Data Compression Systems for Home-Use Digital Video Recording

Peter H. N. de With, *Member, IEEE,* Marcel Breeuwer, and Peter A. M. van Grinsven

Abstract—**Newly developed communication- and information networks offer a large number of services which make use of image data, leading to an increasing demand for image storage systems. This paper focuses on a new emerging technology, namely image data compression techniques for digital recording. Image coding for storage equipment covers a large variety of systems because the applications differ considerably in nature. In this paper, video coding systems suitable for digital TV and HDTV recording and digital electronic still picture storage are considered. In addition, attention is paid to picture coding for interactive systems, such as the compact-disc interactive system. The relation between the recording system boundary conditions and the applied coding techniques is outlined. The main emphasis is on picture coding techniques for digital consumer recording.**

I. Introduction

In the coming years, digital recording will gradually take over from conventional analog recording systems. This is motivated by three distinct developments. First, continuous improvements in the head and tape technology have resulted in a substantial growth in recording densities. Digital recording can benefit more from these density improvements than analog recording, because a high carrier-to-noise ratio is not required [1]. Second, the initial generation of video registration equipment in the professional field is gradually being replaced by digital recording systems, e.g., *D*-1 and *D*-2 recorders [2]. Apparently, the benefits of digital recording, namely a flawless reproduction of the video signal even after multiple copying and its unquestionable timing stability for video and audio, have been recognized. A third argument for the introduction of digital storage equipment is the growth in VLSI technology which has made advanced digital processing feasible for widespread use in practical applications. Only recently has it been proven that for video speeds (e.g., sampling frequency exceeding 10 MHz) VLSI is feasible while still using clock- and throughput rates of 20–40 MHz. This enables advanced coding techniques with considerable computational complexity to be applied on a larger scale.

A major transition to digital recording is still yet to come, namely digital recording for consumer applications. Here, constraints differ substantially from those in recording for professional use. In the latter field, signal processing is avoided because the input signal is preferably registered in its original form. In the consumer market, however, simple mechanics play a key role in order to obtain small-sized devices for portable applications. It is recognized that data compression techniques are the means to achieve that goal. A substantial compression factor can be obtained because the compression system needs not be lossless, since the picture quality constraints for consumer applications are less stringent than in the professional field. Another reason is the progress in the coding efficiency of recently developed algorithms, resulting in considerable bit-rate reduction while maintaining the subjective picture quality. Therefore, investments in VLSI to reduce the complexity and size of recording mechanics are valuable.

This paper reports on the picture coding techniques which have been implemented for experimental digital recording systems in the past decade. The emphasis will be on picture coding for experimental digital home-use systems, especially video tape recorders. Additionally, the first developments on HDTV consumer recording will also be discussed. Finally, coding techniques for special applications, as digital electronic still picture storage (ESP) and compact disc interactive (CD-I), are described.

The processing blocks in digital recording systems resemble a digital communication system. Fig. 1 shows a system block diagram of the video processing part of an experimental digital video recorder. First, the analog TV video signal is digitized. For coding, composite PAL/NTSC or YUV-component signals can be considered. Sampling frequencies for composite signals are taken as an integer multiple of the subcarrier frequency in order to preserve the modulated color information. More recent system proposals refer to YUV-component coding. Since there is no digital component sampling standard for consumer applications yet, the sampling rates are a variable parameter for the system designer. For studio environments, however, a worldwide sampling standard for component TV signals has been agreed upon, namely the CCIR-601 standard. This standard uses 13.5 MHz sampling frequency for the luminance (Y) signal and 6.75 MHz for the color-difference signals U and V. This results in 720 and 360 active samples/line for Y and U/V, respectively. In 50 Hz systems 576 lines are active and in

Manuscript received September 1990; revised June 1, 1991.

The authors are with Philips Research Laboratories, Eindhoven, The Netherlands.

IEEE Log Number 9104239.

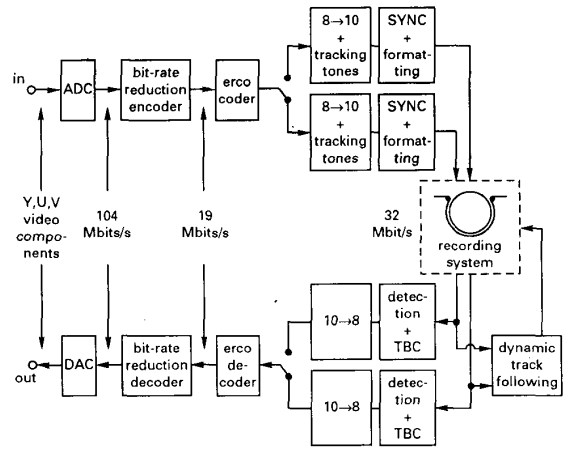


Fig. 1. Block diagram of a digital video recorder (from [3]).

60 Hz systems 485. Mostly, sampling frequencies for component signals are chosen as an "easy-to-handle" fraction of the CCIR-601 sampling rates, e.g., 3/4, 1/2, etc.

Bit-rate reduction is applied to compress the digital video signal, such that the resulting bit rate yields the required playing time on the chosen cassette size. The compression algorithms are described in detail in Sections III–V. It is emphasized here that the combination of sampling and bit-rate reduction determines the subjective image quality of the system. For this reason, one can distinguish several compression techniques with the same output bit rate. Each specific solution is a tradeoff between resolution (sampling frequency) and the applied bit-rate reduction system (attainable compression factor) to achieve the desired output bit rate for recording.

The third step in the diagram of Fig. 1 involves channel coding. It is known that the magnetic recording channel suffers from burst errors. These burst errors are mostly caused by dirt and scratches on head and tape, resulting in temporal signal losses. Another problem is the tape and head wear in the system. Repetitive playback of the same tape may damage the magnetic layer of the tape, which also increases the bit error rate significantly. To overcome this problem, error correcting coding is needed at the expense of some channel capacity. In practice, Reed–Solomon codes [4] are often used in a digital recording system. These codes make efficient use of the extra parity symbols which must be added. Recently, product codes based on Reed–Solomon codes have gained more interest. In such a code, data are organized in two-dimensional blocks, in which the parity symbols are added to both rows and columns. Block-based processing breaks large burst errors into smaller, manageable pieces. Even relatively simple decoding strategies yield a robust system performance. Typical values for the amount of redundancy range from 15 to 30%.

The fourth and last step in the digital recorder is to match the bit stream to the properties of the magnetic recording channel and to add tracking information with a channel code. The head-to-tape system itself has a bandpass character: dc current cannot be recorded and too many flux reversals per length unit result in intersymbol interference (ISI). Both phenomena can be controlled by posing constraints on the minimum and maximum distance between two signal transitions (hence, bit reversals) in the channel. This can be obtained by using run-length limited codes [5] or by using a block code [6], e.g., an 8–10 code as indicated in Fig. 1. Basically, tracking information is added in two ways: frequency division multiplex or time division multiplex (TDM). In the example described (Fig. 1), tracking tones have been added by encrypting them in the channel code. An example of TDM is to reserve a special part of the track for tracking information, as in the R-DAT system.

Section II is devoted to the specific system constraints for recording. Three items are discussed in some detail: the tradeoff between playing time and video bit rate, the consequences of variable-speed playback, and the fluctuating performance of the recording channel.

Section III focuses on data compression for TV consumer recording. A distinction has been made between predictive coding and transform coding systems. The use of predictive coding in storage systems has been investigated for both composite and component signal recording. Recent developments are pointing in the direction of more advanced systems, such as transform coding in combination with variable-length coding. Special attention will be paid to the data buffering when variable-length codes are used. This is a very relevant problem in recording because editing and other special modes must be possible.

Section IV discusses the first attempts of HDTV coding for consumer recording purposes. In this section, a very specific problem will be dealt with, namely compatible coding of HDTV signals. In such a system the HDTV signal is split into a "TV-like" signal and a "surplus," thereby enabling gradual introduction of HDTV systems on the market.

Section V deals with digital electronic still picture storage and its coding. We have included a section about digital still picture storage because it is our opinion that such a facility will become an integrated part of a video camera recorder. The tendency is therefore that techniques which have been exclusively considered for electronic still picture coding will coincide to a large extent with the systems developed for digital video tape recording.

Finally, Section VI considers moving video coding for very low bit rates. The application is digital storage of video signals for interactive systems, such as the compact disc interactive (CD-I) system, in which a bit rate of only 1 Mb/s is available for the compressed video data. It will be shown that motion-compensated interpolation plays a key role to obtain a sufficiently high compression factor.

II. Coding Constraints for Digital Recording Systems

The possible complexity reduction of recording mechanics by using picture coding may be best clarified with the following example. Suppose a composite standard TV signal is sampled with four times the color subcarrier frequency. The required bit rate to store the digital signal is around 100 Mb/s. If per video head 25 Mb/s can be recorded (a reasonable assumption for consumer type magnetic heads), a four-channel system is needed, without considering error correcting coding. A simple data compression system with a compression factor of two would halve the required amount of recording channels, at the expense of a chip containing the additional signal processing. With large scale production and in the longer term, VLSI is less expensive than a plurality of magnetic heads. Hence, the overall costs of the recording system would be reduced, thereby favoring the exchange of mechanical complexity to more advanced signal processing.

In the following, three key aspects of coding for recording are highlighted to indicate that it differs from transmission systems.

The first aspect is related to the aimed recording system size. This results in a more or less predetermined compression factor that must be obtained. This constraint is clearly portrayed by Fig. 2, in which the available video bit rate is plotted as a function of the playing time of an experimental 8 mm digital recording system. In the curve two points are of particular interest. If no data compression is used, the required bit rate for the video signal would be 166 Mb/s (a CCIR-601 signal), resulting in an (unacceptable) playing time of 10–15 min. If, on the other hand, a playing time of 2–3 h must be realized, the available video bit rate is only 15–20 Mb/s. In other words, the chosen mechanical configuration can impose severe limitations on the available bit rate for the video signal, thereby necessitating advanced compression systems.

The second aspect in which recording differs from transmission is that a recorder can be operated at varying speeds. Such an operation mode, often referred to as a multispeed mode, dramatically influences the amount of information that can be recovered from the tape. This is indicated in Fig. 3 with and without head actuation. By head actuation we mean that the position of the head is modulated by some electromechanical system, e.g., by using a piezo crystal, as a function of the search speed. If no head actuation is used, the position of the video heads is fixed on the drum. The drum is a cylinder around which the tape is wrapped. At an increased playback speed, the head(s) cross several tracks during one scan from the lower part to the upper part of the tape. Only in cases where the head sufficiently covers a track and the azimuth (i.e., the angle between head gap and the transversal track direction) of head and track are equal, can data be recovered. If head actuation is applied, a larger

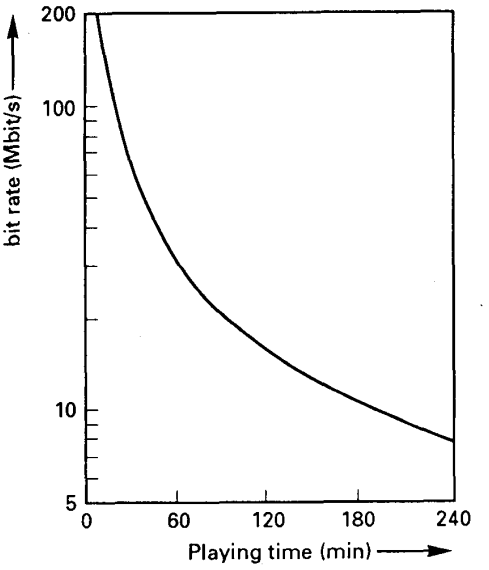


Fig. 2. The net available bit rate of an 8 mm system as a function of the playing time. Track width 10 μm, bit length 0.3 μm, effective tape use 6.7 mm width and 120 m tape length. Overhead: 20% error correction and 8–10 channel modulation.

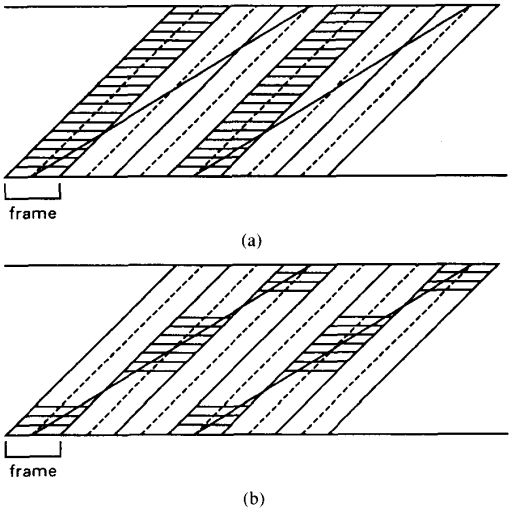


Fig. 3. Recovered data bursts (arced areas) in the case of increased playback speed (three times the normal speed). In the example, two tracks per frame are used. Head actuation enables reading of a complete frame and then skipping of the two subsequent frames. Recovered data bursts without head actuators. (a) Recovered data with head actuation. (b) Recovered data without head actuation.

consecutive area can be read from tape at the cost of an increased area in which no data can be recovered. The conclusion of both situations is similar: the multispeed operation of a recorder leads to data recovery in burst mode and the bursts originate from varying positions on the tracks. If the search speed of the recorder increases, the data bursts will become smaller. If in the described

mode complete images must be reconstructed, the decoder should be able to, at least partially, decode the incoming data bursts. This limitation requires a data compression technique in which coding is limited in time, e.g., it should allow decoding of data parts without knowledge about previous data parts that could not be restored. The limitation in time holds also for editing, i.e., insertion or replacement of other video frames within a previously recorded video sequence.

The third aspect in which the recording process substantially differs from the transmission link is the reliability of the communication channel, especially magnetic recording channels. The performance of this channel in terms of bit error rates can fluctuate greatly or a rapid deterioration may take place for reasons pointed out in the previous section. In order to enhance system robustness, one can notice a substantial amount of redundancy added to the source data prior to storage in the proposed experimental recorders. However, the measures indicated are not enough in the multispeed modes. Since only pieces of information are restored, error correction can only be partially carried out because not all data of the same (product) data block on which the data protection was based could be recovered. Another reason for decreased robustness during multispeed processing is the, mostly, increased tape speed, so that stable data reading from tape is more difficult. The conclusion of these conflicting situations is that the data format after compression must be robust in nature. Also the applied codes can be optimized for system robustness, for example, by giving them special synchronizing properties [7].

III. Picture Coding for Digital TV Consumer Recording

This section deals with video coding techniques for TV that have been proposed for consumer digital video recording. In the last decade, the diversity in compression algorithms has grown extensively. Excellent reviews on picture coding in general have been written by Netravali and Limb [8], Jain [9], and Mussman *et al.* [10]. In this section we focus on the specific systems for digital TV recording for home use. Digital signal processing for analog TV recording will not be considered. The emphasis in this section will be on specific details of data compression systems which make them more suitable for application in digital recording systems. The limitations from the previous section have hampered the use of *interframe* coding techniques, in which the time correlation between frames is exploited, for recording applications. Therefore, only field- or frame-based systems are discussed.

The section is written more or less in chronological order. Section III-A deals with predictive coding systems, in particular differential pulse code modulation (DPCM). The majority of the systems are based on the recording of composite signals. The popularity of DPCM is to a large extent based on its simplicity so that the feasibility of the recording systems could be tested. Section III-B acts as a bridge between predictive coding and transform coding, since it deals with a relatively simple combination of both techniques. Section III-C is devoted to transform coding systems and is split into two parts. First, [Section III-C-1)] the Hadamard transform is discussed, which is primarily implemented with fixed-rate coding systems. Section III-C-2) focuses on the discrete cosine transform and its proposed variable-wordlength coding techniques.

