# MPEG Encoding Basics

SNELL & WILCOX

# MPEG basics

A compression encoder works by identifying the useful part of a signal which is called the entropy and sending this to the decoder. The remainder of the signal is called the redundancy because it can be worked out at the decoder from what is sent. Video compression relies on two basic assumptions. The first is that human sensitivity to noise in the picture is highly dependent on the frequency of the noise. The second is that even in moving pictures there is a great deal of commonality between one picture and the next. Data can be conserved both by raising the noise level where it is less visible and by sending only the difference between one picture and the next.

In a typical picture, large objects result in low spatial frequencies whereas small objects result in high spatial frequencies. Human vision detects noise at low spatial frequencies much more readily than at high frequencies. The phenomenon of large-area flicker is an example of this. Spatial frequency analysis also reveals that in many areas of the picture, only a few frequencies dominate and the remainder are largely absent.

For example if the picture contains a large, plain object, high frequencies will only be present at the edges. In the body of a plain object, high spatial frequencies are absent and need not be transmitted at all.

In MPEG, two-dimensional spatial frequency analysis is performed using the Discrete Cosine Transform (DCT). An array of pixels, typically 8 x 8, is converted into an array of coefficients. The magnitude of each coefficient represents the amount of a particular spatial frequency which is present. Fig.1 shows that in the resulting coefficient block, the coefficient in the top left corner represents the DC component or average brightness of the pixel block. Moving to the right the coefficients represent increasing horizontal spatial frequency. Moving down, the coefficients represent increasing vertical spatial frequency. The coefficient in the bottom right hand corner represents the highest diagonal frequency.
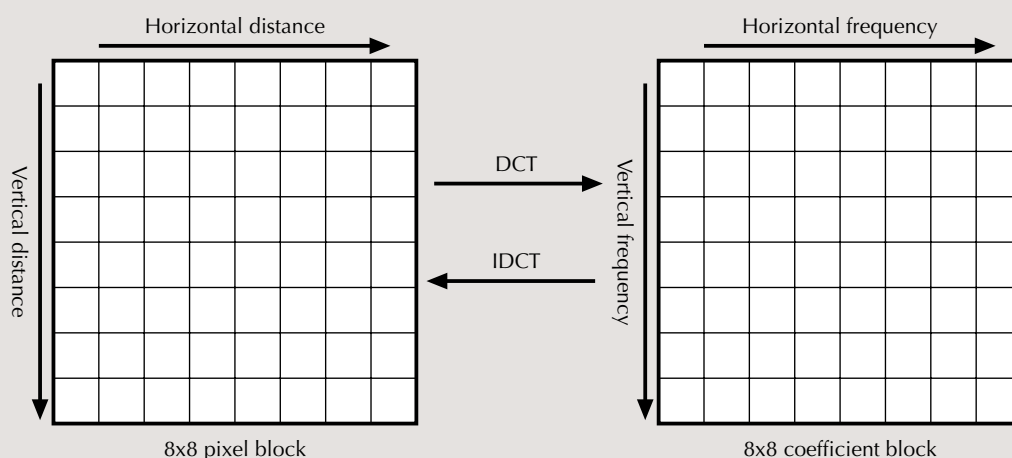


**Figure 1**

In real program material, many of these coefficients will have negligible or zero value yielding enough compression for some purposes. Further compression is obtained by shortening or truncating the wordlength of the remaining coefficients, reducing their resolution and raising noise. If this noise is to be produced in a way which minimizes its visibility, it must vary with spatial frequency. Prior to truncation, the coefficients are weighted, or multiplied by scale factors which are a function of their spatial frequency. At the decoder an equal but opposite weighting process is needed. This multiplies higher spatial frequency coefficients by larger factors, raising the high frequency noise which is less visible without raising low frequency noise.

