

JOHN WATKINSON

# THE MPEG HANDBOOK

MPEG-1 MPEG-2 MPEG-4

THE MPEG HANDBOOK

INCLUDES:

- » THE MPEG-4 STANDARD
- » MPEG AND NETWORKS
- » EXTENSIVE EXAMPLES OF APPLICATIONS

REVIEW OF PREVIOUS EDITION: "...PROVIDES EVERYTHING YOU  
NEED TO KNOW ABOUT COMPRESSION." (MART TECHNOLOGY

Cisco Systems, Inc., IPR2018-01384, Ex. 1017, p. 1 of 31

# The MPEG Handbook

MPEG-1, MPEG-2, MPEG-4


John Watkinson



**Focal Press**

OXFORD AUCKLAND BOSTON JOHANNESBURG MELBOURNE NEW DELHI

Focal Press  
An imprint of Butterworth-Heinemann  
Linacre House, Jordan Hill, Oxford OX2 8DP  
225 Wildwood Avenue, Woburn, MA 01801-2041  
A division of Reed Educational and Professional Publishing Ltd

 A member of the Reed Elsevier plc group

First published 2001

© John Watkinson 2001

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1P 0LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publishers

**British Library Cataloguing in Publication Data**

A catalogue record for this book is available from the British Library

**Library of Congress Cataloguing in Publication Data**

A catalogue record for this book is available from the Library of Congress

For information on all Focal Press publications visit  
our website at [www.focalpress.com](http://www.focalpress.com)

ISBN 0 240 51656 7

Composition by Genesis Typesetting, Rochester, Kent  
Printed and bound in Great Britain



# 1

## Introduction to compression

---

### 1.1 What is MPEG?

MPEG is actually an acronym for the Moving Pictures Experts Group which was formed by the ISO (International Standards Organization) to set standards for audio and video compression and transmission.

Compression is summarized in Figure 1.1. It will be seen in (a) that the data rate is reduced at source by the *compressor*. The compressed data are then passed through a communication channel and returned to the original rate by the *expander*. The ratio between the source data rate and the channel data rate is called the *compression factor*. The term *coding gain* is also used. Sometimes a compressor and expander in series are referred to as a *compander*. The compressor may equally well be referred to as a *coder* and the expander a *decoder* in which case the tandem pair may be called a *codec*.

Where the encoder is more complex than the decoder, the system is said to be asymmetrical. Figure 1.1(b) shows that MPEG works in this way. The encoder needs to be algorithmic or adaptive whereas the decoder is 'dumb' and carries out fixed actions. This is advantageous in applications such as broadcasting where the number of expensive complex encoders is small but the number of simple inexpensive decoders is large. In point-to-point applications the advantage of asymmetrical coding is not so great.

The approach of the ISO to standardization in MPEG is novel because it is not the encoder which is standardized. Figure 1.2(a) shows that instead the way in which a decoder shall interpret the bitstream is defined. A decoder which can successfully interpret the bitstream is said to be *compliant*. Figure 1.2(b) shows that the advantage of standardizing the decoder is that over time encoding algorithms can improve yet compliant decoders will continue to function with them.

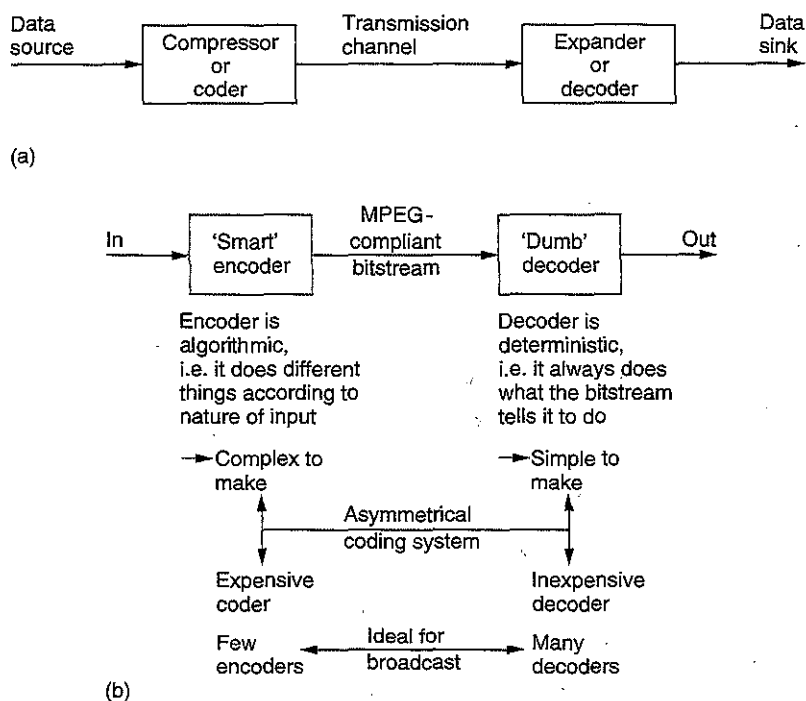
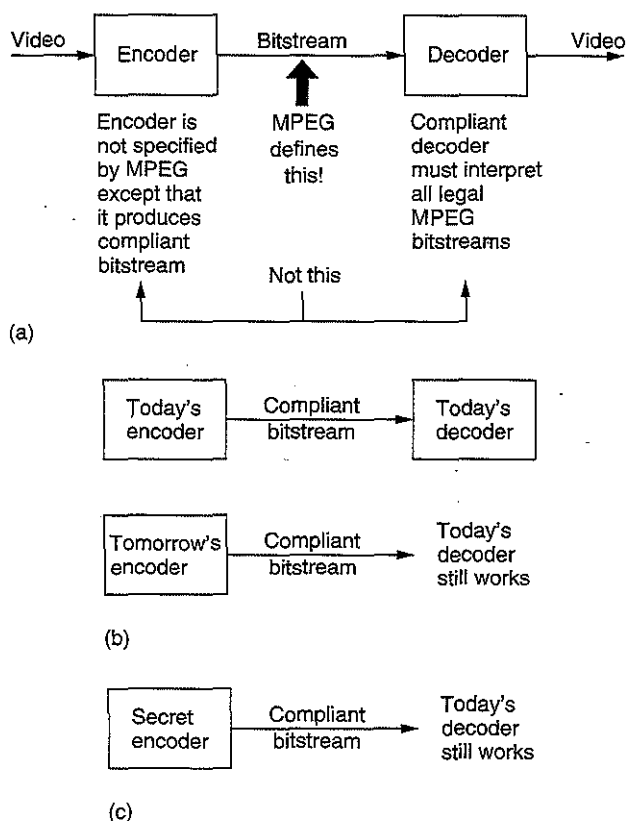


Figure 1.1 In (a) a compression system consists of compressor or coder, a transmission channel and a matching expander or decoder. The combination of coder and decoder is known as a codec. (b) MPEG is asymmetrical since the encoder is much more complex than the decoder.

It should be noted that a compliant decoder must correctly be able to interpret every allowable bitstream, whereas an encoder which produces a restricted subset of the possible codes can still be compliant.

The MPEG standards give very little information regarding the structure and operation of the encoder. Provided the bitstream is compliant, any coder construction will meet the standard, although some designs will give better picture quality than others. Encoder construction is not revealed in the bitstream and manufacturers can supply encoders using algorithms which are proprietary and their details do not need to be published. A useful result is that there can be competition between different encoder designs which means that better designs can evolve. The user will have greater choice because different levels of cost and complexity can exist in a range of coders yet a compliant decoder will operate with them all.

MPEG is, however, much more than a compression scheme as it also standardizes the protocol and syntax under which it is possible to combine or multiplex audio data with video data to produce a digital equivalent of a television program. Many such programs can be



**Figure 1.2** (a) MPEG defines the protocol of the bitstream between encoder and decoder. The decoder is defined by implication, the encoder is left very much to the designer. (b) This approach allows future encoders of better performance to remain compatible with existing decoders. (c) This approach also allows an encoder to produce a standard bitstream while its technical operation remains a commercial secret.

combined in a single multiplex and MPEG defines the way in which such multiplexes can be created and transported. The definitions include the metadata which decoders require to demultiplex correctly and which users will need to locate programs of interest.

As with all video systems there is a requirement for synchronizing or genlocking and this is particularly complex when a multiplex is assembled from many signals which are not necessarily synchronized to one another.

