Design considerations of the video compression system of the new DV camcorder standard

de With, P.H.N.; Rijckaert, A.M.A.

Published in:
IEEE Transactions on Consumer Electronics

DOI:
10.1109/30.642384

Published: 01/01/1997

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. Dec. 2018
DESIGN CONSIDERATIONS OF THE VIDEO COMPRESSION SYSTEM OF THE NEW DV CAMCORDER STANDARD

P. H. N. de With, member IEEE, A. M. A. Rijckaert
Philips Research Labs., Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Abstract — The digital camcorder system DV, which features advanced intra-frame DCT-based video compression, has been successfully introduced for the consumer and semi-professional user. This paper focuses on the design considerations and system issues that have resulted in the final video compression standard applied.

Keywords: DV, digital video, camcorder, video compression, recording, DCT.

1. Introduction

Recently, advanced video compression has become an emerging and enabling technology for new systems in digital transmission and storage. An example of successful video compression supporting digital video broadcasting is the MPEG standard [1], which is widely used in other environments as well.

In this paper, we describe digital home-use recording of high-quality video signals, which is relevant as a high-performance successor of the existing analogue formats (Hi-8 and VHS). We will particularly focus on the video compression system and its relation to the adopted recording system parameters, thereby giving technical insight to the design considerations of this completely new recording standard. The technical insight results from continuous participation in the standardization committees for DV recording and video compression for a number of years.

For consumer recording, the trend is towards small recording mechanics, especially for portable applications, and digital recording of compressed digital video signals. These aspects have been combined into the DV system, which was firstly announced in 1993 [2] and which has been further developed by a substantial group of companies. The main issue of the standard was to introduce attractive small camcorder products in the mid-nineties. Since September 1995, a number of companies have introduced DV camcorders into the market. The small products provide good examples of the original intentions and aims of the standardization group. Based on the DV technology, the professional systems DVCPRO and DVCAM have emerged, which are increasingly popular.

Despite the acceptance of the MPEG and JPEG standards in numerous applications, the development of the DV video compression standard was largely guided and justified by the targeted product and cassette size, power consumption and the consumer price level. Irrespective of the previous issues, a number of recording system constraints have to be satisfied in any case and are rather typical, namely

- editing, preferably on picture basis;
- robust for repeated (de-)compression, resulting from analogue video dubbing;

<table>
<thead>
<tr>
<th>Scanner</th>
<th>2 heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanner diameter</td>
<td>21.7 mm</td>
</tr>
<tr>
<td>Scanner speed</td>
<td>9000 rpm</td>
</tr>
<tr>
<td>Tape width</td>
<td>6.35 mm</td>
</tr>
<tr>
<td>Tape thickness</td>
<td>7 μm</td>
</tr>
<tr>
<td>Track pitch</td>
<td>10 μm</td>
</tr>
<tr>
<td>Recording rate</td>
<td>41.85 Mbit/s</td>
</tr>
<tr>
<td>Tracking</td>
<td>Embedded tracking tones</td>
</tr>
<tr>
<td>Channel modulation</td>
<td>24-25 code</td>
</tr>
<tr>
<td>Error correction code</td>
<td>R-S product code</td>
</tr>
<tr>
<td>Video bit rate</td>
<td>25 Mbit/s</td>
</tr>
<tr>
<td>Video coding</td>
<td>8x8 intraframe DCT</td>
</tr>
<tr>
<td>Trick mode implementation</td>
<td>Video macroblock shuffling</td>
</tr>
</tbody>
</table>

Table 1: Key parameters of digital consumer DV recorder.
2. System architecture

Figure 1 portrays the block diagram of the recording system. Signal processing is divided into two main classes:
- source coding employing
  video compression and source data formatting,
- channel coding using
  error-correcting codes and modulation codes
  for spectrum shaping of the bit stream.

The video compression is based on the Discrete Cosine Transform (DCT), and subsequent quantization and Variable-Length Coding (VLC) of the transformed video data. Since the operations are performed on blocks of $8 \times 8$ samples, the video input signal is stored for line-to-block conversion. Shuffling of data blocks is performed for improving the picture quality under normal and trick play. The sampling of the video signal is derived from the CCIR-601 standard (4:2:2) with 13.5 MHz sampling frequency for the luminance (Y) signal. The colour-difference signals (Cr and Cb) are subsampled either horizontally with a factor 2 (4:1:1) for 60 Hz or vertically with a factor 2 (4:2:0) for 50 Hz systems.

For Error-Correction Coding (ECC), the DV recording format employs the obligatory Reed-Solomon (R-S) product codes, of which the error-correcting properties are defined within a track. After compression, the data are organized in small packets, called sync blocks. Each sync block gets some parity bytes (HECC) assigned to it for random byte errors and for error detection of burst errors. Subsequently, the complete video data block (or audio) is extended with ex-
100% 134.9 100%

tra sync blocks containing vertical parity bytes (VECC). The performance of these codes is very good for removing random errors and general burst errors. As a second step in channel coding, a high-efficiency DC-free 24-25 channel code was adopted, described in [3] [4], to constitute pilot tones for tracking which are embedded in the data.

The recording mechanism is based on a small drum of 21.7 mm which rotates at 9000 rpm. The basic recording mode (2 × 1 head) has a capacity of 41.85 Mbit/s (equiv. to 25 Mbit/s video), but the system can be upgraded mechanically and electronically to 83.7 Mbit/s (50 Mbit/s net video rate) for HDTV recording. Possibilities to go to 12.5 Mbit/s video rate or lower, required for recording of MPEG signals (MP@ML SDTV) have been indicated earlier [3]. Table 1 portrays the key parameters of the experimental recording system.

3. Cassette size

3.1 Cassette capacity

This section addresses the relation between cassette capacity, bit rate and playing time. The well-known analogue 8mm cassette is the reference and the starting point. The relevant numbers of the 8mm cassette specification are listed in Fig. 3. D and d are the maximum diameter of the wound magnetic tape and the core, respectively. These numbers lead to a largest tape length L of around 125 meters.

Some relevant parameters of the 8mm tape format are depicted in Fig. 4. The effective width w of the tape area for analogue video and PCM audio (221° wrap) equals 6.6 mm, which is somewhat lower than the tape width of 8 mm. The edges cannot be used for the helical tracks due to the limited head-tape contact. The product of L and w results in 0.83 m² maximum usable tape area At.

The cassette capacity C in bits, is the ratio of At to the bit area Ab. It was estimated from experiments on ME tape [5] [6] [7], that a track width of 10 µm with a bit length of 0.25 µm would result in sufficient SNR and robustness for a system to be introduced a few years later. The gross 8mm cassette capacity is therefore 330 Gbit.

