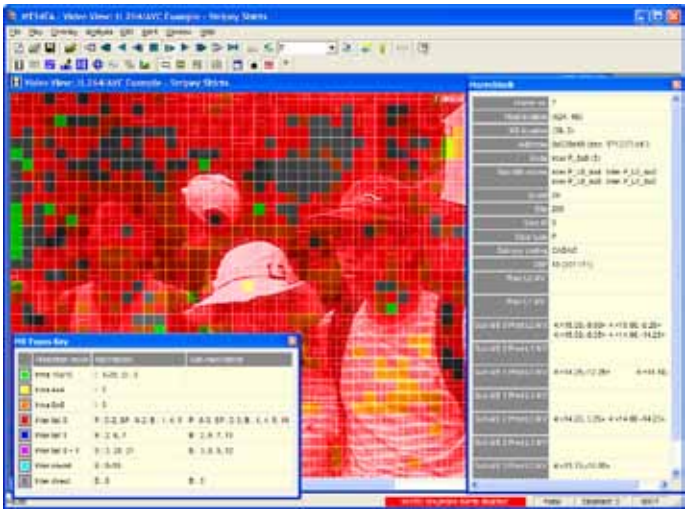


Figure 12-16. Color-coding different types of blocks.

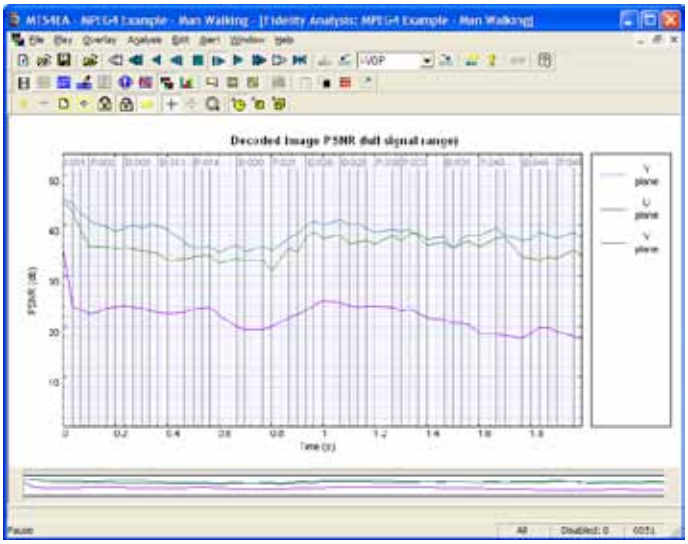


Figure 12-17. PSNR of reference file and test file.

Figure 12-15 shows a bits per macroblock display that illustrates an encoder optimisation problem; many bits are being used to encode the non-changing background information. At a lower level of testing, Figure 12-16 shows a decoded P frame overlaid with macroblock type information. The details of a selected macroblock are also shown.

Elementary stream analysis discussed so far has been single ended; that is, a measurement of the output signal with no regard to the quality of the original uncompressed video data. If the original video is available for reference then the MTS4EA can create a signal to noise ratio graph (Figure 12-17) and visual difference display.

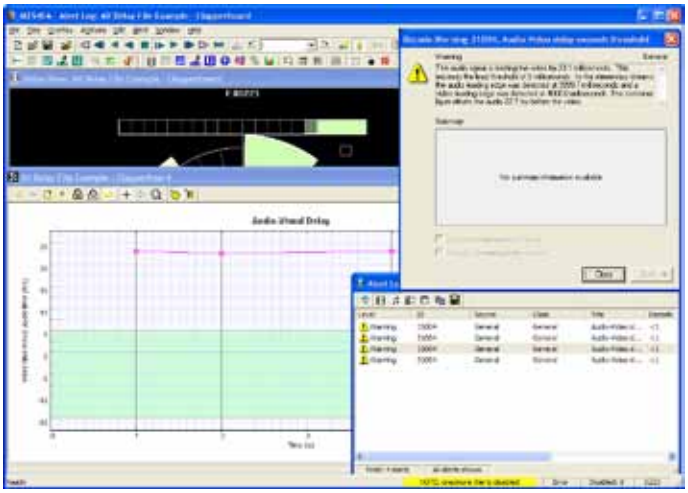


Figure 12-18. Reference audio and video files used to accurately detect A/V delay when encoded into transport stream.

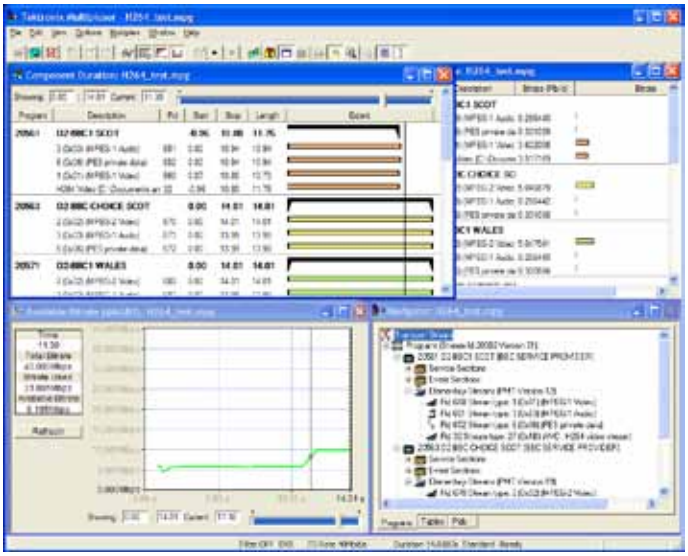


Figure 12-20. Multiplexer application showing hierarchy, timeline, and available bandwidth.

12.10 Audio Visual Synchronization

One of the major problems affecting broadcasting throughout its entire history is synchronising audio and video information, commonly referred to as lip sync. In a production facility, the audio and video information typically takes different routes before they are encoded and multiplexed together, thereby leaving plenty of scope for delays to be introduced. The MTS4EA includes an AV Delay feature that measures the delay introduced by the encoder under test by using an input test pattern, as shown in Figure 12-18. A graph is plotted of the AV Delay, as well as errors generated for each point that exceeds the threshold for the chosen standard.

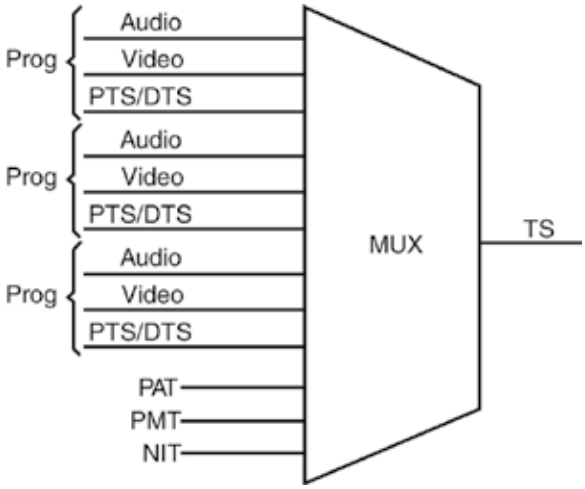


Figure 12-19. Generic multiprogram transport stream multiplexer.

12.11 Creating a Transport Stream

Whenever the decoder is suspect, it is useful to be able to generate a test signal of known quality. Figure 12-19 shows that an MPEG transport stream must include Program Specific Information (PSI), such as PAT, PMT, and NIT describing one or more program streams. Each program stream must contain its own PCR and elementary streams having periodic time stamps.

A DVB transport stream will contain additional service information, such as BAT, SDT and EIT tables. A PSI/SI editor enables insertion of any desired compliant combination of PSI/SI into a custom test bitstream.

Clearly, each item requires a share of the available transport stream rate. The MTS400 Multiplexer provides a graph showing the total bit rate used at each point in the bitstream, as illustrated in Figure 12-20. The overall transport stream bit rate must remain constant, so the remainder of the bit rate is used up by inserting stuffing packets with PIDs that contain all 1s, which a decoder will reject.

