15

Trick play on Audiovisual Information for Tape, Disk and Solid-State based Digital Recording Systems

O. Eerenberg and P.H.N. de With
NXP Semiconductors Research, CycloMedia Technology /
Eindhoven University of Technology
The Netherlands

1. Introduction

Digital video for consumer applications became available in the early 1990s, enabled by the advances in video compression techniques and associated standards and their efficient implementation in integrated circuits. Convergence of transform coding and motion compensation into a single hybrid coding scheme, resulting in various standards in the early 1990s, such as MPEG-1 (ISO/IEC11172-2, 1993), MPEG-2 (ISO/IEC13818-2, 2000) and DV (CEI/IEC61834, 1993). These standards are capable of compressing video at different quality levels with modest up to high compression ratios and differing complexity. Each of the previous standards has been deployed in a digital storage system, based on different storage media.

The above hybrid coding schemes, distinguish intraframe and interframe coded pictures. The former, also labelled as I-type pictures, can be independently decoded so that it can be used for video navigation. The latter coding form result in P-type and B-type pictures requiring surrounding reference pictures for reconstruction. The availability of these reference pictures cannot be guaranteed during fast-search trick play, which makes these pictures unsuitable for certain navigation functions. Digital consumer storage standards are equipped with locator information facilitating fast-search trick play. The locator information considers the data dependencies and enables the system entry points for proper decoding. These mechanisms form the basis for selectively addressing coded data and the associated data retrieval during trick play.

This chapter discusses trick play for push- and pull-based architectures and elaborates on their implementation for tape, optical and disk-based storage systems. Section 2 introduces traditional and advanced video navigation. Section 3 presents the concepts of low-cost trick play. Section 4 elaborates on trick play for tape-based helical-scan digital video recording. Section 5 discusses trick play in relation with three popular optical-storage systems. Section 6 introduces trick play for push- or pull-based personal video recording deploying a hard-disk drive or solid-state disc. The chapter concludes and presents a future outlook in Section 7.

2. Navigation methods

Video navigation is defined as video playback in non-consecutive or non-chronological order as compared to the original capturing order. Video navigation can be divided into

Source: Digital Video, Book edited by: Floriano De Rango,
ISBN 978-953-7619-70-1, pp. 500, February 2010, INTECH, Croatia, downloaded from SCIYO.COM

traditional fast forward or fast rewind playback and advanced search methods, which are modern forms of video navigation. The former is found in analogue and digital video recorders. The latter has become possible for random accessible media such as disc and silicon-based memories. This section covers the basic aspects of traditional video navigation and presents two forms of advanced video navigation. The navigation methods are presented without implementation aspects. Implementation aspects will be discussed in consecutive sections addressing tape, optical disk, silicon-based storage solutions.

2.1 Traditional video navigation

For traditional video navigation a distinction is made between fast search and slow search mode, also known as trick play. Let *Ps* be the relative playback speed, which is unity for normal play, then fast-search trick play is obtained for $Ps > 1$ and slow-motion trick play for $Ps < 1$. Although this is a firm seperation between fast search and slow motion, there is an overlapping area for playback speeds in the vicinity $L_b < Ps < U_b$ of normal play.

Fast-search video navigation is characterized in the sense that the pictures forming the trick-play sequence are derived from the normal-play sequence by applying a temporal equidistant sub-sampling factor, which corresponds to the intended playback speed *Ps*. In practice, *Ps* is limited but not restricted to integer values.

Slow-motion search is obtained by repetitive display of each normal-play picture. The amount of repetitions is equal to the reciprocal of *Ps*. Again, practical values for *Ps* are limited but not restricted to integer values. Although the basic operation to obtain slow motion has low costs, a distinction is made between slow motion on progressive video and interlaced video. When video originates from an interlaced video source, repetition of an interlaced picture may result in motion judder. Such a situation occurs when the spatial area contains an object that is subject to motion between the capture time of the odd and even field, forming a single frame picure. Repetition of such an interlaced picture causes the object repetetively traveling along the trajectory, which is perceived by the viewer as motion judder (van Gassel et al., 2002).