After weighting, the large value coefficients are mostly found in the top left corner; the remainder of the coefficients are often negligible or zero. It is an advantage to transmit the coefficients in a zig-zag sequence starting from the top left corner. When this is done, the non-zero coefficients are typically transmitted first.
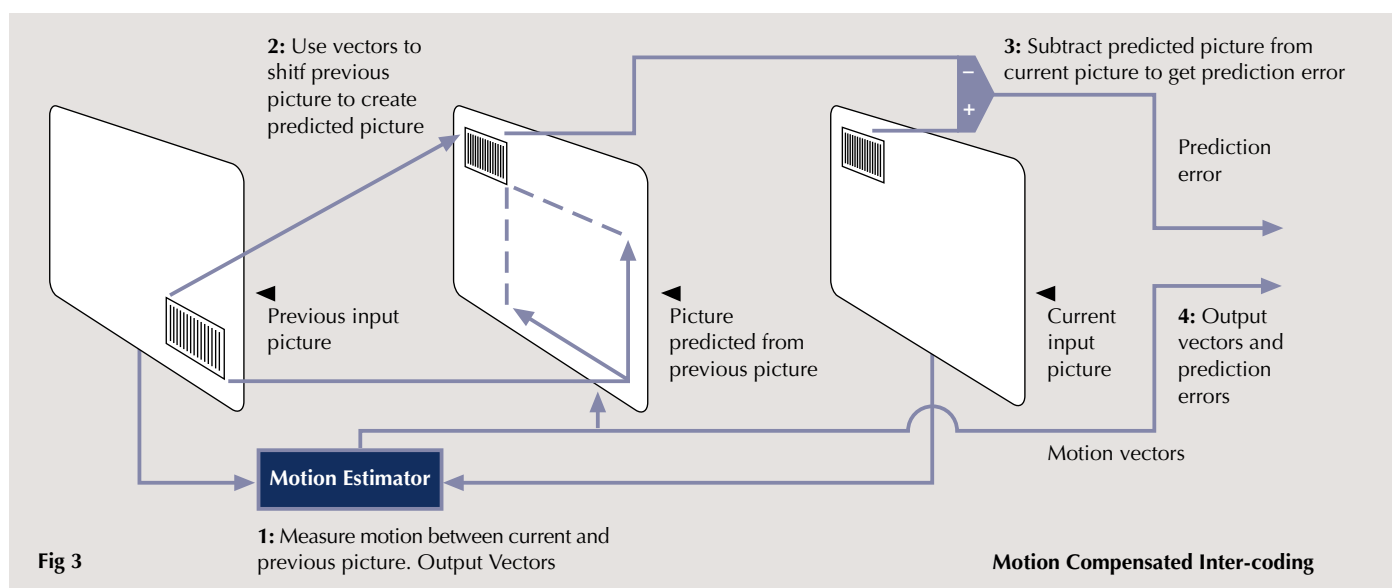
# MPEG I-frame encoding

Fig.2 shows a complete intra-coding scheme which MPEG uses on the first picture in a Group. The input picture is converted from raster scan to blocks. The blocks are subject to a DCT. The coefficients are then zig-zag scanned and weighted, prior to being requantized (wordlength shortening) and subject to run-length coding. Fig.2 also shows the corresponding decoder. The run length coding is decoded and then the coefficients are subject to an inverse weighting before being assembled into a coefficient block. The inverse transform produces a block of pixels which is stored in RAM with other blocks so that a raster scanned output can be produced.