12.12 PCR Inaccuracy Generation

The MPEG decoder has to recreate a continuous clock by using the clock samples in PCR data to drive a phase-locked loop. The loop needs filtering and damping so that jitter in the time of arrival of PCR data does not cause instability in the clock.

To test the phase-locked loop performance, a signal with known inaccuracy is required; otherwise, the test is meaningless. The MTX100 Generator can generate simulated inaccuracies for this purpose. Because it is a reference generator, the MTX100 Generator has highly stable clock circuits and the actual output jitter is very small. To create the effect of jitter, the timing of the PCR data is not changed at all. Instead, the PCR values are modified so that the PCR count they contain is slightly different from the ideal. The modified value results in phase errors at the decoder that are indistinguishable from real jitter.

The advantage of this approach is that jitter of any required magnitude can easily be added to any program simply by modifying the PCR data and leaving all other data intact. Other programs in the transport stream need not have jitter added. In fact, it may be best to have a stable program stream to use as a reference.

For different test purposes, the time base may be modulated in a number of ways that determine the spectrum of the loop phase error in order to test the loop filtering. Square-wave jitter alternates between values which are equally early or late. Sinusoidal jitter values cause the phase error to be a sampled sine wave. Random jitter causes the phase error to be similar to noise.

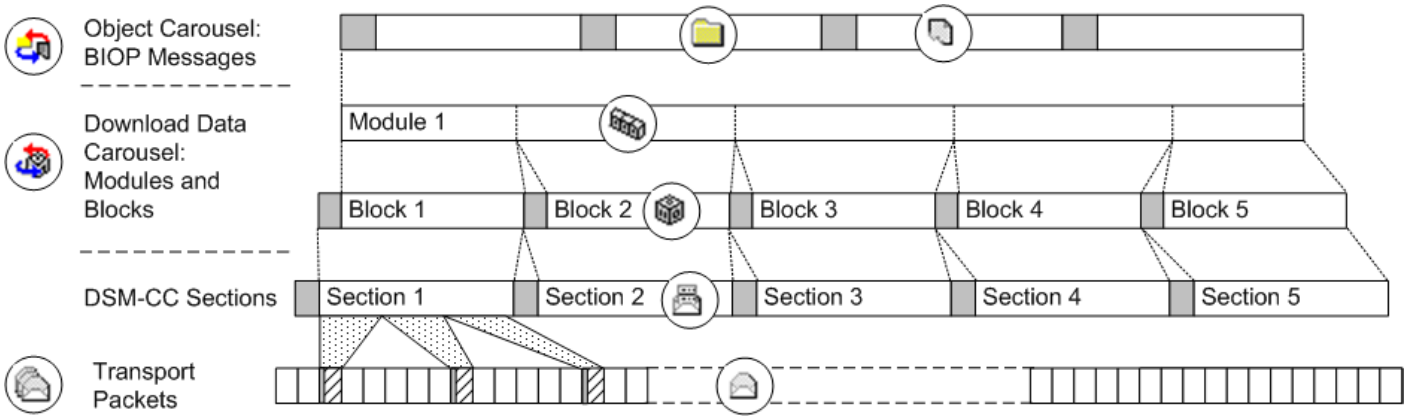


Figure 12-21. MPEG Carousel sections multiplexed into a transport stream.

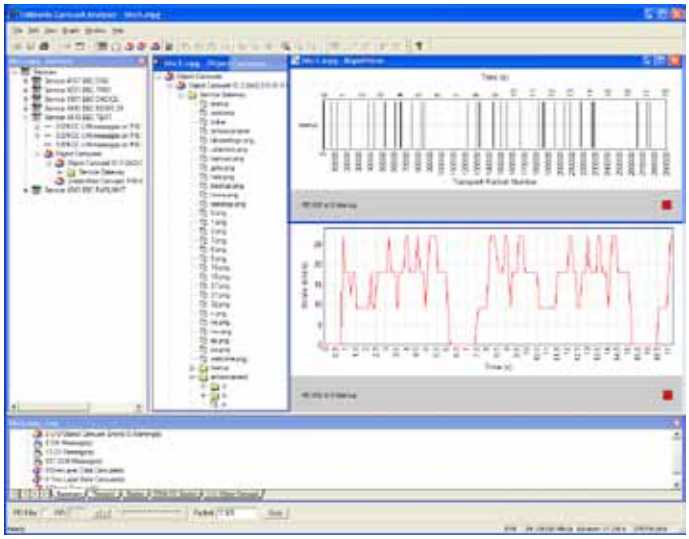


Figure 12-22. Section timing of carousels.

12.13 Data Broadcast Carousel Analysis

The description of carousels in Section 11 should lead most readers to conclude that they are a complex animal. Part of the complexity is driven by the trade-off between transmission bandwidth and responsiveness of the user experience. This results in a necessity to restrict bandwidth by repeating group data at different rates, depending upon its likelihood to change. That is, football score updates will change much more frequently than the weather forecast. Incorporating this flexibility means that carousels typically consist of a large number of components that need to be linked together and signalled correctly to ensure the integrity of the overall interactive application. The main problems faced with transmitting data broadcast applications may be summarized as:

- Timing.
- Bit rates.
- Signaling and component linkages (references).

The Tektronix Carousel Analyzer is designed to help diagnosis of all of these problems, both at the data and object carousel layers. Indeed, it allows analysis of bit rate and repetition interval graphs for each node indicated in the protocol layers shown in Figure 12-21.

Figure 12-22 shows a typical Carousel Analyzer view with Services pane on the left and Message Log pane at the bottom. The contents of an Object carousel are displayed, together with the repetition interval and bit rate graphs for the selected 'Start' object. If multiple nodes are selected for either bit rate or repetition interval then each additional trend is added to the same graph for ease of comparison.

One of the main timing issues is the boot time. That is, the amount of time it takes from the start of transmitting a carousel to the point where the STB can execute it. This is especially important for interactive adverts, where the advert to which the carousel is associated may only last for, say 20 seconds.

The Carousel Analyzer incorporates many timing methods, one of which is Cold Boot Time. This measures the time taken to signal the carousel and download the service gateway, the boot class and all of the assets. This is repeated at regular intervals within the bitstream.

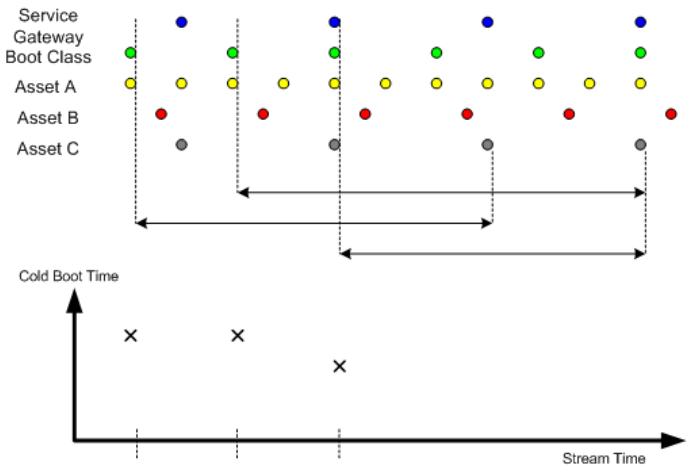


Figure 12-23. Cold-boot times of the carousel across the file.

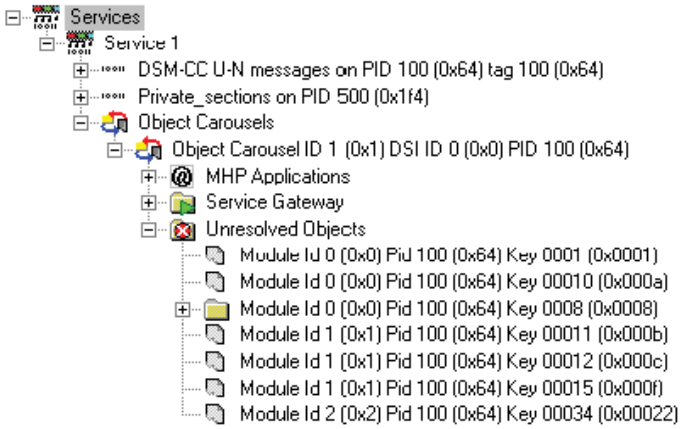


Figure 12-24. Hierarchical view of entire service.