To derive the net capacity for compressed video, we have estimated the contributions for overhead as tabulated in the left column of Table 2. The estimations are educated guesses of the expected overhead, assuming a sync block size of around 1000 bits and a not usable track margin of 5–6°. For clarification, the real values of the DV system have been added in the right columns of Table 2. The overhead for channel modulation and digital audio will be clarified in more detail. A well-known block code like 8-10 modulation needs an overhead of 20%.
form of channel modulation is a must, primarily for tracking and robustness reasons, but the overhead should be limited. The sophisticated 24-25 modulation from [4] with only 4% overhead has proven to be an attractive candidate. The overhead for digital audio is in principle more an absolute value in bit rate and less a relative overhead. However, we have assumed that the bits needed for digital audio are a reasonable 10% fraction of the bits for compressed video, hence 6% of the total.

From the previously discussed overhead inventory it is clear that, even with an economical solution for channel modulation, only around 60% of the gross capacity is available for compressed video. The net cassette capacity for the 8mm system is therefore 200 Gbit. From this number a simple relation is obtained between the compressed video bit rate and the playing time, which is graphically presented in Fig. 5.

**3.2 Outline of the new DV system**

Prior to discussing the principal DV system parameters, we list the basic aims and assumptions for defining the recording system.

- Pocketable camcorders, which are substantially more compact than the analogue 8mm system. Playing time is less important; 60 minutes is regarded as sufficient. Picture quality and editability are of utmost importance.

- Home-use recording of HDTV signals with a minimum playing time of 135 minutes, needed to record almost all movies in one piece. At the time of standardization, HDTV systems like MUSE in Japan, HD-MAC in Europe and ATV in USA were seen as very important for the near future.

- Home recording of SDTV signals, with the same compression as for the camcorder. Ratio of HDTV to SDTV bit rate of 2:1, resulting in a minimum playing time of 270 minutes for SDTV.

- Preferably, a one-cassette system has to be adopted.

It will be clear that the last point can be conflicting with the first two, due to the significantly higher bit rate and playing time demands of the HDTV system, when compared to the camcorder application. This will become more apparent from a further evaluation of Fig. 5. At a playing time of 135 minutes for HDTV, the available video bit rate for an 8mm cassette equals 25 Mbit/s, and consequently 12.5 Mbit/s for SDTV. It is elucidated later in this paper, that this bit rate is considered too low for an economical, well editable and high-quality compression scheme for the camcorder. In fact, the double of this bit rate was regarded as realistic. This leads inevitably to a two-cassette system with a larger cassette for home use and a compact cassette with less playing time for the camcorder.

**3.3 Cassette size for new DV system**

From the preceding sections it can be concluded that the net cassette capacity for home use should be doubled to 400 Gbit, requiring a substantially
larger cassette compared to the analogue 8mm system. It is easily understood that the capacity is roughly proportional to the effective volume, hence to the product of the dimensions, excluding the fixed mouth part of around 13 mm for the cassette lid etc. Assuming a constant thickness results at first instance in a standard DV cassette of around 1.4 times the 8mm cassette size. Using the volume more efficiently reduces this factor to 1.3. The required capacity of the small camcorder cassette is roughly one quarter of the standard cassette, due to the playing time ratio, and has therefore half its size. This is substantially smaller than the 8mm cassette. The relative sizes of the final DV cassettes are depicted in Fig. 6. The DV recording channel is described in more detail in [9].

Thus far the thickness of the cassettes was assumed identical to the 8mm cassette. This is suitable for the larger standard cassette, but a small camcorder cassette with this thickness is not very elegant. Additionally, the compactness of the camcorder is influenced by this parameter. The height of the deck mechanics is proportional to the cassette thickness, which should therefore be reduced as far as possible for a camcorder application. Fortunately, the tape technology improvement provided expectations that the thickness of the ME tape could be reduced to around 7 μm in a few years time. This allows for a longer but narrower tape with the same capacity. A logical choice for the tape width is the well-known quarter inch or 6.35 mm. The final thicknesses for the standard and small cassettes are 14.6 and 12.2 mm, respectively.

4. Drum diameter and speed

4.1 Basic drum

In principle, the drum configuration is a manufacturer's option, as long as the correct tape format is written. In this section we will focus on an economical drum for the camcorder, which is the well-known drum concept having two heads of different azimuth on opposite sides and using a wrap angle of 180°. This not only leads to the lowest drum complexity, but also minimizes the required read/write electronics and therefore power dissipation, because only single recording channel electronics are needed. In the next section some alternative drum configurations and their relations to other applications will be elucidated. As an example taken from the literature, we refer to [8] where an experimental digital VCR on the basis of ME tape and with a small drum is presented.

The analysis commences with the relation between the drum and video signals. The following wish list has to be fulfilled.

- **World-wide recording system.**
  This results in an identical bit rate and therefore drum speed all over the world.
- **Integer number of tracks/frame.**
  Originates from frame-based editability.
  Valid for 525/30 and 625/25 SDTV (camcorder) systems.
- **Scalability.**
  Attractive properties for higher (HD) and lower (LP) bit-rate recording systems.
- **Small drum size.**
  Drum size as small as possible because of compact camcorder.

The product of effective lines and frame rate is identical for 525/30 and 625/25 systems, leading to a common bit rate. This automatically implies that the drum speed and the resulting number of tracks/s, will be the same for both systems. Consequently, the number of tracks/frame must have a ratio of 5 to 6. Only distinct combinations are possible, due to the editability demand of an integer number of tracks/frame.

Let us choose 5 tracks/frame for 525/30 and 6 for 625/25, as an example. This combination leads to a universal track rate of 150 tracks/s. A drum with 2 heads has to rotate with a speed of 75 rps or 4500 rpm. The results thus far are not dependent on the bit rate itself in any way. It is

<table>
<thead>
<tr>
<th>Tracks/frame</th>
<th>Track rate (tr/s)</th>
<th>Drum speed (rpm)</th>
<th>Track length (mm)</th>
<th>Drum diam. (mm)</th>
<th>Track angle (deg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/6</td>
<td>150</td>
<td>4500</td>
<td>67</td>
<td>42</td>
<td>5</td>
</tr>
<tr>
<td>10/12</td>
<td>300</td>
<td>9000</td>
<td>33</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>15/18</td>
<td>450</td>
<td>13500</td>
<td>22</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>20/24</td>
<td>600</td>
<td>18000</td>
<td>17</td>
<td>11</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3: Various drum configurations.
assumed that the compressed video bit rate for SD will be 25 Mbit/s, which is 60% of the channel bit rate of around 40 Mbit/s. Combining this with the bit length of 0.25 μm, indicates that the surface speed of the drum must be 10 m/s, resulting in a track length of 67 mm and a drum diameter of 42 mm. The track angle, which depends on the applied tape width and track length, is somewhat less than 5°, which is a common value. Other possible combinations like 10/12, 15/18 and 20/24 have been evaluated in a similar fashion. The results are given in Table 3.