The above solutions for trick play are based solely on using resampled video information. For playback speeds in the vicinity of unity, there is an alternative implementation for trick play, that builds on also re-using the normal play audio information. For this situation, time-scaled normal-play audio information is used to create an audiovisual trick-play sequence. To maintain audibility of the time-scaled audio information, pitch control is required. The playback speeds *Ps* that can be used for this type of trick play depend heavily on aspects of the normal play sequence such as speed of the oral information and the algorithm used for processing the audio signal. Algorithms for time- and pitch-scaling can be divided into frequency-domain and time-domain methods. Good results are obtained using Pointer Interval Controlled OverLap and Add (PICOLA) (ISO/IEC 14496-3, 1999), an algorithm that operates in the time-domain on mono-channel audio signals for playback speeds in the range $0.8 \leq Ps \leq 1.5$ for audiovisual content with spoken text. This form of trick play indicates that *Ps* does not need to be limited to integer-based values.

2.2 Advanced video navigation

Traditional video navigation methods are adequate for navigation in video sequences with a duration of a few minutes. Navigation becomes already a daunting task when a two-hour movie is searched for a particular scene. Increasing the playback speed *Ps* up to a factor 50 or even higher, does not result in a better navigation performance. There is a twofold reason for this. First, at high playback speeds, pictures forming the trick-play video sequence are less correlated. As a result, it is difficult for the viewer to interpret each individual picture. Lowering the refresh rate by a factor three, which means that each picture is rendered three times and maintaining the playback speed *Ps,* results in an increment of the normal-play temporal sub-sampling factor, thereby effectively tripling *Ps*. For example, if *Ps* is equal to 50, this value will be scaled to 150. A typical scene duration lasts 3–4 seconds. With a *Ps* of 150, the trick-play sequence may not contain information of all scenes, which makes this approach less suitable for fast search. A more effective method to navigate through hours of video is *hierarchical* navigation, which is based on the usage of mosaic screens to obtain an instant overview of a certain time interval (Eerenberg & de With, 2003). When descending the hierarchy, the mosaic screens contain images that correspond to a smaller time interval, increasing the temporal resolution. Images used for the mosaic screen construction can be either based on a fixed temporal sub-sampling factor, or based on the outcome of a certain filter e.g. a scene-change detection filter. Mosaic screens based on fixed temporal sub-sampling, results in selection of normal-play pictures that in traditional video navigation are used to create a trick-play video sequence. When using these pictures for hierarchical navigation, the navigation efficiency (presence of many similar sub-pictures) of the lowest hierarchical layer depends on the applied temporal sub-sample factor. The usage of scene-change algorithms results in the selection of normal-play pictures that are effective from information point of view. This becomes apparent when only a single picture of a scene is used for construction of a mosaic screen, avoiding the navigation efficiency problem of equidistant temporal sub-sampling.

Hierarchical video navigation forms a good extension on traditional video navigation, but involves user interaction for browsing through the individual mosaic screens. Another advanced video navigation method which minimizes user interaction, is obtained when combining normal-play audiovisual information and fast-search video trick-play (Eerenberg et al., 2008). Time- and pitch scaling of audio information to create an audiovisual navigation sequence is only effective for trick-play speeds about unity, as indicated in Section 2.1. Although the human auditory system is a powerful sensory system, concatenation of normal-play audio samples corresponding to the time interval of a picture selected for fast-search trick play, is not an effective approach. The auditory system requires depending on the audio type, e.g. speech or music, hundreds of milliseconds of consecutive audio to become effective with respect to interpretation of the received audio signal. Based on this observation, an audiovisual trick-play method is proposed, in which normal-play audiovisual fragments are combined with fast-search video, resulting in a double window video with corresponding audio. The strength of this navigation method is that the viewer is not only provided with a course overview of the stored video information via the fast-search video trick-play sequence. Additonally, the viewer is also provided with detailed audiovisual information via the normal-play fragments. The duration of the normal-play fragments can be freely chosen obeying the minimal duration required by the auditory system to recognize well-known audio content or interpret new audio information. For the former situation, an audio fragment duration of about 1 second is suitable for recognition the origin of that particular fragment. For the latter situation, a longer duration is required leading to a practical value of 3 seconds.