The boot class must be received before any of the assets. This simulates the STB not knowing which carousel objects the boot class depends upon until it has received and processed. Directories must be received before any child objects, as this timing method assumes no signaling caching. The end of each cycle will be the end of the module containing the last asset - this is an important distinction, as an asset may not be the only one in that module.

The most common signaling problems are unresolved objects and absent objects. Figure 12-24 shows that when an object is found during analysis, but its key is not referred to in the Service Gateway, it is displayed under an Unresolved Objects node. However, when a directory object is found, the keys of its child nodes can be resolved and its tree structure built correctly, even though the directory object itself may still be unresolved.

Figure 12-25 shows that objects referenced, but not found are identified. The path in which the absent object is expected is also marked.

Occasionally, the signaling is sufficiently broken that a carousel present in the Transport Stream is not found during analysis. In this instance the carousel can be identified with Manual Signaling from the related PID node. The Transport Stream will then be re-analyzed with any DSM-CC sections on the selected PID being processed.

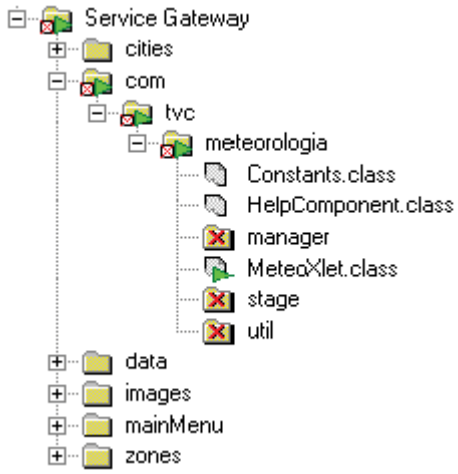


Figure 12-25. Elements of service gateway.

As might be expected, Tektronix also offers a Carousel Generator product for creating carousel contents within an output Transport Stream. This is particularly useful in test situations where the effects of varying parameters, such as individual repetition intervals, may be quickly ascertained. The Generator will create object carousels conforming to the MPEG-2, DVB, DTT (MHEG-5) or MHP standards.

Figure 13-1. Low blockiness from satellite distribution feed.

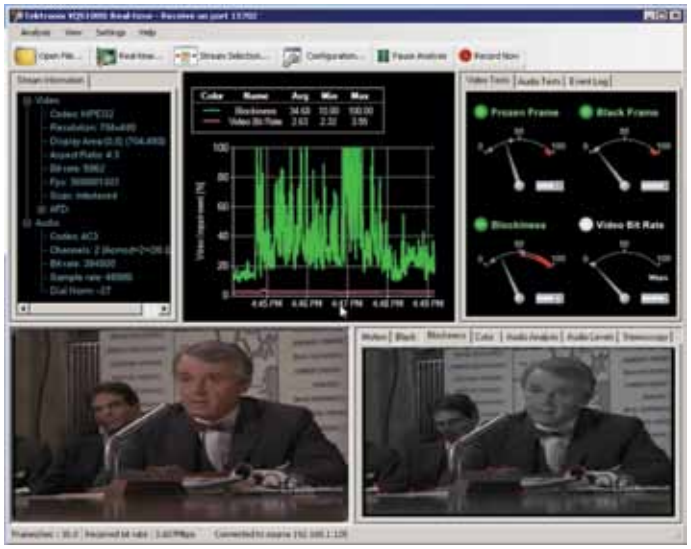


Figure 13-2. Transcoded content with significantly higher blockiness (green plot) over the same timeframe.

Section 13 – MPEG QoE Testing

QoE testing examines the baseband impairments of the Elementary Stream and which are noticeable to viewers and subscribers.

13.1 Testing Requirements - QoE

Quality of Experience (QoE) QoE is a metric that relies much more upon our ability to notice a problem in the midst of a TV program rather than a single bit error. For instance, if single bit error affected the far corner of a B-frame, then that bit error would have a much lower impact to QoE than if the error had landed in the middle of an I-frame. Subsequently, if the single error in the I-frame ended up corrupting an entire slice of 16x16 pixel blocks, then the impact to QoE would be huge. Therefore, to be able to measure the impact of an error upon a TV program, it is critical to know exactly where the error occurred. Another issue is that the transport may be perfectly error free (no syntax or semantic errors), but the video is objectionable because there is not enough bandwidth to clearly portray the high-motion video scene. In this case, there are no protocol errors, but the picture frame or video sequence is made up of large 16x16 solid squares rather than a clean picture. Viewers often refer to this as a blocky video problem. Figure 13-1 shows a low blockiness measurement on a TV program from satellite distribution at 4.1 Mbps. Figure 13-2 shows the same TV program over cable (256QAM), but only after it has been transcoded (or rate-shaped, clamped, etc.) to a lower rate of 3.6 Mbps. The resulting changes cause significantly more blockiness artifacts over the same period of time. In this case, the QoE has deteriorated enough that viewers would recognize the blockiness issues in the video.

Another video QoE issue occurs when the same video frame occurs repeatedly for a long period of time (frozen or black frames). Obviously, video frames repeat when there is no activity, but at some point (e.g., 2 minutes), an alarm is needed to alert the operator if the frozen frames are on purpose, or from a broken link or piece of equipment.

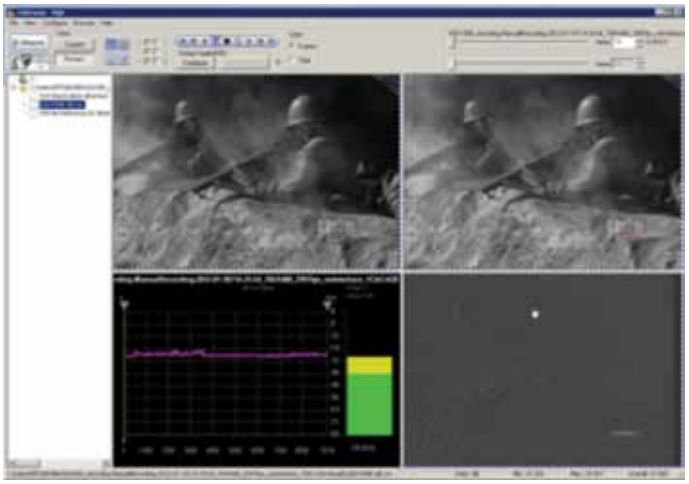


Figure 13-3. The Picture Quality Analyzer performs direct comparisons of video frames before and after a live transcode.

Figure 13-4. Satellite ingest with audio loudness at 2 dB quieter than DialNorm.

When a reference signal is available for direct comparison as in the case shown above from satellite and cable, an objective picture quality rating can be made using the Tektronix PQA600 Picture Quality Analyzer, or from the Picture Quality Analysis software option in the MTS4000. In this case, a stream is recorded from both satellite and cable, and then the Picture Quality Analyzer will convert to YUV, perform spatial and temporal alignments, and then measure the difference between the two video clips. The basic Peak Signal to Noise Ratio (PSNR) measurement is shown, but the Picture Quality Analyzer also performs many more important tests such as Picture Quality Rating, Attention model, etc. Figure 13-3 shows the Picture Quality Analyzer making a frame by frame PSNR comparison between every frame of the satellite and cable video clips.

The Picture Quality Analyzer has been pitted directly against human view trials in a variety of studies. The IRT study[11] (Institut fuer Rundfunktechnik in Munich, Germany) showed extremely high correlation between the PQR results and human viewer results using ITU-R_BT.500-11[12] testing.