**Fig 2**
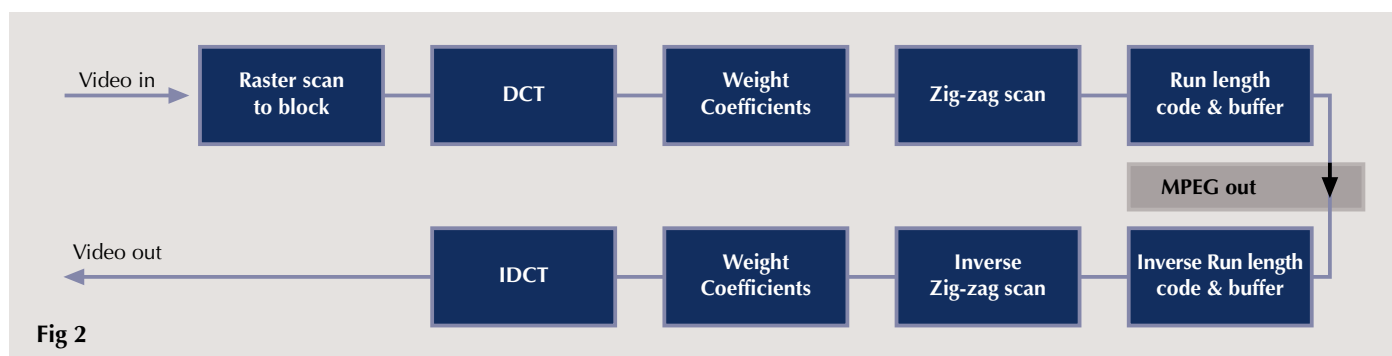
# MPEG long GOP encoding (inter-coding)

For moving pictures, exploiting redundancy between pictures gives a higher compression factor. In a simple inter-coder, after starting with an intra-coded picture, the subsequent pictures are described only by the way in which they differ from the one before. The decoder adds the differences to the previous picture to produce the new one. The difference picture, produced by subtracting every pixel in one picture from the same pixel in the next, is an image in its own right and is compressed with a DCT process.

Simple inter-coding falls down where there is significant movement between pictures and the solution is to use motion compensation. At the coder, successive pictures are compared and the motion of an area from one picture to the next is measured to produce motion vectors. Fig.3 shows that the coder attempts to predict the object in its new position by shifting pixels from the previous picture using the motion vectors. Any prediction errors are eliminated by comparing the predicted picture with the actual picture. The coder sends the motion vectors and the errors. The decoder shifts the previous picture by the vectors and adds the errors to produce the next picture.



**Fig 3**     **1:** Measure motion between current and previous picture. Output Vectors     **Motion Compensated Inter-coding**

In MPEG the picture is broken up into rectangular areas called macroblocks, each of which has its own motion vector. If the edge of a moving object lies across a macroblock, some of the macroblock is moving and some is not. This situation may be handled by setting the motion vectors to zero and handling the moving part with difference data, or by using finite motion vectors and handling the stationary part with difference data. A smart compressor might try both to see which approach yielded the smallest bit rate. As motion in real program material tends to be dominated by overall panning of the picture, the motion vectors are transmitted differentially. Consequently a motion vector parameter of zero is interpreted as the same motion as the neighboring macroblock. The vectors from the previous macroblock are copied. If the motion changes across the picture, vector differences are sent.

There are a number of potential problems with long GOP. If any errors occur in the channel they will propagate into every subsequent picture. It is impossible to decode the signal if it is selected after transmission has started. MPEG overcomes these problems by using different types of transmitted picture.