For the sake of completeness, we report the results of one of the earliest investigations of digital consumer recording, namely PCM video recording by Mita *et al.* [11]. In their proposed system, bit-rate reduction techniques are not used at all. The video is sampled at 10.7 MHz (with 8 b resolution) as a composite signal, or the separate components are sampled at 7.16 MHz for luminance and 1.79 MHz for the color components. Bit rate for the active video part is around 86 Mb/s. A 1 h playing time has been obtained on VHS-like mechanics. The authors conclude that bit-rate reduction techniques provide very effective means for increasing the playing time or reducing the size of the system.

A. Predictive Coding

Because of its simplicity, predictive coding had been investigated extensively and over a long time period. The most interesting system is differential pulse-code modulation (abbreviated as DPCM). The fundamentals of DPCM can be understood easily from the block diagram in Fig. 4.

The basic property of the system is that incoming samples $s(i)$ are predicted by means of previously processed samples. The prediction $p(i)$ is a weighted sum of neighboring samples that can be assumed to correlate with the sample to be coded. The difference $d(i)$ between the predicted and the present sample value is determined, and then quantized. The quantized differences $d_Q(i)$ may be coded prior to storage in order to enhance channel properties. In the decoder, the quantized difference is added to the prediction value, resulting in the reconstructed sample value $\hat{s}(i)$. In order to keep predictions in encoder and decoder equal to each other, the encoder predictor uses quantized difference inputs as well. In a more formal notation, DPCM can be formulated with the following recursive procedure:

$$\begin{aligned} p(i) &= \sum_{k=1}^{K} a_k \hat{s}(i-k), \\ d(i) &= s(i) - p(i), \\ d_Q(i) &= Q(d(i)), \\ \hat{s}(i) &= p(i) + d_Q(i). \end{aligned} \tag{1}$$

In (1), the K different weighting factors a_k are the prediction coefficients. The function Q quantizes the differences to a coarser value. Generally, large difference values are quantized more coarsely than small values. For a more detailed introduction of DPCM systems we refer to [10]

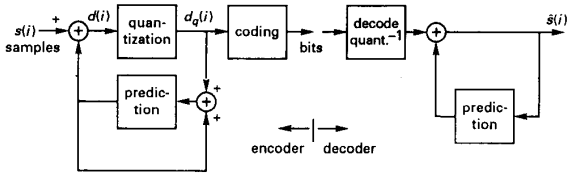


Fig. 4. Block diagram of DPCM system.

in which the most recent improvements are discussed. In the following, we focus on DPCM systems which have been designed exclusively for digital recording.

One of the first investigations using DPCM for experimental digital recording has been reported by Hirota, Hirano, and Higurashi [12]. The developed recorder is based on modified VHS mechanics with a recording time of 2 h. For data compression, a scheme has been implemented which is suited for composite recording. The main steps in the coding system are indicated in Fig. 5. In the first step, sub-Nyquist sampling on a field basis is carried out. Second, the subsampled image is coded with DPCM. The subsampling already reduces the bit rate by a factor of two, thereby enabling a higher overall compression ratio.

The initial sampling frequency is $4f_{sc}$ (the color-subcarrier frequency) with 7 b resolution, yielding 100 Mb/s input bit rate for NTSC. The subsampling pattern is depicted in Fig. 6. As can be noticed, it is a combination of line-quincunx and field-quincunx sampling. The sampling pattern must allow for a good reconstruction of color information modulated on the subcarrier. Due to the combination of line- and field-subsampling, the sampling pattern is a periodic four-field sequence. Interpolation for image reconstruction is performed by using luminance samples of the same line and the corresponding line in the previous field, and the nearest chrominance sample of the previous field (see Fig. 7).

The DPCM system is based on an intrafield prediction technique, in which the subcarrier phase is taken into account. The prediction is defined as follows:

$$X = P_2 + (P_3 - P_5). \tag{2}$$

The location of the sample values is shown in Fig. 8. Quantization of the difference signal is adaptive to the input levels and nonuniform. A fixed-rate coding of 4 b per quantized difference has been proposed. The constant data rate after DPCM is 22 Mb/s, and 28.6 Mb/s in the recording channel.

A similar coding approach for composite PAL television has been followed by Driessen *et al.* [13]. Again as a first step sub-Nyquist sampling on $2f_{sc}$ is employed prior to a DPCM coding technique. The intrafield predictor, which is suitable for both luminance and chrominance parts in the signal, is a weighted sum of 9 samples from the current and last two lines of the same field. The DPCM system is portrayed by Fig. 9 in which the predictor is denoted as P_{NL}. In the following, the use of nonlinear prediction for improving the error response of the system is described. Two nonlinear memoryless elements NL_i are used for two different predictors P_1 and P_2, which are

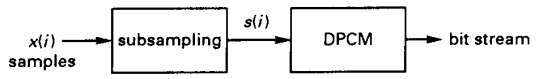


Fig. 5. Encoding diagram using subsampling and DPCM.

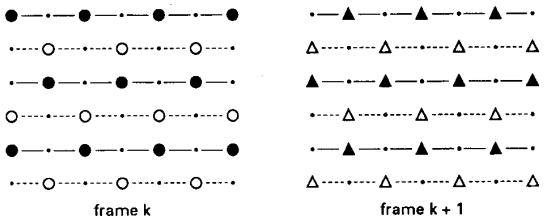


Fig. 6. Frame-alternating subsampling pattern, which is periodic over four fields, for recording of composite NTSC signals. The dotted lines represent the lines of the second field.

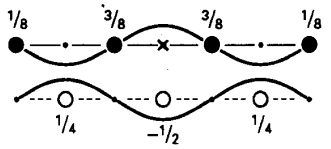


Fig. 7. Samples used for interpolative reconstruction. The sample pattern corresponds to Fig. 6. The numbers are filter coefficients.

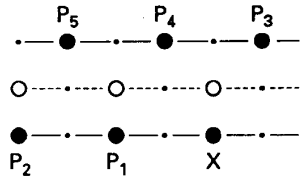


Fig. 8. Intrafield prediction for composite NTSC, based on the sample pattern obtained after the subsampling of Fig. 6.

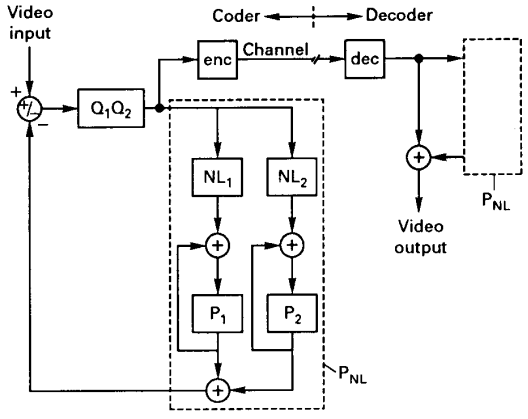


Fig. 9. Architecture of DPCM system with nonlinear prediction (from [13]).

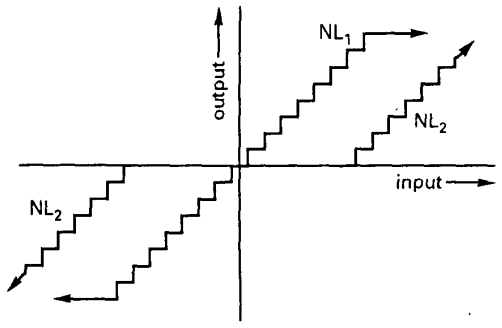


Fig. 10. Characteristics of nonlinear elements NL_1 and NL_2 (from [13]).

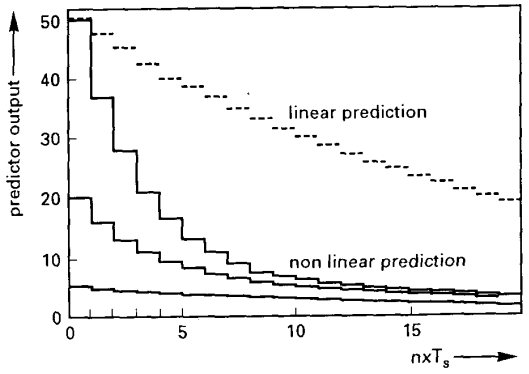


Fig. 11. Error response of nonlinear predictor compared with linear prediction as a function of the sample distance. The solid curves indicate the response for various inputs (from [13]).

working in parallel. The characteristics of the nonlinear elements are shown in Fig. 10. For small quantizer outputs occurring in relatively flat image areas, predictor P_1 is active, whereas predictor P_2 is active in high-contrast picture areas. Channel errors have a limited impact on the output of P_1, since NL_1 clips large variations of the input signal. Predictor P_2 is optimized for high-frequency response combined with rapid error decay (much leakage). The overall result of the predictor is the sum of the two terms, which is substantially more robust against channel errors than a linear predictor (see Fig. 11). The quantization (in Fig. 9) is adaptive to the input levels and switches between two tables Q_1 and Q_2. On signal transitions, the system switches to coarser quantization Q_2 for the next input, but delays the return to fine quantization Q_1 by 3 sample periods, in order to avoid quantizer oscillation. Promising results with this DPCM system have been reported on 22 Mb/s (24 Mb/s in the channel) in comparison with a simple Hadamard transform coding technique (see Section III-C).

A more recent result in DPCM picture coding for recording, that reflects the transition to component coding, is outlined in [14]. In the system proposal, the composite signal is initially sampled at $4f_{sc}$, but then converted to a 230 Mb/s YUV-component signal with $4f_{sc}/2f_{sc}/2f_{sc}$ sampling frequencies. By resampling this signal to $3f_{sc}/0.5f_{sc}/0.5f_{sc}$ and putting the color in the blanking of the luminance, a TDM signal is obtained with 78 Mb/s bit rate. This signal is then compressed to 30 Mb/s with an interframe DPCM system. The system works in fixed-length coding of samples, but operates in various modes to improve the overall picture quality. The basic technique is to code blocks of 8×8 samples, the first sample of each block being coded in PCM to avoid error propagation on block basis. Depending on the total bit cost of a coded field, the correct setting of the modes is chosen. The various modes are indicated in Table I.

TABLE I
DIFFERENT MODES IN BLOCK-ADAPTIVE DPCM SYSTEM WITH SUBSAMPLING (FROM [14])

	Coding Methods	Code Length	Comp. Ratio
Mode 1	In-block DPCM with a	5 b/s	5/8
Mode 2	combination of:		
	—no compression (8 b)	4 b/s	4/8
Mode 3	—previous sample/line prediction		
Mode 4	—two-dimensional prediction	2 b/s	
Mode 5	In-block DPCM using the same method as above after	4 b/s	
Mode 6	2/1 subsampling		2/8
Mode 7	DPCM with a combination of noncompression and interframe prediction	2 b/s	

B. Dynamic Range Coding

Dynamic range coding can be regarded as a combination of block coding techniques and DPCM. The basic ideas have been proposed by Kondo *et al.* in [15]. The technique works as follows. The image is partitioned in small subblocks of, say, 4×4 samples. The minimum sample value inside the block is transmitted as a global prediction for all other sample values inside the same block. The difference between the minimum value in the block and the actual sample value is quantized and coded. As with DPCM, it is attractive to quantize these sample differences adaptively. Therefore, it has been proposed to use the maximum difference, i.e., the difference between the maximum sample value and the minimum value in the block, as a classification parameter. This maximum difference is termed as a dynamic range in the following. Hence, in more formal notation, adaptive dynamic range coding (ADRC) is based on the following steps:

$$\begin{aligned} \text{MIN} &= \min\{s(i,j), && i,j = 0, 1, \cdots, N-1\}; \\ d(i,j) &= s(i,j) - \min, && i,j = 0, 1, \cdots, N-1; \\ DR &= \max\{d(i,j), && i,j = 0, 1, \cdots, N-1\}; \\ d_Q(i,j) &= Q_{DR}(d(i,j)). && \end{aligned} \tag{3}$$

In this expression, a sample from a two-dimensional frame is denoted by $s(i,j)$, MIN stands for the block minimum, the dynamic range parameter is DR, and $d_Q(i,j)$ is the quantized version (controlled by the value of DR) of the difference signal $d(i,j)$. Figs. 12 and 13 portray the fundamentals of dynamic range coding.

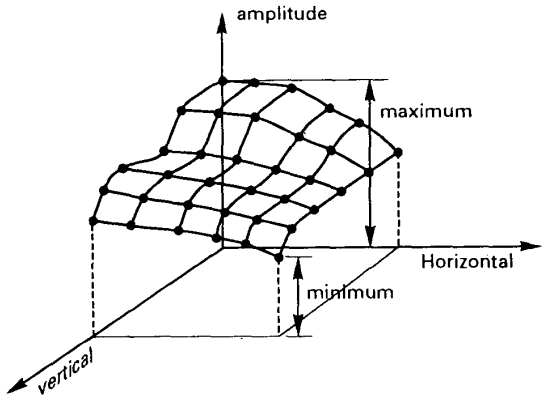


Fig. 12. Example of a sample value grid of a 6 × 6 block with minimum and maximum value indication.

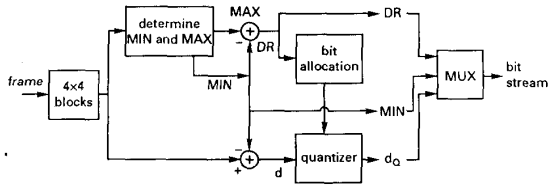


Fig. 13. Encoder block diagram of adaptive dynamic range coding.

The technique described can be extended to three dimensions by considering the corresponding block in the previous frame as well, thereby reducing temporal redundancy. Coding has been improved by distinguishing blocks without motion or with a significant motion component. In the former case, the blocks are averaged

$$s(i,j)' = [s_k(i,j) + s_{k-1}(i,j)]/2 \qquad 0 \le i,j \le N-1 \tag{4}$$

where $s_k(i,j)$ is sample location (i,j) from the kth frame. In order to decide whether (4) can be applied, motion detection is essential. A simple detector based on the maximum of the differences between corresponding samples in the two frames has been worked out. Hence, a 3-D block is coded in the motion mode if

$$\max\{|s_k(i,j) - s_{k-1}(i,j)|,\ 0 \le i,j \le N-1\} > T_m \tag{5}$$

in which T_m denotes the motion threshold. For editing purposes, Kondo *et al.* proposed to code groups of two frames as an entity. However, the averaging of (4) in the nonmotion mode results in motion judder if the motion threshold is increased to obtain a lower bit rate. For this reason, postfiltering has been implemented, which averages the two groups of frames during image reconstruction each time a new group of two frames is started (see Fig. 14).

The next coding step is quantization and coding. It has already been mentioned that the dynamic range parameter is well suited for quantizer classification. In a recent publication [16], it is shown that good results can be obtained by increasing the number of levels when the dynamic range augments. An example of such a strategy is outlined in Table II.

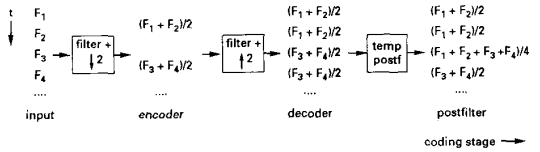


Fig. 14. Temporal subsampling and interpolation diagram during the encoding and decoding process for a sequence of frames F_i. The coding stages are indicated horizontally, while the subsequent frames are shown vertically.

