

# Circuits and Systems Tutorials

*Editor*

**Chris Toumazou**

*Assistant Editors*

**Nick Battersby**

**Sonia Porta**



**IEEE  
PRESS**



IEEE Press  
445 Hoes Lane, P.O. Box 1331  
Piscataway, NJ 08855-1331

Editorial Board  
John B. Anderson, *Editor in Chief*

R. S. Blicq	S. V. Kartalopoulos	R. S. Muller
M. Eden	P. Laplante	I. Peden
R. Herrick	A. J. Laub	E. Sánchez-Sinencio
G. F. Hoffnagle	J. M. F. Moura	D. J. Wells
R. F. Hoyt		

Dudley R. Kay, *Director of Book Publishing*  
Carrie Briggs, *Administrative Assistant*  
Lisa S. Mizrahi, *Review and Publicity Coordinator*  
Denise Phillip, *Book Production Assistant*

# Circuits and Systems Tutorials

Editor  
**Chris Toumazou**  
*Imperial College  
London, U. K.*

Assistant Editors  
**Nick Battersby**  
*Northern Telecom  
Ontario, Canada*

**Sonia Porta**  
*Oxford Brookes University  
Oxford, U. K.*



The Institute of Electrical and Electronics Engineers, Inc., New York

This book may be purchased at a discount from the publisher when ordered in bulk quantities. For more information contact:

IEEE Press Marketing  
Attn: Special Sales  
445 Hoes Lane, P.O. Box 1331  
Piscataway, NJ 08855-1331  
Fax: (908) 981-9334

©1996 by the Institute of Electrical and Electronics Engineers, Inc.  
345 East 47th Street, New York, NY 10017-2394

Originally produced by LTP Electronics Ltd.,  
Oxford UK on behalf of IEEE ISCAS '94.

*All rights reserved. No part of this book may be reproduced in any form,  
nor may it be stored in a retrieval system or transmitted in any form,  
without written permission from the publisher.*

Printed in the United States of America

TK3226  
. I1142  
1994

10 9 8 7 6 5 4 3 2 1

**ISBN 0-7803-1170-1**  
**IEEE Order Number: PP5596**

**Library of Congress Cataloging-in-Publication Data**

Circuits and systems tutorials / editor, Chris Toumazou : assistant  
editors, Nick Battersby & Sonia Porta.

p. cm.

"IEEE order number PP5596"—T.p. verso.

Based on tutorials presented at the 1994 IEEE International  
Symposium on Circuits and Systems held in London.

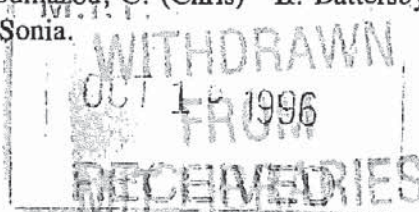
Includes bibliographical references and index.

ISBN 0-7803-1170-1 (pbk.)

1. Signal processing—Congresses. 2. Electronic circuits—  
Congresses. 3. Telecommunication systems—Congresses. 4. Electric  
filters—Congresses. I. Toumazou, C. (Chris) II. Battersby, N. C.  
(Nicholas C.) III. Porta, Sonia.

TK5102.9.C49 1996

621.382—dc20



95-42332  
CIP

## MPEG-2: The Main Profile

John Morris

---

### 3.11.1 MPEG-2 - What is it?

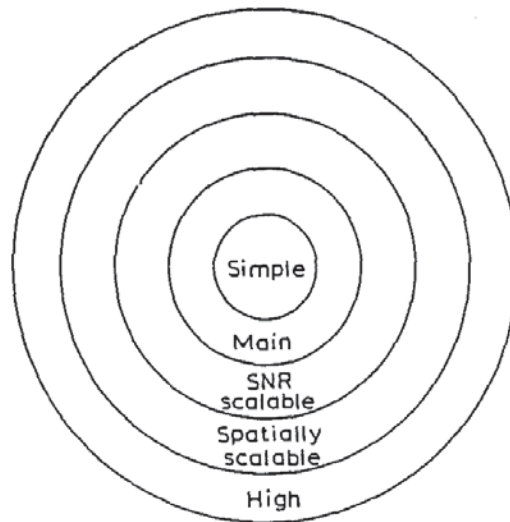
The rapid development of digital technology in the 1980's made compression of digital video signals a practical possibility. This opened up the possibility of a whole range of new services and products. To facilitate these new services, standardization has become a high priority to ensure that interworking between products is possible and to encourage investment in standard technologies so that the technology becomes cost-effective for consumer applications.

The Moving Pictures Expert Group (MPEG) is a working group within ISO/IEC that defines standards for digital compression of audio-visual signals. MPEG's standards are generic, meaning that they are capable of being used by a wide range of applications, covering storage-based multimedia systems, television broadcasting, and communications applications such as Video-on-Demand and face-to-face communication.

MPEG's first project was MPEG-1, published in 1993 as ISO/IEC 11172 [1,2]. It is a three part standard defining audio and video compression coding method and a multiplexing system for interleaving audio and video data so that it can be played back in close synchronization. It has been applied in the CD-i system and Video-CD for publishing full-screen motion video on CD-ROM. It also forms the basis of a number of field trials for Video-on-Demand services, and was used in one of the demonstrations of HDTV systems in the USA. Although MPEG-1 has found many applications, it principally supports video coding at bit-rates only up to about 1.5 Mbit/s using a non-interlaced format. It is not optimal for interlaced signals. During 1990, therefore, MPEG recognized the need for a second, related standard for coding video at higher data rates in an interlaced format (the essential technical difference is that MPEG-1 is designed for coding progressive pictures whereas MPEG-2 is optimized for interlaced pictures - MPEG-1 bit-streams are decodable on MPEG-2 decoders). The MPEG-2 standard [3] is capable of coding standard definition television at bit-rates from about 4-9 Mbit/s and high definition television at 15-25 Mbit/s.