## 1.2 Why compression is necessary

Compression, bit rate reduction, data reduction and source coding are all terms which mean basically the same thing in this context. In essence the

same (or nearly the same) information is carried using a smaller quantity or rate of data. It should be pointed out that in audio *compression* traditionally means a process in which the dynamic range of the sound is reduced. In the context of MPEG the same word means that the bit rate is reduced, ideally leaving the dynamics of the signal unchanged. Provided the context is clear, the two meanings can co-exist without a great deal of confusion.

There are several reasons why compression techniques are popular:

- (a) Compression extends the playing time of a given storage device.
- (b) Compression allows miniaturization. With fewer data to store, the same playing time is obtained with smaller hardware. This is useful in ENG (electronic news gathering) and consumer devices.
- (c) Tolerances can be relaxed. With fewer data to record, storage density can be reduced making equipment which is more resistant to adverse environments and which requires less maintenance.
- (d) In transmission systems, compression allows a reduction in bandwidth which will generally result in a reduction in cost. This may make possible a service which would be impracticable without it.
- (e) If a given bandwidth is available to an uncompressed signal, compression allows faster than real-time transmission in the same bandwidth.
- (f) If a given bandwidth is available, compression allows a better-quality signal in the same bandwidth.

### 1.3 MPEG-1, 2 and 4 contrasted

The first compression standard for audio and video was MPEG-1.<sup>1,2</sup> Although many applications have been found, MPEG-1 was basically designed to allow moving pictures and sound to be encoded into the bit rate of an audio Compact Disc. The resultant Video-CD was quite successful but has now been superseded by DVD. In order to meet the low bit requirement, MPEG-1 downsampled the images heavily as well as using picture rates of only 24–30 Hz and the resulting quality was moderate.

The subsequent MPEG-2 standard was considerably broader in scope and of wider appeal.<sup>3</sup> For example, MPEG-2 supports interlace and HD whereas MPEG-1 did not. MPEG-2 has become very important because it has been chosen as the compression scheme for both DVB (digital video broadcasting) and DVD (digital video disk). Developments in standardizing scaleable and multi-resolution compression which would have become MPEG-3 were ready by the time MPEG-2 was ready to be

standardized and so this work was incorporated into MPEG-2, and as a result there is no MPEG-3 standard.

MPEG-4<sup>4</sup> uses further coding tools with additional complexity to achieve higher compression factors than MPEG-2. In addition to more efficient coding of video, MPEG-4 moves closer to computer graphics applications. In the more complex Profiles, the MPEG-4 decoder effectively becomes a rendering processor and the compressed bitstream describes three-dimensional shapes and surface texture. It is to be expected that MPEG-4 will become as important to Internet and wireless delivery as MPEG-2 has become in DVD and DVB.

## 1.4 Some applications of compression

The applications of audio and video compression are limitless and the ISO has done well to provide standards which are appropriate to the wide range of possible compression products.

MPEG coding embraces video pictures from the tiny screen of a videophone to the high-definition images needed for electronic cinema. Audio coding stretches from speech-grade mono to multichannel surround sound.

Figure 1.3 shows the use of a codec with a recorder. The playing time of the medium is extended in proportion to the compression factor. In the case of tapes, the access time is improved because the length of tape needed for a given recording is reduced and so it can be rewound more quickly. In the case of DVD (digital video disk aka digital versatile disk) the challenge was to store an entire movie on one 12 cm disk. The storage density available with today's optical disk technology is such that consumer recording of conventional uncompressed video would be out of the question.

In communications, the cost of data links is often roughly proportional to the data rate and so there is simple economic pressure to use a high compression factor. However, it should be borne in mind that implement-

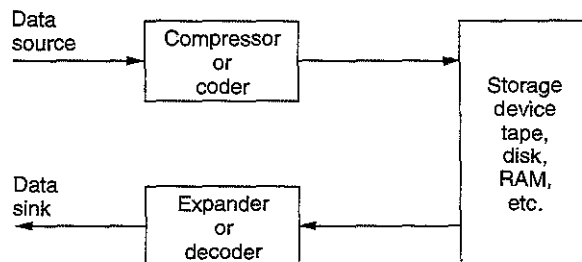


Figure 1.3 Compression can be used around a recording medium. The storage capacity may be increased or the access time reduced according to the application.

As interlaced systems have very poor vertical resolution, down-sampling to CIF actually does little damage to still images, although the very low picture rates damage motion portrayal.

Although MPEG-1 appeared rather rough on screen, this was due to the very low bit rate. It is more important to appreciate that MPEG-1 introduced the great majority of the coding tools which would continue to be used in MPEG-2 and MPEG-4. These included an elementary stream syntax, bidirectional motion-compensated coding, buffering and rate control. Many of the spatial coding principles of MPEG-1 were taken from JPEG. MPEG-1 also specified audio compression of up to two channels.

## 1.9 MPEG-2: Profiles and Levels

MPEG-2 builds upon MPEG-1 by adding interlace capability as well as a greatly expanded range of picture sizes and bit rates. The use of scaleable systems is also addressed, along with definitions of how multiple MPEG bitstreams can be multiplexed. As MPEG-2 is an extension of MPEG-1, it is easy for MPEG-2 decoders to handle MPEG-1 data. In a sense an MPEG-1 bitstream is an MPEG-2 bitstream which has a restricted vocabulary and so can be readily understood by an MPEG-2 decoder.

MPEG-2 has too many applications to solve with a single standard and so it is subdivided into Profiles and Levels. Put simply a Profile describes a degree of complexity whereas a Level describes the picture size or resolution which goes with that Profile. Not all Levels are supported at all Profiles. Figure 1.11 shows the available combinations. In principle there

		Profiles					
		Simple	Main	4:2:2	SNR	Spatial	High
Levels	High		4:2:0 1920 × 1152 90 Mb/S				4:2:0 or 4:2:2 1920 × 1152 100 Mb/S
	High 1440		4:2:0 1440 × 1152 60 Mb/S			4:2:0 1440 × 1152 60 Mb/S	4:2:0 or 4:2:2 1440 × 1152 80 Mb/S
	Main	4:2:0 720 × 576 15 Mb/S NO B	4:2:0 720 × 576 15 Mb/S	4:2:2 720 × 608 50 Mb/S	4:2:0 720 × 576 15 Mb/S		4:2:0 or 4:2:2 720 × 576 20 Mb/S
	Low		4:2:0 352 × 288 4 Mb/S		4:2:0 352 × 288 4 Mb/S		

Figure 1.11 Profiles and Levels in MPEG-2. See text for details.

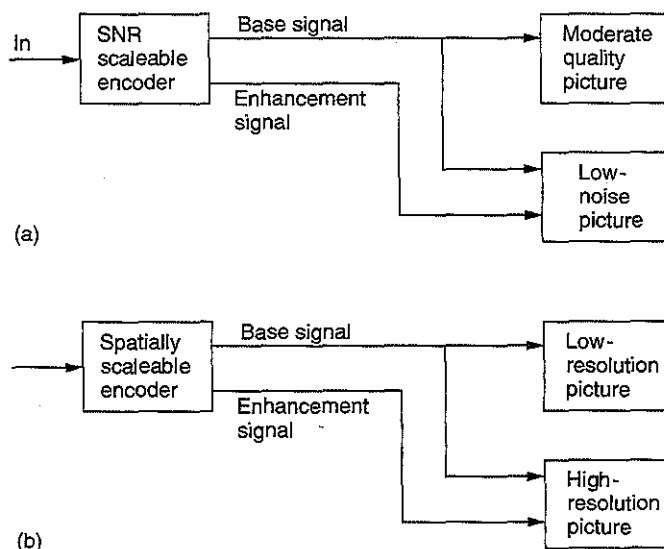
are twenty-four of these, but not all have been defined. An MPEG-2 decoder having a given Profile and Level must also be able to decode lower Profiles and Levels.

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 (SPML).

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 MPML (Main Profile at Main Level) subset of MPEG which supports SDTV (standard definition television). The High-1440 Level is a high-definition scheme which doubles the definition compared to 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 scaleable signals. A scaleable 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.