TABLE II
EXAMPLE OF A CLASSIFICATION AND THE CORRESPONDING BIT COST AS A FUNCTION OF THE DYNAMIC RANGE (BIT-PLANE INTERVALS)

Dymanic Range	Level Bit Cost
0 · · · 4	0
5 · · · 13	1
14 · · · 35	2
36 · · · 103	3
104 · · · 255	4

Quantization has been optimized for multiple encodings resulting in a slightly modified linear quantizer. The starting point is a linear quantizer with the reconstruction levels in the middle of the quantizer intervals

$$d_Q = (d + 0.5)M/DR \tag{6}$$

where d and DR are defined according to (3), and M ($M > 1$) stands for the number of output levels. In the next step, the minimum and maximum output value are redefined by taking all input values d which lead to the minimum or maximum quantizer output, and averaging these input levels

$$\text{MIN}' = \frac{1}{K}\sum_i\sum_j s(i,j), \qquad \text{for } d(i,j) = 0,$$

$$\text{MAX}' = \frac{1}{L}\sum_i\sum_j s(i,j), \qquad \text{for } d(i,j) = M-1 \tag{7}$$

where K and L are the number of samples having $d(i,j) = 0$ or $M-1$, respectively. The modified dynamic range of the block DR' and the modified quantized output are

$$DR' = \text{MAX}' - \text{MIN}'$$

$$d_Q = (s(i,j) - \text{MIN}')/(M-1) + 0.5. \tag{8}$$

The specified quantizer preserves the dynamic range very closely (only the averaging of (7) results in a minor change) after multiple encodings. On the other hand, the new quantizer benefits from the regular linear quantizer since this one yields the highest signal-to-noise ratio.

The last point to be discussed is the buffer regulation part of the system. It may be noticed that the ADRC ap-

proach is very close to coding with M-functions [10]. It is known that the attainable compression ratio of such as system is limited, as with DPCM. Therefore, coding in groups of two frames has been adopted. To keep the data constant in such a period, the data are stored prior to coding and the dynamic ranges of all blocks are analyzed in a histogram of occurrences. On the basis of quantization Table II, thresholds TH_i can be defined (see Fig. 15) leading to the number of blocks S_i that have equal or more than i b/sample. The total number of bits (globally) per two frames I can be easily calculated by

$$I = 16S_1 + N^2 \left(\sum_{i=1}^{4} S_i \right). \tag{9}$$

The first term of (9) denotes the minimum amount of information coded for each block of $N \times N$ samples, namely 8 b for the minimum value and 8 b for the dynamic range. If the thresholds TH_i are modified, the number of blocks with a particular bit allocation changes, thereby varying the total bit cost I. The buffer control adjusts the levels in such a way that the bit cost fits in the desired bit rate. It can now be understood why variable-length coding is not attractive in this system: the bit cost computation would be dramatically more expensive, since all the quantized samples would be required to carry out the calculation.

The key aspect of the ADRC system is its simplicity and the small block size. This enables a simple feedforward analysis of two frames prior to coding. It also ensures multiple encodings without further picture quality degradation. However, the simplicity mentioned is at the same time the major drawback of the system: only a limited compression can be achieved, even if the three-dimensional processing over two frames is included. The authors report good subjective image qualities at 25 Mb/s with a 3:1:0 sampling standard (2.5 b/sample).

C. Transform Coding

Similar to the proposals for transmission, transform coding has become one of the most attractive systems for data compression in storage applications. In spatial transform coding, the input signal is segmented into small blocks of, say 8 × 8 or 16 × 16 samples. The transform attempts to remove spatial correlation by converting the block to more or less uncorrelated components, from now on referred to as *coefficients*. The transform itself does not fundamentally alter the signal, apart from changes due to finite precision calculation, but it yields only a different representation. The modified representation can be seen as a rotation of the coordinate system [17]. In the two-dimensional case, the unit vectors along the coordinate axes are basis pictures of which examples can be found in the following section. In most transforms, the aim is to concentrate the signal energy into as few coefficients as possible. This phenomenon paves the way for significant data reduction, since the nonrelevant coefficients may be skipped for coding. In Fig. 16 a block diagram is depicted of a transform encoder which indicates the major steps in

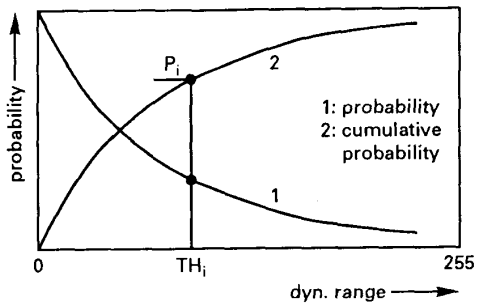


Fig. 15. Histogram of occurrences of various dynamic range values and the corresponding cumulative distribution for the number of blocks. The threshold TH_i specifies the fraction $(1 - P_i)$ of the total number of blocks that is coded with $\geq i$ bits/sample.

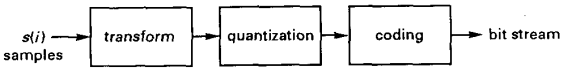


Fig. 16. Basic steps in a transform encoder.

the total system. Prior to transformation, a line-to-block converter enables the required block construction. A distinction can be made between intrafield transformation or intraframe transformation. In the former, blocks are constructed from samples within one field, whereas in the latter, the samples are taken within one frame (see Fig. 17). In intraframe coding, the interlacing of fields can play a substantial role in the coding algorithm. This influence will be discussed later in this section. After transformation, the coefficients are quantized and coded (this explains the term transform coding). Whereas the transformation yields another signal representation, a major step in data reduction is obtained in the quantizer by means of the suppression of irrelevant signal components. In the last step, coding of the coefficients ensures that the allocation of transmitted bits to the coefficients is based upon their relevance (e.g., statistical variance). For recording applications, it is attractive to use special codes with improved synchronization properties in order to enhance decoder robustness.

Let us consider the transformation in a more formal notation. The input sample block $f(i, j)$ of size $N \times N$ is transformed to the output block of coefficients $F(u, v)$ by means of the *forward transformation kernel* $T(i, j, u, v)$. The coordinates i and j refer to the sample domain, while the parameters u and v are active in the transform domain. The two-dimensional transform T is defined by

$$F(u, v) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) T(i, j, u, v). \tag{10}$$

Of all commonly applied transforms the transform kernel is *separable*, so that $T(i, j, u, v) = T_1(i, u)T_2(j, v)$. This means that the two-dimensional transform can be performed by carrying out subsequently two one-dimensional transforms. This property greatly simplifies the hardware realization of such a system. Evidently, a forward transformation only makes sense if its inverse also exists, the

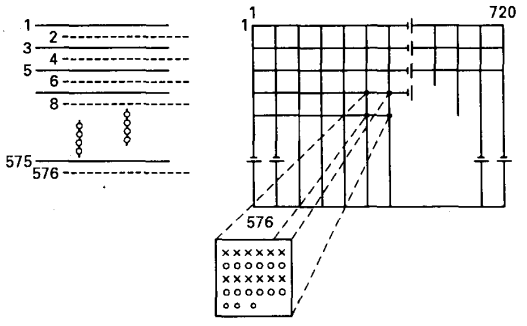


Fig. 17. Intraframe block segmentation prior to transformation for a 625-line 50 Hz interlaced television signal.

latter one being performed at the decoder

$$f(i, j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) T^{-1}(i, j, u, v) \qquad (11)$$

resulting in the sample block $f(i, j)$ again. The transformation kernel can be complex valued, such as, for example, the Fourier transform, or real valued. Numerous transforms have been proposed for transform coding, such as the Fourier, Hadamard, Haar, Slant, sine, and cosine transform, among which the Hadamard and the cosine transform have been considered most extensively for recording. The Hadamard transform is particularly interesting because of its simplicity in hardware realization, whereas the discrete cosine transform (DCT) is very efficient in energy compaction. In the following we will deal with these two separable transforms in particular. The one-dimensional kernels are defined by

$$\text{Hadamard: } T(i, u) = \frac{1}{\sqrt{N}} (-1)^{\alpha(i,u)},$$

$$\text{DCT: } T(i, u) = \sqrt{\frac{2}{N}} C(u) \cos [(2i + 1)u\pi/2N] \qquad (12)$$

where $\alpha(i, u)$ is an integer function of the sample index i and the transform domain index u. The constant $C(u)$ equals unity for $1 \leq u \leq N - 1$ and $C(0) = 1/\sqrt{2}$. In the system proposals published for digital recording, investigations deal with both component and composite NTSC and/or PAL coding systems.

1) Hadamard Transform: Hadamard transform coding has been investigated for composite PAL [13] and NTSC coding [18]. The main step in both approaches is to perform sub-Nyquist sampling prior to Hadamard transformation. Yamamitsu *et al.* [18] investigated the Hadamard coding technique by using field-quincunx subsampled NTSC as an input of the transform coder. The original sampling frequency is $4f_{sc}$, and the sampling pattern is orthogonal within each field. After subsampling, the sampling distance within each field is $1/2f_{sc}$ and the pattern is still orthogonal. However, in the second field the sampling pattern is shifted over $1/4f_{sc}$ (see also Fig. 18). For

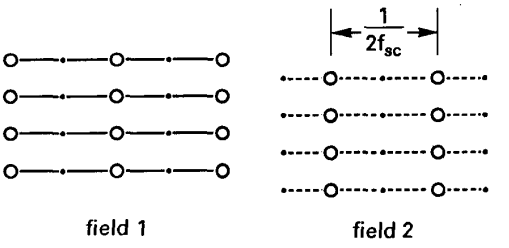


Fig. 18. Field-quincunx subsampling results in orthogonal sampling patterns on field basis, which is suitable for Hadamard coding.

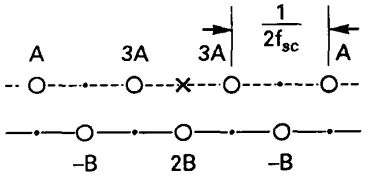


Fig. 19. Samples and filter coefficients for interpolation of sample X, using (NTSC) lines of the actual and previous fields (dotted lines). A feasible example for filter coefficients is $A = 1/8$ and $B = 2A$.

picture reconstruction after Hadamard decoding, an interfield interpolation filter is required using intraline interpolation in the same field and interpolation with the corresponding line of the previous field. The interpolation filter coefficients are shown in Fig. 19.

For Hadamard transformation, blocks of 2×4 samples are proposed, which are extracted within a field. If the order of the samples is taken as shown in Fig. 20, the two-dimensional block can be written as a one-dimensional vector $\boldsymbol{f} = (f_{(0)}, f_{(1)}, \cdots, f_{(7)})$, and the output vector $\boldsymbol{F}$ is specified according to $\boldsymbol{F} = H\boldsymbol{f}$, where

$$H = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \end{bmatrix}. \qquad (13)$$

Quantization and coding have been kept fairly simple in the system described. The bit cost per 2×4 block has been chosen at 4.5 b/sample resulting in a fixed bit rate per Hadamard block. The 36 b are not equally distributed over the coefficients (see Table III), since there is a significant difference in their importance. The principal coefficients are F_0 and F_4 because they represent the average value and the color subcarrier pattern, respectively. The latter value contains the low-frequency content of the color after demodulation. All ac coefficients ($F_{(1)} \cdots F_{(7)}$) are

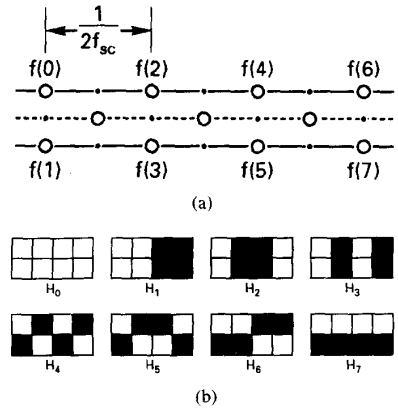


Fig. 20. Block structure of Hadamard transform and the Hadamard 2 × 4 basis pictures. (a) Samples for block construction after the field-quincunx subsampling of Fig. 18. (b) Hadamard basis pictures (white corresponds to +1 in the transform matrix).

TABLE III
EXAMPLE OF BIT ALLOCATION FOR SPECIFIC HADAMARD COEFFICIENTS $F_{(i)}$

$F_{(0)}$	$F_{(1)}$	$F_{(2)}$	$F_{(3)}$	$F_{(4)}$	$F_{(5)}$	$F_{(6)}$	$F_{(7)}$
8	5	4	3	6	3	3	4

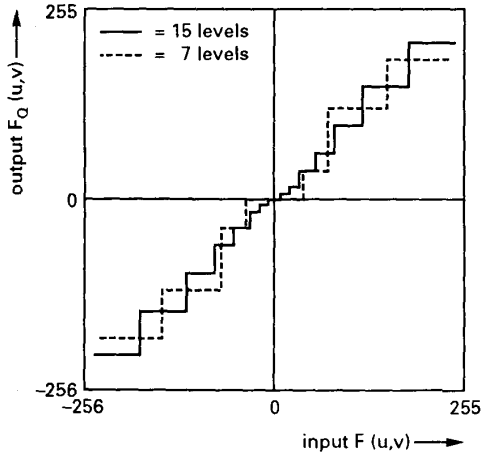


Fig. 21. Example of symmetric nonuniform quantizers having 7 and 15 output levels.

quantized with nonuniform quantizers (see Fig. 21) for which the minimum mean square error has been one of the optimization criteria. The authors report that exact symmetry in the quantizers is of vital importance in order to avoid color flicker components since the subcarrier has periodic phase reversals. In [19], [20] it is stated that the subjective image quality of the system can be further improved by using two-dimensional vector quantization. In contrast with scalar quantization, in which each coefficient is quantized independently, vector quantization exploits the mutual dependency between coefficients by quantizing *combinations* of coefficients. For hardware simplicity, pairs of coefficients have been used. The bit rate after subsampling, 44 Mb/s, is further compressed by the Hadamard techniques mentioned to approximately 25 Mb/s.

A Hadamard coding system for PAL composite signals has been reported by Driessen *et al.* [13]. Again, prior to coding subsampling is performed in which the initial sampling rate at $4f_{sc}$ (17.7 MHz) is reduced to $2f_{sc}$ by means of line-quincunx sampling in each field. Although it is recognized that transform coding is not very effective in general for composite signals, because the subcarrier energy is spread out over a large set of coefficients, the situation for coding of PAL at twice the subcarrier frequency is relatively beneficial. Fig. 22(a) portrays the output of a comb-filtered PAL-U and PAL-V signal with constant input amplitude. The sampling phase pattern shifts over $\pi/4$ and $-\pi/4$ for U and V, respectively, for subsequent lines in the actual field. This pattern matches well with the basis pictures H_6 and H_7 of the 4 × 4 Hadamard transform, thereby concentrating the color information in two coefficients. Since the block size is only 4 × 4, this situation occurs regularly. Another interesting feature of the 4 × 4 system is its adaptive quantization strategy [13], that will be briefly resumed here because it resembles the quantizers in more advanced systems discussed in the next section. The coefficient pattern is analyzed prior to quantization and classified into one of four categories. A control table is chosen, corresponding to the derived class, which defines two parameters P_i and Q_i for each coefficient $F_{(6)}(0 \leq i \leq 15)$. Each class has a predetermined set of quantizers used for the various coefficients, with characteristics similar to Fig. 21 and bit allocations as shown in Table III. The quantizer sets are suited for processing low to highly detailed blocks. The parameter P_i denotes a preemphasis factor (always a power of 2) for normalizing the input [21]. The figure Q_i selects one out of 16 quantizers in a particular class. A 2-b classifier Cl indicates the chosen class for decoding. The output bit rate is fixed per block (40 b) at 2.5 b/sample, resulting in 22 Mb/s bit rate for the compressed video signal. The quantizer block diagram is shown in Fig. 23.

Comparisons between DPCM and Hadamard coding of composite signals have been reported [13], [19]. The conclusion is that DPCM outperforms the Hadamard technique under the system design constraints described. The DPCM system yields better picture quality with textured input signals and it is simpler to implement in real-time hardware. In the following, we will focus on the more advanced DCT coding schemes.