Because MPEG has taken a generic approach to standardization, its standards provide techniques to support different applications with differing requirements. Analyzing these differing requirements led MPEG to realize that it is not possible to provide a single, unique method for all the different problems. Instead it has followed a "tool-kit" approach in which an extensive set of algorithmic "tools" are defined. For instance coding modes are provided both for scalable and non-scalable coding systems. The coding syntax that MPEG has defined provides tools to cover all the different applications it considered, and parameters can be chosen to allow working at different bit-rates and picture sizes and resolutions.





**Figure 1** MPEG-2 profiles.

It is neither cost effective nor an efficient use of bandwidth to support all the features of the standard in all applications. In order to make the standard practically useful and enforce interoperability between different implementations of the standard, MPEG has defined *profiles* and *levels* of the full standard. Roughly speaking, a profile is a sub-set, suitable for a particular application, of the full possible range of algorithmic tools, and a level is a defined range of parameter values (such as picture size for instance) that are reasonable to implement and practically useful. Figure 1 illustrates the relationship between the different profiles that MPEG has adopted. The more complex profiles are super-sets of the less complex ones. Therefore a decoder capable of decoding the scalable profiles will also be able to decode data generated for the Main Profile (for non-hierarchical, non-scalable coding).

This chapter describes only the Main Profile of MPEG. Other papers in the series will consider the scalable options, and methods of providing low-delay performance typically needed for conversational or interactive telecommunication services [4, 5].

### 3.11.2 Schedule and method of working

MPEG began work on MPEG-2 in the summer of 1990. The Main Profile algorithm was frozen in March 1993 so that no further changes would take place, and the whole specification completed in November 1993. Currently it is going through the ISO/IEC approval process of balloting and possible revision. It is intended that this process will be completed by November 1994 and that the final text will be published during 1995. Early implementations of the standard are already available.

The MPEG group worked in three phases. Firstly, requirements of different applications were analyzed. A consequence of these requirements, and the technical implications of the conclusions reached, was MPEG's decision to adopt a tool-kit approach with profiles and levels. Once the requirements were clearer, the detailed definition of the algorithm began with a competitive phase during which many proposals from different laboratories were put through subjective tests aimed at identifying promising techniques. Finally, a convergence phase of collaborative work, doing well-defined experiments on a single *Test Model* (a full description of an encoder/decoder), progressively enhanced the performance of the system as the Test Model was updated based on the experimental results. At the end of this phase the Test Model was an embodiment of an MPEG-2 coding

and decoding system on which consensus had been reached and from which the final specification is derived.

### 3.11.3 Requirements

MPEG-2 aims to be applicable to a wide range of applications and services covering many bit-rates, resolutions, and picture qualities. The applications range from interactive storage-based multimedia systems, to television broadcasting by cable, satellite or terrestrial networks, or personal communication systems. These applications require that a range of features should be supported:

**Picture quality** - The aim was to achieve normal TV picture quality at around 5 Mbit/s and studio quality at around 9 Mbit/s.

**Delay** - For some applications the encoder-decoder end-to-end delay is a very critical parameter. For others it is not at all important. A system with low-delay has a penalty in picture bandwidth efficiency, and so MPEG has introduced optional modes that an encoder can use to support low-delay.

**Random access** - Most applications require that it should be possible to start decoding either at defined points in the stream or a short time after acquiring the channel (channel hopping).

**Error resilience** - All applications need the system to recover if there are residual bit-errors after error correction. In different applications the residual error rate may differ by many orders of magnitude. MPEG has chosen to build in a certain amount of error robustness, and to allow a much higher degree to be used optionally by the encoder.

**Data security** - Many applications will use encryption to control access. MPEG has not defined encryption methods. It has provided methods for transporting information that allows unscrambling of scrambled data. This information is carried in the system multiplex, and is not discussed further in this paper.

**Trick Modes** - Applications that store video, either on tape or on disc, need to be able to play fast forwards or backwards. MPEG has provided optional support for these modes.

### 3.11.4 What is standardized

The Main Profile of the MPEG-2 standard allows encoders and decoders from different manufacturers to operate together. The interface between the two is the compressed bit-stream that represents the coded video. To achieve interoperability it is only necessary to standardize the structure, content and meaning of the bit-stream and the way that it should be decoded to reconstruct the desired picture. There is no need to standardize the encoder. This is the approach that MPEG has taken as it has the advantage of leaving considerable freedom to encoder manufacturers to improve their encoding strategies as more is learnt about encoding, or to address different market segments with different tradeoffs of cost and complexity against picture quality.

### 3.11.5 Compression methods

The full MPEG-2 Video standard can be divided into two broad categories. The non-scalable syntax, which is structured as a super-set of the MPEG-1 standard, forms the Main



Profile. It provides a set of compression methods for coding interlaced video efficiently. The second major category is the scalable coding system, which has the key property of making it possible to recover useful pictures at a reduced quality or sampling rate from a part of the whole bit-stream. This is done by structuring the data in a number of hierarchical layers. This subject will be covered in more detail in another paper [4]. The Main Profile has the advantage that it is less complex to implement and is more efficient than scalable coding. Scalability is nevertheless important for some applications.