A comparison between Table 3 and the wish list at the beginning of this section shows that 10/12 tracks/frame is a well-balanced choice, taking into account the following additional remarks.

- An even number of tracks/frame is preferred for both systems, due to azimuth recording.
- A large number of tracks/frame has a negative effect on the picture quality in trick play.
- Higher drum speeds generally yield a higher power dissipation of the drum motor.
- Small drum diameters are difficult to realize due to the limited internal space for transformers, motor and head mounting.
- Large track angles should be avoided because of the magnetic properties of the tape and the compact construction of the deck.

Fine tuning of the tape format resulted in 41.85 Mbit/s channel rate, 34.024 mm track length, 21.7 mm drum size and 9.1668° track angle for the DV system. The effective track length is 32.890 mm which corresponds to an effective wrap of 174°.

4.2 Alternative drum concepts

At this point, it is interesting to check some alternative drum concepts. The basic drum is depicted in Fig. 7a. In Fig. 7b, one head is shifted to the opposite side to create a compatible drum with one double head. This drum is used in a two-channel burst mode with one active scan per revolution, resulting in an average bit rate identical to the basic drum. The problem of the drum in Fig. 7b is that it requires two-channel electronics and additional delay to compensate for the head shift, leading to more circuitry and a higher power dissipation. For camcorders this would certainly be disadvantageous, but it can be interesting for other applications. Let us now briefly discuss such applications and special modes of operation.

LP modes

An advantage of the concept in Fig. 7b is the ability for LP burst modes with not only odd but also even dividers of the bit rate, due to a suitable azimuth configuration on the drum. For example, when recording every second scan at half the tape speed, the bit rate is halved (burst-mode recording, see [3]). This opens up the possibility of an LP mode for home use with halved bit rate and double playing time.

HD modes

Another option is to add a second double-head as indicated in Fig. 7c for a continuous two-channel recording, allowing for an HD mode with two times the bit rate of SD. Full compatibility to the SD and LP modes is realized by using only one head pair. For further details see [22].

Data recording

The second double-head of concept 7c but with a slightly modified height position can also be used for read-after-write in generic data recording applications. Crosstalk from write-to-read circuitry, including rotary transformers and heads, is an important issue in this case, due to the required damping of over 100 dB. A good crosstalk separation is simply achieved, because the write and read phase are interleaved in time.

Although this listing of alternative concepts is certainly not complete, it will be clear that a drum of 21.7 mm rotating at 9000 rpm is a good choice for the basic scanner with interesting possibilities for extensions to other applications.
5. Video blocks and sync block format

5.1 Macroblocks and tracks

In this section, the division of video MacroBlocks (MB) on the tracks and the relation to the recording Sync Blocks (SB) will be discussed, under the assumptions of 10/12 tracks/frame and 25 Mbit/s for compressed video. After sampling of the analogue signal, the effective area of the video frame consists of 720 pixels horizontally by 480 or 576 lines vertically for the 525/30 and 625/25 systems, respectively. This is depicted in Fig. 8. The pixels are grouped in DCT blocks of $8 \times 8$ pixels, resulting in 5400 or 6480 luminance DCT blocks/frame. With 10/12 tracks/frame, this leads to an identical number of 540 luminance DCT blocks/track for both systems. Although the subsampling of the chrominance signals is different for 525/30 and 625/25, the ratio of luminance to chrominance is 4:2 in both cases. Macroblocks are formed, consisting of 4 luminance and 2 chrominance DCT blocks, resulting in 135 MB's per track.

5.2 Restrictions on macroblock mapping

For reasons of convenience (hardware complexity), robustness and trick-play performance, a fixed mapping of a limited number of MBs (a unit) onto an integer number of SBs is preferable. Moreover, a multiple of such mapping units should exactly fit into one track for obvious reasons. As a result, the number of MBs per unit must be a divider of 135. Possibilities are 1, 3, 5, 9, 15 etc. Similarly, the same units, in the sequel called segments, will be used as independent coding units for the video compression. The video segment size was fixed to 5 MBs, because it was found from simulations that no essential improvement in picture quality could be obtained for larger segments (see Fig. 17 in Section 6.5). This results in 27 video segments per track.

The remaining key parameter in the mapping is the number of SBs per segment. It is clear that this number will vary the total amount of video SBs in a track, which as a consequence of the previous result, should be a multiple of 27. Let us now investigate the relation between this number and the SB length. In many cases, recording SBs have a size of around 1000 channel bits, as a good compromise between synchronization aspects (short SB) and a limited overhead for the synchronisation pattern, ID, etc. (requiring a long SB). With an estimated overhead of 15% for Sync, ID, SB error-correction coding (ECC) and channel modulation, 850 bits are available for the video information itself. The total amount of video bits in a track can be calculated from the video bit rate, 25 Mbit/s, and the track frequency, 300 tracks/s, and is equal to 83 Kbit. With the example of 850 video bits per SB, around 100 SBs are needed in a track for video information. As is clear from this example, there is a direct relation between the amount of video SBs per track and the SB length. The variation in this length will be limited from 700 to 1400 bits leading to a range of 70-140 SBs per track. When combined with the "multiple of 27" demand, it results in only 3 possible mappings as indicated in Table 4. As an example, a system based on the 5-3 mapping is presented in [8]. One of the important factors for making an optimal
choice are the trick-play aspects, which will be described in detail in Sections 5.3 and 5.4.