3. Low-cost trick play

The basic principles of traditional trick play have been presented in Section 2. This section discusses low-cost trick play for fast search as well as slow motion in the context of the applied storage architecture. Section 3.1 will discuss two types of architectures. In digital consumer recording systems, trick play is a low-cost feature, resulting in the need for simple high-efficient signal processing algorithms fulfilling the cost requirement. Low-cost trick play is in general characterized by signal processing in the compressed domain, avoiding expensive transcoding. This involves selection and manipulation of coded pictures, which is covered in succeeding sections.

3.1 Impact of pull and push-based architectures on trick play

The chosen architecture of the storage system influences the involved trick-play signal processing. A digital storage system can have either a pull- or a push-based architecture. In a pull-based architecture, the video decoder pulls the data from the storage medium, whereas in a push-based architecture the video decoder receives the audiovisual information at a certain rate. For traditional trick play, there is a difference in performance due to the different architectures. Moreover, a push-based architecture allows two approaches. The first approach is based on a solution provided by the MPEG-2 system standard (ISO/IEC13818-2, 2000). This involves trick-play signalling information provided by the Packetized Elementary Stream (PES) header. This information is used to control the output of the video decoder during trick play. The second solution applies signal processing in the compressed domain, avoiding the usage of previous MPEG-2 system signalling information (van Gassel et al., 2002). Although the first solution is described by MPEG-2, its usage is optional and not obligatory (ETSI_TS_101154, 2009). The second solution is a generic approach, which delivers a compliant MPEG-2 video stream. This makes it an attractive solution from the video decoder point of view.

3.2 Influence of video encoding parameters on trick play

Modern compression schemes such as MPEG-2 or H.264, achieve high compression ratios by exploiting spatial and temporal correlation in the video signal. Compression of pictures exploiting only spatial information are intraframe compressed, whereas pictures that exploit temporal correlation are interframe compressed. The latter can be split in two categories: uni-directional (P-type) and bi-directional (B-type) predictive pictures. In a compressed video sequence, the distance between two successive intraframe compressed pictures is expressed by N, which is also known as the Group-Of-Pictures (GOP) length, whereas the distance between P-type predictive pictures is expressed by M. For the situation that $M>1$, the number of B-type pictures preceding a P-picture is equal to $M-1$. For the common situation that N is an integer multiple of M, the construction of trick play sequences in simplified because P-pictures are a sub-integer fraction of the GOP length N.

In general, trick play for digital consumer storage equipment is a low-cost feature, which limits the amount of involved signal processing. Low-cost trick-play algorithms deploy pictures that are selected from the compressed normal-play sequence. For fast-search trick play, the minimum fast-forward search speed is equal to M, whereas other fast-forward speeds are obtained for speed-up factors equal to N, or an integer multiple of N. Note that there is basically a gap between speed-up factor M and N if $N >> M$. This gap is caused by the fact that

P-type pictures can only be decoded if the reference (anchor) picture has been decoded. For typical video compression applications such as Digital Video Broadcast (DVB), or digital recording, the GOP size *N=12* and P-picture distance *M=3*, resulting in fast-search speed-up factors *Ps={ 3, 12, 24, …, 12n }*, with *n={1, 2, 3… }*. Fast-search trick play in reverse direction is obtained in a similar way, but for a speed-up factor equal to *M*, buffering is required to store the decompressed pictures to facilitate reordering. This is required to match (potential) motion with the reverse playback direction, as the decoding is always performed in the positive time direction. This forward decoding direction introduces an extra delay, which occurs only once when switching to the reverse search mode with speed-up factor equal to *M*.

From the concept point of view, trick play in the compressed domain is equal for push or pull-based architecture. However, from a compliancy point of view, they are different. In a pull-based architecture, the decoder retrieves either fragments containing the intended intraframe compressed pictures, or the whole compressed video sequence at a higher rate. For either case, there will most probably be a frame-rate and bit-rate violation. From trick-play point of view, a high quality is perceived by the viewer both for low and high search speeds.