Figure 1.12(a) shows that in a conventional MPEG coder, by heavily requantizing coefficients a picture with moderate signal-to-noise ratio



**Figure 1.12** (a) An SNR scaleable encoder produces a 'noisy' signal and a noise cancelling signal. (b) A spatially scaleable encoder produces a low-resolution picture and a resolution-enhancing picture.

and encode the new data entirely as an unweighted prediction error. In this case it is better to encode that material as an *I* macroblock because then weighting can be used and this will require fewer bits.

Requantizing increases the step size of the coefficients, but the inverse weighting in the decoder results in step sizes which increase with frequency. The larger step size increases the quantizing noise at high frequencies where it is less visible. Effectively the noise floor is shaped to match the sensitivity of the eye. The quantizing table in use at the encoder can be transmitted to the decoder periodically in the bitstream.

## 5.12 Intra-coding in MPEG-1 and MPEG-2

Study of the signal statistics gained from extensive analysis of real material is used to measure the probability of a given coefficient having a given value. This probability turns out to be highly non-uniform suggesting the possibility of a variable-length encoding for the coefficient values. On average, the higher the spatial frequency, the lower the value of a coefficient will be. This means that the value of a coefficient tends to fall as a function of its radius from the DC coefficient. DC coefficients also have certain characteristics. As they represent the average brightness of a block, in many picture areas, adjacent blocks will have similar DC coefficient values. This can be exploited using differential coding which will be dealt with in section 5.14. However, errors in the magnitude of DC coefficients cause wide area flicker in the picture and so the coding must be accurate.

Typical material often has many coefficients which are zero valued, especially after requantizing. The distribution of these also follows a pattern. The non-zero values tend to be found in the top-left-hand corner of the DCT block, but as the radius increases, not only do the coefficient values fall, but it becomes increasingly likely that these small coefficients will be interspersed with zero-valued coefficients. As the radius increases further it is probable that a region where all coefficients are zero will be entered.

MPEG uses all these attributes of DCT coefficients when encoding a coefficient block. By sending the coefficients in an optimum order, by describing their values with Huffman coding and by using run-length encoding for the zero-valued coefficients it is possible to achieve a significant reduction in coefficient data which remains entirely lossless. Despite the complexity of this process, it does contribute to improved picture quality because for a given bit rate lossless coding of the coefficients must be better than requantizing, which is lossy. Of course, for lower bit rates both will be required.

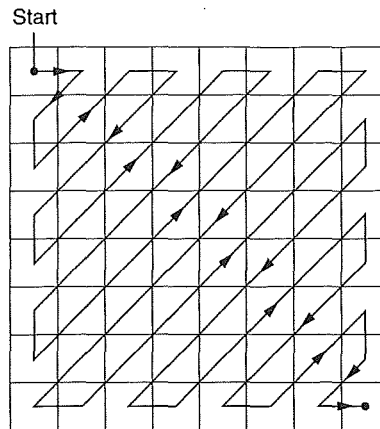


Figure 5.32 The zig-zag scan for a progressively scanned image.

It is an advantage to scan in a sequence where the largest coefficient values are scanned first. Then the next coefficient is more likely to be zero than the previous one. With progressively scanned material, a regular zig-zag scan begins in the top-left corner and ends in the bottom-right corner as shown in Figure 5.32. Zig-zag scanning means that significant values are more likely to be transmitted first, followed by the zero values. Instead of coding these zeros, a unique 'end of block' (EOB) symbol is transmitted instead.

As the zig-zag scan approaches the last finite coefficient it is increasingly likely that some zero value coefficients will be scanned. Instead of transmitting the coefficients as zeros, the *zero-run-length*, i.e. the number of zero-valued coefficients in the scan sequence, is encoded into the next non-zero coefficient which is itself variable-length coded. This combination of run-length and variable-length coding is known as RLC/VLC in MPEG.

The DC coefficient is handled separately because it is differentially coded and this discussion relates to the AC coefficients. Three items need to be handled for each coefficient: the zero-run-length prior to this coefficient, the wordlength and the coefficient value itself. The wordlength needs to be known by the decoder so that it can correctly parse the bitstream. The wordlength of the coefficient is expressed directly as an integer called the *size*.

Figure 5.33(a) shows that a two-dimensional run/size table is created. One dimension expresses the zero-run-length; the other the size. A run length of zero is obtained when adjacent coefficients are non-zero, but a code of 0/0 has no meaningful run/size interpretation and so this bit pattern is used for the end-of-block (EOB) symbol.

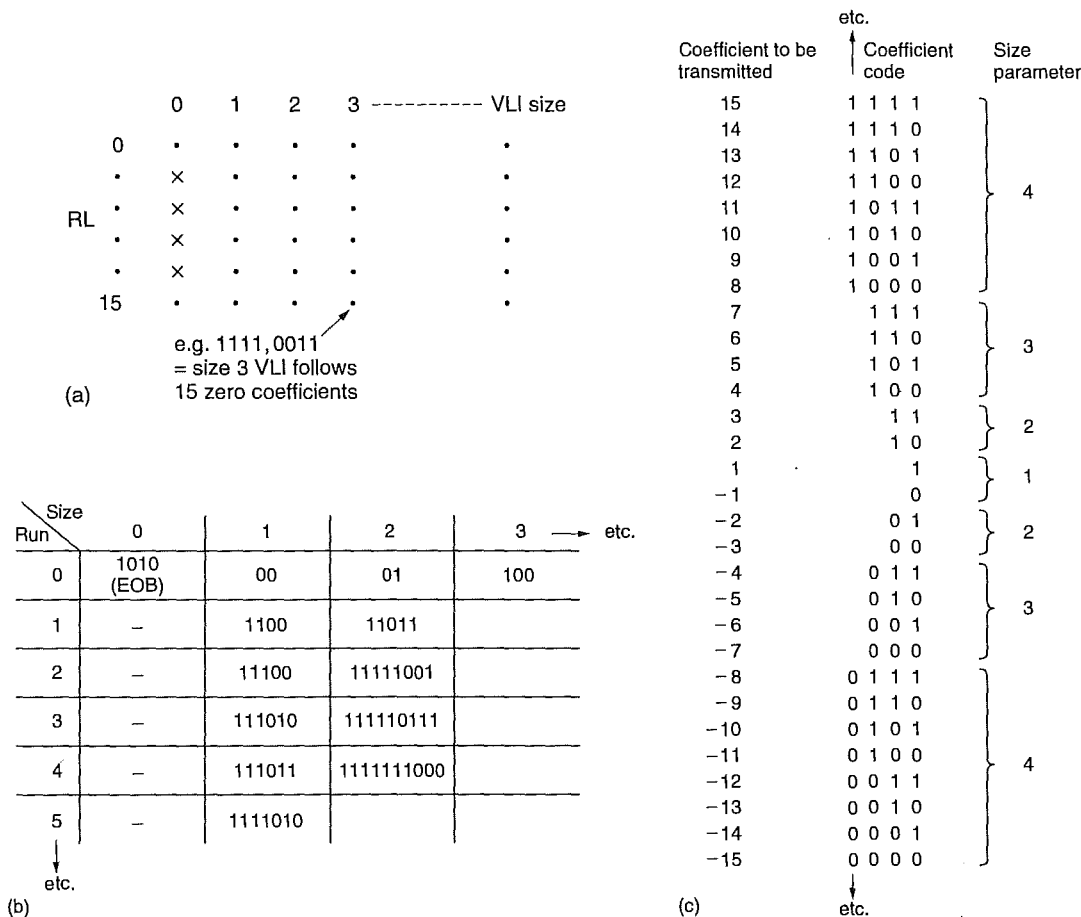


Figure 5.33 Run-length and variable-length coding simultaneously compresses runs of zero-valued coefficients and describes the wordlength of a non-zero coefficient.

In the case where the zero-run-length exceeds 14, a code of 15/0 is used signifying that there are fifteen zero-valued coefficients. This is then followed by another run/size parameter whose run-length value is added to the previous fifteen.

The run/size parameters contain redundancy because some combinations are more common than others. Figure 5.33(b) shows that each run/size value is converted to a variable-length Huffman codeword for transmission. As was shown in section 1.5, the Huffman codes are designed so that short codes are never a prefix of long codes so that the decoder can deduce the parsing by testing an increasing number of bits until a match with the look-up table is found. Having parsed and decoded the Huffman run/size code, the decoder then knows what the coefficient wordlength will be and can correctly parse that.