5.3 Trick-play aspects

An important factor in deriving the SB length is the behaviour in a trick-play situation, where the heads are scanning across a number of recorded tracks, due to the deviating tape speed [lo]. During each track crossing with the correct azimuth, a data burst can be read, as indicated in Fig. 9. Two different concepts have to be distinguished. In the first concept the drum speed is constant at 150 rps. In this case, the data burst length is given by the following expression

\[ D = Q \times 2 \left| \frac{kT_L}{n-1} - C - A \right|, \]

in which
- \( D \) = Data burst length,
- \( T_L \) = 180 deg. track length,
- \( n \) = Tape speed factor,
- \( C \) = Track-track shift = \( T_p / \tan(\theta) \),
- \( A \) = Azimuth effect = \( T_p \tan(\alpha) \),
- \( Q \) = Quality factor = \( T_{e_{max}} / T_p \),
- \( k \) = Drum configuration

The parameter \( T_{e_{max}} \) is the maximum tracking error at which an adequate bit detection is still possible, \( T_p \) is the track pitch of 10 \( \mu \)m, and \( \theta \) and \( \alpha \) are the track and azimuth angle, respectively. The factor \( k \) depends on the drum configuration and equals unity for the basic scanner. Despite the constant drum speed, the bit rate recovered from tape is varying, because the relative speed is resulting from a combination of drum and tape speed. The bit rate relative to normal play is

\[ m = 1 - \frac{(n-1)(C+A)}{kT_L}, \]  \hspace{1cm} (2)

A graphical representation of this relationship is given by the solid lines in Fig. 10. For low trick-play speeds, this variation can be acceptable but at high speeds, problems occur in the head, transfer, preamplifier, equalizer and PLL circuitry. For such high speeds, it is better to change the drum speed in relation to the tape speed in such a way that the average bit rate for both heads is constant. This second concept leads to

\[ m = 1 - \frac{(n-1)A}{kT_L + C}, \]  \hspace{1cm} (3)

and

\[ D = Q \times 2 \left| \frac{kT_L + C}{n-1} - A \right|, \]  \hspace{1cm} (4)

The bit-rate variation for the "constant bit rate" technique is given by the dashed lines in Figure 10. The dashed lines show that the percentual bit rate variation is reduced by a factor 20 when compared to the solid lines. Note that in general \( A \ll C \ll T_L \). The recovered data burst length as a function of the trick-play speed is depicted in Fig. 11.

5.4 Maximum search speed

For trick play, the highest possible trick-play speed is a decisive factor. As a consequence, the constant bit-rate concept has been adopted. As-
Assuming $Q = 1/2$ and $n \gg 1$, the data burst length $D$ can be approximated by:

$$D = \frac{T_L}{n}.$$  (5)

Because the position of the track crossings with respect to the SBs is arbitrary at these speeds, the data burst must be at least two SBs in order to retrieve one correctable SB from tape. The maximum trick-play speed $n_{\text{max}}$ is then given by

$$n_{\text{max}} = \frac{T_L}{D_{\text{min}}} = \frac{T_L}{2},$$  (6)

with $T_L$ in SBs being the total length of a track for $180^\circ$ wrap. The total overhead in a $180^\circ$ track is assumed to be 40%, of which 15% is SB-based. The remaining 25% of the overhead can be expressed as SBs on top of the video SBs. Therefore, the total number of SBs in a track ($T_L$) is $4/3$ times the number of video SBs. Results for $T_L$ and $n_{\text{max}}$ are given in Table 4. In the preceding calculation, the effect of PLL run-in has been neglected. In practice, the value of $n_{\text{max}}$ will be somewhat lower.

### 5.5 Conclusion on SB format

Table 4 indicates that the highest possible trick-play speed is realized with the 5-5 mapping. It should be noticed that with this mapping, in contrast to the other possible mappings, one MB is stored in one SB, thereby enabling a fixed allocation of the most important data. For trick play, this enables the decoding on SB basis, which was assumed implicitly for the calculation of $n_{\text{max}}$. As a bonus, the 5-5 mapping results in the shortest SB length. This property, combined with the fixed data allocation, proves to be very beneficial for the robustness of the system.

Summarizing, the 5-5 mapping was chosen as the best solution for the DV system. This being fixed, the size of an SB can be determined. With 83 Kbit and 135 MBs in a track for video, a 77-Byte data area is required to store one macroblock. The addition of a 2-Byte Sync pattern, a 3-Byte ID and 8-Byte parity for the inner (horizontal) ECC, results in a total SB length of 90 Bytes. For completeness, it is mentioned that this number must be a multiple of 3 Bytes, because of the 24-25 channel coding. The SB structure of the DV system is given in Fig. 12.

### 6. Video compression

In this section, we consider a feedforward-buffered bit-rate reduction scheme, based on DCT coding, in which the pictorial data is analyzed prior to coding. The aim is to define a compression system with fixed-length coding of a relatively small group of macroblocks (see Section 5.1), because this is beneficial for recording applications, such as trick play and robustness.

Given the system constraints from a.o. Section 1, the target system is based on fame-based DCT coding with VLC. It is known by video experts in the field that such a system operates well using compression factors 5–8. This results globally in a bit rate after compression of 20–25 Mbit/s. Numerous subjective quality experiments during development of the DV standard provided evidence that about 25 Mbit/s yielded the desired picture quality.

#### 6.1 Feedforward coding concept

An established feature of a video recorder is to search at various forward and backward speeds in a video sequence (trick-play modes). During the search, the video heads cross tracks at

![Figure 11: Data burst length versus search speed.](image)

![Figure 12: Format of a single sync block (SB).](image)
nearly random positions, resulting in a burst-wise data recovery (see Fig. 2), from which—at least recognizable—pictures must be reconstructed. In most transform coding systems using VLC techniques, the variable-rate output is buffered and monitored by feedback quantizer control [11] [12] to obtain—a constant rate—although the bit rate is locally varying. With feedback compression systems having a locally varying output bit rate, the relation between the recovered data bits and the location of the data in the picture is lost, resulting in a more complicated picture reconstruction and, regularly, a local loss of picture quality.

The major advantage of the feedforward coding system is that relatively small groups of DCT blocks, henceforth termed segments are coded independently and, in contrast with a feedback system, as a fixed entity. This property makes the segments independently accessible on the tape, while the fixed code length ensures a unique relation between the segment location on tape and its location in the reconstructed picture. The latter property, in combination with the 1 macroblock-per-SB data allocation (see Section 5.5) will be exploited for optimizing the picture quality during trick modes (see Section 8).

Because of its robust features, feedforward coding was proposed earlier for experimental recording systems. In [13] a DCT coding system is proposed in which a horizontal stripe of blocks from a picture is coded as a fixed segment, using zonal coding at discrete bit rates. In [14] we notice a feedforward scheme with adaptive zonal coding having a variable bit cost for each DCT block within the segment. The DV standard is based on a threshold coding scheme with a more robust variable-length coding technique and simplified quantization strategies, thereby paying attention to low complexity.

In the feedforward coding system (see Fig. 13), video is first organized into blocks and subsequently in groups of blocks, called segments. Each segment is then compressed with DCT coding techniques into a fixed code length (bit cost), despite the application of variable-length codes (VLC). Fixed-length coding of a small group of DCT blocks (several tenths only) can only be realized if the transformed data is analyzed prior to coding, requiring temporal data storage. During the storage of a segment, several coding strategies are carried out simultaneously, from which only one is chosen for final quantization and coding. This "analysis of the limited future" explains the term feedforward buffering.