For push-based architectures, fast-search trick play based on concatenation of normal-play intraframe compressed pictures may cause a bit-rate violation. A low-cost method to overcome this bit-rate violation is the usage of repetition pictures, i.e. interframe compressed pictures, of which the decoded result is identical to the reference image (anchor picture), which is the last transmitted intraframe or P-type interframe picture, see Fig. 1. A repetition picture precedes the transmission of an intraframe coded picture. Multiple repetition pictures are involved when the transmission time exceeds two or more display periods, where one display period is equal to the reciprocal of the frame rate. For the situation that the normal-play video is MPEG-2 coded, a repetition picture has an extreme small size of 342 Bytes for a picture size 720×576 and 4:2:0 sampling format. The transmission time for such a compressed picture is negligible compared to the display period, leaving the remaining time for transmission of intraframe compressed data. This concept introduces fast-search trick play with a reduced refresh rate, so that the frame rate is not jeopardized. This concept yields a proper result for progressive as well as interlaced video, although motion judder may be introduced for interlaced video. Motion judder is avoided when the two fields that form the repetition pictures are field-based coded, where

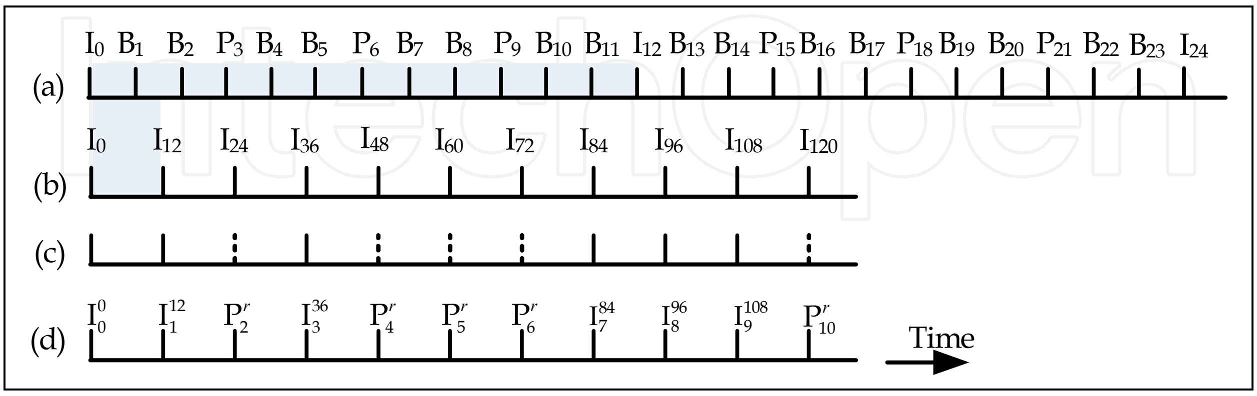


Fig. 1. Low-cost MPEG-compliant fast-search trick play. (a) Normal play compressed pictures with absolute picture index, (b) Intraframe compressed normal-play pictures, (c) Sub-sampling of the normal-play sequence, using a speed-up factor of 12, where the dashed lines indicate the skipped pictures, (d) MPEG-compliant trick play for speed-up factor 12 equipped with repetition pictures to avoid bit-rate violation.

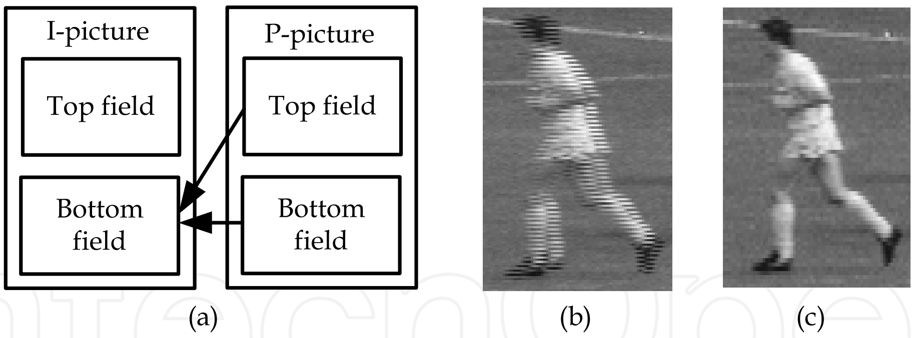


Fig. 2. Effect of "interlace kill". (a) Predictive picture causing top and bottom field to have equal content. (b) Snapshot of I-picture region constructed of two fields from different time instances. (c) Snapshot of the same region after applying the "interlace kill" feature.

both fields refer to the same reference field, resulting in an "interlace- kill" operation see Fig. 2 (van Gassel et al., 2002). In this operation, one field is removed and replaced by the remaining field.