Audio is another element to QoE as it occasionally increases to an unbearable level, or maybe drops out altogether over a long period of time. To accurately track these problems, loudness measurements have been defined for instantaneous peaks (ITU-R BS.1770/1771[13]), as well as short term and long term filters for ATSC[14], ITU, and EBU[15]. These measurement algorithms allow test equipment to track audio over long periods of time and trigger on deviations outside allowable levels. One method of tracking this is to use a DialNorm level as a target for normal dialog levels. Averaged or filtered levels are allowed to rise or fall within a few dB of this set level. Many governments have adopted this requirement and now impose

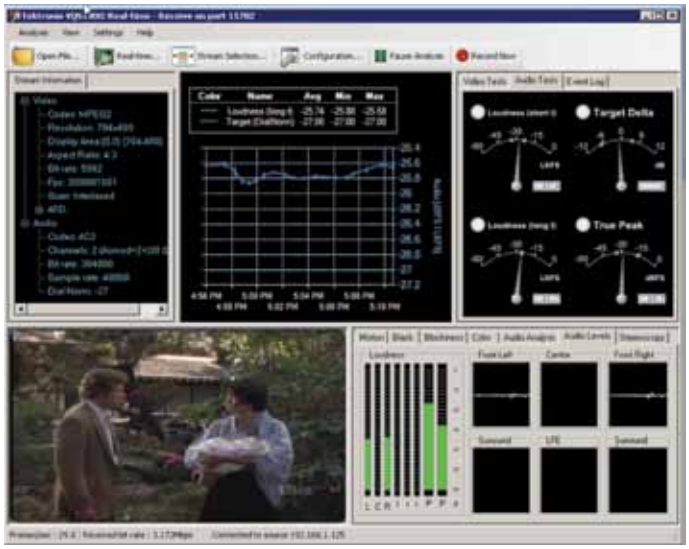


Figure 13-5. Cable broadcast with audio at 2 dB louder than DialNorm. This is a 4 dB increase over the satellite feed.

fines on broadcast operators who deviate beyond the agreed limits. With this agreement in place, TV viewers should now be able to switch from channel to channel without having to adjust the audio levels. The same goes for the commercials or ad-inserts between programming. The loudness of commercials must also stay within the agreed upon levels. As an example of audio levels being altered, Figure 13-4 shows the satellite distribution signal at one audio level (-29 LKFS[13]) while its cable-broadcast counterpart (Figure 13-5) is at a different audio level (-25 LKFS). Both signals use a DialNorm reference level of -27 LKFS.

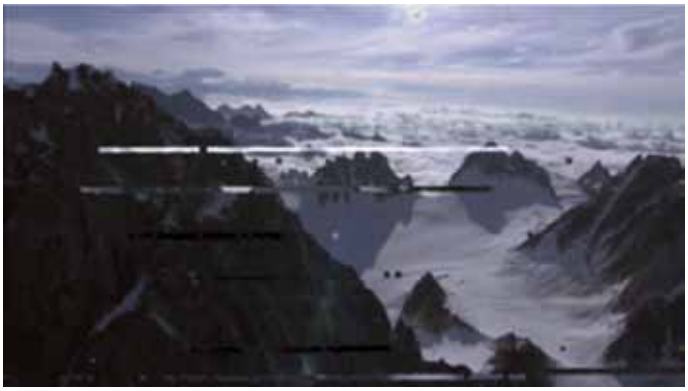


Figure 13-6. Significant loss of RF or IP packets cause blockiness and slice errors.

13.2 How are QoS and QoE interrelated?

Both QoS and QoE are interrelated, but their relationships are not equal. Case in point:

1) A good QoE usually means a good QoS
2) But a good QoS does not always mean you will have a good QoE. For example, Encoder bandwidth starvation or over-compression.
3) A bad QoS, or at least frequently occurring QoS issues usually lend toward a poor QoE. With enough bad bits randomly distributed into all video and audio elements, poor quality of video and audio is bound to occur (see Figure 13-6).
4) A bad QoE is not always related to a bad QoS, but it is always worth validating the QoS performance. Poor QoE can come from misbehaving ad-inserted and any other equipment (not related to TR 101 290). Poor QoE can come from over compression (this is perfectly legal). Poor QoE can come from dropped IP packet, corrupt RF packets, or other transmission problems.

Therefore, it is common to measure both QoE and QoS in order to quickly identify the root cause of a problem and in turn correct the issue.

One problem with QoS issues is that they can be easily hidden or masked when the program is handed off from one network to another (e.g., dropped video packets). Once the program is decoded back to baseband for manipulation and re-encoding, all previous QoS errors are lost forever. After encoding a second time, the transport and video protocol is now error free, but visually, there may be large slice errors seen in the middle of the video frames. At this point, then only thing available to catch these artifacts is to decode the video and audio content and look for anomalies. A common problem in sporting events is when the remote transmission link is lost for a fraction of a second. The video breaks apart, but the receiver is able to catch this QoS issue due to Transport Error Indicator flags (TR 101 290 Priority 2.1 – Also bit #9 in each transport packet). Although, if the decoded and blocky or broken video is passed on to the network as a re-encoded program , then the program will often have a perfect QoS, as well as its compressed video syntax being perfect due to its decode and then second encode process. The video may still look like garbage (as seen in Figure 13-6), but according to QoS and video syntax, the results are error free. To find an artifact, we must actually look at the decoded picture content and determine if the frames contain objectionable artifacts (blockiness, slice errors, frozen frames, etc.).

If we have a high BER or poor QoS, then it is inevitable that the QoE will deteriorate also (as in Figure 13-6). In this case, we do not care which frames the errors are landing in because the rate is high enough to land in every frame. Therefore, the QoE rating becomes highly objectionable. In this case, the QoS measurement is the key to isolating the problem.

When the QoS is quite good (no transmission errors), then there is a very low correlation between QoS and QoE. In this case, it is more important to focus on the QoE.

There is an important case where the QoS can be perfect, but the QoE is highly objectionable. This is due to the common case with networks running constant bit rate (CBR) video. Normal video with low motion may look decent, or even very good. But the scene occasionally changes from slow moving objects (e.g., talking head) to something with a high rate of change (e.g., sporting event, quick successive scene cuts, etc.). If the encoder does not have enough bandwidth to accommodate the transition to the high-motion events, then the only recourse for the encoder is to start decimating the 8x8 blocks of video. At its worst, the 8x8 blocks may only contain a single value representing the video luminance and color. This scenario is important for QoS and QoE testing because the end-user notices a very objectionable TV program segment, although the difference between over compression and near total loss of signal are almost indistinguishable. Figures 13-1 and 13-2 show a program at both 4.1 and 3.6 Mbps with the lower rate having much more blockiness due to high motion in a bandwidth starved stream. Figure 13-6 shows blockiness too, but due to a completely different reason (missing packets). Therefore, QoE testing is extremely important, but testing QoS helps to determine the difference between malfunctioning equipment causing a problem verses a poor transmission link.

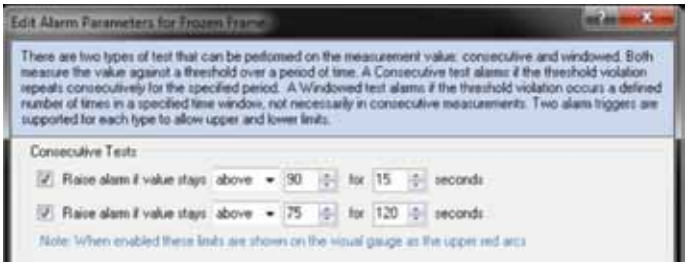


Figure 13-7. Dual trigger levels for Frozen and Black video.

13.3 Using QoS and QoE to identify "real" problems

Most facilities use an ingest, manipulate, and egress architecture. Cable headends use a wide variety of video equipment to capture, manipulate, and ultimately, broadcast. Most all test points would be in the form of QPSK/8PSK, 8VSB, QAM, IP (GigE or 10G), or ASI.

Monitoring QoE aspects on live programming falls into several different audio and video categories.