2) Discrete Cosine Transform: The DCT is one of the most popular transforms because it provides efficient energy compaction and lately it has become feasible for VLSI realization. For this reason, we will find relatively recent results in this section. Another noticeable trend is the transition to YUV-component coding instead of coding the composite NTSC or PAL signal.

An advanced intraframe DCT coding system has been applied in an experimental digital 8 mm recorder by Bor-

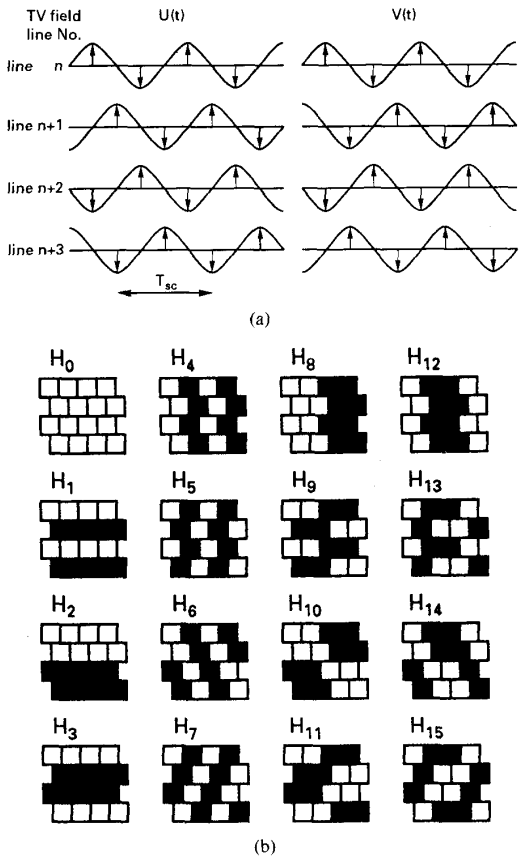


Fig. 22. The correspondence of the two-dimensional phase patterns of the subsampled color-difference signals U and V and the basis pictures H_6 and H_7 (from [13]). (a) Phase pattern of U and V at $2f_{sc}$ line-quincunx subsampling. (b) The 16 Hadamard transform basis pictures of a 4 × 4 transform with line-quincunx subsampling.

gers *et al.* [3]. Improvements with respect to motion have been reported by de With [22]. A similar intraframe DCT architecture has also been developed for an alternative recording technique, termed matrix-scan recording, by Platte *et al.* [23]. The coding technique which has been investigated may best be identified as a motion-adaptive intraframe DCT threshold coding algorithm. In this system, variable-length coding of coefficients has been employed for a significant improvement in coding efficiency. Fig. 24 portrays the block scheme of the encoder. The frame memories for luminance (Y) and chrominance (U/V) gather two fields to construct one video frame, which is subsequently segmented into small blocks of 8 × 8 samples. The luminance and chrominance blocks are multiplexed into one data stream, thereby saving hardware, and improving overall coding efficiency because chrominance blocks usually require less bits for coding. A motion detector analyzes every incoming sample block, and governs the transform computation and, partially, the subsequent processing blocks, so that motion-adaptive processing is achieved. After reordering (scanning), the coefficients are weighted individually to obtain frequency-dependent quantization. In addition, the quantization adapts to local picture statistics. Variable-length coding of the significant coefficients is performed in three steps. First, the coefficient amplitudes are ranked by their magnitude and the difference between successive amplitudes is constructed. Second, the address information, required to indicate the place of a coefficient inside a block, is minimized. Finally, variable-length codes are assigned to both coefficient differences and addresses. A buffer in combination with a control unit smooths the variable bit-rate output, and provides the required interface to the subsequent fixed-rate processing blocks in a recorder, such as error correcting coding and/or channel coding.

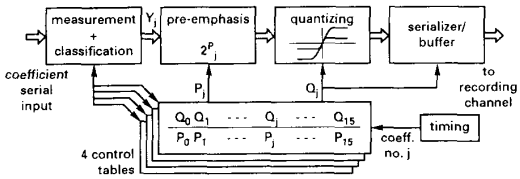


Fig. 23. Adaptive quantizer block diagram for 4 × 4 Hadamard transform coding (from [21]).

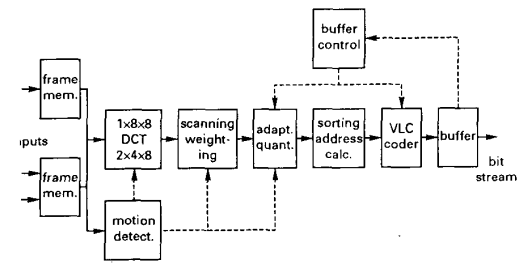


Fig. 24. Motion-adaptive DCT encoder block diagram (from [22]).

Now we will describe some details about the algorithm. An important aspect is the implementation of the DCT itself. Similar to the FFT, fast DCT algorithms have been developed. The main emphasis has been put on reducing the total amount of multiplications and additions. In [24] also the numerical accuracy of the transformer is discussed. A comparison is being made between the algorithm of Lee [25] and a new algorithm that can be found by subsequently adding pairs of terms that will be multiplied with the same cosine factor. The butterfly diagram showing the flowchart of calculations is depicted in Fig. 25. In this figure, two joining branches stand for an addition (or subtraction if indicated), while the circles with the factors C_j^i denote a multiplication with a factor cos $(i\pi/j)$. The authors claim a better accuracy with the same wordlength as compared to the algorithm of Lee.

Another feature of interest is the chosen bit-assignment technique. Instead of the well-known run-length coding procedure of Chen and Pratt [26], it has been proposed to rank the coefficients according to their magnitude. Along with each coefficient amplitude, an address is being transmitted indicating the spectral coordinates of a coefficient inside the 8 × 8 block. Sorting of coefficients in bit planes has first been proposed by Keesen [27]. This idea has been

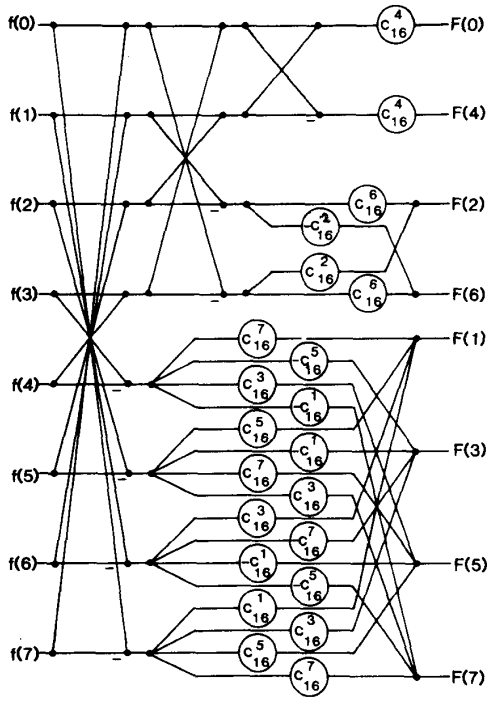


Fig. 25. 8-point DCT butterfly diagram (from [3]).

modified to a full sort of the coefficients and coding of the difference signal resulting from subtracting subsequent amplitudes after sorting. In addition, the address values can be further reduced by considering earlier transmitted amplitudes inside the same DCT block. A formal description of the algorithm now follows. The two-dimensional block of coefficients $F(u, v)$, with $0 \le u, v \le N-1$, is scanned prior to coding, which results in a serial block of coefficients. Each coefficient after scanning, denoted by $F(A)$, is uniquely described by its amplitude F and its address A. The absolute address A ($0 \le A \le N^2 - 1$) of a coefficient is defined as the ranking number after scanning. We assume l ac coefficients (of course, with $A > 0$) to be of nonzero amplitude.

Step 1 (Full Sort): Select the coefficients with $A > 0$ and order them in an amplitude stack. Let k be the index of the coefficients after sorting [if allowed, the notation of the kth coefficient $F_k(A_k)$ is abbreviated to $F_k(A)$]. For the kth selected coefficient $F_k(A_k)$ the following holds:

$$|F_{k-1}(A)| \ge |F_k(A)| \cdots, |F_k(A)| > 0, \qquad 2 \le k \le l. \tag{14}$$

The corresponding addresses A_k are stored in an address stack. If $|F_j(A_j)| = |F_k(A_k)|$ and $A_j < A_k$, then $j < k$.

Step 2 (Construct Difference Signal): Calculate the amplitude differences $D_k(A)$ defined by

$$D_k(A) = |F_{k-1}(A)| - |F_k(A)|, \qquad 2 \le k \le l,$$
$$D_1(A) = |F_1(A)|, \qquad k = 1. \tag{15}$$

This difference signal $D_k(A)$ is a nonnegative sequence for each block. The original amplitude signs of $F_k(A)$ are encoded separately. There are no changes in the address stack during this step.

Step 3 (Address Calculation): Recalculate from each address A_k an address T_k according to

$$T_k = A_k - N_k(0, A_k), \qquad D_k > 0,$$
$$T_k = A_k - A_{k-1} - N_k(A_{k-1}, A_k), \qquad D_k = 0 \tag{16}$$

where $N_k(A_1, A_2)$ refers to the number of (previously encoded) coefficients $F_j(A_j)$, $1 \le j \le k-1$, which have addresses satisfying $A_1 < A_j < A_2$. The stack with amplitude differences is not changed during this step.

Step 4 (Bit Mapping): Coding of the sequence D_k and the corresponding addresses T_k for $1 \le k \le l$ is performed here with two variable-length code tables, viz. $VLC_D(K_k)$ and $VLC_T(T_k)$. Experimental VLC tables for the signals T and D can be found in [24]. The major advantage of the algorithm described, except for a small gain in coding efficiency, is the predetermined order of transmission of signal energy. This can be best understood by considering Fig. 26. This phenomenon makes selective error protection of the most important coefficients relatively easy. Acceptable-to-good picture quality results have been reported with this system between 1–2 b/sample, while an experimental recorder has been realized with 19 Mb/s video bit rate (1.8 b/sample).

A more conservative DCT coding approach has been followed by Doi *et al.* [28]. Intrafield DCT coding results with the same block size are reported for 40 Mb/s (2–3 b/s). The quantizer is nonuniform (Max–Lloyd type [29]) and the adaptivity and bit allocation (using fixed-length codes) are according to the activity within the block [30]. The buffering of this system is of the feedforward type and enables multispeed modes and high error robustness. This is caused by the fixed-length coding of a group of DCT blocks, e.g., the amount of blocks within 8 video lines. The bit cost of each DCT block can have three different values only, namely G-1, G-2, or G-3 (see Fig. 27). Bit cost G-2 is the average bit rate, while the "overflow" of G-3 equals the "underflow" of G-1, hence, $G\text{-}1 + G\text{-}3 = 2G\text{-}2$. In the filling procedure of the group of DCT blocks, termed the reference region, the unused data bits resulting from G-1 coded blocks are used for data bits of G-3 coded blocks. This procedure is carried out going from the first to the last DCT block (left to right) in the reference region. If some G-3 coded blocks remain, then they are quantized according to the G-2 bit allocation. The result is a fixed bit cost per group of DCT blocks. The authors report that for a reference region covering 96 DCT blocks, reallocation to G-2 for remaining G-3 blocks occurs in less than 10% of the cases for typical images.

One of the most recent results in DCT coding for home-use digital video recording has been published by Yamamitsu *et al.* [31]. This is also a feedforward buffering system over a small group of DCT blocks. However, a

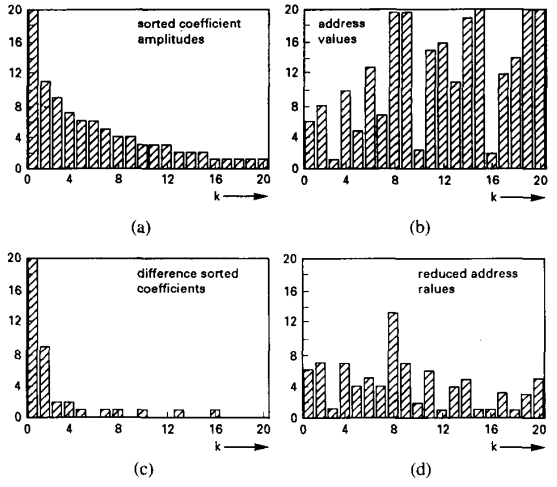


Fig. 26. (a) Example of sorted coefficient amplitudes and the derived amplitude differences, together with corresponding address values and the reduced addresses after coding of an 8 × 8 DCT block. (a) Sorted coefficient amplitudes. (b) Corresponding subsequent differences. (c) Original address values. (d) Reduced address values.

8	4	2	0	0	0	0	0
4	3	2	2	2	0	0	0
4	3	2	0	0	0	0	0
3	2	2	2	2	0	0	0
3	2	0	0	0	0	0	0
2	2	0	3	3	2	0	0
2	2	0	2	2	0	0	0
2	4	2	5	5	4	0	0

(a)

8	5	3	2	2	0	0	0
5	4	3	3	3	3	0	0
4	3	3	3	3	0	0	0
4	3	3	3	3	3	0	0
3	3	3	3	3	2	0	0
3	3	3	4	4	3	0	0
3	3	3	3	3	3	0	0
3	4	3	6	6	5	0	0

(b)

8	5	4	3	3	2	2	2
6	5	4	4	4	3	2	0
5	4	4	4	3	3	0	0
5	4	4	4	4	4	2	0
4	4	4	4	4	3	2	0
4	4	4	5	5	4	2	0
4	4	4	4	4	4	2	0
4	5	4	6	7	6	2	2

(c)

Fig. 27. Bit allocation tables *G*-1, *G*-2, and *G*-3 for 8 × 8 intrafield DCT coding (from [28]). (a) *G*-1 (96 bits). (b) *G*-2 (160 bits). (c) *G*-3 (224 bits).

more flexible approach has been adopted for bit allocation. The system block diagram is shown in Fig. 28. First, a group of DCT blocks is temporarily stored in a buffer memory, while at the same time the blocks are quantized with different quantizers. Then the bit cost of each block is calculated for these quantizers and accumulated in order to select the best quantizer. The advantage of this system is the fixed bit cost per group of DCT blocks, so that error propagation due to the use of variable-length coding is strictly limited.

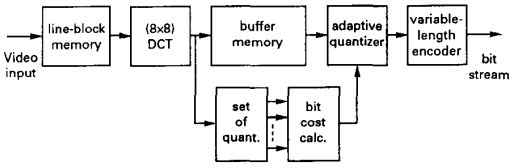


Fig. 28. Feedforward DCT coding scheme with parallel quantization and bit cost calculation prior to final quantization and coding.

Before concluding this section, two details of the scheme are highlighted, i.e., the fast DCT algorithm and the bit-assignment technique. The proposed DCT algorithm has not been optimized for accuracy within the limits of the definition as indicated previously, but for a significant reduction of the amount of multiplications. This has been obtained by sacrificing the exact definition of the DCT [32] and using a DCT with a spectral weighting function inside the transform. This can be modeled as multiplying each coefficient $F(u)$ by a weighting factor $w(u)$. If the one-dimensional DCT is applied, specified by

$$F(u) = \sqrt{\frac{2}{N}}\, C(u) \sum_{i=0}^{N-1} f(i) \cos\left[(2i + 1)u\pi/2N\right] \quad (17)$$

for $0 \le u \le N - 1$, then the spectrally weighted DCT and its weighting factors $w(u)$ have been chosen according to

$$F_w(u) = w(u) \cdot F(u), \qquad (0 \le u \le N - 1)$$

$$w(0) = 1$$

and

$$w(u) = D(u) \frac{\sin(\pi/4)}{\sin(u\pi/2N)} \qquad (1 \le u \le N - 1). \quad (18)$$

The constant $D(u)$ is specified by $D(1) = 1/4$, $D(2) = 1/2$, $D(3) = 3/4$, and $D(u) = 1$ elsewhere. In [31], (18) is derived for block size $N = 8$. The use of including a spectral weighting inside the transform has been reported earlier (see, e.g., [33]) in order to obtain a visually optimized DCT, but in the current proposal it has been explicitly used for reduction of hardware complexity. The result is a butterfly diagram based upon a very limited amount of cosine terms. As an example, the inverse DCT butterfly diagram is shown in Fig. 29.