The reference field depends on the search direction and is bottom field for forward and top field for the reverse search direction. The result of the required decoding operation is that after decoding, both fields are equal, eliminating the possibility of motion portrayal. Using the MPEG-2 coding format, the bit cost for such a predictive-field coded picture is 954 Bytes, with a picture size of 720×288 pixels and 4:2:0 sampling format. The coding of two involved field pictures results in a total of 1,944 Bytes, including the required picture header and corresponding extension header.

3.3 Slow-motion for push-based storage systems

Slow-motion trick play is based on repetitive display of a picture. The number of display periods that a normal-play picture is displayed, is defined by Eqn. (1), where D_p indicates the number of display repetitions and Ps the slow-motion speed factor, giving

$$D_p = \frac{1}{Ps}. \tag{1}$$

Also here, there is a difference between push- and pull-based architectures. In a pull-based architecture, the video decoder outputs the decoded image data D_p times. In a push-based architecture, a slow-motion sequence is derived on the basis of the normal-play compressed video sequence. The slow motion operation requires that newly created display periods, are filled up with the repetition of normal-play pictures. For progressive video, all newly created frame periods are inserted after each original frame. Repetition of normal-play I-type and P-type pictures is achieved by uni-directional coded B-type repetition pictures. Any normal play B-type picture is repeated via repetitive transmission and decoding of the original B-picture. Special attention is required for the repetition of normal-play anchor pictures when they are originating from an interlaced video source. To avoid corruption of the decoder reference pictures, only uni-directional B-type coded "interlace-kill" pictures can be used for repetition of the normal-play anchor pictures, as they do not modify the anchor picture memory.

Anchor pictures maintaining their proper content is a basic requirement for the slow-motion decoding process, as each predictive-coded normal-play picture may use reference data from two or more fields. The basic problem is that an "interlace-kill" operation can only be applied to the anchor pictures, which are intraframe or P-type predictive normal-play pictures. There is no low-cost algorithm for removing interlacing to process the B-type coded pictures, when they are frame-based coded. For the situation that B-type pictures are field based, only one field out of two B-type field pictures is used. As a result, frame-based B-type coded pictures can only be used once, if the video originates from an interlaced source.

Slow motion results in $N \cdot D_p$ pictures per normal-play GOP. For the situation that the video originates from an interlaced video source, a display error D_e occurs as indicated in Eqn. (2), the display error per normal-play GOP. From Eqn. (2), it becomes clear that for $M=1$, which conforms to MPEG-2 simple profile, no display error occurs, as indicated by

$$D_e = N(D_p(1-\frac{1}{M})-(1-\frac{1}{M})). \tag{2}$$

In order to avoid a speed-error, normal-play anchor pictures which are displayed D_p times, must be displayed an additional A_r times, as specified in Eqn. (3), resulting in

$$A_r = D_p(M-1)-(M-1). \tag{3}$$

4. Digital tape-based helical-scan video recording

This section discusses Digital Video (DV) and Digital Video Home System (D-VHS), two tape-based Video Cassette Recording (VCR) standards developed in the 1990s. The DV standard differs from the D-VHS standard in the sense that video coding, on tape storage and trick play form an integral approach. The standardization of D-VHS followed a two-step approach. First, a basic recording engine was established capable of storing an MPEG-2 transport stream. In the second stage of the standardization process, trick play was added to this system. This section is divided in three sections. Section 4.1 discusses the impact of helical-scan recording on trick play. Section 4.2 elaborates on the DV trick-play solution. Finally, Section 4.3 presents trick play for the D-VHS STD mode.

4.1 Helical-scan video recording

Tape has been a popular storage medium for many decades due its good price/storage capacity ratio and the high capacity storage per volume unit. Tape-based video recorders apply helical-scan recording, where the magnetic heads are mounted on a rotary head wheel inside a cylindrical drum, and the tape is helically wrapped around it. Such a recording solution provides a large recording bandwidth, due to the fast rotation speed of the scanner and the relatively slow tape speed. This concept creates high-density recording with modest sized cassettes and sufficient playing time. The heads mounted on the rotary wheel have a different azimuth and write the information in slanted tracks on tape, resulting in a high track density, due to partial overlapped writing. Moreover, each track forms a fixed bit-rate channel, see Fig. 3 (a).