Another distinct aspect of feedforward coding is the data analysis part, which is absent in most feedback systems. Because of the speed requirements resulting from a luminance sample frequency exceeding 10 MHz, all different coding strategies must be carried out in parallel, whereas in principle, a feedback system uses only one quantizer strategy. However, in contrast with the increased system complexity of feedforward coding, two important advantages remain, namely a fixed relation between the data on tape and the reconstructed image, and a high intrinsic robustness as the channel error propagation is principally limited within a video segment.

6.2 Motion-adaptive DCT

The Discrete Cosine Transform (DCT) has become the most popular transform in picture compression, since it has proven to be the most efficient transform for energy compaction, whereas the implementation can be of nowadays limited complexity.

The definition of the DCT used in the DV sys-
where a block of samples \( f(i,j) \) has size \( N \times N \). The two constants \( C(u) \) and \( C(v) \) are defined by \( C(w) = 1/2 \) for \( w > 0 \) and \( C(0) = 1/2 \sqrt{2} \).

In the DV standard, we have adopted a block size of \( 8 \times 8 \) samples because it provides the best compromise between compression efficiency and complexity and robustness.

One of the first main parameters for coding efficiency is to choose between intrafield and intraframe coding. In the latter system, the odd and even fields are first combined into a complete image frame prior to block coding. Computer simulations showed that intraframe coding is about 20–30\% more efficient than intrafield coding, or, for the available 25 Mbit/s bit rate, a considerable better quality can be obtained. For this reason, intraframe coding was adopted in the standard. However, it was found in earlier investigations [15] that local motion in sample blocks leads to complicated data structures after DCT transformation, which usually are particularly difficult to code. The solution for this problem is to split the vertical transform into two field-based transforms of length \( N/2 \). Hence, first an \( N \)-point horizontal transform (HDCT) is performed yielding intermediate data \( F_h(i,v) \), and subsequently, two vertical \( (N/2) \)-point transforms (VDCT), specified by

\[
F(u,v) = C(u)C(v) \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} f(i,j) \cos \left( \frac{(2i+1)\omega \pi}{N} \right) \cos \left( \frac{(2j+1)\omega \pi}{2N} \right),
\]

with

\[
g_s(i,j) = [f(2i,j) + f(2i+1,j)], \quad g_d(i,j) = [f(2i,j) - f(2i+1,j)]. \quad (10)
\]

Note that in the first field-based output coefficient block, the sum of the two fields is taken as an input, while in the second coefficient block the difference of the two fields is considered. Hence, this corresponds to a two-point transform in the temporal domain.

The required DCT processor architecture is depicted in Fig 14. One additional feature is the application of a motion detector which detects – a priori – the data structures resulting from local motion, so that the VDCT can switch on a block basis to be in either the stationary \( N \)-point mode, or in the motion mode (M) using two \( (N/2) \)-point transforms.

### 6.3 Adaptive quantization

Most of the conventional techniques to define the quantization of the coefficients \( F(u,v) \) from feedback coding systems apply for feedforward coding as well. The primary elements for quantization are frequency-dependent weighting, adaptivity to local image statistics and global bit rate control. Let us now address those elements individually.

Since a number of quantization strategies must be performed in parallel in a feedforward system, it would be attractive to parameterize the weighting function in order to optimize its usage in various strategies. After extensive experiments in [16], the transfer function \( H \) of the Human Visual System (HVS) was modelled by

\[
H(f_r) = H_1(H_2 + H_3 f_r) \exp(-H_4 f_r^2), \quad (11)
\]

where \( f_r \) denotes the radial frequency in cycles per degree (c/d) and \( H_i \) are constants, e.g. \( H_1 = \)
2.2, \( H_2 = 0.192 \), \( H_3 = 0.114 \) and \( H_4 = 1.1 \). Nil1 [17] proposed to include the sensitivity of the HVS in the DCT calculation by using a multiplicative weighting function that matches the HVS. As an example, the HVS model of [16] has been plotted in Fig. 15, in which the transfer function of the HVS has been normalized.

It can be computed with a simple model [18], assuming that DCT frequency coefficients resemble real frequencies (which is not the case), which radial frequencies correspond with the spatial frequencies inside a DCT block. For example, when we use a high-quality display with a pixel distance of 0.4 mm and a viewing distance of 1 m, the spatial frequency covers roughly the range 5-20 cycles per degree. In Figure 15 we find that for the area specified, the HVS decreases exponentially from globally 1.0 to 0.4, so that a reasonably strong weighting function can be adopted. Basically, the weighting is multiplicative from nature, hence \( F_W(u,v) = W(u,v)F(u,v) \). The weighting function can be simplified using special multiplying factors (see Table 6). For simplicity, the matrix of factors \( W(u,v) \) are not different for each \((u,v)\) combination. Instead, groups of \((u,v)\) combinations apply the same weighting factor according to Table 5.

The second element of quantization is the adaptivity to local image statistics. In [19], a number of possible measures are given to analyze the contents of a coefficient block (all coefficients \( F(u,v) \) with \((u,v) \neq (0,0)\)). One of the most well-known metrics for local activity is the "ac energy" of the block, \( \sum_{u,v} F(u,v)^2 \). However, simpler metrics, such as the maximum of all \( F(u,v) \) perform satisfactorily as well. The DV system allows that any metric in the encoder is acceptable: the decoder simply follows the two decision bits reserved for the quantizer classification. This freedom also allows different quantization of luminance (Y) and colour (Cr, Cb) blocks. Generally, more activity or information content results in more coarse quantization.

For the global final block quantization, a linear division with a step size \( S \) was adopted. The advantage of this approach is its simplicity. The division in the encoder and the multiplication in the decoder leads to uniform quantization. The variable \( S \) defines the accuracy of the global block quantization. When taking into account the elements previously discussed, the overall quantization is specified by

\[
F_Q(u,v) = W(u,v)F(u,v)/S. \tag{12}
\]

Since the quantization is performed for a number of strategies in parallel, low implementation cost is of utmost importance. A particular simple system is obtained by taking \( W(u,v)/S = 2^{-p} \) with \( p \) being an integer that is controlled by all three elements [20] discussed in this section. The final quantization table is shown in Table 6. The weighting area numbers in Table 6 refer to Table 5.

6.4 Variable-length coding

Since a plurality of coding strategies must be performed in the analysis part, a simple bit-assignment technique is essential for limiting the hardware complexity of the bit-cost calculation. Within the class of threshold coding systems, one-table coding techniques have the highest potential to be implemented with simple hardware.
Table 6: Table of step sizes using area indication for weighting, and strategy number for global uniform quantization.