Finally, we summarize the bit-assignment technique, which can be termed as an adaptive zonal coding system. In the first step a zone (H_{max}, V_{max}) is identified by searching the largest column value H_{max} for which nonzero coefficients occur. In the same way V_{max} is defined by considering the rows of the quantized coefficient matrix. Then the zone (H_{max}, V_{max}) is selected for transmission and the remaining part of the block is omitted. The zone parameters must also be sent for unique block decoding. The coefficients inside the selected area are coded with a variable-length code having length $L(i)$ for the ith codeword. If the wordlength for zero values is 1 b, the bit cost B for

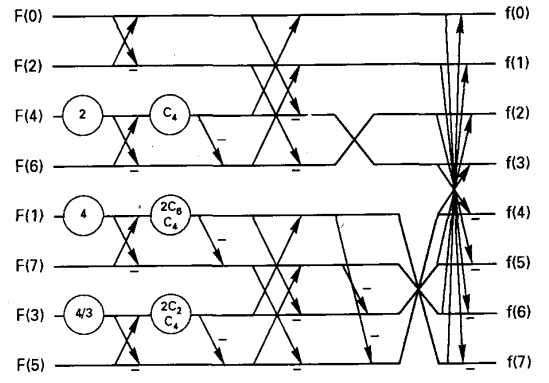


Fig. 29. Butterfly diagram of inverse eight-point DCT with internal spectral weighting function for reduction of multiplications. Minus signs denote a subtraction (from [31]).

DC	0	1	2	3	(4)	5	6	7
0	91	25	5	0	1	0	0	0
1	33	15	0	1	0	0	0	0
2	8	9	0	0	0	0	0	0
3	1	2	3	1	0	0	0	0
4	0	0	0	1	0	0	0	0
(5)	1	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0

(v →, u ↓)

Fig. 30. Example of an 8 × 8 quantized coefficient block indicating the selected zone (H_{max}, V_{max}) = (5, 6) for transmission. Note that $H_{max} = v_{max} + 1$, and $V_{max} = u_{max} + 1$.

a DCT block is easily computed by

$$B = \sum_{i=0}^{N^2-1} (L(i) - 1) + H_{max} V_{max}. \quad (19)$$

An example block indicating the selected zone is depicted in Fig. 30.

IV. Picture Coding for Digital HDTV Consumer Recording

A. Introduction

Digital communication networks (B-ISDN) and digital storage media with a large capacity are emerging. This development opens the possibility to introduce high-definition video signals (HDTV) that have a significantly higher spatial and/or temporal resolution than the standard-definition television signals described in the previous section. Although the HDTV signal has not yet been introduced as a standard, research into HDTV transmission and storage systems started several years ago. In Europe, for example, this has led to the standardization of an HDTV transmission system called HDMAC [34], while in Japan it has resulted in the development of an HDTV transmission system called MUSE [35].

In this section, ongoing research into systems for magnetic recording of HDTV is described. As in the previous section (Section III), the attention is focused mainly on home-use recording systems. First, in Section IV-B, some existing experimental HDTV recording systems without picture coding will be described. However, for portability and compactness (see Section II), home-use digital HDTV recorders will require picture coding. Section IV-C is devoted to picture coding for home-use digital HDTV recorders. Special attention is paid to the important issue of compatibility between HDTV and TV signals.

B. Existing HDTV Recording Systems

A limited number of publications is available about experimental digital HDTV recorders. The publications that will be mentioned here have in common that no bit-rate reduction is applied to the HDTV signal, which means that the recorders are mainly meant for professional use.

Eto *et al.* [36]–[38] describe an experimental digital HDTV recording system that records the luminance component of the HDTV signal with a bandwidth of 20 MHz, and each of the color difference components with a 5.5 MHz bandwidth. The total bit rate of the signal that is recorded is equal to about 460 Mb/s. Hashimoto [39], Thorpe *et al.* [40], and Przybyla [41] report on digital HDTV recorders that record HDTV at a much higher bit rate, namely about 1.2 Gb/s. The bandwidth of the recorded signal is 30 MHz for the luminance component and 15 MHz for each of the color difference components. Finally, in [42] a system is reported that records analog HDTV with a bandwidth of 20 MHz in a digital form on tape (54 MHz sampling frequency, 8 b/sample).

C. HDTV Coding

This section discusses bit-rate reduction for magnetic recording of the HDTV signal. Section IV-C-1) is about direct HDTV coding, i.e., about bit-rate reduction techniques that do not supply compatibility between HDTV and TV signals. Section IV-C-2) is about techniques that do supply this compatibility. As an example, Section IV-C-3) deals with one of these systems in greater detail.

1) Direct HDTV Coding: A definite standard for the HDTV signal has not yet been chosen. It will probably be a component-based signal with a format similar to that of the CCIR-601 TV signal, except that it will have more lines per frame, and more samples per line. In Europe, the number of lines will probably be 1250, and the number of luminance and chrominance samples per line could be 1440 for consumer applications, and 1920 for studio applications. In the United States, the number of lines could be 1050. The bit rates of these types of signals are about 650–900 Mb/s, respectively. An important difference between the TV and the HDTV signal is the aspect ratio (ratio of screen width and height), which is 4:3 for TV and 16:9 for HDTV signals. The 16:9 aspect ratio of HDTV improves its suitability for the display of movie material.

For consumer magnetic recording of HDTV signals, bit-rate reduction is highly beneficial because it can sim-

plify the complexity of the recording mechanics considerably. The net bit rate that can be spent on the HDTV signal can be as low as 50–100 Mb/s, which means that compression ratio of about 7–14 are needed. The constraints on the bit-rate reduction of HDTV signals are basically the same as those for the reduction of TV signals. Therefore, in principle the bit-rate reduction techniques that have been explained in Section III can also be applied directly to HDTV signals. For example, consider the intraframe DCT coding scheme described in Section III-C-2). With that system, the bit rate of CCIR 601 standard-definition TV signals can be reduced from 166 Mb/s to about 20 Mb/s, while maintaining high picture quality. With this coding system, the bit rate of HDTV signals (1250 lines, 1440 samples/line, frame rate 25 Hz, 2:1 interlaced) can be reduced from 664 to about 80 Mb/s.

If bit rates much lower than 80 Mb/s are needed, it will be necessary to use coding strategies with a much higher coding efficiency. One possibility could be to take advantage of the redundancy in time in the HDTV signal, by applying interframe coding techniques [10]. However, interframe coding techniques, such as those used in videophone and video-conferencing systems [43], [44], can be applied only in a very restricted way. Similar to the TV recorder, the HDTV recorder must have a visible playback facility operating at various playback speeds. In Section II it has been explained that at higher tape speeds, only small data blocks can be recovered.

2) Compatible HDTV Coding: An important issue for HDTV coding is that of compatibility between HDTV and standard-definition TV signals. The problem of compatibility is as follows. A possible scenario for the introduction of digital video recorders could be to introduce a digital TV recorder first, and in a later phase a digital HDTV recorder. For a successful introduction of an HDTV recorder, it is highly favorable that this recorder is compatible with the TV recorder introduced earlier. More precisely, the question is how to enable tapes recorded on a TV recorder to be played back on an HDTV recorder, and vice versa. Fig. 31 depicts the problem of compatibility in a schematical way.

A solution could be to split the HDTV signal, prior to encoding and recording, into a standard-definition "compatible" TV signal and one or more additional signals. The additional signals will henceforth be called surplus. The compatible TV signals and the surplus can now be coded separately, and the resulting bit streams can be distributed on the magnetic tape. For the compatible TV signal, the same coding techniques can be applied as for an ordinary TV signal (Section III). For the surplus signals, a special coding method has to be developed.

Two well-known coding techniques exist that can split a high-resolution picture into a low-resolution picture and surplus, namely pyramidal coding [45] and subband coding [46]–[50]. These techniques will be discussed in some detail in the remainder of this section. Subsequently, in Section IV-C-3) an example system using subband coding is described. For the sake of simplicity, it will be assumed

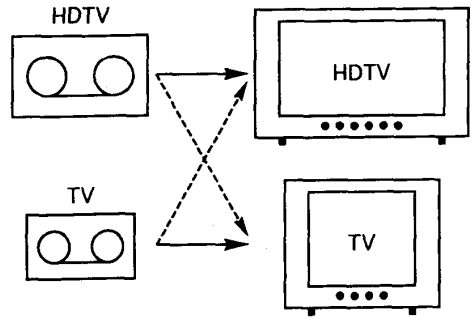


Fig. 31. The problem of compatibility.

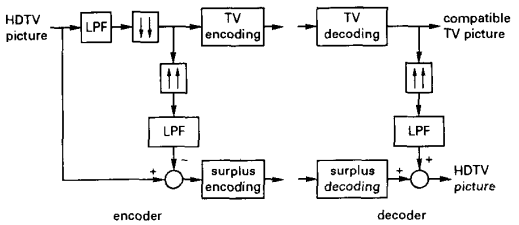


Fig. 32. Encoder and decoder block diagram of a pyramidal coding system.

that the compatible TV picture has half the resolution horizontally and vertically of the high-resolution HDTV picture.

Fig. 32 depicts the encoder and decoder of a pyramidal coding system. First, the high-resolution HDTV picture is passed through a two-dimensional low-pass filter, which preserves the lower half of the horizontal and vertical spatial frequencies. Subsequently, the low-pass filtered signal is subsampled by a factor of two horizontally and a factor of two vertically (indicated by $\downarrow\downarrow$). The resulting compatible TV picture is coded separately. A surplus picture is created by upsampling ($\uparrow\uparrow$) and interpolating the compatible picture, and by subtracting the result from the input HDTV picture. In the decoder, the decoded TV signal and the decoded surplus are added to reconstruct the HDTV signal. Fig. 33 shows how the spatial frequency spectrum of the HDTV picture is divided into two parts by pyramidal filtering. By assuming that the HDTV picture has a resolution of M lines with N samples per line, the resolution of the compatible TV picture is equal to $M/2$ by $N/2$, and that of the surplus picture M by N. So, in total $5MN/4$ samples have to be coded and recorded, which is one quarter more than the amount of samples in the input HDTV picture.

The second approach that splits the HDTV spectrum is subband filtering. As an example, Fig. 34(a) depicts the encoder of a simple subband coding system that splits the HDTV picture into a compatible TV picture and three surplus components, and Fig. 35 shows how it divides the spatial frequency spectrum into four parts. In the encoder, the HDTV picture is first divided into two components, by low-pass and high-pass filtering in the horizontal di-

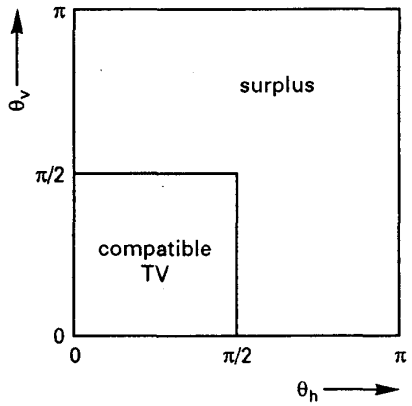


Fig. 33. Division of the spectrum of HDTV by pyramidal coding ($\theta = 2\pi f / f_s$).

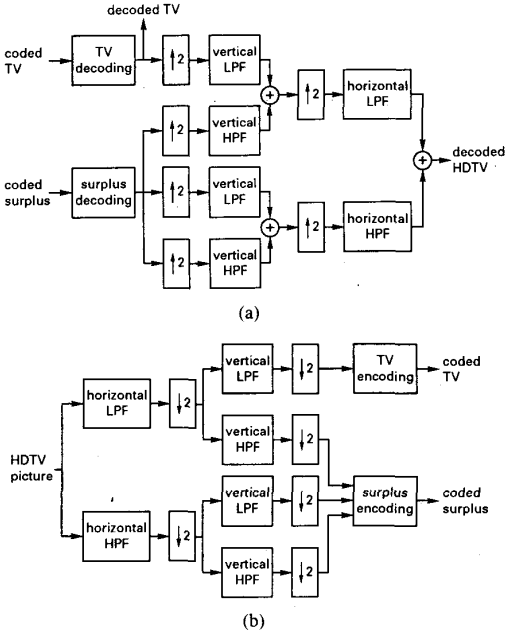


Fig. 34. Example of a subband coding system. (a) Encoder. (b) Decoder.

rection, each followed by subsampling with a factor of two (indicated by $\downarrow 2$). Each of the two components is further split into two parts, by performing the same filtering and subsampling procedure in the vertical direction. The compatible TV signal and the surplus are coded separately. In the decoder, the HDTV signal is reconstructed by upsampling and interpolating the compatible TV signal and the surplus as is indicated in Fig. 34(b). Assuming again that the HDTV picture has a resolution of M by N, the resolution of the compatible TV picture and that of the surplus components is equal to $M/2$ by $N/2$ each. So, in total MN samples have to be coded and recorded, which means that with subband coding the number of samples is not increased. This is a clear advantage of subband coding compared to pyramidal coding.

The bit rate for the compatible TV picture must be such that the coding artifacts in the compatible TV signal do not lead to visible distortions in the HDTV signal. From experiments [see Section IV-C-3)] it has become clear that the bit rate needed for the compatible TV picture can be considerably higher than that needed for ordinary TV pictures (see also [51]). If the bit rate for the compatible TV picture is too low, it may be required to feed back the coding artifacts of the compatible TV picture into the surplus signal, or to transmit these coding artifacts as a separate surplus signal (for the latter solution, see [52]).

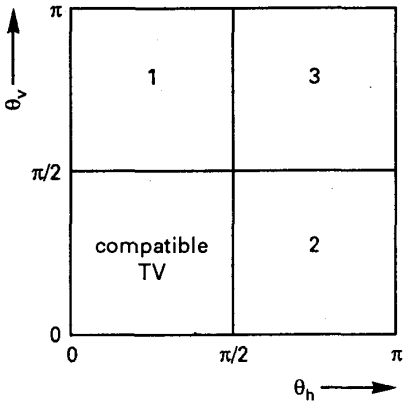


Fig. 35. Division of HDTV spectrum by subband coding.

Two important issues concerning compatibility have not yet been discussed, namely interlacing and the different aspect ratios for TV and HDTV. First, the HDTV signal is likely to be a 2:1 interlaced signal, i.e., each HDTV picture (called a frame) consists of two interlaced "half pictures" (called fields), which have been sampled at different time moments. Both pyramidal and subband coding of the HDTV frames will not give a high-quality 2:1 interlaced compatible TV signal (it will contain motion artifacts, since information from different time moments is merged before the filtering and subsampling). This problem is addressed in Section IV-C-3).

Second, the HDTV signal will have a format that is meant for display on monitors with a 16:9 aspect ratio, while the TV signal is for 4:3 monitors. Without claiming that they are complete, two solutions are indicated here. The first, and probably most simple, approach is to convert a 16:9 HDTV picture to a 16:9 compatible TV picture and 16:9 surplus, and to record them separately. Then, in the decoder the 16:9 TV picture must be further converted to a 4:3 picture. This can be done in at least two ways, namely conversion to a picture called side-panel format and to the letter-box format (see Fig. 36). The second method to convert a 16:9 HDTV picture to a 4:3 compatible TV picture is to divide the 16:9 HDTV picture into two parts in the encoder: the 4:3 middle part and two side panels. From the 4:3 middle part of the HDTV picture a 4:3 compatible TV picture can be derived. The two HDTV side panels have to be transmitted as surplus signals.