The variable-length coefficient code has to describe a bipolar coefficient, i.e. one which can be positive or negative. Figure 5.33(c) shows that for a particular size, the coding scale has a certain gap in it. For example, all values from  $-7$  to  $+7$  can be sent by a size 3 code, so a size 4 code only has to send the values of  $-15$  to  $-8$  and  $+8$  to  $+15$ . The coefficient code is sent as a pure binary number whose value ranges from all zeros to all ones where the maximum value is a function of the size. The number range is divided into two, the lower half of the codes specifying negative values and the upper half specifying positive.

In the case of positive numbers, the transmitted binary value is the actual coefficient value, whereas in the case of negative numbers a constant must be subtracted which is a function of the size. In the case of a size 4 code, the constant is  $15_{10}$ . Thus a size 4 parameter of  $0111_2$  ( $7_{10}$ ) would be interpreted as  $7 - 15 = -8$ . A size of 5 has a constant of 31 so a transmitted coded of  $01010_2$  ( $10_2$ ) would be interpreted as  $10 - 31 = -21$ .

This technique saves a bit because, for example, 63 values from  $-31$  to  $+31$  are coded with only five bits having only 32 combinations. This is possible because that extra bit is effectively encoded into the run/size parameter.

Figure 5.34 shows the entire MPEG-1 and MPEG-2 spatial coding subsystem. Macroblocks are subdivided into DCT blocks and the DCT is calculated. The resulting coefficients are multiplied by the weighting matrix and then requantized. The coefficients are then reordered by the zig-zag scan so that full advantage can be taken of run-length and variable-length coding. The last non-zero coefficient in the scan is followed by the EOB symbol.

In predictive coding, sometimes the motion-compensated prediction is nearly exact and so the prediction error will be almost zero. This can also happen on still parts of the scene. MPEG takes advantage of this by sending a code to tell the decoder there is no prediction error data for the macroblock concerned.

The success of temporal coding depends on the accuracy of the vectors. Trying to reduce the bit rate by reducing the accuracy of the vectors is false economy as this simply increases the prediction error. Consequently for a given GOP structure it is only in the the spatial coding that the

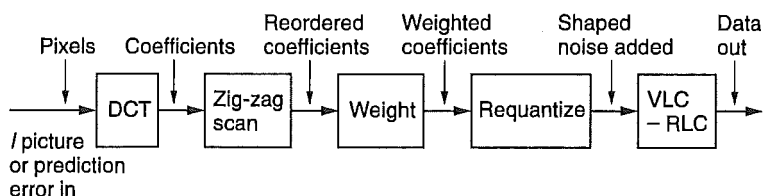


Figure 5.34 A complete spatial coding system which can compress an *I* picture or the prediction error in *P* and *B* pictures. See text for details.

overall bit rate is determined. The RLC/VLC coding is lossless and so its contribution to the compression cannot be varied. If the bit rate is too high, the only option is to increase the size of the coefficient-requantizing steps. This has the effect of shortening the wordlength of large coefficients, and rounding small coefficients to zero, so that the bit rate goes down. Clearly if taken too far the picture quality will also suffer because at some point the noise floor will become visible as some form of artifact.

### 5.13 A bidirectional coder

MPEG does not specify how an encoder is to be built or what coding decisions it should make. Instead it specifies the protocol of the bitstream at the output. As a result the coder shown in Figure 5.35 is only an example. Conceptually MPEG-2 Main Profile coding with a progressive scan input is almost identical to that of MPEG-1 as the differences are to be found primarily in the picture size and the resulting bit rate. The coder shown here will create MPEG-1 and MPEG-2 bidirectionally coded bitstreams from progressively scanned inputs. It also forms the basis of interlaced MPEG-2 coding and for MPEG-4 texture coding, both of which will be considered later.

Figure 5.35(a) shows the component parts of the coder. At the input is a chain of picture stores which can be bypassed for reordering purposes. This allows a picture to be encoded ahead of its normal timing when bidirectional coding is employed.

At the centre is a dual-motion estimator which can simultaneously measure motion between the input picture, an earlier picture and a later picture. These reference pictures are held in frame stores. The vectors from the motion estimator are used locally to shift a picture in a frame store to form a predicted picture. This is subtracted from the input picture to produce a prediction error picture which is then spatially coded.

The bidirectional encoding process will now be described. A GOP begins with an *I* picture which is intra-coded. In Figure 5.35(b) the *I* picture emerges from the reordering delay. No prediction is possible on an *I* picture so the motion estimator is inactive. There is no predicted picture and so the prediction error subtractor is set simply to pass the input. The only processing which is active is the forward spatial coder which describes the picture with DCT coefficients. The output of the forward spatial coder is locally decoded and stored in the past picture frame store.

The reason for the spatial encode/decode is that the past picture frame store now contains exactly what the decoder frame store will contain, including the effects of any requantizing errors. When the same picture is

# 7

## MPEG applications

---

### 7.1 Introduction

It should be borne in mind that however exciting it may be, MPEG is only a technology. Technology *per se* is not useful and only becomes useful when it is incorporated into an appropriate and affordable product or service. If the use is not appropriate or affordable, there is no point in blaming the technology.

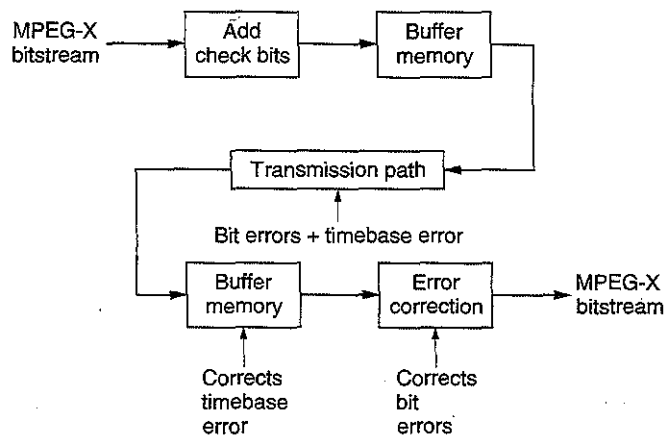
In this chapter a number of useful applications of MPEG are outlined. In each case designers have followed the rules of high technology to design systems which maximize the potential of the technology whilst protecting against its weaknesses.

MPEG is an information technology. It cannot deliver anything physical: you get the movie but no popcorn. The strength of MPEG is that it delivers audio-visual material with economy of bandwidth. Its weaknesses are sensitivity to bit errors and the requirement for near-real-time transmission. The variable-length coding of MPEG can lose synchronization in the presence of a single bit error. Decoders can take a considerable time to recover from a buffer underflow or overflow.

Wherever bandwidth or storage capacity is in short supply, MPEG is useful. Where bandwidth is plentiful, as in optical fibre networks, MPEG is pointless. As the capacity and economics of storage devices continue to increase, MPEG becomes less important.

MPEG will remain important in any area where there will be a permanent shortage of bandwidth; this includes any mobile equipment or service which must use radio communication.

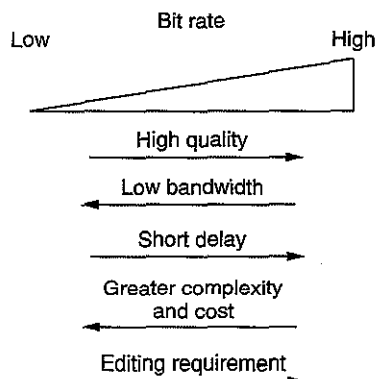
Figure 7.1 shows that well-engineered applications of MPEG will always include an error-correction system appropriate for the characteristics of the channel and suitable mechanisms to ensure that decoder



**Figure 7.1** Any real-time MPEG delivery system must include means to maintain a low error rate and buffering to deliver data within the decoder's time window.

buffers neither overflow nor underflow. Constant-bit rate channels having constant delay are easy to implement because the buffering mechanisms of the standard MPEG encoders and decoders are able to operate with no further assistance. Digital TV transmitters provide essentially constant bit-rate because this is related to the standardized (and regulated) channel bandwidth allowed.