### 3.11.5.1 Colour sampling

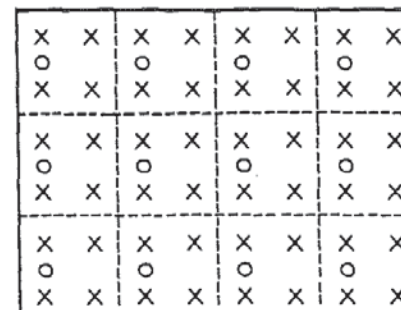
A number of different colour sampling structures are supported by MPEG-2. They are all based on the luminance and two colour difference model of colour representation ( $Y$ ,  $C_r$ ,  $C_b$ ) that is used in television systems. The 4:2:2 and 4:4:4 colour sub-sampling pattern defined in CCIR 601 are both supported [5]. However, the so-called 4:2:0 format (see Figure 2) is expected to be more widely used for consumer applications. 4:2:0 colour sub-sampling reduces the chrominance sampling rate by a factor of 2:1 with respect to the luminance in both the horizontal and vertical directions. This leads to a significant reduction in the size of the frame-stores needed in the decoder.



X Represent luminance pels

O Represent chrominance pels

The position of luminance and chrominance samples. 4:2:2 data.



X Represent luminance pels

O Represent chrominance pels

The position of luminance and chrominance samples. 4:2:0 data.

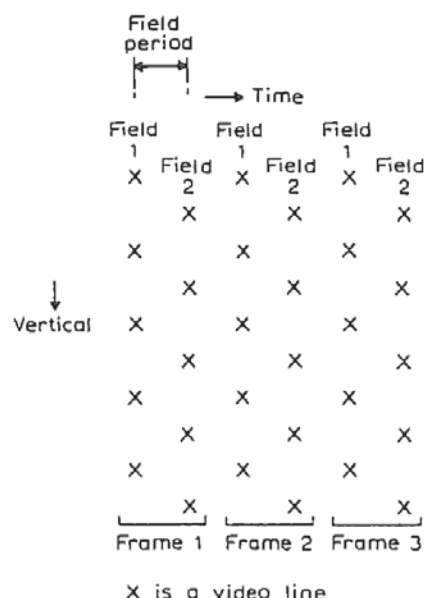
**Figure 2** 4:2:2 and 4:2:0 colour sub-sampling patterns.

### 3.11.5.2 Interlace

The most important difference between MPEG-1 and MPEG-2 is full support for coding interlaced video material. Each frame of interlaced video consists of two fields separated in time by a field period (20 ms in 50 Hz countries and 16.67 ms in 60 Hz countries). The vertical position of the lines in field one fall midway between the lines of field two. Together the two fields sample the image at the full vertical resolution of the frame, though at different time periods (see Figure 3).

On slowly moving video material the difference between the information on each field is small. In this case it is most efficient to combine the two fields into one frame and code the combined frame as one unit. However, when the motion is large, the information on each field is very different and it is more efficient to encode each field independently. For this reason it is possible in MPEG-2 to switch adaptively between field coding and frame coding modes on a frame-by-frame basis. Further, in the frame coding mode it is possible to switch into field coding modes locally within blocks. These coding modes form the basis of the system for highly efficient coding of interlaced video.





**Figure 3** Vertical-temporal position of video data in an interlaced system.

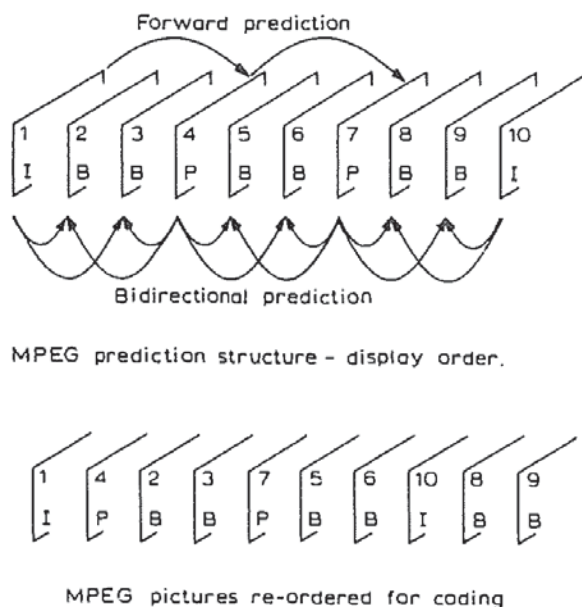
### 3.11.5.3 Overview of the algorithm

MPEG achieves a high compression ratio, while preserving good picture quality and the necessary features of random accessibility and robustness to errors, by reducing the redundancy in the signal and also discarding information that is irrelevant to the observer. It is not a lossless coding scheme - some data is irreversibly discarded during encoding. It is possible to trade-off the subjective picture quality on one hand and the bit-rate on the other. The final bit-rate is a free parameter in MPEG that the encoder is able to select.

In the design of the algorithm it is necessary to balance the conflicting requirements of high compression and random access. High compression means that we need to remove the temporal redundancy in sequences of pictures by using motion compensated prediction from frame to frame. However, the need for random accessibility means that there must be breaks in the sequence of frame-to-frame prediction. Therefore there has to be a careful balance between coding intra-frame information without using any knowledge of previous pictures, and using recursive and non-recursive motion compensated coding.