3) Subband Coding of 2:1 Interlaced HDTV: As an example, this section briefly describes a subband coding system for 2:1 interlaced HDTV signals [51]. First, it is

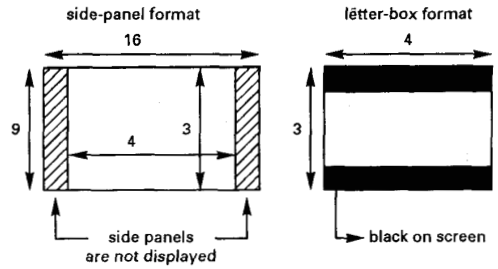


Fig. 36. Side-panel and letter-box format.

indicated how the conversion from a 2:1 interlaced HDTV picture to a 2:1 interlaced compatible TV picture can be performed. Then, it is described how this method of conversion can be incorporated into a subband coding system. Finally, some coding results are given.

Fig. 37(a) shows the sampling pattern of a 2:1 interlaced HDTV signal in the time versus vertical direction. Fig. 37(b) shows the required sampling pattern of the 2:1 interlaced compatible TV picture. In principle, a compatible TV field can be derived from an HDTV field by vertical and horizontal low-pass filtering, and subsampling by a factor of two (to avoid phase distortion, the use of symmetrical finite impulse response (FIR) filters is preferred). If, however, vertically the same filter is applied to every HDTV field, then the lines of the resulting TV frame (equal to two interlaced fields) will not be equidistant. Fig. 38(a) illustrates this problem for the case that an even-length vertical FIR low-pass filter of length 4 is used. A very simple solution to this problem is to use an odd-length vertical low-pass filter for every odd field, and an even-length low-pass filter for every even field, or vice versa. The frequency responses of these filters should be about equal. Fig. 38(b) illustrates this solution in the case where filters of lengths 3 and 4 are used.

The question now is how to incorporate this solution into a subband coding system. Basically, the following procedure has been used. A subband coding system such as the one in Fig. 34 has been developed which operates on HDTV fields. Each HDTV field is split into a compatible TV field and three surplus fields. Horizontally, even-length filters are used; vertically, odd-length filters are used for every odd field, and even-length filters for every even field.

The splitting technique used is that of quadrature-mirror filtering (QMF) [53]. Fig. 39 depicts the basic cell of a QMF splitting and reconstruction unit. The filters $h_l(n)$, $h_h(n)$, $g_l(n)$, $g_h(n)$, are all derived from a common symmetrical FIR low-pass filter with impulse response $h(n)$, $n = 0, \cdots, N-1$, where N is the length of the filter, by

$$\begin{aligned} h_l(n) &= h(n), \\ h_h(n) &= (-1)^{n+1} \cdot h(n), \\ g_l(n) &= 2 \cdot h(n), \\ g_h(n) &= 2 \cdot (-1)^{n+1} \cdot h(n), \end{aligned} \tag{20}$$

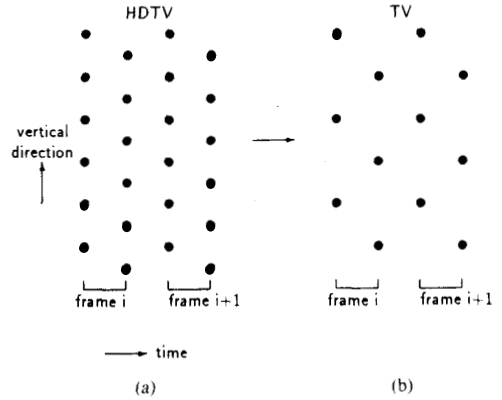


Fig. 37. Line sampling patterns of subsequent fields. (a) Sampling pattern HDTV. (b) Required sampling pattern TV.

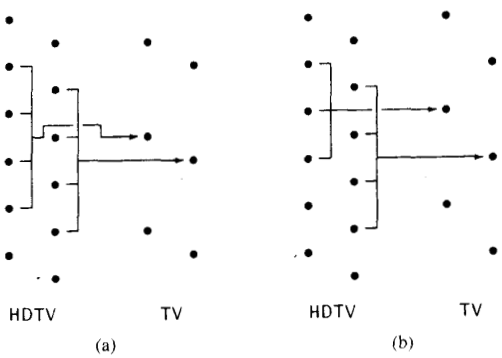


Fig. 38. (a) Conversion from HDTV to TV with even-length filters. (b) Conversion with odd-length and even-length filters.

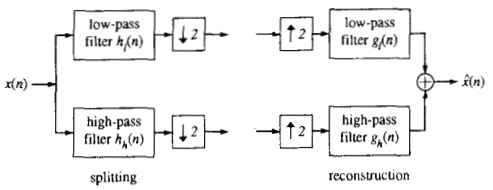


Fig. 39. Even-length QMF splitting and reconstruction unit.

where $h(n)$ is chosen such that

$$|H(e^{j\theta})|^2 + |H(e^{j(\theta+\pi)})|^2 \simeq 1 \tag{21}$$

in which $H(e^{j\theta})$ is the Fourier transform of $h(n)$. It can be easily derived [53] that with this structure a close replica of the input signal can be obtained only with even-length filters. Now, by changing the scheme slightly to the one shown in Fig. 40, odd-length filters can also be applied [54]. In this odd-length QMF, the filters $h_l(n)$, $h_h(n)$, $g_l(n)$, $g_h(n)$, are derived from an odd-length FIR symmetrical low-pass filter $h(n)$, by

$$\begin{aligned} h_l(n) &= h(n), \\ h_h(n) &= (-1)^n \cdot h(n), \\ g_l(n) &= 2 \cdot h(n), \\ g_h(n) &= 2 \cdot (-1)^n \cdot h(n). \end{aligned} \tag{22}$$

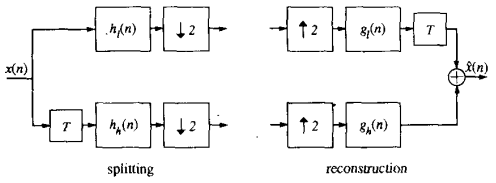


Fig. 40. Odd-length QMF splitting and reconstruction unit.

The compatible TV signal has been coded with the intraframe DCT coding system described in Section III-C-2). The three surplus signals have been coded with a more simple technique, namely nonuniform quantization followed by variable-length coding (VLC) and/or run-length coding (RLC). For more details the reader is referred to [51]. Computer simulation experiments showed that, in order to obtain a high-quality HDTV picture, the bit rate needed for the compatible TV picture is about 50–60 Mb/s. This is considerably more than needed for ordinary TV pictures. For the three surplus components a total of 30–40 Mb/s is needed. So, in total 80–100 Mb/s is needed for the HDTV picture.

A comparison has been made between the subband coding system described and direct (noncompatible) intraframe DCT coding of the HDTV signal. The subband coding system appeared to require about 20% more bits than direct intraframe DCT coding.

V. Coding for Digital Still Picture Storage

This section deals with coding techniques for still picture recording. Digital recording of still images is a rather new subject and not much about this subject has been published yet in literature. Nevertheless, the interest in still picture recording for electronic photography, archiving, desk top publishing, etc., is growing and several coding techniques have been tested for these applications. Although the basic part of the algorithms is the same as for moving video coding, the user and system requirements are different, so that new tests and developments have been carried out for digital still picture storage systems. In addition to these typical recording applications, the Joint Picture Experts Group (JPEG), a subgroup of ISO and CCITT, is developing an international standard for general-purpose, continuous-tone, still-image compression [55]–[58].

A. JPEG

JPEG started its activities in 1986 and aimed initially at compression techniques for a wide variety of image communication services. Later on, JPEG widened its scope to all kinds of photographic image applications, for both storage and transmission. The primary requirement is to develop a compression technique for transmission of digital images (sampled according to the CCIR-601 standard) over an ISDN basic channel of 64 kb/s. At 1 b/pixel, such an image can be sent within 6.5 s, which seems an acceptable waiting time for occasional image viewing. However, for browsing in a database facility, this time delay is much too long and therefore the function of progressive build-up, starting at a substantially lower number of bits per pixel, has been added to the list of requirements. For other applications, the possibility of lossless coding also has to be provided. Within JPEG, 12 different algorithms have been evaluated for functionality, complexity, and subjective image quality. An adaptive DCT algorithm has been finally adopted for its best overall performance and is being further refined. Now the overall algorithm structure consists of: 1) a baseline system; 2) an extended system; and 3) an independent lossless method. The algorithm of the baseline system will be described in more detail.

The DCT is performed on 8×8 blocks, using integer arithmetic. Next the two-dimensional block of coefficients $F(u, v)$ is quantized by means of linear scalar quantization. Quantization is accomplished via use of an 8×8 quantization matrix $Q(u, v)$, containing step sizes dependent on the spectral position of the coefficients. The step sizes $Q(u, v)$ are determined by psychovisual experiments, and should stand for the perceptual threshold of the cosine basis function of frequency (u, v). Thus, the quantization function is defined as

$$F_Q(u, v) = \text{integerround}\,[F(u, v)/Q(u, v)]. \quad (23)$$

The dc coefficient, $F_Q(0, 0)$, of block i is differentially encoded with respect to the dc term from the previous block $i - 1$. Next, the differential-dc coefficient is encoded using two codes. Code 1, called the size-code, indicates the number of bits needed to represent code 2, the amplitude-code, as a signed integer. Prior to Huffman coding, the quantized ac coefficients are rearranged from their two-dimensional matrix into a one-dimensional sequence, according to a "zig-zag" scan. Again two codes are used for coding this one-dimensional sequence: one code for entropy encoding of the nonzero ac coefficients and one for the run-length of the zero-valued ac coefficients preceding them.

The baseline system is a general-purpose algorithm, able to encode images of up to four components and with any pixel dimensions. This system only supports sequential build-up. The extended system contains some extra capabilities not provided by the baseline system, such as arithmetic coding [59] instead of Huffman coding, progressive build-up, and "progressive lossless" coding. The independent lossless coding is a simple ADPCM-based, stand-alone coding method for those applications which require lossless coding and which do not justify a complex DCT-based nonlossless option.

B. Archiving

In archiving systems, the storage capacity is more important than the image transfer rate. If the recording medium falls short in storage capacity, video source coding can offer a solution to increase the storage capacity and at the same time improve the image transfer rate.

Kageyama *et al.* [60] describe a digital still TV picture recorder utilizing an ordinary audio cassette deck. The input NTSC signals are sampled at 10.7 MHz, 8 b/pixel, and stored in an image memory. The image of 486 (horizontal) × 192 (vertical) pixels is compressed by means of a simple one-dimensional DPCM coding technique to 4 b/pixel and recorded on tape, using (1, 0, −1) class-IV partial response signaling. As a result, about 300 still pictures can be stored on a normal C-90 cassette (120 Mb). Recording or playback takes about 20 s/image (data rate 28.8 kb/s).

A digital still picture system for HDTV is presented in [61] by Takahashi. With this system, an uncompressed video signal is recorded on a disc and reproduced by a disc player connected to an HDTV processor. The digital still images of 1024 × 1280 active pixels are recorded at a bit rate of 1.41 Mb/s. At this data rate it takes 15 s to recover one HDTV frame.

C. Electronic Photography

Recently, electronic still-video cameras have been introduced in the consumer market. In 1981 a first generation of still-video cameras was announced, the Magnetic Video Camera (Mavica), that could store 25 to 50 images on a 2 in magnetic floppy (often called "video floppy"). In this system, images are stored in analog form using frequency modulation and can be displayed on a standard television set. Technological improvements have resulted in a new (analog) still-video standard in 1988, namely the "Hi-Band Mavica." At this moment, several companies are already studying a next generation of still-video cameras, offering improved picture quality and storage in digital form. The digital images can be recorded on magnetic or optical storage media or stored in solid-state memories. The mechanics of a recorder together with the linear density on the recording medium, define an upper bound for the data transfer rate and thus limit the image transfer rate. Solid-state memories do not have this transfer rate limitation, but only offer a limited storage capacity. Within certain constraints, video compression can offer a solution for both problems.

To get a system overview, we now briefly describe an experimental digital camera system (see [62]), from which the block diagram is depicted in Fig. 41. The system comprises three modules: the sensor module, the image-processing module, and the digital-recording module.

The sensor module contains a CCD image sensor with HD resolution, preprocessing electronics, and an A/D converter to generate a digital color image. The images are transferred to the buffer in the image-processing module. During recording and display, this buffer adapts the data rate of the recorder to the peripheral equipment. It also allows for different mains frequencies (50 Hz and 60 Hz) during playback. The image-processing module also contains the electronics for the compression and expansion of image information and both digital and analog interfaces for connection to printers and monitor output.

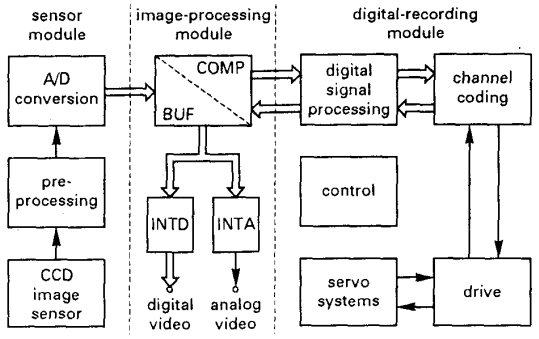


Fig. 41. Block diagram of a camera system.

The digital-recording module comprises electronics for error correction, data formatting, modulation and data recovery, plus the optical recorder. This optical recorder is a miniaturized WORM recorder using a 50 mm optical disc with a storage capacity of some 640 Mb and a net data rate of about 12 Mb/s.

One of the extra constraints on the image compression algorithm for this electronic photography is the size and power consumption of the electronics for digital signal processing, which has to be battery fed and packed in a compact camera. Therefore, at this moment ADPCM-based algorithms might be more attractive then the rather complex transform coding methods. Additional differences with video tape recording are that still pictures will be full-frame noninterlaced, and will not only be displayed on a monitor but also be reproduced on a hard-copy. Hard-copies, made with modern high-quality color printers, offer a much higher resolution than monitors. The improved resolution and the reduced viewing distance introduce additional constraints on the picture quality after compression. In the following, some experimental systems will be outlined.

Izawa *et al.* [63] describe a prototype of a digital still picture system utilizing a 20 Mb SRAM memory card as a storage medium. The analog RGB signals from the CCD sensor are converted into 8 b digital data, comprising a luminance signal Y and two color-difference signals U and V. The digital image of 774 (horizontal) × 494 (vertical) pixels would normally require a storage capacity of 6 Mb, 3 Mb for Y and 1.5 Mb for U and V. The data compression for this camera system is carried out in two steps. First, the number of samples is halved by means of a line offset subsampling (see Fig. 42). Second, the number of data bits of each remaining sample is reduced from 8 b to 4 b by an ADPCM encoder. Fig. 43 shows the pixel arrangement used for the prediction. The prediction of the luminance signal Y is given by $X_p = (A + B)/2$. The prediction error $X_e = X - X_p$ is encoded in 4 b after nonlinear quantization. To reduce the visibility of coding errors, a level variation detector is used to switch between 8 different quantizers. The variation P_{lum} is calculated from surrounding pixels according to

$$P_{lum} = |A - B| + |A - C|. \tag{24}$$

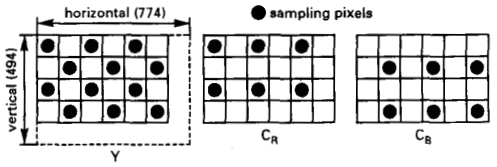


Fig. 42. Subsampling pixel arrangement for luminance and chrominance (from [63]).

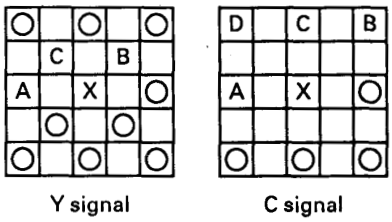


Fig. 43. Pixels used for prediction (from [63]).