Several proposals exist [14] [18] for one-table coding, but the technique from [21] later used in MPEG, was adopted finally, because of its simplicity using runlength counting of zeros only and the use of an EOB (end-of-block) codeword.

The principle of the algorithm is that first the block of quantized coefficients is scanned using diagonal zigzag scanning in order to create a one-dimensional stream of numbers. The scanning is adapted to motion (see e.g. [18]). The purpose of the scanning is to cluster zero coefficients (see Fig. 16), so that they can be coded efficiently. Secondly, from the start of the string, zeros are counted until a non-zero coefficient is noticed. The magnitude of this non-zero coefficient is combined with the preceding length of zeros into a single event \( (\text{runlength}, \text{coefficient}) \), which is jointly coded with a single codeword. The sign of the coefficient is appended at the end of the codeword. The individual steps of the algorithm are listed below.

**Step 1:** load a coefficient and test if it is zero.  
**Step 2:** if zero coefficient, increment zero counter and go to step 4.  
**Step 3:** if nonzero coefficient then: \( a. \) jointly code the value \( (\text{runlength}, \text{coefficient}) \) \( b. \) append sign and reset the runlength counter.  
**Step 4:** if the actual coefficient is not the last one in the block, then go to step 1.  
**Step 5:** terminate the block with EOB codeword; ignore the runlength value.

An excerpt of the encoding table showing the variable wordlengths and codewords is given in Table 7. Special care has been taken to prevent the occurrence of very long codewords. For example, in Table 7a, the "X" denotes the very unlikely event \( (5, 7) \) which is mapped onto the concatenation of two more probable events having a shorter codeword, namely \( (4, 0) + (0, 7) \). It can be verified that the decoder is transparent for receiving this modified sequence. Similarly, large amplitudes are mapped onto a sequence of shorter codewords, consisting of an escape code followed by a fixed-length coded amplitude bit pattern.

### 6.5 Total video performance

In this section we briefly address the control and overall performance of the compression system. Finding the optimal quantization strategy can be formulated as follows. The bit cost \( B_C \) of the kth block of a segment of size K is

\[
B_C = \sum_{k=0}^{K-1} VLC_L[F_Q(u, v)],
\]

where \( F_Q \) results from (12) and \( VLC_L \) denotes the encoding technique of the previous section, but only considering the wordlengths. The optimal quantization strategy \( m_{opt} \) is the strategy...
Table 7: Table of wordlengths (a) and codewords (b) of DV system.

<table>
<thead>
<tr>
<th>run</th>
<th>amplitude (abs.)</th>
<th>event codeword</th>
<th>event codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>(0, 1) 00s</td>
<td>(4, 1) 110001s</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>(0, 2) 010s</td>
<td>(0, 7) 110010s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EOB 0110</td>
<td>(0, 8) 110011s</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>(1, 1) 0111s</td>
<td>(5, 1) 1101000s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0, 3) 1000s</td>
<td>(6, 1) 1101001s</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>(0, 4) 1001s</td>
<td>(2, 2) 1101010s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2, 1) 10100s</td>
<td>(1, 3) 1101011s</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>(1, 2) 10101s</td>
<td>(1, 4) 1101100s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0, 5) 10110s</td>
<td>(0, 9) 1101101s</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>(0, 6) 10111s</td>
<td>(0, 10) 1101110s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3, 1) 110000s</td>
<td>(0, 11) 1101111s</td>
</tr>
</tbody>
</table>

Table 7: Table of wordlengths (a) and codewords (b) of DV system.

that yields the desired bit rate $R_d$ such that

$$| R_d - R_{\text{opt}} | \approx 0, \quad \text{with} \quad R_{\text{opt}} = \frac{1}{K N^2} B_C(Q_{m, \text{opt}}). \quad (15)$$

The choice of $m_{\text{opt}}$ can vary between 0 and $M - 1$ when $M$ different quantization strategies are used.

We have simulated the picture quality of the compression system described in this section for various segment sizes $K$ and number of quantization strategies $M$. The resulting picture quality is expressed in SNR (dB) which refers to the Mean Squared Error (MSE) with the original picture compared to squared peak white ($255^2$). The results of the measurements are shown in Figures 17 and 18.

Let us now discuss the optimal choice of the segment size $K$ and the number of strategies $M$. Evidently, the recording system designer opts for the smallest possible value of $K$ (small segments), because it yields a high robustness and it enables higher search speeds. However, Fig. 17 shows that if the size becomes too small, i.e. $K < 30$, the picture quality deteriorates rapidly. For $K = 30 - 60$ DCT blocks, the resulting SNR remains nearly constant. For this reason, a segment size of $K = 30$ was adopted as being the best compromise.

The 30 DCT blocks are not arbitrarily chosen from the picture, but clustered in groups, called macroblocks. A macroblock (see Fig. 8) is a group consisting of $2 \times 2$ DCT blocks of $8 \times 8$ Y samples each and the two corresponding $8 \times 8$ colour blocks $C_R$ and $C_B$. In order to improve the picture quality of the compression system, macroblocks are selected from different areas of the picture, so that the influence of local image statistics is smoothed. The result is a stable and high picture quality for a large class of images. However, after compression, the macroblocks are redistributed in order to improve the visual performance during search. Further details on this topic will be supplied in Section 7.

A second conclusion is that $M = 16$ gives a substantial improvement in picture quality, compared to $M = 8$. The quality improvement can be fully explained by a more efficient use of the available bit rate, which becomes particularly important for small segment sizes. In other words, for a larger $M$, the X-axis resolution of Fig. 18 increases, so that it becomes easier to select a quantizer strategy which satisfies equation (15) more accurately.

The subjective picture quality of the system is
known as excellent and regarded as very close to levels for professional use. For normal, easy-to-code, pictures, the SNR approaches 40 dB, and the resulting subjective image quality of the system comes rather close to the quality of existing professional recording systems. For complex and detailed imagery, the SNR is a few dBs lower.

It has also been verified that the performance of the system is comparable to intraframe feedback coding systems under the same conditions [18], whereas the proposed feedforward compression technique, gives a very robust performance for trick play and error concealment. Last but not least, the fixed-length coding of small segments allows to design a very robust tape format.

6.6 Macroblock-based SB format

In this section we address globally the construction of the compressed data format in a Sync Block (SB). The format needs to be robust, particularly for the special modes (trick play) of the recording system, where a part of the error-correction coding (ECC) cannot be applied, because only data portions of a track are recovered (see Fig. 9). This multiplies considerably the chance of having errors in the signal. Secondly, at higher tape speeds, the head-to-tape contact is reduced and less stable, which also leads to a lower robustness. It is therefore required to construct a tape format which enables the compression decoder to cope with residual errors in the video data.