Fig.4 shows an example of a long GOP picture sequence used in MPEG. The sequence begins with an I (intra-coded) picture as an anchor, and this and all pictures before the next I picture are called a Group of Pictures (GOP). Within the GOP are a number of forward predicted or P pictures. The first P picture is decoded using the I picture as a basis, using motion compensation and adding difference data. The next and subsequent P pictures are decoded using the previous P picture as a basis. The remainder of the pictures in the GOP are B pictures.
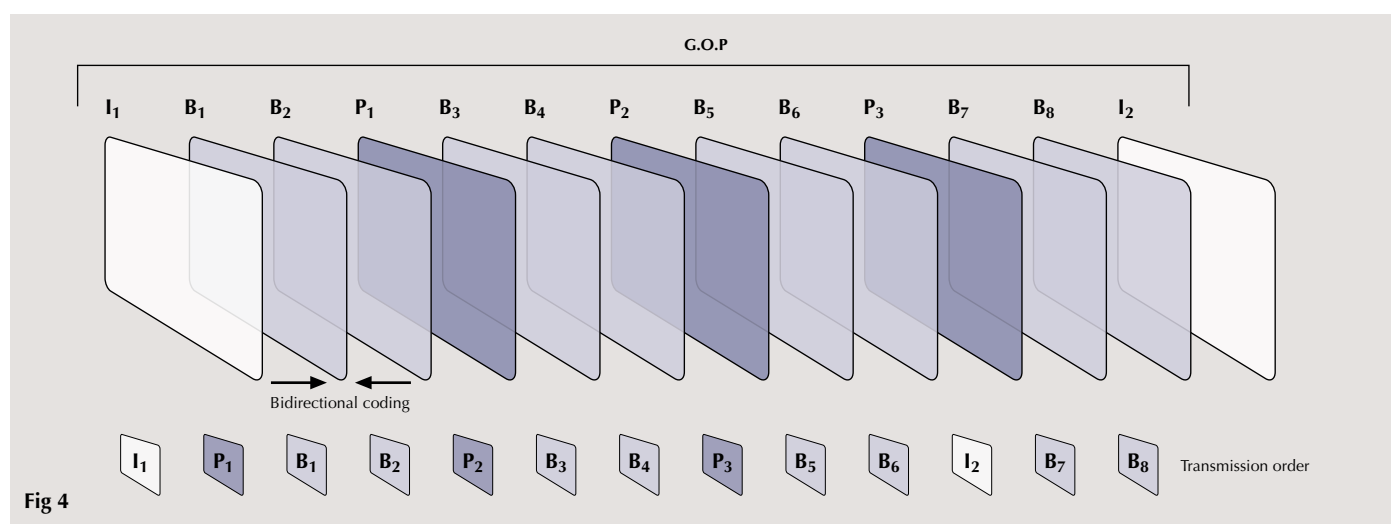


**Fig 4**

B pictures may be decoded using vectors and difference data from I or P pictures immediately before or afterwards. As the pictures are divided into macroblocks, each macroblock may be forward and/or backward predicted on an individual basis. Obviously backward prediction cannot be used if the decoder does not yet have the frame being projected backwards. The solution is to send the pictures out of display order. After the I picture, the first P picture is sent next. Once the decoder has the I and P pictures, the B pictures in between can be decoded by moving motion compensated data forwards or backwards. A bidirectional system needs more memory at encoder and decoder so that pictures can be re-ordered and this causes a greater coding delay. The simple profile of MPEG dispenses with B pictures to make a complexity saving, but will not be able to achieve such good quality at a given bit rate.

# What is the standard?

MPEG does not define how the encoder should carry out compression. Instead MPEG defines precisely how a decoder should attribute meaning to a variety of compressed bitstreams. This approach allows a great deal of flexibility whilst retaining compatibility. The manufacturers can use proprietary encoding algorithms yet produce a standard bitstream. As algorithms improve with time, the signals will still be compatible with existing decoders.

The uses of video compression go beyond the requirements of broadcast television. Different applications can afford different investments in encoder and decoder complexity. In addition to video, MPEG caters for associated audio and data, and means to multiplex audio, video and data together in a way which retains synchronization.

The wide range of performance and complexity required is handled by subdividing MPEG into profiles and then dividing each profile in level. A profile defines a set of techniques whereas a level is a constraint such as picture size or bitrate used with a technique. The main level is appropriate for standard definition (SD). Consequently the majority of broadcast and production interest is in MP@ML (main level at main profile) for 4:2:0 video and in 4:2:2P@ML (known as professional profile at main level) for 4:2:2 video. MPEG-2 supports both interlace and progressive video.

Most of the hierarchy of MPEG is based on 4:2:0 format video where the color difference signals are vertically subsampled. Most digital broadcast equipment works on 4:2:2 and the separate 4:2:2 profile is used. However, downconversion to 4:2:0 and interpolation back to 4:2:2 with properly engineered hardware gives visually indistinguishable results on a single codec. Unfortunately not all available equipment is made to a sufficiently high standard.

In many applications, DVB and ATSC included, MPEG compression will be used to allow several television channels and their associated sound channels to be time-multiplexed into one common digital telecomms system. The output of a single channel compressor is called an elementary stream and a number of these can be multiplexed together to form a transport stream. Multiplexing involves buffering each elementary stream and squirting the data out at a higher bit rate during the multiplex time slot. An equivalent time expansion process is needed at the demultiplexer.