Video QoE

Video QoE includes monitoring video for frozen frames, black, and blockiness. In the case of frozen frames, a long series of nearly identical frames should trigger a problem since live TV programming is usually made up from moving video. We know that there are many occasions where it is acceptable for a frame to be repeated many times, but at some point, the operator should be alerted to excessive repetitive frames. It could be the sign of a broken encoder or multiplexer, or any of a number of pieces of hardware. Figure 13-7 shows the VQS1000 QoE software setup for Frozen Frames that will be triggered if 90% of the frame is frozen for more than 15 seconds, or when 50% of the frame is frozen for more than 120 seconds. The requirements are similar for excessively long periods of black. Blockiness is a little different in that it can occur in a small portion of the picture, or over the entire picture. Blockiness will often occur during short scenes where the video content is moving too quickly for the encoder to faithfully compress all of the details. Therefore, the encoder tends to throw away high frequency details rather than failing completely. In this case, the monitor must be set to trigger once the blockiness level has crossed a threshold, and then

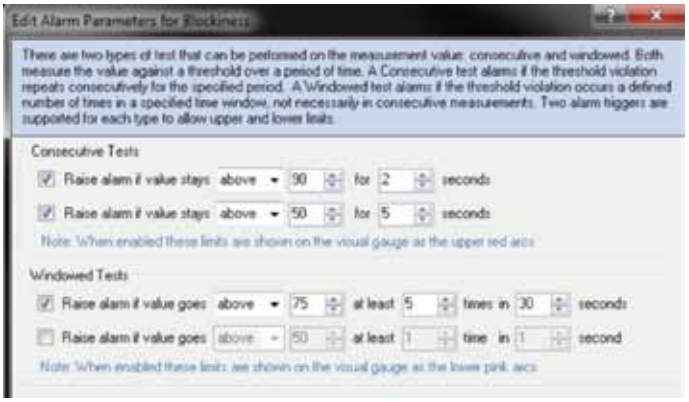


Figure 13-8. Dual trigger levels and gating for Blockiness.

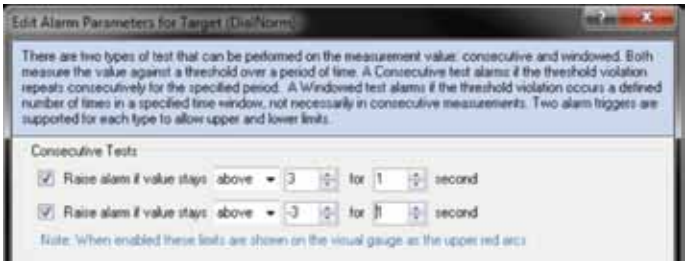


Figure 13-9. VQS Target DialNorm settings.

maintain that level for a significant amount of time. Otherwise, the monitor would trigger an alert every time the scene was overly blocky, even for one single frame. Figure 13-8 shows the VQS1000 QoE software setup for Blockiness triggering when

- level of 90% is reached for over 2 seconds, or
- level of 50% is reached for over 5 seconds, or
- level of 75% is reached at least 5 times within 30 seconds.

Audio QoE

Audio QoE includes monitoring audio for loudness levels. With such a high emphasis on audio levels today, it has become critical to measure overall loudness levels according to the new guidelines. Just as with video QoE monitoring, triggers for levels, deviations, and durations are important to minimize the many false positives than can occur. Figure 13-9 shows audio loudness limits must be within 3 dB of DialNorm.

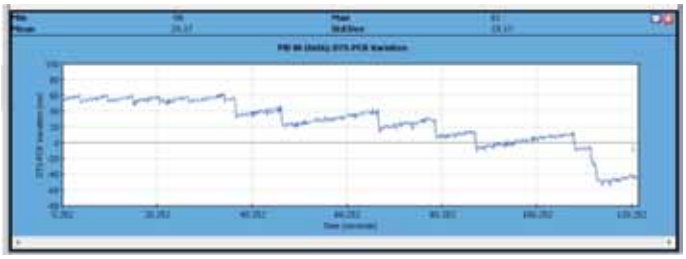


Figure 13-10. Terrestrial encoder generating negative DTS-PCR delta values (should be impossible).

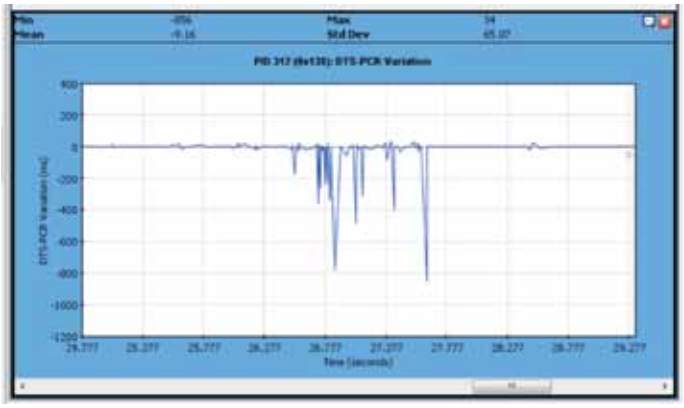


Figure 13-11. Cable headend encoder generating negative DTS-PCR delta values (should be impossible).

Troubleshooting Over-compression

The VQS1000 QoE software allows for real-time monitoring of multiple RF and IP signals in both the transport layer as well as the video and audio layer. When blockiness is found to be excessive on in a TV program, the MTS4000 MPEG Analyzer can measure the broadcast program as well as the ingest distribution feed.

When the QoE of a TV program is called into question, the MTS4000 MPEG Analyzer with its VQS1000 QoE software can quantify the magnitude of the audio and video issues. Once the levels are proved to be unacceptable (as seen in Figure 13-2), the MTS4000 MPEG Analyzer can quantify the ingest program (as seen in Figure 13-1), and then a comparison can be made between the two programs. In the example here, the cable broadcast program has a stable QoE or blockiness as long as the content does not contain too much action. Once the scene changes quickly, the blockiness rating spikes dramatically. In comparison, the ingest content is much less blocky and only varies slightly when the scenes change quickly. It can noted that the ingest content is coming into the facility at about 4.1 Mbps whereas the broadcast program is leaving at only 3.6 Mbps. Therefore, the blockiness issue can be identified as a result of decoding and re-encoding without providing sufficient bandwidth to maintain the quality of the original picture.

Troubleshooting Interoperability

The MTS4000 can also be used to identify non-compliance issues when televisions and set top boxes begin to react abnormally to specific programming. In a local broadcast case, several viewers complained about inaudible audio on the SAP channel (audio sounding like motorboats). Other viewers had no problem at all on the SAP channel. It sounded like an interoperability issue, so the MTS4000 MPEG Analyzer was tuned to the local off-air channel to look at its TR 101 290 results. The TR 101 290 tests, and virtually every other test came up clean except for one. The MTS4000 MPEG Analyzer can graph the audio or video Decode Time Stamp (DTS) value as it arrives, against its current PCR value. This delay is always a positive number (representing buffer delay time) and by definition it is required to be between zero and one second. The newly added SAP audio had its DTS-PCR delta ranging above and below zero, which is by definition, never allowed to happen. Figure 13-10 shows the local broadcaster SAP audio delay issue.

This same terrestrial broadcast feed was also sent to both local cable companies for redistribution.

Another similar audio example was found at the cable headend where the audio buffer delay was a negative value, which is impossible for a set top box to support. The MTS4000 MPEG Analyzer once again showed the values going below zero. Figure 13-11 shows the same audio problem, but at a different facility.

In both cases, reconfiguring the encoder fixed the problem.

The MTS4000 MPEG Analyzer with its unique DTS-PCR measurement was able to quickly pinpoint the problem to the recent configuration change to the encoder, and thus take the blame off the small set of failing decoders throughout the city.