In order to construct a robust format, it is absolutely essential to limit the propagation of errors that emerge from erroneous variable-length decoding. The propagation of errors is limited in three ways which are briefly discussed below.

1. Fixed-length coding of segments. This is fulfilled by the choice of a feedforward coding scheme. Every segment is compressed into a fixed bit cost, so that the decoder should periodically reset itself at a segment border. A proper numbering of SBs allows to identify the start of a new segment without using the compressed data. Error propagation from segment to segment is therefore impossible. Care has been taken to ensure that the decompression does not rely on any coding parameters of previous segments whatsoever.

2. Identification of individual macroblocks. A segment consists of 5 full-colour macroblocks as indicated in the previous section. In Section 5 it was elaborated that this optimally fits on 5 SBs. For robustness, a single macroblock is put into a single SB. Note that we are considering a data block within a segment, which is variable-length encoded, meaning that the macroblock is sometimes smaller than a SB and sometimes larger. In the former case, all data of the macroblock is contained so that full decoding is enabled. In the latter case, the most important data, i.e. low-frequency DCT coefficients, should be inside the SB. Note that every SB has a fixed unique location on tape. As a result, each macroblock—at least the low-frequency information—can be addressed.

3. Identification of individual DCT blocks. Within a SB (one macroblock) 6 DCT blocks are located. Each compressed DCT block is of variable length. Similarly to the macroblocks, by putting the low-frequency data of each DCT block on a fixed position (see e.g. [23]), each DCT block can be addressed and partially decoded, and error propagation is limited to high-frequency components of DCT blocks only.

The internal SB format is depicted in Fig. 19. A group of 5 SBs forms a fixed-length segment, preventing error propagation. As a bonus, the fixed-length segment compression allows replace-
Figure 19: Inner SB data format showing the fixed predetermined positions of low-frequent DCT coefficients of each block, and the construction of segments using these sync blocks.

The robustness of the format is best illustrated by an example. Let us suppose that an uncorrectable error occurs in SB0 and consider decoding of SB1 of the same segment. If MB1 is smaller than 77 Bytes, there will be no decoding errors in MB1. If the MB1 is larger than 77 Bytes, some of the high-frequency information will be in another SB (e.g. SB0), so that generally some DCT blocks (luminance and/or colour) of MB1 will show high-frequency noise after decoding. The robustness of the system is enhanced by putting as much as possible high-frequency information of the actual DCT blocks of a MB in the corresponding SB.

6.7 Comparison with MPEG

Having presented the compression system in considerable detail, it becomes possible to highlight a number of significant differences with the popular MPEG standard.

Firstly, the variable bit cost per macroblock and data slice in MPEG has little robustness against residual channel errors when compared to the feedforward coding system presented. Firstly, the data is not protected against error propagation, whereas in the DV system, intrinsic synchronization is coupled to the sync blocks in the channel.

Secondly, the MPEG data need decoding in a predetermined time order, which is particularly problematic for implementing trick modes (e.g. fast picture search, especially in reverse). In contrast with this, the DV system allows independent manipulation of individual video segments, or even macroblocks (when some high-frequency noise is accepted). The decoding of data slices in MPEG also relies on previous slices, which is not the case in the DV system.

Thirdly, local error concealment is difficult in MPEG and well possible in the DV system. The independent (de)coding of segments in the DV standard allows replacement of individual segments, whereas in MPEG full decoding of the data is required.

Fourthly, editing of sequences is in MPEG only possible on Group-of-Picture (GOP) borders (a reasonable independent unit of compression), whereas the DV system enables editing of individual frames. It should be noted that even editing on a GOP basis in MPEG is already very complicated, because the GOP size is not fixed (a part of the information on a track where a GOP transition occurs, cannot be overwritten). Instead, the DV system has a fixed track cost for each video frame and the border of a frame always coincides with a track border.

Fifthly, there is a great difference in complexity between MPEG compression and the DV system. Since a compression encoder and decoder are both required in the camcorder application, a very complex encoder such as in MPEG heavily based on (memory-expensive) motion compensation, should be avoided, taking into account the desire to enter in the market in the mid nineties.

The aforementioned reasons have provided substantial motivation to implement a different compression system.
7. Data shuffling for trick play

7.1 Introduction

This section concentrates on the highlights of the macro block shuffling for video encoding and the mapping of MBs on tape. It was elucidated in preceding sections that a coding unit, called video segment, consists of 5 MBs only. Taking neighbouring MBs to construct a video segment leads to the unattractive situation where most segments contain the same (local) type of image statistics. This leads to the phenomenon that some are easy to compress (e.g. flat area) and others are difficult to code. It is a well-known technique in picture coding to combine easy and difficult to code picture areas in one segment in order to smooth statistics and obtain stable quantization (see e.g. [18]). This results in the best overall performance. A regular distribution over the picture area of the MBs for one segment results in the highest average picture quality for all practical video material. Such shuffling of 5 MBs is globally described in [8]. We will describe macroblock shuffling in more detail, taking the 625/25 system as an example.

7.2 Macroblock shuffling

As depicted in Fig. 8, a picture consists of 480/576 lines of 720 pixels wide. It was already elaborated that the number of lines is linked to the number of tracks to record one video frame. Although the distribution of the information in the tracks will be described later, it is clear that the data of 48 lines or 135 MBs is stored in one track. In the case of the 625/25 system, the picture of 36 by 45 MBs is divided into 12 horizontal rows of 3 MBs high and 45 MBs wide, where each row corresponds to one track. In the case of the 525/30 system, the picture of 36 by 45 MBs is divided into 12 horizontal rows of 3 MBs high and 45 MBs wide, where each row corresponds to one track.

On the other hand, one segment consists of 5 MBs which should originate from distributed parts of the picture, preferably with the highest distance. A maximum horizontal distance is achieved when they are distributed regularly, leading to a horizontal pitch of 9 MBs. Consequently, the picture is divided in 5 columns of 9 MBs wide. The row/column structure is depicted in Fig. 20.

A unit of 3 by 9 MBs is called a superblock. A picture consists of superblocks $S_{i,j}$ with $i,j$ being the row and column number, respectively. The numbering of the macroblocks within the superblock can be found in Fig. 21. The construction of a superblock for the 525/30 system is somewhat different due to modified macroblock dimensions (see Fig 8). Despite this difference, the 525/30 picture is divided similarly in 10 rows and 5 columns. The horizontal distance between the MBs constituting a segment has been discussed already. In the vertical direction, a regular distribution with maximum distance is also the optimal choice. Taking into account the 10/12 rows for both systems and a universal algorithm, a distance of 2 rows is the best option.