A number of different compression methods are combined in sequence. The first set of methods uses block based motion compensation. Motion compensation is a process where pictures that have already been coded are combined with data representing motion vectors to form a prediction of the current picture. The encoder searches the previous picture to find a block of data that is the best match of the block that is currently being coded. The offset to this block is coded as a motion vector, and the best matching block itself is used as a prediction of the current block by subtracting the prediction from the sample values of the block. The residual prediction error is then quantized and encoded. At the decoder the process is inverted to recover the picture. In order to capture the different levels of redundancy that can be present in video signals MPEG has resorted to several different prediction modes. The basic block size for motion compensation is 16 pixels by 16 lines. These are known as macroblocks. Three main picture types have been defined. Intraframe coded pictures (I-pictures) are coded without reference to any previous pictures and are provided to allow access to the sequence of coded pictures at the cost of a reduction in the coding efficiency. Two types of motion compensated pictures are defined - predictive pictures (P-pictures) and bi-directional pictures (B-pictures). P-pictures use causal

prediction from previously decoded pictures. Both P-pictures and previous I-pictures may be used as the basis for prediction of P-pictures. Thus P-pictures use recursive prediction. B-pictures on the other hand use non-causal prediction from pictures that will be displayed in either the past or the future and never form the basis of further predictions. To use bi-directional prediction the pictures must be re-ordered so that they are transmitted in a different order from the order they are displayed in (see Figure 4).



**Figure 4** MPEG picture reordering.

As discussed previously, pictures may consist either of single fields or single frames (the two fields combined together). If the picture is a frame picture then the encoder may choose to form the prediction by using a vector for the entire block, or by using separate vectors for each field. Thus in B-pictures, where the motion compensation may be made from both past and future pictures in either field or frame mode, it may be necessary to transmit information for up to four motion vectors. It is important therefore to encode the motion vector information as efficiently as possible.

Motion compensation generates a prediction error signal that has eliminated most of the temporal redundancy. The error signal is then encoded by a set of techniques to reduce the spatial redundancy of the information. Whereas motion compensation is based on a 16 pixel by 16 line macroblock, spatial redundancy reduction is based on an 8 by 8 block structure. The data in the 8 by 8 blocks are transformed by the discrete cosine transform (DCT), quantized and then variable length coded. This coded data is then multiplexed with the side information - the prediction modes, motion vectors and other necessary information - to give the compressed bit-stream that is transmitted or stored.

A decoder reads the compressed bit-stream, separates out the different types of data, uses the mode and motion vector information to create a prediction of the current frame from previously decoded pictures, and combines this prediction with the information it recovers by decoding the compressed DCT data to form the final picture.

#### 3.11.5.4 Motion compensation

Of the three picture types defined by MPEG-2, I-pictures do not use motion compensation. P-pictures use causal prediction from previous pictures in a recursive manner. B-pictures



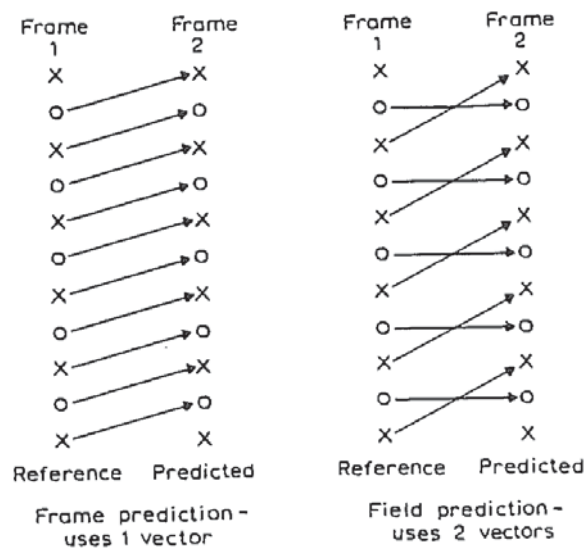
use non-recursive, non-causal prediction from pictures that have been decoded but will be displayed both before and after the current picture.

I-pictures may be used to break the recursive sequence of pictures, and so can be used to enable random access. Both P- and B-pictures are coded using motion compensation techniques. B-pictures are more efficiently coded than P-pictures for three main reasons. The bi-directional motion compensation improves prediction of areas of the picture uncovered by the motion; the averaging of data from two pictures has some noise reduction properties; and because the prediction of B-pictures is not recursive, it is possible to reduce the quality of the B-pictures relative to P-pictures, saving bits, and hence increasing the overall quality of other pictures, leading to a net increase in the quality of the sequence. Within these three picture types a range of coding modes are defined.

### (i) Prediction modes

In the case of I-pictures and intra coded macroblocks, no prediction is required. However intra coded macroblocks may carry motion vector information. In normal circumstances it is not used. Its function is to provide a means of concealing decoding errors when data errors in the bit-stream make it impossible to decode the data correctly. This is one of the optional error concealment strategies that has been defined.

All pictures may either be structured as fields or frames. In field structure, predictions are made independently for each field of data from one or more of the previously decoded fields. In frame structure, the predictions are made either from one or more previously decoded frames, or fields. The fields and frames that are used to construct the predictions may themselves have been decoded either as field structured pictures or frame structured pictures (Figure 5).