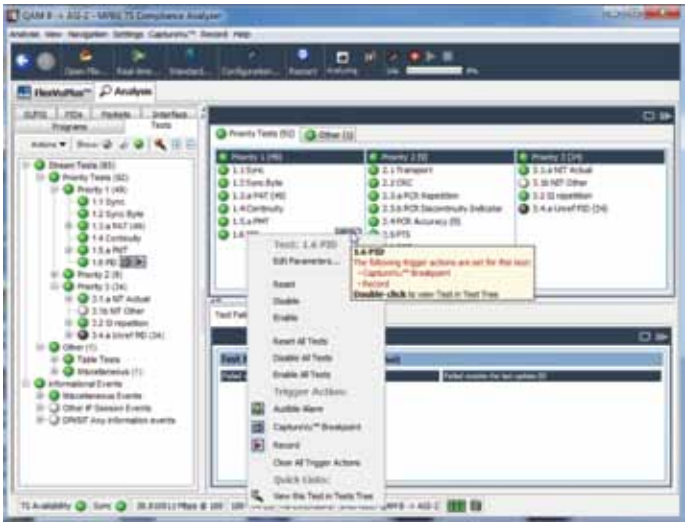


Figure 13-12. MTS4000 Transport Stream Compliance Analyzer: Trigger and capture enabled for missing encryption key packet.

Troubleshooting Transport Streams with CaptureVu

A key part of delivering a quality experience is finding the root cause of problems in the transport stream. It's no longer good enough just to spot a problem, reboot the set-top box and hope that it goes away. The MTS4000 can uniquely go as deep as is needed into the transport and elementary streams to track down sources of picture anomalies, like discovering that a closed caption stream contained too much information, causing set-top box buffers to overflow and precipitate automatic reboots.

While the MTS4000 provides real-time analysis, the instrument's ability to capture events for deep analysis is critical to identifying the root cause of problems. Especially useful are the pre-triggers in CaptureVu that not only capture an error, but also provide information leading up to the error.

This last troubleshooting example occurred with a DVR from a major consumer electronics manufacturer that was misbehaving whenever it was tuned to a Switched Digital Video (SDV) channel. The MTS4000 discovered the device was dropping the encryption key whenever a service was added to or removed from the multiplex. Figure 13-12 shows the MTS4000 TSCA setup screen for enabling a trigger on a specific test or missing PID. By documenting the scenario with screen shots and test reports, the DVR manufacturer was contacted, and they provided a software update that solved the problem and eliminated many expensive truck rolls for the Cable Operator.

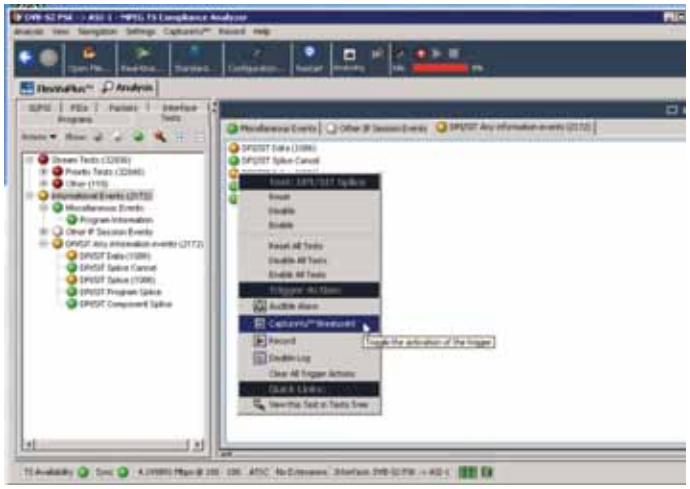


Figure 13-13. MTS4000 Transport Stream Compliance Analyzer enabling the CaptureVu and Recording for SCTE35 DPI packets.

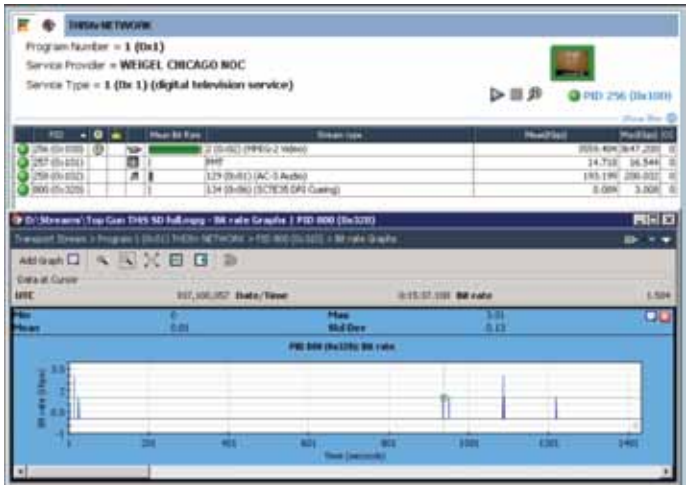


Figure 13-14. PID 800 DPI packets arrive about every 15 minutes.

Troubleshooting Ad Insertion

All broadcast companies fund their businesses based upon selling advertising time between programs. The ads need to be inserted at very specific time, and often timed by SCTE35[16] Digital Program Insertion (DPI) cueing. If these cue tones do not make it through, then the national ads go through and local revenue is lost. The MTS4000 MPEG Analyzer can trigger on the STCE35 packets as well as record programming to disk before and after the event (pre-trigger). Figure 13-13 shows the MTS4000 Transport Stream Compliance Analyzer enabling the CaptureVu and Recording feature based upon SCTE35 packets. The trigger/record function allows the recorded file to be as small as possible.

In the case of much larger recordings, the MTS4000 can record for many hours. In the example shown in Figure 13-14 running at 4 Mbps, the MTS4000 could hold over 11 days of continuous programming. The figure also shows the PID 800 DPI packets arriving around 15 minutes apart, as you would expect for local ad-inserts.

With these SCTE35 DPI tools, troubleshooting and debugging missing DPI packets become relatively simple.

13.4 Summary

Troubleshooting video signals in a broadcast or cable facility requires the use of an MPEG Analyzer that provides multiple input signals all running in parallel. The minimum set of test points should be at ingress, ASI/IP switch, and egress. It is critical to measure QoE at both egress and ingest, but in the case of audio and video problems, QoS testing at multiple points in the facility may be needed to pinpoint the source of the impairments. Critical capabilities in an MPEG Analyzer should include:

- Transport stream generation, modification and analysis
  - Real and deferred time
  - Including MPEG-2, H.264 & MPEG-4 AAC
  - Multiplexer
  - Automatic error triggered recording and capture
- A range of physical interfaces
  - Multi-port ASI
  - RF (DVB-S2, 8-VSB & QAM-B)
  - Video over IP analysis and generation (1Gbps and 10Gbps)
- Comprehensive suite of software tools for analyzing all layers of video
  - High accuracy RF layer analysis
  - Transport, Program & Elementary Streams
  - Video & Audio Quality of Experience (QoE)
  - Picture Quality

A facility equipped with an analyzer having the feature set listed above should be able to resolve problems in minutes rather than hours or days using alternative methods.

Glossary

AAC – Advanced Audio Coding.

AAU – Audio Access Unit. *See Access unit.*

AC-3 – The audio compression scheme invented by Dolby Laboratories and specified for the ATSC Digital Television Standard. In the world of consumer equipment it is called Dolby Digital.

ACAP – Advanced Common Application Platform.

Access Unit – The coded data for a picture or block of sound and any stuffing (null values) that follows it.

A/D – Analog to digital converter.

AES – Audio Engineering Society.

Anchor Frame – A video frame that is used for prediction. I frames and P frames are generally used as anchor frames, but B frames are never anchor frames.

ANSI – American National Standards Institute.

API – Application Program Interface.

ARIB – Association of Radio Industries and Businesses.

ASF – Advanced Systems Format (formerly "Advanced Streaming Format"), a proprietary Microsoft streaming format.

Asynchronous Transfer Mode (ATM) – A digital signal protocol for efficient transport of both constant-rate and bursty information in broadband digital networks. The ATM digital stream consists of fixed-length packets called “cells,” each containing 53 8-bit bytes – a 5-byte header and a 48-byte information payload.

ATM – *See asynchronous transfer mode.*

ATSC – Advanced Television Systems Committee.

ATVEF – Advanced Television Enhancement Forum. AU – Access Unit.