In packet networks the delivery delay is not constant because other types of messages are multiplexed with the audio-visual message of interest. Variable delay channels cause MPEG decoders difficulty. Not only is there a greater danger of buffer overflow or underflow, but the program clock reference information is no longer correct if a packet is



**Figure 7.2** There are a large number of conflicting demands in the design of a compression system. Here the effect of a number of choices is shown to alter the balance between economy and bandwidth.

received with a time shift. In this case additional buffering is needed at both ends of the channel so that the variable delay characteristic of the channel is invisible to the MPEG layer.

There are now many members of the MPEG family and it can be difficult to decide which type of MPEG coding to use. Figure 7.2 shows some general guidelines. As a decoder becomes more complex, the bit rate which can be used for a given quality will fall. However, this will be accompanied by an increase in power consumption at the encoder and decoder as well as an increased delay. The designer will have to balance the cost of the bandwidth with the cost of the encoding and decoding equipment.

## 7.2 Video phones

The telephone is a classic example of limited bandwidth. If there is a market for video phones, then compression is the only way forward. The facial animation coding developed for MPEG-4 is an obvious contender as it allows a reasonable reproduction of the caller's face with a data rate of only a few kilobits per second. This is considerably less than the data rate needed for the speech. Video telephony is relatively easy between desk-based equipment, where the camera can be mounted on top of the display, but it is not clear how this can be implemented in a cellular telephone which is held against the user's ear. Another important issue for cellular telephones is the amount of signal processing needed to decode and display facial animation data. Essentially the decoder is a real-time rendering engine and this is only practicable in a portable product if the current consumption is within the range of available battery technology.

## 7.3 Digital television broadcasting

Digital television broadcasting relies on the combination of a number of fundamental technologies. These are: MPEG-2 compression to reduce the bit rate, multiplexing to combine picture and sound data into a common bitstream, digital modulation schemes to reduce the RF bandwidth needed by a given bit rate and error correction to reduce the error statistics of the channel down to a value acceptable to MPEG data.

MPEG is a compression and multiplexing standard and does not specify how error correction should be performed. Consequently a transmission standard must define a system which has to correct essentially all errors such that the delivery mechanism is transparent.

DVB (digital video broadcasting)<sup>1</sup> is a standard which incorporates MPEG-2 picture and sound coding but which also specifies all the additional steps needed to deliver an MPEG transport stream from one place to another. This transport stream will consist of a number of elementary streams of video and audio, where the audio may be coded according to MPEG audio standard or AC-3. In a system working within its capabilities, the picture and sound quality will be determined only by the performance of the compression system and not by the RF transmission channel. This is the fundamental difference between analog and digital broadcasting. In analog television broadcasting, the picture quality may be limited by composite video encoding artifacts as well as transmission artifacts such as noise and ghosting. In digital television broadcasting the picture quality is determined instead by the compression artifacts and interlace artifacts if interlace has been retained.

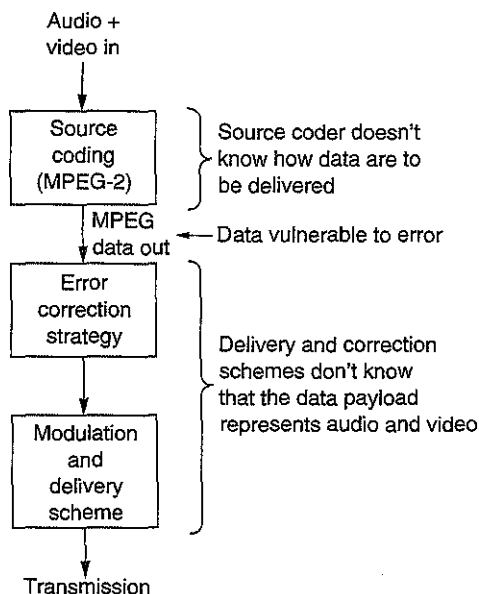
If the received error rate increases for any reason, once the correcting power is used up, the system will degrade rapidly as uncorrected errors enter the MPEG decoder. In practice decoders will be programmed to recognize the condition and blank or mute to avoid outputting garbage. As a result digital receivers tend either to work well or not at all.

It is important to realize that the signal strength in a digital system does not translate directly to picture quality. A poor signal will increase the number of bit errors. Provided that this is within the capability of the error-correction system, there is no visible loss of quality. In contrast, a very powerful signal may be unusable because of similarly powerful reflections due to multipath propagation.

Whilst in one sense an MPEG transport stream is only data, it differs from generic data in that it must be presented to the viewer at a particular rate. Generic data are usually asynchronous, whereas baseband video and audio are synchronous. However, after compression and multiplexing audio and video are no longer precisely synchronous and so the term *isochronous* is used. This means a signal which was at one time synchronous and will be displayed synchronously, but which uses buffering at transmitter and receiver to accommodate moderate timing errors in the transmission.

Clearly another mechanism is needed so that the time axis of the original signal can be re-created on reception. The time stamp and program clock reference system of MPEG does this.

Figure 7.3 shows that the concepts involved in digital television broadcasting exist at various levels which have an independence not found in analog technology. In a given configuration a transmitter can radiate a given payload data bit rate. This represents the useful bit rate and does not include the necessary overheads needed by error correction, multiplexing or synchronizing. It is fundamental that the transmission system does not care what this payload bit rate is used for. The entire



**Figure 7.3** Source coder doesn't know delivery mechanism and delivery doesn't need to know what the data mean.

capacity may be used up by one high-definition channel, or a large number of heavily compressed channels may be carried. The details of this data usage are the domain of the *transport stream*. The multiplexing of transport streams is defined by the MPEG standards, but these do not define any error correction or transmission technique.

At the lowest level in Figure 7.3 is the source coding scheme, in this case MPEG-2 compression results in one or more elementary streams, each of which carries a video or audio channel. Figure 7.4 shows that elementary streams are multiplexed into a transport stream. The viewer then selects the desired elementary stream from the transport stream. Metadata in the transport stream ensure that when a video elementary stream is chosen, the appropriate audio elementary stream will automatically be selected.

A key difference between analog and digital transmission is that the transmitter output is switched between a number of discrete states rather than continuously varying. The process is called channel coding which is the digital equivalent of modulation.<sup>2</sup> A good code minimizes the channel bandwidth needed for a given bit rate. This quality of the code is measured in bits/s/Hz.

Where the SNR is poor, as in satellite broadcasting, the amplitude of the signal will be unstable, and phase modulation is used. Figure 7.5 shows that phase-shift keying (PSK) can use two or more phases. When four

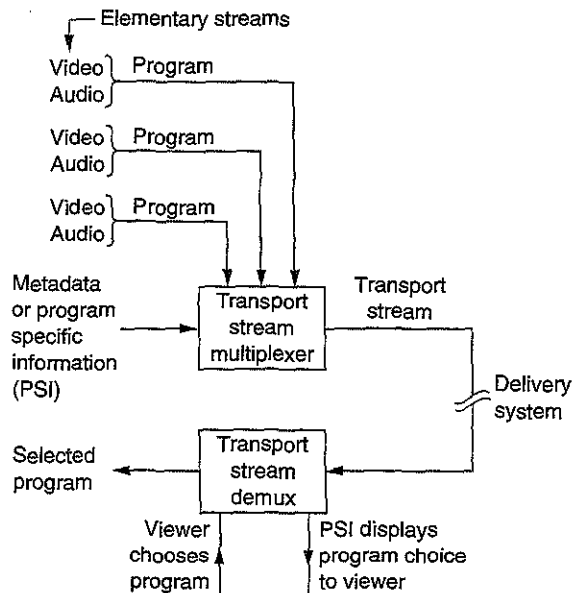


Figure 7.4 Program Specific Information helps the demultiplexer to select the required program.