For the color signals, the prediction is calculated by $X_p = (A + C)/2$. Again a level variation P_{chr} is calculated to switch between 8 quantizers, namely

$$P_{chr} = |A - B| + |A - C| + |A - D|. \quad (25)$$

The image size after compression is 1.5 Mb, enabling storage of 13 full-frame compressed images on one memory card.

Caronna *et al.* [64], [62] report on an ADPCM algorithm for a prototype high-definition still picture camera. This camera should offer a picture quality close to conventional 35-mm photography, and should be able to store at least 24 full-frame images in digital form. Furthermore, the camera must allow an image transfer rate of about 2 images/s. To meet the requirements on picture quality, HDTV parameters have been adopted. A high-definition image of 1152 lines with 1440 pixels/line in 4:2:2 YUV format, 8 b/sample, requires 26.5 Mb storage capacity. Consequently, about 640 Mb storage capacity would be needed to store 24 images. Compression with at least a factor 16 is necessary to enable digital data storage in solid-state memory cards. Only very complex compression methods based on transform coding yield these compression factors. To solve the storage problem, a miniaturized optical recorder is used, offering a storage capacity of 640 Mb on a 50 mm disc. Recorder constraints limit the data transfer rate to about 16 Mb/s (including overhead for error correction). With a resulting net data transfer rate of 12 Mb/s, recording or playback would take about 2.2 s/image. A factor 4 or 5 data compression is needed to meet the requirements on image transfer rate (at least 2 images/s). This increases the storage capacity to around 100 images per disc.

Let us now focus on the DPCM picture coding technique in the camera. Experiments showed that a rather simple switching scheme such as median prediction [65], [66] provides a good compromise between complexity and performance. The median predictor selects the median of

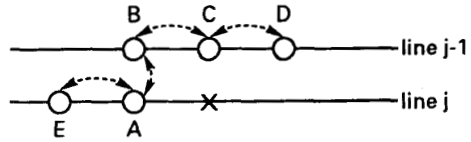


Fig. 44. Pixels used for adaptive prediction with sample differences used for masking.

three elementary predictors, a horizontal (P_h), a vertical (P_v), or a diagonal predictor (P_d) for the actual prediction. The predictions are calculated from previously coded pixels according to

$$P_h = A,\ P_v = C,\quad P_d = (2A + C + D)/4 \quad (26)$$

where the pixels are arranged as shown in Fig. 44. The advantage of median prediction is that it always selects the best or second best prediction, whereas the worst prediction is always rejected. The quantization is adaptive and controlled by a "masking function," which exploits the properties of the human eye, in particular, the masking effect of luminance edges [67], [68]. A masking function M is calculated from previously coded pixels to indicate the visibility of coding errors [69], according to

$$M = |A - E| + |B - A| + |C - B| + |D - C| \quad (27)$$

where the pixels are arranged as shown in Fig. 44. The given masking function divides the pixels into 4 classes (Y), using one quantizer per class.

In the case that the prediction error after quantization is (almost) zero, "conditional" variable-length coding is employed, thereby enhancing the coding efficiency with 5–20% (total compression 3–4.5 b/pixel). The conditional coding technique is controlled by the masking function.

Since the images are coded at a variable output rate, some buffering and feedback quantizer control is required to adjust the data rate from the compression system to the data rate of the optical recorder, so that data underflow or overflow is prevented. A fixed amount of disc space (8 Mb) is reserved per picture, so that even a "difficult" image fits in this area after compression. This enables a quick image access, while the image transfer rate remains close to 2 images/s. For robustness in the case of a disk sector loss, unique sync words (every line) and reset information are added to the bit stream produced by the compression system. The predictor uses only a horizontal predictor every 4 lines to limit error propagation.

VI. MPEG Video Coding for Digital Storage Media (Low Bit Rate)

A. Introduction

In 1988, the International Organization for Standardization (ISO) formed the Motion Picture Experts Group (MPEG). The task of this group was to develop a standard for coding of video and associated audio, on digital stor-

age media such as CD-ROM, magnetic tape, and optical disk. MPEG focused on a bit rate of 1.5 Mb/s in total, for video and audio. Recently, MPEG agreed upon a draft standard for video coding, which specifies the syntax of the bit stream produced by an MPEG encoder, and the MPEG decoding process. The proposed standard covers video coding at various resolutions and bit rates. However, to guarantee interoperability of equipment using the MPEG standard, a constrained parameter set has been defined for video coding at 1.5 Mb/s. This constraint set is indicated as the MPEG core; all MPEG decoders should be able to decode the core bit stream.

A detailed description of the MPEG bit-stream syntax is beyond the scope of this paper. For specific information about MPEG, the reader is referred to [70], [71].

B. MPEG Video Coding Algorithm

This section will generally explain the MPEG video coding algorithm. In order to obtain bit rates as low as 1.5 Mb/s, both the temporal and the spatial redundancy in the video signal must be removed. For this reason, the MPEG proposal is an interframe coding algorithm. It includes two interframe coding techniques, namely temporal predictive coding and frame interpolation.

In the predictive coding part, the picture to be coded is forward predicted from one of the pictures previously coded. The encoder of the predictive coding part is depicted in Fig. 45.

Comparison of this figure to Fig. 4 of Section III-A shows that it is in fact a temporal DPCM scheme. Since it is assumed that the picture to be coded is a translation of the previous picture on a local basis, motion-compensated prediction [10] can be applied. The motion-compensated prediction is performed on a block basis, where a block has the size of 16 × 16 luminance samples (this luminance block, together with the corresponding chrominance samples, is often called a macroblock). Per macroblock one motion vector is transmitted to the decoder, indicating the position of the best-matching block in the previous picture. This has been indicated schematically in Fig. 46.

Since the prediction is almost never perfect (e.g., it is on a block basis) the prediction error signal, which is the difference between the actual and the predicted signal, is transmitted to the decoder. For improving the coding efficiency, this information is compressed by using DCT coding techniques. The algorithm to code the prediction error signal is discussed at the end of this section. After inverse transformation, the prediction error block is then added to the best-matching block of the previously decoded picture. Hence, the picture memory in Fig. 45 contains the reconstructed picture.

In the interpolative coding part, the picture to be coded is "bidirectionally" predicted from some previous and some "future" picture. This prediction is again performed with block-based motion estimation techniques, as is illustrated in Fig. 47.

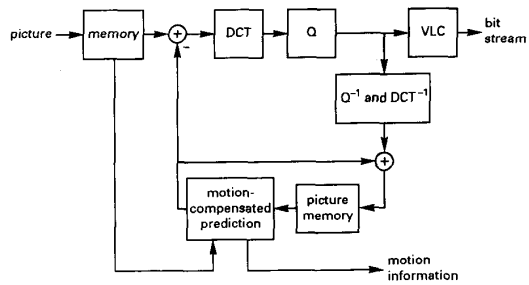


Fig. 45. Motion-compensated predictive encoder block diagram.

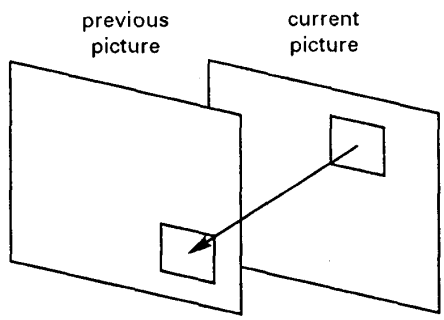


Fig. 46. Indication of the position of the best-matching block in the previous picture with a motion vector.

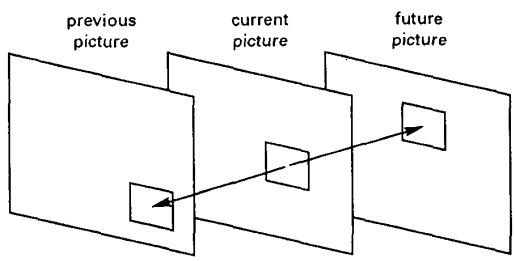


Fig. 47. Bidirectional prediction using the previous and future picture.

Now per macroblock, two motion vectors are transmitted to the decoder, indicating the positions of the best-matching blocks in the previous and the future picture, respectively. Furthermore, the prediction error signal is transmitted. The motion-compensated bidirectional prediction is also known as motion-compensated (frame) interpolation. In the decoder, the prediction error signal is added to an average of the best-matching block in the previous and the future picture.

Fig. 48 portrays how the predictive and the interpolative coding are combined. In this figure, which gives a "side view" of a video sequence, three types of pictures can be distinguished: *P* pictures, *B* pictures, and *I* pictures. First, one picture is compressed independently using intraframe coding (this is the *I* picture). Second, the fourth picture, called a *P* picture, is predicted in the forward direction, using the *I* picture. Third, the bidirectionally predicted pictures, called *B* pictures, are derived on the basis of the already coded *I* and *P* pictures. The *I* pic-

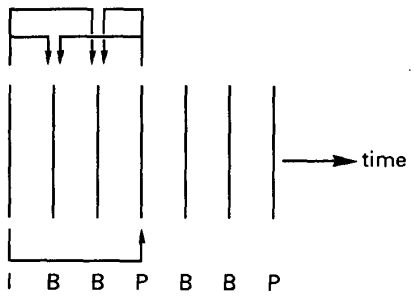


Fig. 48. The combination of intraframe coding (*I* pictures), temporal predictive coding (*P* pictures) and bidirectional interpolative coding (*B* pictures) into one coding strategy.

tures are of great importance as they provide random-access points at which the decoding of the bit stream can start. Evidently, in order to be able to decode the *B* pictures, one has first to decode the previous and the future pictures. In the example of Fig. 48, the previous and future pictures are either *I* or *P* pictures.

We now briefly summarize the spatial compression technique. Both the *I* pictures, and the prediction error signals of the *B* and *P* pictures contain a considerable amount of spatial redundancy. This spatial redundancy is exploited by coding these signals with a block-based DCT transform coding. After transforming a block of 8 × 8 samples, the corresponding 64 DCT coefficients are quantized, zig-zag scanned, and coded with run-length coding (see, for example, [26]). Le Gall [72] explained that the picture quality can be improved by applying different methods of quantization for the *I* pictures and for the *B* and *P* pictures (e.g., see [72]).

Table IV shows some of the parameters for the MPEG core algorithm. It can be noticed that the MPEG core parameters impose restrictions on the size of a picture, and on the number of pictures per second. For example, the video signal could be a signal with a frame rate of 25 Hz, where each frame (picture) consists of 288 lines with 352 luminance samples per line (this would exactly yield the maximally allowed number of 396 macroblocks). The bit rate of this signal is equal to about 30 Mb/s. This means that for compression to 1.5 Mb/s (1.2 Mb/s video), a reduction factor of 25 is needed. The picture quality obtained at 1.5 Mb/s is sometimes referred to as "VHS-like" quality.

The MPEG standard is suited for a variety of storage and transmission applications: it allows for different parameter settings which can be optimal for the application in mind. For example, on random-access digital storage media, it may be required to have many access points in the bit stream, whereas for transmission purposes less access points may be needed. The flexibility to use a variable number of access points has been realized with the help of header information. In this header information, which precedes the coded video signal, data that describes the coding parameters, such as the picture size, frame rate, bit rate, and access points, are stored.

TABLE IV
SOME CODING PARAMETERS OF THE MPEG CORE ALGORITHM

Parameter	Value
horizontal picture size	≤ 720 pixels
vertical picture size	≤ 576 lines
macro blocks/picture	≤ 396
macro blocks/second	≤ 396 × 25 (or 330 × 30)
picture rate	≤ 30 pictures/s
bit rate	≤ 1.86 Mb/s

C. Discussion

The MPEG core algorithm has been developed for video signals with a limited resolution. The MPEG algorithm, however, is also suited for coding of video signals having higher resolutions. For example, high-quality coding of CCIR-601 signals is possible with a bit rate of, say, 5–10 Mb/s. If we compare this bit rate to the rate obtained by the intraframe DCT coding algorithm proposed for magnetic consumer TV recording (Section III-C-2), we see that it is a factor of 2–3 lower. Apparently, the use of temporal interpolative and predictive coding gives a considerable extra reduction of the bit rate. For consumer TV recording, however, the MPEG algorithm is less suited than intraframe DCT coding. For consumer recording, frame-based editing is preferred, and it should be possible to obtain a high image quality during visible search (replay at different playing speeds). For low-rate coding, the MPEG bit stream should not contain too many intraframe coded pictures, because they increase the bit rate significantly. As a consequence, editing might be possible on groups of pictures (e.g., 8) only, depending on the bit rate variation per group, resulting in a relatively poor visible search quality.

VII. CONCLUSIONS

Picture coding techniques of the past decade for digital TV and HDTV recording systems for consumer applications have been reviewed.

In this field of application, digital consumer TV recording has grown to a more or less mature status. Especially in the last five years, a strong increase in the number of system proposals can be noticed. The functional processing blocks in a digital consumer recorder are bit-rate reduction, error correction coding, and channel coding. From these steps, the bit-rate reduction system is a key aspect of home-use recorders, since it enables the use of smaller recording mechanics. This requirement is very important for an increasing number of recording applications. The relation between specific recording aspects and the signal coding has been elaborated. In particular, the search facility on a recorder substantially constrains the applicability of coding algorithms for digital recording. For this reason, motion-compensated interframe coding has not been used in recording applications.

Three methods suitable for TV compression have been discussed in detail. In the early 1980's, DPCM systems

were adopted for simplicity. Results for composite recording using DPCM have been dealt with at bit rates in the order of 20–30 Mb/s. Improvements for system robustness by making use of nonlinear prediction are outlined. A possibility for higher coding efficiency when using fixed-rate coding is to use a set of compression factors. An alternative technique, called dynamic range coding, has been identified. In such a system, small blocks are coded by adaptively quantizing the differences between the actual sample values and the minimum sample value of the current data block. Most of the recent proposals are based on discrete cosine transform coding with variable-length wordmapping. Coding techniques using ranked coefficient data are identified to give selective error protection features. Another promising development that enables the different modes in the recorder is the use of special buffering techniques, such as feedforward buffering over a group of DCT blocks which can be independently accessed. Similarly, this technique avoids large error propagation. Results using DCT coding have been reported at 20–40 Mb/s, depending on the coding algorithm applied. The popularity of DCT-based compression systems, first noticeable for transmission purposes, is also valid for recording applications. This does not hold for vector quantization (VQ) which has attracted little attention. For a substantial gain in coding efficiency, the complexity of VQ is too high for consumer applications.

In the past decade, two distinct developments can be noticed. The first one is the transition to YUV-component coding instead of using composite signals. A second development is the change to more complex compression systems, guided by the continuous progress in VLSI implementation. The improvement in coding efficiency using complex systems has not only been exploited to further reduce the bit rate, but more to increase the resolution (sampling frequency) and the overall picture quality (coding noise).

The situation for HDTV systems is still in a very preliminary stage. Basically, only experimental digital HDTV recorders without bit-rate reduction have been reported in literature. There are two basic approaches for dealing with HDTV recording. First, by direct coding of the signal or second, by compatible coding so that a part of the compressed data can be used as a TV signal. For the first approach, it has been found that the same bit-rate reduction techniques can be used as for TV coding. However, for a successful introduction of a HDTV recorder it is very important that compatibility with TV is assured. Two techniques have been discussed briefly to obtain compatibility, namely pyramidal coding and subband coding. A more detailed example has been put forward in which subband coding for the coding of 2:1 interlaced HDTV is applied. Compared to direct intraframe DCT coding of the HDTV signal, the compatible subband coding system requires about 20% more bits. It appears that the extra constraint of compatibility cannot be implemented without sacrificing coding efficiency. It is very likely that more experiments will be performed in this area in the near future, to find the best system architecture for compatible coding and to bridge the efficiency gap indicated.