Since the distances are known now, a suitable algorithm has to be defined such that the 5 MBs forming one segment are spread out over the picture area. The following algorithm has been adopted:

<table>
<thead>
<tr>
<th>Track</th>
<th>MB 1</th>
<th>MB 2</th>
<th>MB 3</th>
<th>MB 4</th>
<th>MB 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 20: Selection of MBs for segment construction and assignment of picture areas on tracks.

Figure 21: Ordering of macroblocks in repetitive clusters, called superblocks.
Figure 22: Reorganization of macroblocks on tape for trick play so that the picture is stored in coherent form on tape.

\[ V_{i,k} = \sum_{p=0}^{4} MB[(i + 2p) \mod n, 2p \mod 5, k] \]  
with \(0 \leq i \leq n - 1, 0 \leq k \leq 26\) and \(n = 10\) or \(12\). Furthermore, \(M[i, j, k]\) denotes the macroblock number \(k\) in superblock \(S_{i,j}\). The MBs forming segment \(V_{0,0}\) are indicated in Fig. 20. The order in which the individual macroblocks from a segment (the black blocks in Fig. 20) are processed by the video compression system is from the center of the picture to the edges. The motivation for this different order is beyond the scope of the paper.

### 7.3 Macroblock mapping

One might expect that the 5 MBs of one video segment are stored in 5 subsequent SBs on tape. However, editing and picture reconstruction in normal play are performed on a picture basis, so that the mapping can be optimized for trick play. Consequently, the data of one picture is recorded in an alternative order within the boundaries of the set of tracks assigned to the complete picture.

The choice of the optimal mapping is based on the following observations.

1. **Coherency**
   The best subjective trick-play picture quality is obtained if the largest possible coherent picture parts are refreshed in one piece.

2. **Speed adaptivity**
   The picture quality should be adaptive to the speed, which means that the highest quality should be achieved for the lowest search speeds. As described in the Section 5.3, such an option is intrinsically available, because the recovered data burst length from tape increases for lower speeds (see Fig. 11). As a logical consequence, neighbouring pictorial information should be recorded side by side in a track.

The previous insight is in contradiction with the demands on macroblock shuffling for efficient video encoding as described in Section 7.2, where the pictorial information should be spread out over the picture. As indicated in Fig. 19, the majority of the information for one MB is stored in one SB. This is an attractive property for trick play. It enables decoding on a SB rather than on a segment basis resulting in a slightly reduced but still adequate picture quality for trick play. This feature allows for a block mapping on macroblock basis instead of segments, thereby solving the paradox between macroblock shuffling and actual mapping on tape.

The selection of the optimal mapping for a large set of trick-play speeds is a difficult issue, especially if the possibilities for alternative scanners have to be considered simultaneously. The highest flexibility in all situations is achieved by a more or less direct projection of the picture on the tape. This is realized by recording the superblocks in one row of Fig. 20 one after the other in a track, with the row number equal to the track number. MBs within a superblock are stored in the order indicated in Fig. 21. A schematic representation of the final mapping is given in Fig. 22.

Despite the flexibility of the mapping proposed, certain trick-play speeds still perform better than others. Specific parts of the picture...
must be refreshed as often as possible. This is strongly influenced by the relation between the speed factor and the number of tracks required for one picture. Let us consider as an example a speed of 12 times the nominal speed for the 625/25 system with 12 tracks per frame. Examining the parts read from tape will quickly reveal that, irrespective of the mapping chosen, a substantial part of the picture (around 95%) will never be refreshed. This leads to static or even black picture parts on the display, which is of course unacceptable. Similarly, it is concluded that all dividers of the number of tracks should be avoided for the same reason [18]. More detailed analysis shows that in fact all integer speed factors are unusable, but that a fractional speed like 11.5 should be used. In general, it can be concluded that the search speed should be such that all other picture parts are refreshed between the updates of a specific part.

8. Conclusions

We have presented a number of design considerations for the video compression system in the DV recording system standard. It has been clarified that the adopted cassette size and tape thickness have determined the required compression factor and the resulting bit rate of 25 Mbit/s after compression. At this bit rate, sufficient playing time can be obtained, both for a ultra-small camcorder cassette and the larger desktop cassette.

The compression system has been optimized for low-cost intraframe coding because of the use in portable equipment. This can be seen from e.g. simple quantization techniques using uniform quantizers based on shifting only and from the single-table runlength coding algorithm. The system yields a high picture quality, even in the case of motion, due to a special motion adaptivity.

The independent compression and coding of segments, based on five macroblocks only, allows for high search speeds during trick play and provides a very robust tape format. In normal play, the powerful R-S product ECC ensures an error-free and flawless reproduction of the recorded images. A special macroblock format mapped onto single channel sync blocks enables data recovery at very high search speeds. Even at high search speeds, the format limits error propagation severely and shows only errors in high-frequency information. Moreover, a special data shuffling scheme enables a relatively high perceptual performance during trick-play modes.

References


Peter H.N. de With was born in Lexmond, The Netherlands, in 1958. He graduated in electrical engineering from the University of Technology in Eindhoven. In 1992 he received the Ph.D. degree from the University of Technology Delft, The Netherlands, for his work on video bit-rate reduction for recording applications. He joined Philips Research Laboratories Eindhoven in 1984, where he became a member of the Magnetic Recording Systems Department. From 1985 to 1993 he has been involved in several European RACE projects (a.o. 1001 and 2026) in which digital SDTV, HDTV and data recording was studied. In the early nineties he contributed as a video coding expert to the DV standardization committee. Since 1994 he became a member of the TV System group where he is working on advanced video processing architectures for various enhancements and compression systems. Regularly, he is a teacher of the Philips Centre for Technical Training. In 1995 he co-authored the paper that received the IEEE CES Transactions Paper award. In 1996, he received a company Invention Award.

Albert M.A. Rijckaert was born in Eindhoven, The Netherlands, in 1945. He received the Ing. degree in electrical engineering from the Polytechnic College in Eindhoven in 1966. He joined Philips Research Laboratories Eindhoven in 1968 and has been a member of the Magnetic Recording Systems Department since then. He has worked on numerous subjects related to magnetic recording, such as the physical recording process, and tracking and skew correction systems for linear recording. Since 1982, he is involved in system research on helical-scan recording and he contributed to various recording system standards among which the DV and the D-VHS system. Mr. Rijckaert has participated in several European projects on digital video recording. In 1996, he received a company Invention Award.