BAT – Bouquet Association Table.

BCH – Bose, Ray-Chaudhuri, Hocquenghem Code. A BCH code is a multilevel, cyclic, error-correcting, variable-length digital code used to correct multiple random error patterns.

BER – Bit Error Rate.

BFSK – Binary Frequency Shift Keying.

BIOP – Broadcast Inter-ORB Protocol. Bit rate – The rate at which the compressed bit stream is delivered from the channel to the input of a decoder.

Block – A block is an array of pixel values or DCT coefficients, usually 8-by-8 (8x8), representing luminance or chrominance information.

Bouquet – A group of transport streams in which programs are identified by combination of network ID and PID (part of DVB-SI).

BPSK – Binary Phase Shift Keying.

CA – Conditional Access. Information indicating whether a program is scrambled.

CAT – Conditional Access Table. Packets having PID (see Section 8 – Transport Streams) codes of 1 and that contain information about the scrambling system. *See ECM and EMM.*

CD – Compact disc.

CELP – Code Excited Linear Predictive. Channel Code – A modulation technique that converts raw data into a signal that can be recorded or transmitted by radio or cable.

CIF – Common Interchange Format. A 352x240 pixel format for 30 fps video conferencing.

Closed GOP – A Group of Pictures in which the last pictures do not need data from the next GOP for bidirectional coding. Closed GOP is used to make a splice point in a bit stream.

C/N – Carrier to Noise ratio.

Coefficient – A number specifying the amplitude of a particular frequency or basis function in a transform.

CORBA – Common Object Request Broker Architecture.

COFDM – Coded Orthogonal Frequency Division Multiplex, a modified form of OFDM. A digital modulation scheme using a very large number of carriers, each carrying a very low data rate. Used by DVB-T.

Compression – Reduction in the number of bits used to represent an item of data.

CRC – Cyclic Redundancy Check.

DAB – Digital Audio Broadcast.

DAB-IP – Digital Audio Broadcast - Internet Protocol.

DAC – Digital-to-Analog Converter.

DASE – Digital TV Application Software Environment.

DAVIC – Digital Audio Visual Council.

DCT – Discrete Cosine Transform.

DDB – DownloadDataBlock.

DET – Data Event Table.

DFT – Discrete Fourier Transform.

DII – DownloadInfoIndication.

DMB-T/H – Digital Multimedia Broadcast-Terrestrial/Handheld.

Dolby Digital – *See AC-3.*

DRM – Digital Rights Management.

DSI – DownloadServerInitiate.

DSMCC – Digital Storage Media Command and Control.

DST – Data Services Table.

DTS – Decoding Time Stamp. Part of PES header indicating when an access unit is to be decoded.

DVB – Digital Video Broadcasting. Generally refers to the European-initiated consortium of broadcasters, manufacturers, regulatory bodies and others that created standards for the delivery of digital television and data services. Includes DVB-C (cable), DVB-S (satellite) and DVB-T (terrestrial) versions.

DVB-H – Digital Video Broadcasting - Handheld.

DVB-SI – DVB Service Information. Information carried in a DVB multiplex describing the contents of different multiplexes. Includes NIT, SDT, EIT, TDT, BAT, RST and ST *(see Section 10 – Introduction to DVB & ATSC).*

DVC – Digital Video Cassette.

DVD – Digital Versatile Disk or Digital Video Disk. Elementary Stream – The raw output of a compressor carrying a single video or audio signal.

ECM – Entitlement Control Message. Conditional access information specifying control words or other stream-specific scrambling parameters.

ECS – Enhanced Content Specification.

EIT – Event Information Table. Part of DVB-SI.

EMM – Entitlement Management Message. Conditional access information specifying authorization level or services of specific decoders. An individual decoder or a group of decoders may be addressed.

ENG – Electronic News Gathering. Term used to describe use of video-recording instead of film in news coverage.

Entropy Coding – Variable length lossless coding of the digital representation of a signal to reduce redundancy.

EOB – End of Block.

EPG – Electronic Program Guide. A program guide delivered by data transfer rather than printed paper.

ETSI – European Telecommunication Standard Institute.

FEC – Forward Error Correction. System in which redundancy is added to the message so that errors can be corrected dynamically at the receiver.

FFT – Fast Fourier Transform.

GEM – Globally Executable MHP.

GOP – Group of Pictures. In transmission order a GOP starts with an I-picture and ends with the last picture before the next I-picture.

HAVI – Home Audio Video Interoperability.

HDTV – High Definition Television.

HTML – HyperText Markup Language.

Huffman coding – A type of source coding that uses codes of different lengths to represent symbols which have unequal likelihood of occurrence.

IEC – International Electrotechnical Commission.

Inter-coding – Compression that uses redundancy between successive pictures; also known as temporal coding.

Interleaving – A technique used with error correction that breaks up burst errors into many smaller errors.

Intra-coding – Compression that works entirely within one picture; also known as spatial coding.

IOR – Inter-operable Object Reference.

IP – Internet Protocol.

I-pictures – Intra-coded Pictures.

IRD – Integrated Receiver Decoder. A combined RF receiver and MPEG decoder that is used to adapt a TV set to digital transmissions.

ISDB – Integrated Services Data Broadcasting, the digital broadcasting system developed in Japan.

ISDB-Tss – Single Segment ISDB-T.

ISO – International Organization for Standardization.

ITU – International Telecommunication Union.

JPEG – Joint Photographic Experts Group.

JTC1 – Joint Technical Committee of the IEC.

JVT – Joint Video Team.

LDPC – Low-Density Parity Check Code.

Level – The size of the input picture in use with a given profile *(see Section 2 – Compression in Video).*

MAC – Media Access Control.

Macroblock – The screen area represented by several luminance and color-difference DCT blocks that are all steered by one motion vector.

Masking – A psycho-acoustic phenomenon whereby certain sounds cannot be heard in the presence of others.

MDCT – Modified Discreet Cosine Transform.

MGT – Master Guide Table.

MHP – Multimedia Home Platform.

Motion Vector – A pair of numbers which represent the vertical and horizontal displacement of a region of a reference picture for prediction.

MP@HL – Main Profile at High Level.

MP@LL – Main Profile at Low Level.

MP@ML – Main Profile at Main Level.

MPE – Multi-protocol Encapsulation.

MPEG – Moving Picture Experts Group ISO/IEC JTC1/SC29/WG11, and the Standards developed by this Group.

MPEG-LA – MPEG License Agreements.

MPTS – Multi Program Transport Stream.

MXF – Material eXchange Format.

NIT – Network Information Table. Information in one transport stream that describes many transport streams.

NPT – Normal Play Time.

NRT – Network Resources Table.

Null Packets – Packets of "stuffing" that carry no data but are necessary to maintain a constant bit rate with a variable payload. Null packets always have a PID of 8191 (all ones). *(See Section 8 – Transport Streams.)*

OCAP – OpenCable™ Applications Platform.

OFDM – Orthogonal Frequency Division Multiplexing.

ORB – Object Request Brokerage.

PAL – Phase Alternate Line.

PAT – Program Association Table. Data appearing in packets having PID *(see Section 8 – Transport Streams)* code of zero that the MPEG decoder uses to determine which programs exist in a Transport Stream. PAT points to PMT, which, in turn, points to the video, audio and data content of each program.

PCM – Pulse Code Modulation. A technical term for an analog source waveform, for example, audio or video signals, expressed as periodic, numerical samples. PCM is an uncompressed digital signal.

PCR – Program Clock Reference. The sample of the encoder clock count that is sent in the program header to synchronize the decoder clock.

PES – Packetized Elementary Stream.

PID – Program Identifier. A 13-bit code in the transport packet header. PID 0 indicates that the packet contains a PAT PID. *(See Section 8 – Transport Streams.)* PID 1 indicates a packet that contains CAT. The PID 8191 (all ones) indicates null (stuffing) packets. All packets belonging to the same elementary stream have the same PID.

PLL – Phase Locked Loop.

PMT – Program Map Tables. The tables in PAT that point to video, audio and data content of a transport stream.