In the field of digital electronic still picture recording, system constraints are somewhat different from those for moving video coding. Data compression in still picture recording helps to achieve a sufficiently high image transfer rate and to increase the storage capacity. One of the major differences with moving video coding is the printing of images with a high-resolution facility. This setup, in combination with the variable viewing distance, requires a very high subjective image quality after coding. A DPCM coding technique satisfying this requirement has been described in considerable detail. Key features of this system are adaptive prediction and quantization, based on human visual masking functions, and conditional variable-length coding. In the near future, electronic photography is likely to be implemented with memory cards, obeying the system considerations from Section V. Furthermore, it is felt that digital electronic still picture storage will become an integrated part of a camera recorder in the near future.

For low bit rate coding of video on digital storage media, such as CD-I and CD-ROM, motion-compensated hybrid DCT coding has proven to be an efficient technique. Recently, the Motion Picture Experts Group (MPEG) has proposed a draft standard for such a video coding scheme. Important features of this standard are the use of predictive and interpolative coding with motion compensation in the temporal domain. This results in a substantially higher compression factor than intraframe coding schemes. A disadvantage of the MPEG algorithm is, however, that due to the extensive use of interframe coding techniques, the algorithm is less suited for applications where random access of individual pictures is required. For example, the algorithm may be less suited for home-use digital video recording. For other applications, such as digital broadcasting, this algorithm appears to be very promising.

References

[1] K. Sadashige, "Transition to digital recording: An emerging trend influencing all analog signal recording applications," *SMPTE J.*, pp. 1073-1078, Nov. 1987.

[2] S. Gregory, *Introduction to the 4:2:2 Digital Video Tape Recorder*. London: Pentech, 1988.

[3] S. M. C. Borgers, W. A. L. Heijnemans, E. de Niet, and P. H. N. de With, "An experimental digital VCR with 40 mm drum and DCT-based bit-rate reduction," *IEEE Trans. Consum. Electron.*, vol. 34, pp. 597–605, Aug. 1988.

[4] W. Wesley Peterson and E. J. Weldon, Jr., *Error-Correcting Codes*. Cambridge, MA: M.I.T. Press, 1972, 2nd ed., chapter 5.

[5] K. A. S. Immink, "Runlength-limited sequences," *Proc. IEEE*, vol. 78, pp. 1745–1759, Nov. 1990.

[6] —, *Coding Techniques for Digital Recorders*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

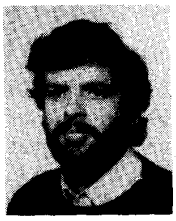
[7] B. L. Montgomery and J. Abrahams, "Synchronization of binary source codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 849–854, Nov. 1986.

[8] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE*, vol. 68, pp. 366–406, Mar. 1980.

[9] A. K. Jain, "Image data compression: A review," *Proc. IEEE*, vol. 69, pp. 349–389, Mar. 1981.

[10] H. G. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, pp. 523-548, Apr. 1985.
[11] S. Mita *et al.*, "Digital video recording techniques using 1/2-inch metal particle tape," *IEEE Trans. Consum. Electron.*, vol. CE-31, pp. 386-397, Aug. 1985.
[12] A. Hirota, S. Hirano, and S. Higurashi, "Picture coding for home VTR," in *Proc. Pict. Coding Symp.*, Tokyo, Apr. 1986, pp. 136-137.
[13] L. M. H. E. Driessen, W. A. L. Heijnemans, E. de Niet, J. H. Peters, and A. M. A. Rijckaert, "An experimental digital video recording system," *IEEE Trans. Consum. Electron.*, vol. CE-32, pp. 362-371, Aug. 1986.
[14] S. Itoi, R. Kawanaka, K. Ohshima, and M. Ashibe, "A half inch consumer-use digital VCR," in *IEEE Tech. Dig.*, ICCE-89, Chicago, IL, June 1989, pp. 124-125.
[15] T. Kondo, N. Shirota, K. Kanota, Y. Fujimori, J. Yonemitsu, and M. Nagai, "Adaptive dynamic range coding scheme for future consumer digital VTR," in *IERE Proc. 7th Int. Conf. Video, Audio and Data Recording*, York, UK, Mar. 1988, pp. 219-226.
[16] T. Kondo, Y. Fujimori, H. Nakaya, A. Yada, K. Takahashi, and M. Uchida, "New ADRC for consumer digital VCR," in *IEE Proc. 8th Int. Conf. Video, Audio and Data Recording*, Birmingham, UK, Apr. 1990, pp. 144-150.
[17] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984, ch. 12.
[18] C. Yamamitsu, K. Suesada, I. Ogura, and A. Iketani, "High-density recording and bit-rate reduction for a 2-hour digital VTR," in *IERE Proc. 6th Int. Conf. Video, Audio and Data Recording*, Sussex, UK, Mar. 1986, pp. 113-120.
[19] C. Yamamitsu, A. Ide, and T. Juri, "An experimental digital VTR capable of 12-hour recording," *IEEE Trans. Consum. Electron.*, vol. 33, pp. 240-248, Aug. 1987.
[20] C. Yamamitsu, A. Ide, A. Iketani, and T. Juri, "An experimental study on bit-rate reduction and high-density recording for home-use digital VTR," *IEEE Trans. Consum. Electron.*, vol. 34, pp. 588-596, Aug. 1988.
[21] J. H. Peters and J. T. Kanters, "Hadamard transform of composite video for consumer recording," *IEEE Proc. Int. Zurich Sem. Dig. Comm.*, Mar. 1984, pp. 46-47.
[22] P. H. N. de With, "Motion-adaptive intraframe transform coding of video signals," *Philips J. Res.*, vol. 44, nos. 2/3, pp. 345-364, 1989.
[23] H.-J. Platte, W. Keesen, and D. Uhde, "Matrix scan recording, a new alternative to helical scan recording on videotape," *IEEE Trans. Consum. Electron.*, vol. 34, pp. 606-611, Aug. 1988.
[24] P. H. N. de With and S. M. C. Borgers, "On adaptive DCT coding techniques for digital video recording," in *IERE Proc. 7th Int. Conf. Video, Audio and Data Recording*, York, UK, Mar. 1988, pp. 199-204.
[25] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal, Processing*, vol. ASSP-32, pp. 1243-1245, Dec. 1984.
[26] W.-H. Chen and W. K. Pratt, "Scene adaptive coder," *IEEE Trans. Commun.*, vol. COM-32, pp. 225-232, Mar. 1984.
[27] W. Keesen, "A new bit-assignment for transform coding," in *Proc. Pict. Cod. Symp.*, Tokyo, Apr. 1986, pp. 148-149.
[28] N. Doi, H. Hanyu, M. Izumita, and S. Mita, "Adaptive DCT coding for home digital VTR," *IEEE Proc. Global Telecomm. Conf.*, Hollywood, CA, Nov. 1988, vol. 2, pp. 1073-1079.
[29] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-6, pp. 7-12, Mar. 1960.
[30] W.-H. Chen and C. H. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. Commun.*, vol. COM-25, pp. 1285-1292, Nov. 1977.
[31] C. Yamamitsu, A. Ide, A. Iketani, T. Juri, S. Kadono, C. Matsumi, K. Matsushita, and H. Mizuki, "An experimental study for home-use digital VTR," *IEEE Trans. Consum. Electron.*, vol. 35, pp. 450-457, Aug. 1989.
[32] N. Ahmed, T. Natarjan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90-93, Jan. 1974.
[33] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. COM-33, pp. 551-557, June 1985.
[34] F. W. P. Vreeswijk, M. R. Haghiri, and C. M. Carey-Smith, "HDMAC coding for compatible broadcasting of high definition television signals," *16th Montreux Int. TV Sympos., Sympos. Record on Broadcasting Sessions*, June 1989, pp. 37-52.
[35] Y. Ninomiya, Y. Ohtsuka, Y. Izumi, S. Gohshi, and Y. Iwadate, "An HDTV broadcasting system utilizing a bandwidth compression technique—MUSE," *IEEE Trans. Broadcast.*, vol. BC-33, pp. 130-160, Dec. 1987.
[36] Y. Eto, M. Umemoto, S. Mita, and S. Nagahara, "An experimental digital VTR for HDTV," *SMPTE J.*, vol. 95, pp. 215-219, Feb. 1986.
[37] M. Umemoto, S. Mita, and Y. Eto, "High data rate recording for a high-definition digital VTR," in *IERE Proc. 6th Int. Conf., Video, Audio, Data Recording*, Sussex, UK, Mar. 1986, pp. 81-85.
[38] Y. Eto, M. Umemoto, and T. Kawamura, "Considerations for improvement of an HDTV digital VTR," *SMPTE J.*, vol. 96, pp. 177-179, Feb. 1987.
[39] Y. Hashimoto, "Experimental HDTV digital VTR with a bit rate of 1 Gbps," *IEEE Trans. Magnet.*, vol. MAG-23, pp. 3167-3172, Sept. 1987.
[40] L. Thorpe, T. Yoshinaka, and K. Tsujikawa, "HDTV digital VTR," *SMPTE J.*, pp. 738-747, Oct. 1988.
[41] H. Przybyla, "Digital HDTV recording," *Fernseh- und Kino-Technik*, vol. 44, no. 3, pp. 133-142, 1990.
[42] W. Schiffler, "HDTV recording for the studio: Analog recording within digital signal processing," *Funkschau*, No. 18, pp. 49-52, 1989.
[43] J. A. Roese, W. K. Pratt, and G. S. Robinson, "Interframe cosine transform image coding," *IEEE Trans. Commun.*, vol. COM-25, pp. 1329-1338, Nov. 1977.
[44] "Special Issue on Visual Communications Systems," *Proc. IEEE*, vol. 73, pp. 549-827, Apr. 1985.
[45] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532-540, Apr. 1983.
[46] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1278-1288, Oct. 1986.
[47] P. H. Westerink, "Subband coding of images," Ph.D. dissertation, Tech. Univ. Delft, Delft, The Netherlands, Oct. 1989.
[48] M. Pecot, P. J. Tortier, and Y. Thomas, "Compatible coding of television signals, Part 1: Coding algorithm," *Image Commun.*, vol. 2, pp. 245-258, Oct. 1990.
[49] M. Pecot, P. J. Tortier, and Y. Thomas, "Compatible coding of television signals, Part 2. Compatible system," *Image Commun.*, vol. 2, pp. 258-268, Oct. 1990.
[50] M. Vetterli, J. Kovacevic, and D. J. Le Gall, "Perfect reconstruction filter banks for HDTV representation and coding," *Image Commun.*, vol. 2, pp. 349-364, Oct. 1990.
[51] M. Breeuwer and P. H. N. de With, "Source coding of HDTV with compatibility to TV," *SPIE*, vol. 1360; also in *Proc. 5th Visual Commun. and Image Proc. '90*, Lausanne, Oct. 1990, pp. 765-776.
[52] J. Biemond, F. Bosveld, and R. L. Lagendijk, "Hierarchical subband coding of HDTV in BISDN," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 2113-2116.
[53] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[54] C. R. Galand and H. J. Nussbaumer, "New quadrature mirror filter structures," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 522-531, June 1984.
[55] H. Yoshida, "Standardization activities on multimedia coding in ISO," *Signal Processing: Image Commun.*, vol. 1, pp. 3-16, 1989.
[56] G. P. Hudson, "The international standardization of a still picture compressing technique," *Brit. Telecom Technol. J.*, vol. 7, pp. 41-47, July 1989.
[57] G. K. Wallace, "Overview of the JPEG (ISO/CCITT) still image compression standard," ISO/JTC1/SC2/WG8 N932, Nov. 1989.
[58] ISO coded representation of picture and audio information, ISO/JTCI/SC2/WG8 N933, Jan. 16, 1990.
[59] W. B. Pennebaker *et al.*, "An overview of the basic principles of the Q-coder adaptive binary arithmetic coder," *IBM J. Res. Develop.*, vol. 32, pp. 717-726, Nov. 1988.
[60] S. Kageyama *et al.*, "Digital still picture recorder utilizing an ordinary audio cassette deck," *IEEE Trans. Consum. Electron.*, vol. CE-31, pp. 96-107, May 1985.
[61] N. Takahashi, "Digital still HDTV system," *IEEE Trans. Consum. Electron.*, vol. 34, pp. 64-71, Feb. 1988.
[62] G. Caronna and P. A. M. van Grinsven, Philips Research Labs, Eindhoven, The Netherlands, personal communication.

[63] F. Izawa *et al.*, "Digital still video camera using semiconductor memory card," *IEEE Trans. Consum. Electron.*, vol. 36, pp. 1-9, Feb. 1990.
[64] G. Caronna, "Adaptive DPCM with conditional coding," in *Proc. 3rd Int. Workshop HDTV*, Torino, Italy, Aug.-Sept. 1989.
[65] H. Murakami, S. Matsumoto, Y. Hatori, and H. Yamamoto, "15/30 Mbit/s universal digital TV codec using a median adaptive predictive coding method," *IEEE Trans. Commun.*, vol. COM-35, pp. 637-645, June 1987.
[66] J. Salo and Y. Neuvo, "A new two-dimensional predictor design for DPCM coding of video signals," presented at the 2nd Int. Workshop on Signal Process. of HDTV, L'aquila-Italy, Feb.-Mar. 1988.
[67] G. A. Fry and S. H. Bartley, "The effect of the one border in the visual field upon the threshold of another," *Amer. J. Physiol.*, vol. 112, pp. 414-421, 1935.
[68] A. Forentini, "Further measurements of the differential threshold in the presence of a spatial illumination gradient," *Atti Ford-Ronchi*, vol. 11, pp. 67-71, 1956.
[69] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, vol. 65, pp. 536-548, Mar. 1977.
[70] "Coding of moving pictures and associated audio," Committee Draft of Standard ISO 11172, ISO MPEG 90/176, Dec. 1990.
[71] "MPEG proposal package description," Document ISO/WG8/MPEG/89-128, July 1989.
[72] D. Le Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, pp. 46-58, Apr. 1991.



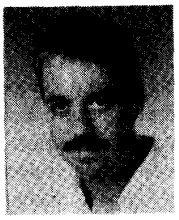
Peter H. N. de With (S'81-M'84) received the Ir. degree in electrical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1984.

He joined Philips Research Laboratories, Eindhoven, The Netherlands, in 1984, where he became a member of the Magnetic Recording Systems Department. Since then he has been working on video-data compression techniques and their implementation for digital video recording. He is currently working towards the Ph.D. degree. From 1985 to 1988 he participated in a European research project on digital video recording. Presently, he is involved in the European RACE 1001 project in which recording of HDTV signals is studied. He has written several papers dealing with data compression techniques for digital recording.

Mr. de With is a member of the Working Group for Information and Communication theory in the Benelux.

Marcel Breeuwer received the Ir. degree in electrical engineering from the University of Technology, Delft, The Netherlands, in 1982, and the Ph.D. degree from the Free University of Amsterdam, Amsterdam, The Netherlands, in 1985. His Ph.D. work was on supplementing speechreading with auditory information.

He joined the Digital Signal Processing Department of the Philips Research Laboratories, Eindhoven, The Netherlands, in 1985. His main area of interest is algorithms for data reduction of digital video and audio signals. Currently, he is involved in the European RACE 100, and in the Eureka 95 projects, in which digital recording of TV and HDTV for consumer and professional applications is studied.



Peter A. M. van Grinsven received the Ir. degree in mathematics from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1986.

He joined Philips Research Laboratories, Eindhoven, The Netherlands, in 1987, where he first studied error-correcting codes for recording channels. Later he was involved in video compression for electronic still picture storage. He is currently working on network aspects for broadband communication.