The telecomms system is purely concerned with sending the right data to the right place at the right time without corrupting its content. There are numerous housekeeping signals added to the Transport stream, generally called System Information. Program Identifiers (PIDs) describe each of the different elementary streams in a Transport stream. Data packets are numbered contiguously so that they can be assembled correctly at their destination. The way in which all of these requirements are encoded and handled is known as a protocol.

In order to give each Elementary Stream the effect of a realtime contiguous transmission, the encoder and decoder are synchronized by a technique which can withstand the multiplexing process. The encoder and decoder clocks are locked by a periodic transmission of a clock count called the Program Clock Reference (PCR) which the decoder can turn into a stable clock using a numerically locked loop. The exact time at which the pictures or audio samples should be presented to the viewer is conveyed by transmitting periodic time stamps which in conjunction with PCR allow the decoder to rebuild the time axis of the original video and audio.

# Is it data or is it television?

There are two quite distinct aspects to the transmission of television in an MPEG transport stream and imperfections could be present in either or both.

The first aspect is that there is a telecommunication system which is carrying a number of multiplexed data streams from one place elsewhere under a standardized protocol. It is really not concerned with the meaning of the data.

The second aspect is that the data represent compressed audio and video whose accurate transmission is taken for granted, but whose quality is defined by the encoding process.

Considering the wrong aspect in the case of a problem could result in considerable wasted time. For example, depending on the location of the failure we could receive an awful picture perfectly or we could receive a perfect picture badly. How do we know whether a poor quality compressor produced the bad picture which the telecomms system accurately delivered, whether the data from a good compressor was corrupted or whether the data was good but some of it was lost because of incorrect protocol?

# Why do I need pre-processing?

Compression works best if the source signals are of very high quality. In the real world this is not always the case. There are a number of defects which a video signal can exhibit, and the result is invariably that the quality after compression is impaired or a higher bit rate is required.

Timebase error is common in signals from analog VCRs and in signals which have been transmitted long distances. One result of timebase error is random horizontal displacement of the television line. Fig.5a) shows a pixel block containing an edge without timebase error, whereas Fig.5b) shows that timebase error produces a high vertical frequency that was absent in the original signal. The compressor will attempt to code this frequency in order to recreate what it believes is the original picture. The result will be additional significant coefficients which take up space in the bitstream.



**Fig 5**

**a)** Object edge without timebase error      **b)** Object edge with timebase error producing spurious spatial frequency
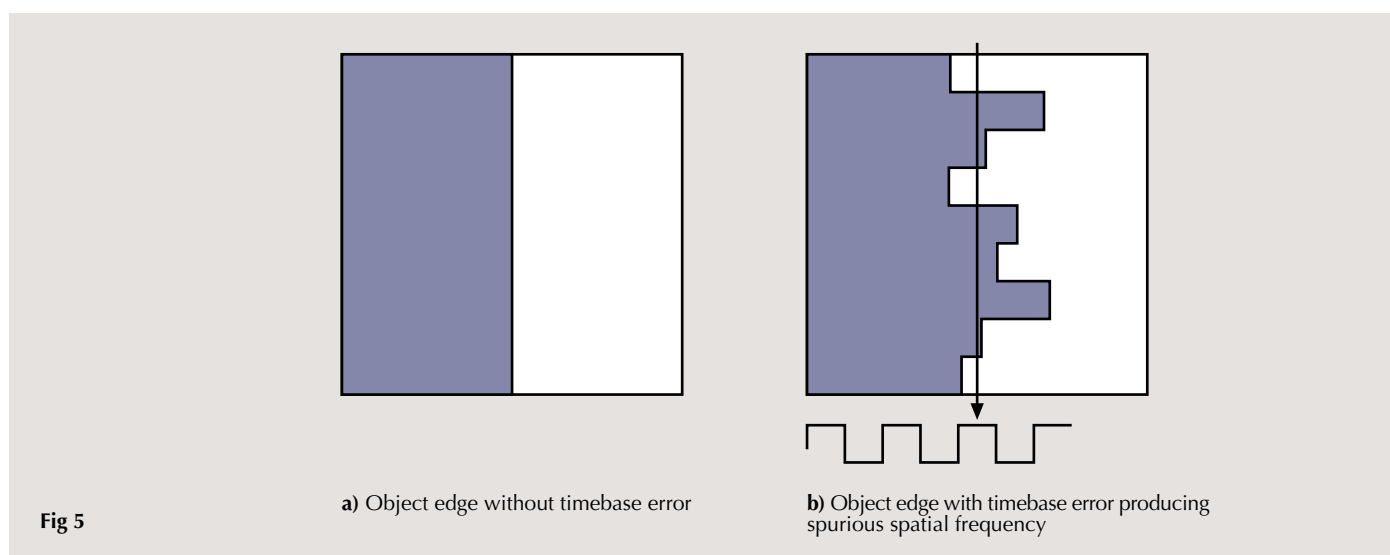
Timebase error also causes difficulties with motion compensation because the horizontal position of stationary objects changes from picture to picture. This means that any macroblock predicted from an earlier picture will differ more from the actual picture than it should and more prediction error data will be required to compensate.