**Figure 5** Frame prediction.

Generally for image sequences where the motion is slow, frame-prediction coding is more efficient. However, when motion speed increases, field structured coding becomes more attractive. Field structure also can be used to reduce the intrinsic delay of the system because in this case the basic period of operation of the system is a number of field periods rather than frame periods.

In addition to the two basic modes of frame prediction and field prediction, two further modes have been defined.

16 by 8 motion compensation uses two motion vectors for each macroblock. The first vector is used for the upper 16 by 8 region of the macroblock and the second vector for the lower half. In the case of B-pictures this may lead to the use of a total of four vectors for coding a macroblock (since there may be two references each to both past and future pictures). This mode is only permitted in field structured pictures. It is provided in field structured pictures so that the spatial area referred to by a vector is comparable to the area that a 16 by 16 block covers in a frame picture (fields have half the vertical sampling density of frames, therefore 8-field lines cover the same picture area as 16-frame lines). This mode is defined to improve the efficiency of field structured coding.

The so-called *dual-prime* mode is only permitted in P-pictures, though it may be used for both field and frame structured coding. For both field and frame structured pictures one vector is transmitted in the bit-stream, together with a small differential vector. In the case of field pictures, two motion vectors are derived from this data and are used to form two predictions from two reference fields (one of even and one of odd parity). These two predictions are averaged to form the final prediction. In the case of frame pictures this process is repeated for each of the two fields. Each field is predicted separately. This gives rise to a total of four field predictions that are combined to form the final predictions (see Figure 6). Dual prime prediction is relatively complex but has been found to be nearly as efficient as the bi-directional coding modes. Since it refers only to past pictures it is possible to use this mode to achieve nearly the same quality as can be achieved with B-pictures without the need for re-ordering the sequence and therefore saving a frame store in the decoder, possibly saving memory, and reducing the intrinsic system delay.

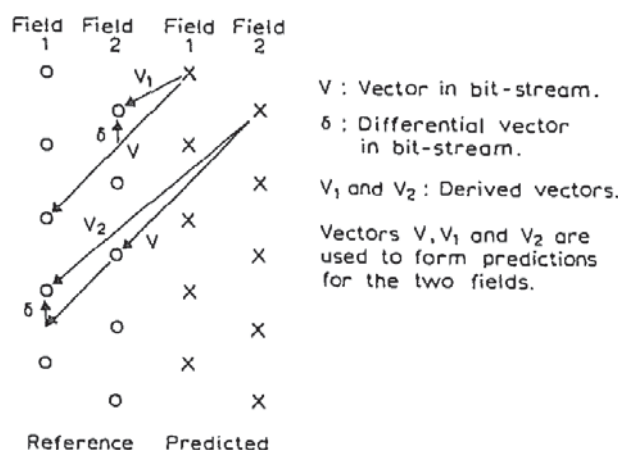


Figure 6 Dual prime predictions.

## (ii) Motion estimation

Motion vectors are explicitly coded and transmitted to the decoder in the data. Therefore there is no need to standardize the methods used to find the vectors. MPEG simply defines how to code the vectors and how the decoder should interpret them. It is the responsibility of the encoder to choose the motion vectors within the constraints defined in the standard. Block matching techniques and various hierarchical schemes have been widely used. They are computationally expensive, and it is important to search a large enough range to capture the motion in the scene. The complexity of motion vector estimation is one of the major reasons why encoders are significantly more complex than decoders.



### (iii) Motion representation

The basic coding unit for representing motion and other side information is the 16 by 16 macroblock. The choice of 16 by 16 is a trade-off between the need to limit the amount of overhead data in the stream and the desire to have a small basic unit for motion representation and to allow local control of the adaptation of the algorithm.

All motion vectors are coded differentially with respect to previously coded vectors. A decoder therefore has to maintain a set of predictors that record the last value of the appropriate motion vector that was decoded. Each vector is represented as the combination of the prediction and an update value. The update value is coded by variable length code for efficiency. One of the most important requirements for the choice of motion vectors is that their range is large enough to cover the actual motion present in the scene. Again, it is the task of the encoder to search a wide enough area to find good vectors.

#### 3.11.5.5 Spatial coding

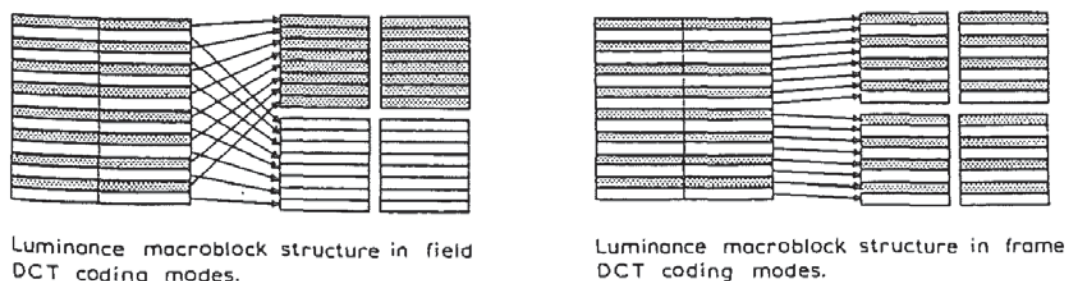
I-pictures and the residual pictures in P- and B-pictures after motion compensated prediction contain a lot of redundancy. The methods for removing this redundancy are block-based and use the DCT (discrete cosine transform).