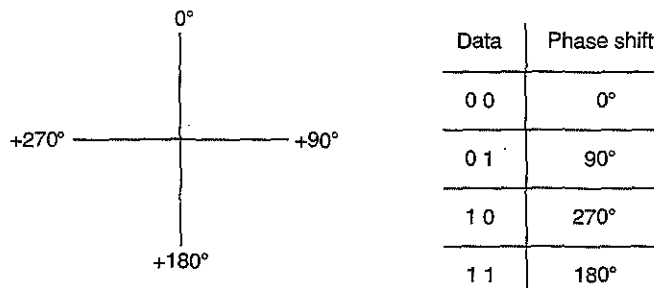
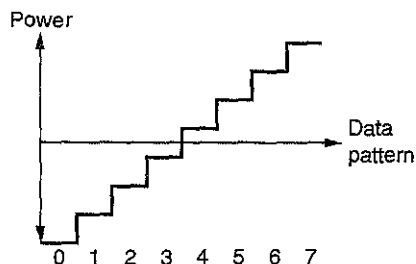


Figure 7.5 Differential quadrature phase shift keying (DQPSK).

phases in quadrature are used, the result is quadrature phase shift keying or QPSK. Each period of the transmitted waveform can have one of four phases and therefore conveys the value of two data bits. 8-PSK uses eight phases and can carry three bits per symbol where the SNR is adequate. PSK is generally encoded in such a way that a knowledge of absolute phase is not needed at the receiver. Instead of encoding the signal phase directly, the data determine the magnitude of the phase shift between symbols.

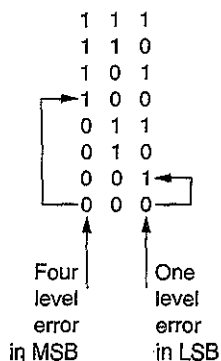
In terrestrial transmission more power is available than, for example, from a satellite and so a stronger signal can be delivered to the receiver. Where a better SNR exists, an increase in data rate can be had using



**Figure 7.6** In 8-VSB the transmitter operates in eight different states enabling three bits to be sent per symbol.

multi-level signalling instead of binary. Figure 7.6 shows that the ATSC system uses an eight-level signal (8-VSB) allowing three bits to be sent per symbol. Four of the levels exist with normal carrier phase and four exist with inverted phase so that a phase-sensitive rectifier is needed in the receiver. Clearly the data separator must have a three-bit ADC which can resolve the eight signal levels. The gain and offset of the signal must be precisely set so that the quantizing levels register precisely with the centres of the eyes. The transmitted signal contains sync pulses which are encoded using specified code levels so that the data separator can set its gain and offset.

Multi-level signalling systems have the characteristic that the bits in the symbol have different error probability. Figure 7.7 shows that a small noise level will corrupt the low-order bit, whereas twice as much noise will be needed to corrupt the middle bit and four times as much will be needed to corrupt the high-order bit. In ATSC the solution is that the lower two bits are encoded together in an inner error-correcting scheme so that they represent only one bit with similar reliability to the top bit. As a result the 8-VSB system actually delivers two data bits per symbol even though eight-level signalling is used.



**Figure 7.7** In multi-level signalling the error probability is not the same for each bit.

Multi-level signalling can be combined with PSK to obtain multi-level quadrature amplitude modulation (QUAM). Figure 7.8 shows the example of 64-QUAM. Incoming six-bit data words are split into two three-bit words and each is used to amplitude modulate a pair of sinusoidal carriers which are generated in quadrature. The modulators

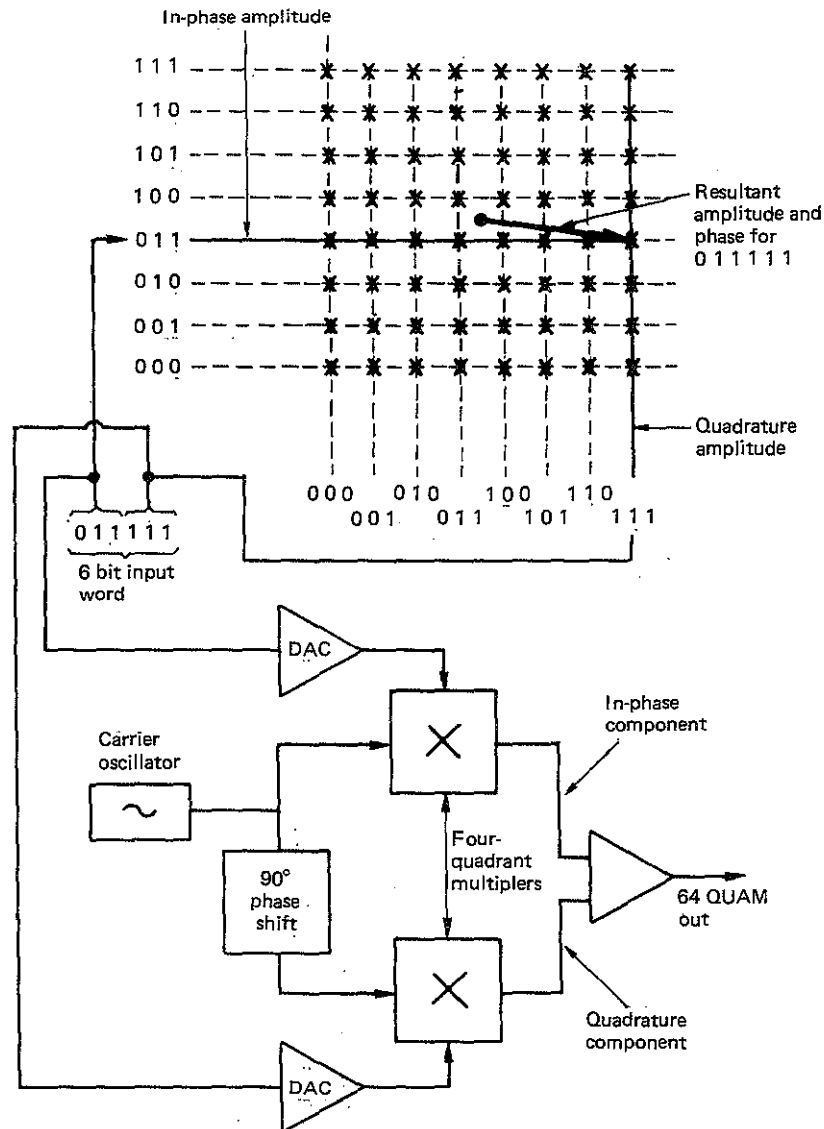
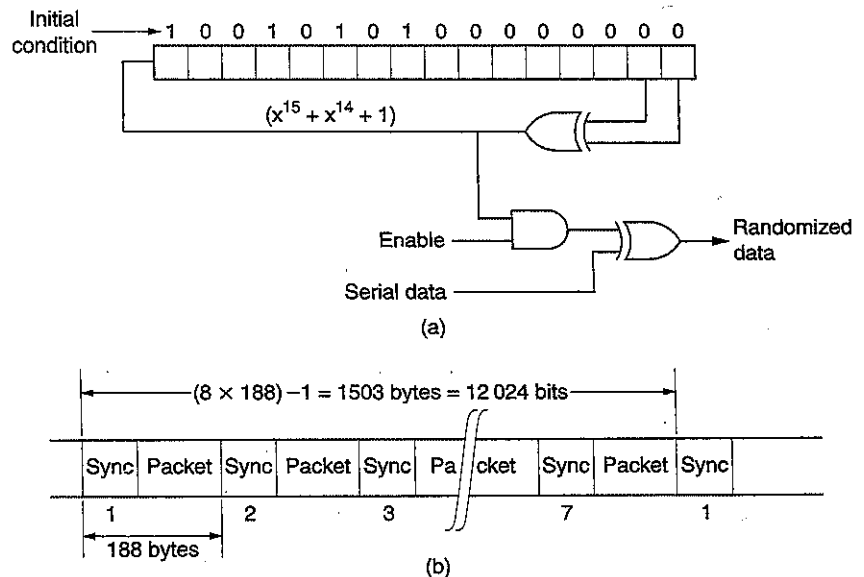


Figure 7.8 In 64-QUAM, two carriers are generated with a quadrature relationship. These are independently amplitude modulated to eight discrete levels in four quadrant multipliers. Adding the signals produces a QUAM signal having 64 unique combinations of amplitude and phase. Decoding requires the waveform to be sampled in quadrature like a colour TV subcarrier.

are four-quadrant devices such that  $2^3$  amplitudes are available, four which are in phase with the carrier and four which are antiphase. The two AM carriers are linearly added and the result is a signal which has  $2^6$  or 64 combinations of amplitude and phase. There is a great deal of similarity between QUAM and the colour subcarrier used in analog television in which the two colour difference signals are encoded into one amplitude- and phase-modulated waveform. On reception, the waveform is sampled twice per cycle in phase with the two original carriers and the result is a pair of eight-level signals. 16-QUAM is also possible, delivering only four bits per symbol but requiring a lower SNR.