In many cases signals to be compressed have been decoded from composite. Unless the decoding process is first rate, the component outputs may have residual subcarrier in them. This is extremely difficult for a compressor to handle as subcarrier deliberately generates high spatial frequencies in both horizontal

and vertical axes to reduce visibility. A DCT will produce spurious coefficients which the compressor will interpret as wanted picture detail.

In composite systems the frequency of subcarrier is chosen so that the same point on the screen is made alternately too dark and too bright on alternate pictures so that persistence of vision reduces visibility. A compressor interprets residual subcarrier as a difference from one picture to the next spread over the whole picture area requiring substantial prediction error data to convey it.

Noise is a particularly difficult problem in compression. Compressors are designed to eliminate the predictable content of the input video in order to leave only the entropy or useful content. Noise is unpredictable and the compressor finds it indistinguishable from entropy. In the spatial domain after DCT noise causes coefficients which should have a negligible value to become significant. In the time domain noise increases the differences between pictures as well as making motion estimation more difficult. Dropouts in VCRs and film dirt in telecines also produce spurious picture differences which need to be coded.

If bandwidth is plentiful, all of the above problems recede because the compressor is able to convey the extra apparent detail. However, compression is generally used where bandwidth is at a premium and so in most cases the extra bandwidth will not be available. The result of imperfect inputs will then be that space has to be found to transmit the spurious coefficients and difference data by sacrificing some of the genuine data. The bit allocator in the compressor will coarsen the requantizing after the DCT in order to get down to the allowable amount of data. The result is an increased level of compression artifacts in the decoded picture.

MPEG uses the periodic insertion of intra-coded information to allow channel switching and to speed recovery from transmission interruptions or errors. Most compression hardware inserts I pictures without reference to picture content. This is a sub-optimal approach because cuts in the source material will be interpreted as enormous prediction errors requiring a great deal of data to convey. If the data rate is limited, as it will be in practice, picture quality will suffer immediately after a cut. A superior solution is to detect cuts and steer the compressor so that the generation of an I picture coincides with the cut.

All of the real-world compression problems listed above will only get worse when compression systems are cascaded. It is easy to see how such cascading will occur in real life. A news team may shoot material on DVCPRO and transfer it to a non-linear editor for production. When the material is aired it will suffer a third compression to pass it down a telco line to the transmitter. Tandem codecs of this kind can cause serious quality loss because the artefacts of the first compression are treated as noise by subsequent coders. Consequently the quality of the first compression in a chain is critical.

Many practical problems can be eliminated or at least offset by the adoption of a suitable MPEG pre-processor containing noise reduction, timebase correction and cut detection. The Snell & Wilcox Prefix is a compression pre-processor specifically designed to extract the best performance from an MPEG system by reducing input artefacts and steering the coder with detected cuts.