#### (i) DCT

The DCT is an orthogonal transform which is exactly invertible and has a fast algorithm. Experiments have shown that for natural pictures it is a close approximation to the optimal transform that compacts the energy present in pictures in the most efficient way. It has been widely used for image compression. When the DCT is applied to pictures it is found that nearly all the energy is concentrated in a small number of low frequency coefficients.

The choice of a block size of 8 by 8 provides a trade-off between the need to use a large area for the transform so that the energy compaction is most efficient with the fact that the characteristics of a picture are spatially varying, which tends to suggest that a smaller block size should be used.

In frame structured coding MPEG-2 gives the encoder a choice of DCT modes. Data from the two fields may be combined into frames, so that the 8-lines of data in a DCT block are drawn from both fields before the DCT is performed, or the data from the two fields may be kept separate (see Figure 7). The choice depends on the speed of motion and how effective the prediction has been.



**Figure 7 Luminance macroblock structures.**

#### (ii) Quantization

Quantization of the DCT coefficients is an essential operation where much of the bit-rate reduction is achieved. The DCT concentrates the energy but does not in itself reduce the

amount of data needed to represent the picture. The amount of data needed is only reduced when the accuracy of the representation of the DCT coefficients is limited and the resulting values encoded. After quantization the number of non-zero DCT coefficients is usually greatly reduced.

Experiments have shown that the subjective perception of quantization errors introduced into the coefficients of block DCT data depends strongly on the frequency of the coefficient being quantized. It is necessary to quantize the low frequency coefficients accurately, but the high frequency ones can be quantized with greater errors because these errors are much less visible to the human visual system. MPEG therefore provides *weighting matrices* that allow the encoder to control the relative accuracy of the quantization of the different coefficients. Because I-pictures have signal energy at most frequencies and motion compensated pictures (P- and B-pictures) tend to have more high frequency energy, different weighting matrices can be used for different picture types. In all cases the encoder can choose appropriate matrices and download the values to the decoder as part of the compressed bit-stream.

The quantization levels are uniformly spaced except in the case of motion compensated pictures around zero, where there is a dead-zone (see Figure 8). The effect of the dead-zone is to suppress small deviations around zero. The deviations are generally caused by noise in the signal and so suppressing these values improves the overall picture quality.

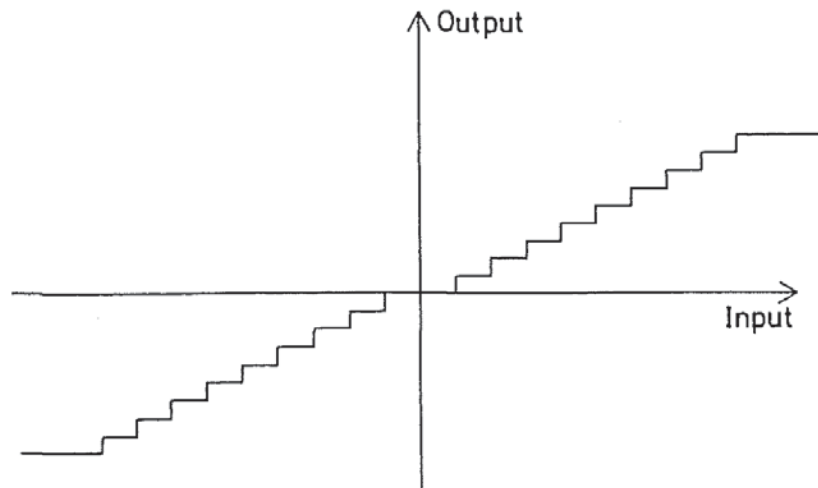


Figure 8 Quantization with a dead zone.

### (iii) Run length coding and VLC coding

After quantization the 8 by 8 blocks are scanned in one of two possible zig-zag orders to turn the two-dimensional array into a linear string of quantized coefficients. Two scan orders are provided because of the differences in video material that is observed. Some material contains a lot of vertical frequencies, some of it generated artificially by the interlace picture structure. For this material there is an advantage in biasing the scan order to prefer the vertical direction. For other material it is better not to include this bias. The encoder chooses the scan order and signals its choice to the decoder in the bit-stream.

The string of coefficients is run-length coded by counting the number of zero coefficients preceding each non-zero coefficient. The combination of the run of zeros and the value of the non-zero coefficient is encoded using a table of variable length codes. Common events use a short code-word and less frequent events use longer code-words. Very unlikely events are coded by the combination of an escape length code and a fixed length code to indicate the event (see Figure 9).