The data bit patterns to be transmitted can have any combinations whatsoever, and if nothing were done, the transmitted spectrum would be non-uniform. This is undesirable because peaks cause interference with other services, whereas energy troughs allow external interference in. A randomizing technique known as energy dispersal is used to overcome the problem. The signal energy is spread uniformly throughout the allowable channel bandwidth so that it has less energy at a given frequency.

A pseudo-random sequence (PRS) generator is used to generate the randomizing sequence. Figure 7.9 shows the randomizer used in DVB. This sixteen-bit device has a maximum sequence length of 65 535 bits, and is preset to a standard value at the beginning of each set of eight transport



**Figure 7.9** The randomizer of DVB is pre-set to the initial condition once every eight transport stream packets. The maximum length of the sequence is 65 535 bits, but only the first 12 024 bits are used before resetting again (b).

stream packets. The serialized data are XORed with the sequence, which randomizes the output that then goes to the modulator. The spectrum of the transmission is now determined by the spectrum of the PRS.

On reception, the de-randomizer must contain the identical ring counter which must also be set to the starting condition to bit accuracy. Its output is then added to the data stream from the demodulator. The randomizing will effectively then have been added twice to the data in modulo-2, and as a result is cancelled out leaving the original serial data.

The way that radio signals interact with obstacles is a function of the relative magnitude of the wavelength and the size of the object. AM sound radio transmissions with a wavelength of several hundred metres can easily diffract around large objects. The shorter the wavelength of a transmission, the larger objects in the environment appear to it and these objects can then become reflectors. Reflecting objects produce a delayed signal at the receiver in addition to the direct signal. In analog television transmissions this causes the familiar ghosting. In digital transmissions, the symbol rate may be so high that the reflected signal may be one or more symbols behind the direct signal, causing intersymbol interference. As the reflection may be continuous, the result may be that almost every symbol is corrupted. No error-correction system can handle this. Raising the transmitter power is no help at all as it simply raises the power of the reflection in proportion.

The only solution is to change the characteristics of the RF channel in some way to either prevent the multipath reception or to prevent it being a problem. The RF channel includes the modulator, transmitter, antennae, receiver and demodulator.

As with analog UHF TV transmissions, a directional antenna is useful with digital transmission as it can reject reflections. However, directional antennae tend to be large and they require a skilled permanent installation. Mobile use on a vehicle or vessel is simply impractical.

Another possibility is to incorporate a ghost canceller in the receiver. The transmitter periodically sends a standardized known waveform known as a training sequence. The receiver knows what this waveform looks like and compares it with the received signal. In theory it is possible for the receiver to compute the delay and relative level of a reflection and so insert an opposing one. In practice if the reflection is strong it may prevent the receiver finding the training sequence.

The most elegant approach is to use a system in which multipath reception conditions cause only a small increase in error rate which the error-correction system can manage. This approach is used in DVB. Figure 7.10(a) shows that when using one carrier with a high bit rate, reflections can easily be delayed by one or more bit periods, causing interference between the bits. Figure 7.10(b) shows that instead, OFDM

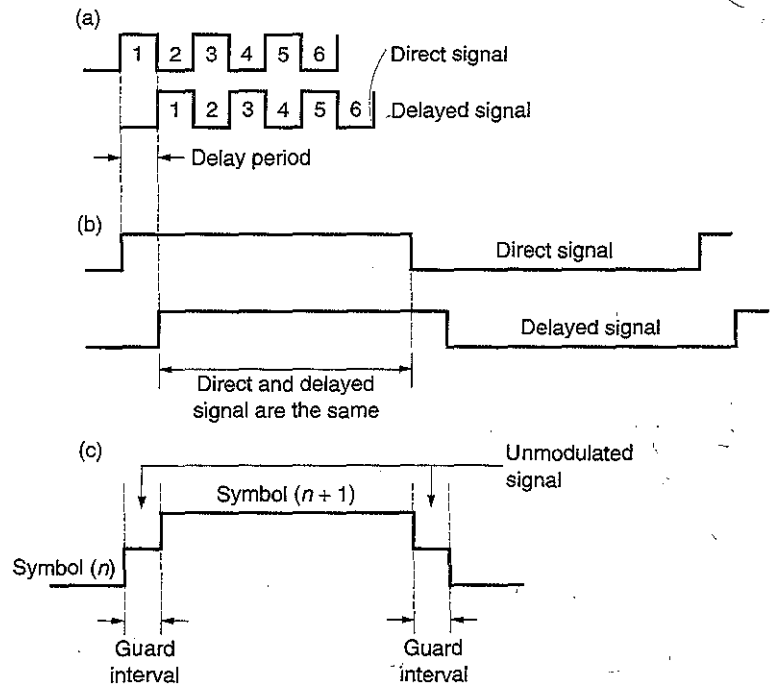


Figure 7.10 (a) High bit rate transmissions are prone to corruption due to reflections. (b) If the bit rate is reduced the effect of reflections is eliminated, in fact reflected energy can be used. (c) Guard intervals may be inserted between symbols.

sends many carriers each having a low bit rate. When a low bit rate is used, the energy in the reflection will arrive during the same bit period as the direct signal. Not only is the system immune to multipath reflections, but the energy in the reflections can actually be used. This characteristic can be enhanced by using guard intervals shown in Figure 7.10(c). These reduce multipath bit overlap even more.

Note that OFDM is not a modulation scheme, and each of the carriers used in a OFDM system still needs to be modulated using any of the digital coding schemes described above. What OFDM does is to provide an efficient way of packing many carriers close together without mutual interference.

A serial data waveform basically contains a train of rectangular pulses. The transform of a rectangle is the function  $\sin x/x$  and so the baseband pulse train has a  $\sin x/x$  spectrum. When this waveform is used to modulate a carrier the result is a symmetrical  $\sin x/x$  spectrum centred on the carrier frequency. Figure 7.11(a) shows that nulls in the spectrum appear spaced at multiples of the bit rate away from the carrier.

Further carriers can be placed at spacings such that each is centred at the nulls of the others as is shown in Figure 7.11(b). The distance between

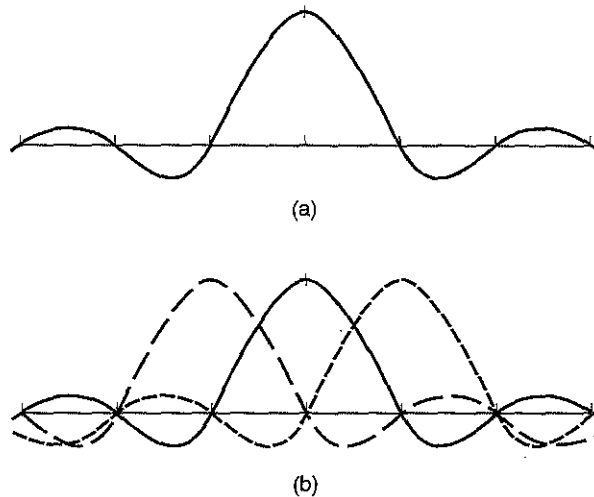


Figure 7.11 In OFDM the carrier spacing is critical, but when correct the carriers become independent and most efficient use is made of the spectrum. (a) Spectrum of bitstream has regular nulls. (b) Peak of one carrier occurs at null of another.

the carriers is equal to  $90^\circ$  or one quadrant of  $\sin x$ . Owing to the quadrant spacing, these carriers are mutually orthogonal, hence the term orthogonal frequency division. A large number of such carriers (in practice several thousand) will be interleaved to produce an overall spectrum which is almost rectangular and which fills the available transmission channel.

When guard intervals are used, the carrier returns to an unmodulated state between bits for a period which is greater than the period of the reflections. Then the reflections from one transmitted bit decay during the guard interval before the next bit is transmitted. The use of guard intervals reduces the bit rate of the carrier because for some of the time it is radiating carrier, not data. A typical reduction is to around 80 per cent of the capacity without guard intervals.