Packets – A term used in two contexts: in program streams, a packet is a unit that contains one or more presentation units; in transport streams, a packet is a small, fixed-size data quantum.

Pixel – Picture element (sometimes pel). The smallest unit of an image, represented by one sample, or a set of samples such as GBR or YC_rC_b.

Preprocessing – The video signal processing that occurs before MPEG Encoding. Noise reduction, downsampling, cut-edit identification and 3:2 pulldown identification are examples of preprocessing.

Profile – Specifies the coding syntax used.

Program Stream – A bit stream containing compressed video, audio and timing information.

PS – Parametric Stereo.

PS – Program Stream.

PSI – Program Specific Information. Information that keeps track of the different programs in an MPEG transport stream and in the elementary streams in each program. PSI includes PAT, PMT, NIT, CAT, ECM and EMM.

PSI/SI – A general term for combined MPEG PSI and DVB-SI.

PSIP – Program and System Information Protocol.

PTS – Presentation Time Stamp. The time at which a presentation unit is to be available to the viewer.

PU – Presentation Unit. One compressed picture or block of audio.

QAM – Quadrature Amplitude Modulation, a digital modulation system.

QCIF – One-quarter-resolution (176x144 pixels) Common Interchange Format. *See CIF.*

QMF – Quadrature Mirror Filter.

QPSK – Quaternary Phase Shift Keying (also known as Quadrature Phase Shift Keying), a digital modulation system particularly suitable for satellite links.

QSIF – One-quarter-resolution Source Input Format. See SIF.

Quantization – a processing step that approximates signal values by allocating one of a number of pre-defined values.

RLC – Run Length Coding. A coding scheme that counts number of similar bits instead of sending them individually.

RRT – Rating Region Table.

RST – Running Status Table.

R-S – Reed-Solomon is a polynomial function used by DVB for protecting up to 8 bytes within each transport packet.

RTP – Real-Time Protocol.

SBR – Spectral Band Replication.

Scalability – A characteristic of MPEG-2 that provides for multiple quality levels by providing layers of video data. Multiple layers of data allow a complex decoder to produce a better picture by using more layers of data, while a more simple decoder can still produce a picture using only the first layer of data.

SDI – Serial Digital Interface. Serial coaxial cable interface standard intended for production digital video signals.

SDK – Software Development Kit.

SDT – Service Description Table. A table listing the providers of each service in a transport stream.

SDTV – Standard Definition Television.

Shannon Limit – Stated by Claude Shannon in 1948, the theorem describes the maximum possible efficiency of error-correcting methods versus levels of noise interference and data corruption.

SI – *See DVB-SI.*

SIF – Source Input Format. A half-resolution input signal used by MPEG-1.

Slice – A sequence of consecutive macroblocks.

SMPTE – Society of Motion Picture and Television Engineers.

SNR – Signal-to-Noise Ratio.

SP@ML – Simple Profile at Main Level.

SPTS – Single Program Transport Stream.

ST – Stuffing Table.

STB – Set Top Box.

STC – System Time Clock. The common clock used to encode video and audio in the same program.

STT – System Time Table.

Stuffing – Meaningless data added to maintain constant bit rate.

Syndrome – Initial result of an error checking calculation. Generally, if the syndrome is zero, there is assumed to be no error.

TCP/IP – Transmission Control Protocol/Internet Protocol.

TDAC – Time Domain Aliasing Cancellation. A coding technique used in AC-3 audio compression.

T-DBM – Digital Multimedia Broadcasting -Terrestrial.

TDT – Time and Date Table. Used in DVB-SI.

TOT – Time Offset Table.

Transport Stream (TS) – A multiplex of several program streams that are carried in packets. Demultiplexing is achieved by different packet IDs (PIDs). *See PSI, PAT, PMT and PCR.*

Truncation – Shortening the wordlength of a sample or coefficient by removing low-order bits.

T-STD – Transport Stream System Target Decoder. A decoder having a certain amount of buffer memory assumed to be present by an encoder.

TVCT – Terrestrial Virtual Channel Table.

UDP – User Datagram Protocol.

VAU – Video Access Unit. One compressed picture in program stream.

VBV – Video Buffer Verify.

VCO – Voltage Controlled Oscillator.

VLC – Variable Length Coding. A compressed technique that allocates short codes to frequency values and long codes to infrequent values.

VOD – Video On Demand. A system in which television programs or movies are transmitted to a single consumer only when requested.

VSB – Vestigial Sideband Modulation. A digital modulation system used by ATSC.

Wavelet – A transform using a basis function that is not of fixed length but that grows longer as frequency reduces.

Weighting – A method of changing the distribution of the noise that is due to truncation by pre-multiplying values.

XHTML – Extended HTML.

Y/C – Luminance and chrominance.

References:

1. ISO/IEC 13818 MPEG-2 System, Video, and Audio
2. ISO/IEC 144961-10 MPEG-4 Part 10 Video, also known as H.264
3. EN 300 421 DVB-S QPSK Satellite
4. EN 302 307 DVB-S2 Second Generation Satellite
5. ITU-T J.83B QAM-B, also known as SCTE 07
6. ATSC A/53 Digital Television Standard
7. ETSI TR 101 290 DVB Measurement Guidelines
8. SCTE 07 Digital Transmission for Cable Television
9. EN 50083-9 DVB-PI ASI and parallel
10. RFC 0791 Internet Protocol
11. IRT / Tektronix Comparison of Subjective and Objective Picture Quality Rating of MPEG-2 Video Report
12. ITU-R_BT.500-11 Methodology for the subjective assessment of the quality of television pictures
13. ITU-R BS.1770/1771 Algorithms to measure audio program loudness and true-peak audio level
14. ATSC A/85 Techniques for Establishing and Maintaining Audio Loudness for Digital Television
15. European Broadcast Union R 128 Loudness normalization and permitted maximum level of audio signals
16. SCTE35 Digital Program Insertion Cueing Message for Cable

Contact Tektronix:

ASEAN / Australasia (65) 6356 3900
Austria* 00800 2255 4835
Balkans, Israel, South Africa and other ISE Countries +41 52 675 3777
Belgium* 00800 2255 4835
Brazil +55 (11) 3759 7627
Canada 1 (800) 833-9200
Central East Europe and the Baltics +41 52 675 3777
Central Europe & Greece +41 52 675 3777
Denmark +45 80 88 1401
Finland +41 52 675 3777
France* 00800 2255 4835
Germany* 00800 2255 4835
Hong Kong 400-820-5835
India 000-800-650-1835
Italy* 00800 2255 4835
Japan 81 (3) 6714-3010
Luxembourg +41 52 675 3777
Mexico, Central/South America & Caribbean 52 (55) 56 04 50 90
Middle East, Asia and North Africa +41 52 675 3777
The Netherlands* 00800 2255 4835
Norway 800 16098
People's Republic of China 400-820-5835
Poland +41 52 675 3777
Portugal 80 08 12370
Republic of Korea 001-800-8255-2835
Russia & CIS +7 (495) 7484900
South Africa +27 11 206 8360
Spain* 00800 2255 4835
Sweden* 00800 2255 4835
Switzerland* 00800 2255 4835
Taiwan 886 (2) 2722-9622
United Kingdom & Ireland* 00800 2255 4835
USA 1 (800) 833-9200

*** If the European phone number above is not accessible, please call +41 52 675 3777**

Contact List Updated 10 February 2011

For Further Information
Tektronix maintains a comprehensive, constantly expanding collection of application notes, technical briefs and other resources to help engineers working on the cutting edge of technology. Please visit **www.tektronix.com**

Copyright © 2012, Tektronix. All rights reserved. Tektronix products are covered by U.S. and foreign patents, issued and pending. Information in this publication supersedes that in all previously published material. Specification and price change privileges reserved. TEKTRONIX and TEK are registered trademarks of Tektronix, Inc. All other trade names referenced are the service marks, trademarks or registered trademarks of their respective companies.

05/12 EA/ 25W-11418-10

Tektronix®