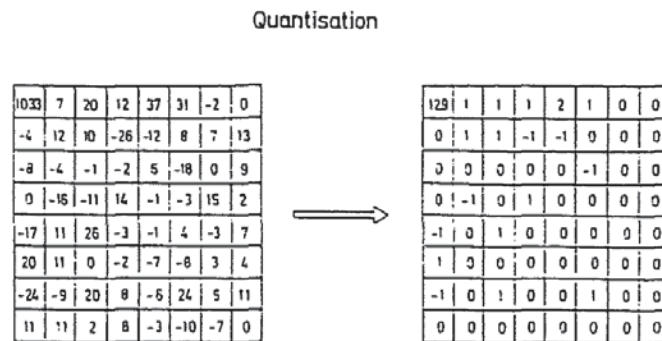
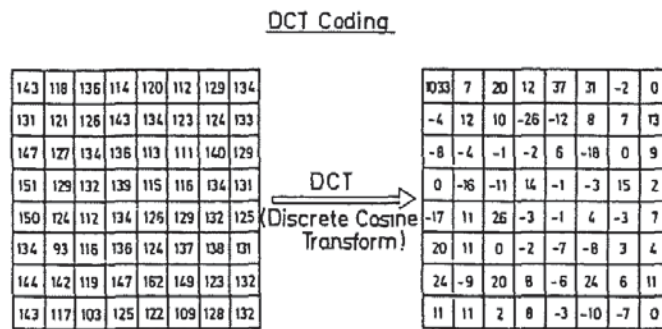
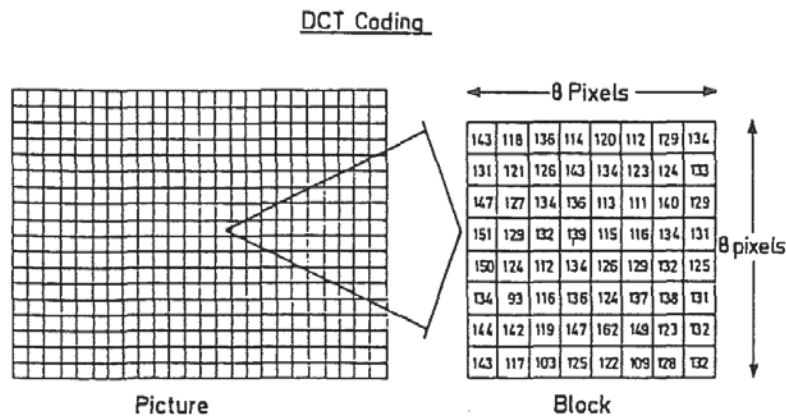


Figure 9 DCT coding and quantization.

*(iv) Constant and variable bit-rate coding and buffering*

The MPEG-2 system may be used for applications using either a constant bit-rate or variable bit-rate. The coding algorithm produces a variable number of bits for each picture. The number varies because the length of the run-length code depends on how many non-zero coefficients are produced, and the variable length code generates a different number of bits depending on the values of the non-zero coefficients and the lengths of the runs. In effect

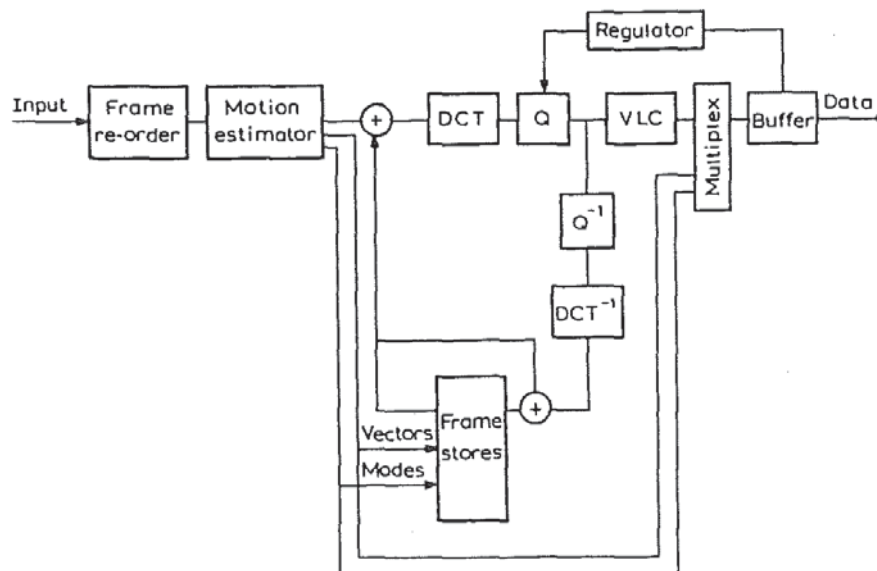
the number of bits generated per picture depends on the complexity and motion in the scene. Simple scenes with regular motions and little detail will produce very few bits. Scenes with many objects, at a high resolution with complex motion will generate many more bits.

If the system is used on a device or network operating at a constant bit-rate, the varying rate produced by the encoder must be interfaced to the fixed bit-rate of the network by a suitably sized FIFO buffer to smooth out the variations in rate. The buffer is filled by the encoder as it generates data and emptied by the channel at a constant rate. The encoder monitors the status of this buffer and uses feedback to control the number of bits generated by the encoder to prevent overflow or underflow of the buffer. There are a number of parameters that an encoder can use to control the bit-rate. Quantization for example gives a direct method that is frequently used. To achieve a good picture quality it is important to combine the feedback from the FIFO buffer with information derived from analyzing the pictures being coded.

Variable bit-rate systems allow a degree of variation about an average bit-rate up to some defined peak. This makes it possible to increase the bit-rate above its average value during sections of the video that are hard to encode for a period of time, and to compensate for this by reducing the bit-rate below its average value during easy sections. The aim is to achieve a more constant picture quality. Furthermore if the medium itself runs at a varying rate then the buffer store can be reduced and so the total delay between the input to the encoder and the output from decoder may be reduced.

### 3.11.6 Encoding

MPEG does not specify an encoding method. It only specifies the syntax and semantics of the bit-streams and how to decode it. Therefore many options are left to the encoder designer. Figure 10 gives a block diagram of the main processing blocks that will be present.



**Figure 10** Simplified block diagram of an MPEG video encoder.

The input signal must first be digitized and converted to the desired colour space ( $Y, C_r, C_b$ ) and colour sub-sampling ratio (4:2:0 for the Main Profile). It may improve the picture quality if some prefiltering is applied to the signal at this point.