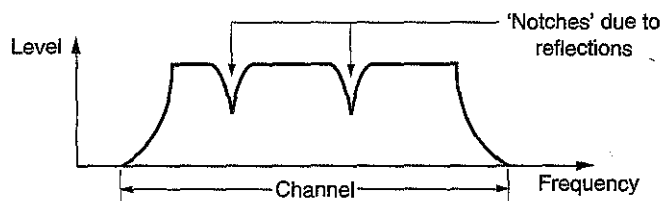
This capacity reduction does, however, improve the error statistics dramatically, such that much less redundancy is required in the error-correction system. Thus the effective transmission rate is improved. The use of guard intervals also moves more energy from the sidebands back to the carrier. The frequency spectrum of a set of carriers is no longer perfectly flat but contains a small peak at the centre of each carrier.

The ability to work in the presence of multipath cancellation is one of the great strengths of OFDM. In DVB, more than 2000 carriers are used in single transmitter systems. Provided there is exact synchronism, several transmitters can radiate exactly the same signal so that a single-frequency network can be created throughout a whole country. SFNs require a variation on OFDM which uses over 8000 carriers.

With OFDM, directional antennae are not needed and, given sufficient field strength, mobile reception is perfectly feasible. Of course, directional antennae may still be used to boost the received signal outside of normal service areas or to enable the use of low-powered transmitters.

An OFDM receiver must perform fast Fourier transforms (FFTs) on the whole band at the symbol rate of one of the carriers. The amplitude and/or phase of the carrier at a given frequency effectively reflects the state of the transmitted symbol at that time slot and so the FFT partially demodulates as well.

In order to assist with tuning in, the OFDM spectrum contains pilot signals. These are individual carriers which are transmitted with slightly more power than the remainder. The pilot carriers are spaced apart through the whole channel at agreed frequencies which form part of the transmission standard. Practical reception conditions, including multipath reception, will cause a significant variation in the received spectrum and some equalization will be needed. Figure 7.12 shows what the



**Figure 7.12** Multipath reception can place notches in the channel spectrum. This will require equalization at the receiver.

possible spectrum looks like in the presence of a powerful reflection. The signal has almost been cancelled at certain frequencies. However, the FFT performed in the receiver is effectively a spectral analysis of the signal and so the receiver computes for free the received spectrum. As in a flat spectrum the peak magnitude of all the coefficients would be the same (apart from the pilots), equalization is easily performed by multiplying the coefficients by suitable constants until this characteristic is obtained.

Although the use of transform-based receivers appears complex, when it is considered that such an approach simultaneously allows effective equalization the complexity is not significantly higher than that of a conventional receiver which needs a separate spectral analysis system just for equalization purposes. The only drawback of OFDM is that the transmitter must be highly linear to prevent intermodulation between the carriers. This is readily achieved in terrestrial transmitters by derating the transmitter so that it runs at a lower power than it would in analog service.

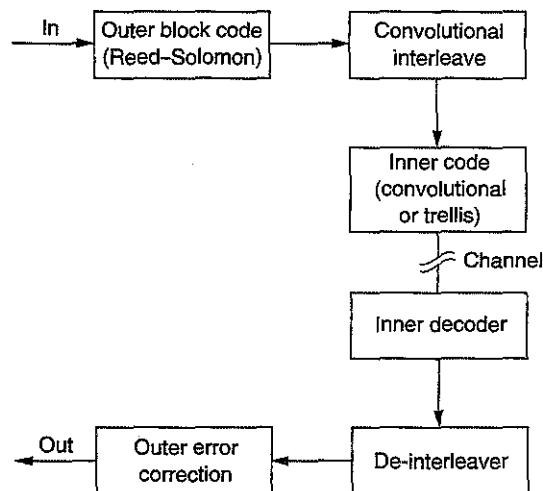


Figure 7.13 Error-correcting strategy of digital television broadcasting systems.

This is not practicable in satellite transmitters which are optimized for efficiency, so OFDM is not really suitable for satellite use.

Broadcast data suffer from both random and burst errors and the error-correction strategies of digital television broadcasting have to reflect that. Figure 7.13 shows a typical system in which inner and outer codes are employed. The Reed-Solomon codes are universally used for burst-correcting outer codes, along with an interleave which will be convolutional rather than the block-based interleave used in recording media. The inner codes will not be R-S, as more suitable codes exist for the statistical conditions prevalent in broadcasting. DVB uses a parity based variable-rate system in which the amount of redundancy can be adjusted according to reception conditions.

Figure 7.14 shows a block diagram of a DVB-T (terrestrial) transmitter. Incoming transport stream packets of 188 bytes each are first subject to

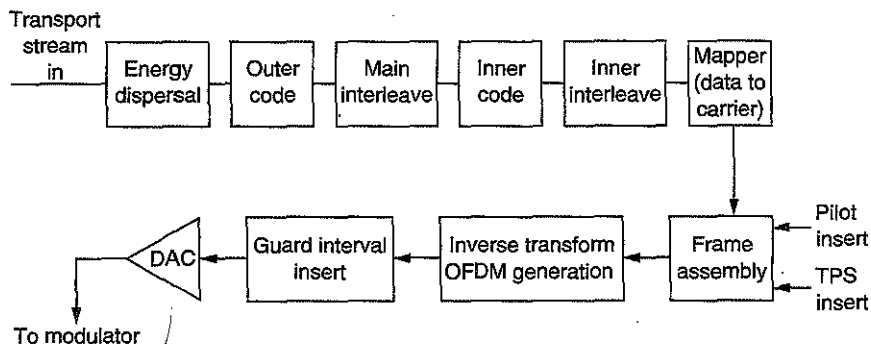
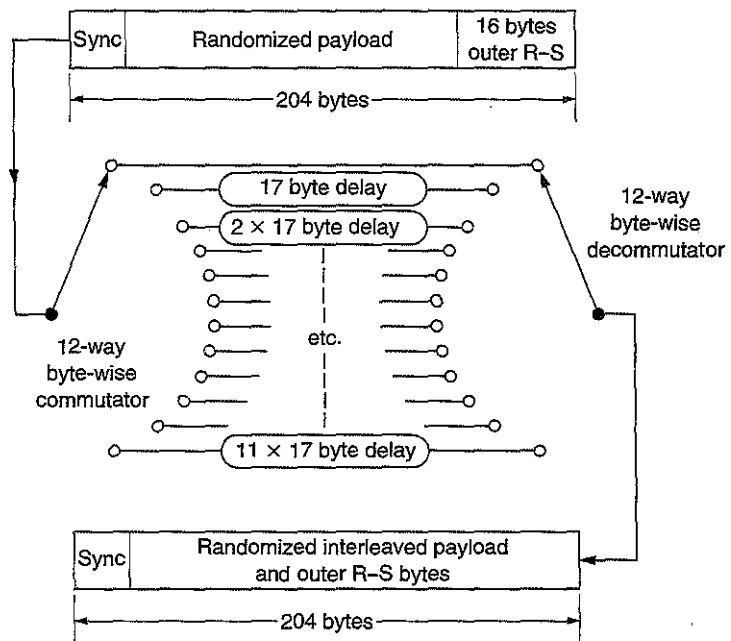


Figure 7.14 DVB-T transmitter block diagram. See text for details.



**Figure 7.15** The interleaver of DVB uses 12 incrementing delay channels to reorder the data. The sync byte passes through the undelayed channel and so is still at the head of the packet after interleave. However, the packet now contains non-adjacent bytes from 12 different packets.

R-S outer coding. This adds 16 bytes of redundancy to each packet, resulting in 204 bytes. Outer coding is followed by interleaving. The interleave mechanism is shown in Figure 7.15. Outer code blocks are commutated on a byte basis into twelve parallel channels. Each channel contains a different amount of delay, typically achieved by a ring-buffer RAM. The delays are integer multiples of 17 bytes, designed to skew the data by one outer block ( $12 \times 17 = 204$ ). Following the delays, a commutator reassembles interleaved outer blocks. These have 204 bytes as before, but the effect of the interleave is that adjacent bytes in the input are 17 bytes apart in the output. Each output block contains data from twelve input blocks making the data resistant to burst errors.

Following the interleave, the energy-dispersal process takes place. The pseudo-random sequence runs over eight outer blocks and is synchronized by inverting the transport stream packet sync symbol in every eighth block. The packet sync symbols are not randomized.

The inner coding process of DVB is convolutional. The output bit rate of the inner coder is twice the input bit rate. Under the worst reception conditions, this 100 per cent redundancy offers the most powerful correction with the penalty that a low data rate is delivered. However, a