The encoder chooses a picture type for each frame and whether it is to be coded as a field-structured picture or a frame-structured picture. Having defined this basic structure it has to estimate motion vectors for each 16 by 16 macroblock. The number of vectors depends on the picture type and the coding mode that will be chosen for each block. If B-pictures are being used the pictures must be re-ordered.

The basic coding unit in a picture is the macroblock. Each macroblock in the picture is coded in sequence from left to right and top to bottom. The macroblock consists of four 8 by 8 blocks of luminance data and blocks of chrominance data covering the same picture area. In the case of the 4:2:0 colour sub-sampling used for the Main Profile there are two chrominance blocks in a macroblock.

For each macroblock the coding mode is chosen. The choices available depend on the picture type, the picture structure and the effectiveness of the different prediction modes. Having chosen the coding mode, the motion compensated prediction of the contents of the block is made from past and/or future reference pictures. The prediction is subtracted from the actual data in the macroblock to form a prediction error signal. In a frame structured picture the encoder now has to choose whether to do a frame-based DCT or a field-based DCT. This mainly depends on how much motion there is between the two fields and how well the motion compensation has managed to predict it. Depending on this decision, the prediction error signal may have to be re-ordered. The error signal is then separated into 8 by 8 blocks and a DCT is applied to each block. Each resulting block of coefficients is quantized and scanned to form a one-dimensional string of coefficients. The side information needed by the decoder to reconstruct the block (coding mode, motion vectors etc.) is encoded and then the quantized coefficients are run-length using a variable length code table.

The rate control procedures monitor the state of the output FIFO that interfaces the encoder to the channel. This information is used to control the number of bits that the encoder will generate for subsequent blocks and pictures.

At this stage in the process a complete bit-stream has been generated that can be used by a decoder to recover the picture. The encoder has one further task to perform. It needs to store the I- and P-pictures that the decoder will recover so that it can use them as reference pictures for further motion compensated prediction. Therefore the quantized DCT coefficients are dequantized and passed through an inverse DCT transform. The prediction error signal is then added back to the prediction signal and limited to the required range to give the reconstructed reference picture.

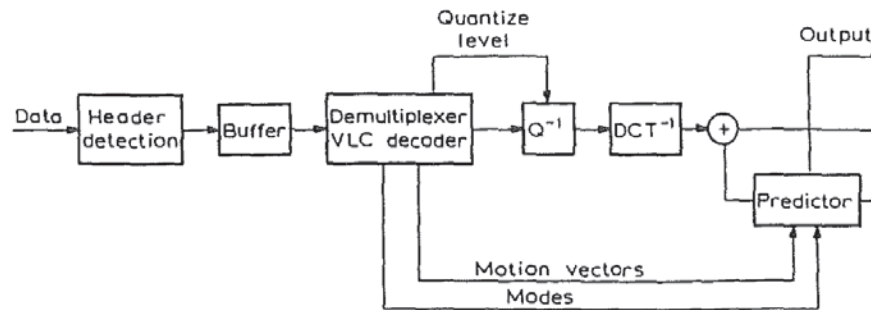
### 3.11.7 Decoding

Decoding is the inverse operation. It is much simpler than encoding as there is no need to do any motion estimation and there are many fewer options. Figure 11 illustrates a block diagram of a basic system.

The input buffer is filled from the channel. The decoder reads the bit-stream until it finds the start of a coded picture and the type and structure of the picture. It decodes each macroblock in turn, using the macroblock type and motion vectors to construct a prediction of the block based on the reference pictures that are stored in the decoder. The coefficient data are read, decoded and dequantized. Each 8 by 8 block of data is transformed by the inverse DCT and the result is added to the prediction.

Once all the macroblocks have been processed, the picture has been reconstructed. If it is an I-picture or a P-picture it may be reference picture for future predictions and so must be stored, replacing the older reference picture stored in the decoder (only two reference

picture need to be stored at any one time). The final stage is to reorder the pictures into the correct sequence before any post-processing and display.



**Figure 11** Block diagram of a decoder.

### 3.11.8 Conclusion

The MPEG-1 standard is becoming established as the method for publishing digital video on disc-based multimedia systems. Several products are now available using the principal capabilities of MPEG-1.

The MPEG-2 standard is now finalized in draft form and the path to publication as a full International Standard is well defined. It is expected that final approval will be obtained at the end of 1994.

Chip-sets and products based on MPEG-2 are now being developed, and early versions are already available. Trial services are being planned, and in some cases are already being started. It is expected that MPEG-2 will have a major impact on industry. The amount of compression that can be achieved means that bandwidth can be used very much more efficiently. This will lead to a huge increase in the number of TV channels available using conventional means of distribution, will allow new industries to enter the TV market offering new services, and open up the possibility of starting a digital HDTV service.

### References

- [1] ISO/IEC 11172, "Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s".
- [2] D. Le Gall, "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, Vol. 34, No. 4, April 1991.
- [3] ISO/IEC CD 13818, "Generic coding of moving pictures and associated audio," (MPEG-2).
- [4] N. Wells and P. Tudor, "Standardization of scalable coding schemes", in *Proc. IEEE ISCAS 94 Tutorials*, May 1994.
- [5] G. Bjøntegaard, "MPEG-2 coding schemes for low delay requirements", in *Proc. ISCAS 94 Tutorials*, May 1994.
- [6] CCIR Recommendation 601-2, "Encoding parameters of digital television for studios".