During playback, the heads with the proper azimuth scan the corresponding tracks on tape, a process that is controlled by the tracking servo. During normal play, see Fig 3 (b), all information that was recorded on tape is read from tape. A different situation arises for trick play. During fast-search trick play, the tape travels at a higher speed along the scanner. As a result, the heads that scan the tape, under control of the tracking servo, follow a different path when compared to normal play, see Fig 4 (a). The tracking servo controls the scanner scan-path via adaptation of the tape speed and allows for two concepts. The first concept is based on speed-lock. Here the heads scan the fast travelling tape, where there is no guarantee which head scans the tape at a particular moment. The second concept is phase-lock. In this concept, the scanner is positioned such that at the beginning of the scan path, the head with a particular azimuth scans the track with the corresponding azimuth. The impact of the different servo approaches has its influence on the available trick-play bit rate. For the situation that the more complex phase-lock servo system is deployed, there is no need for duplicating the trick-play data, which results in a higher available trick-play bit rate because the head starts at the track with the corresponding azimuth, see Fig. 4 (b). For the situation that a speed-lock servo system is used, trick-play data needs to be stored multiple times, such that it reads once during fast search, see Fig. 4 (c).

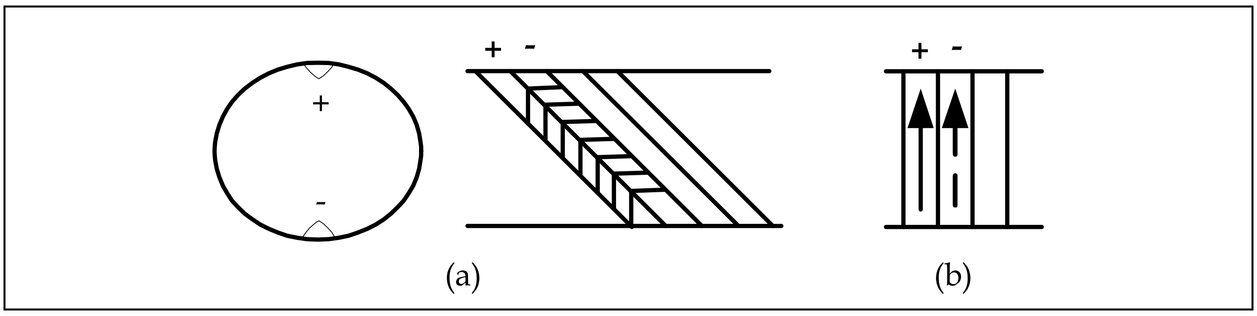


Fig. 3. Helical-scan recording. (a) Two-head scanner and corresponding azimuth tracks, (b) Simplified track diagram showing normal scan path, the non-dashed arrow indicates scan path of `+` head and dashed arrow indicates scan path of `–` head.

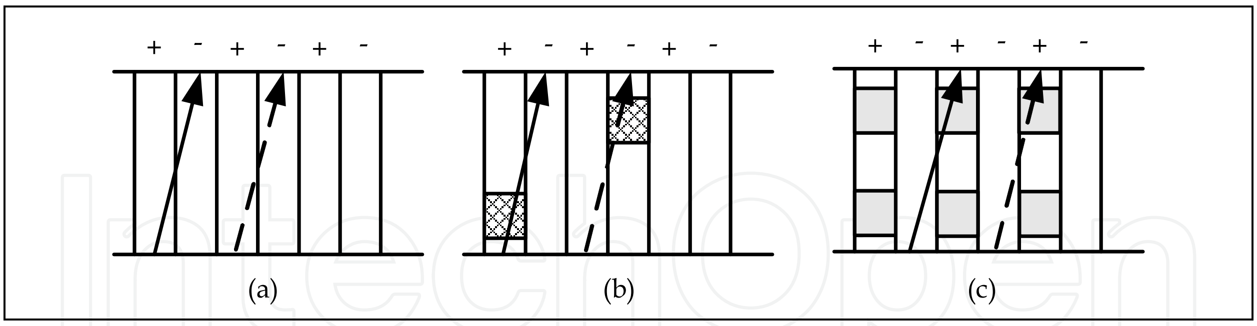


Fig. 4. Helical-scan fast search scan path. (a) Fast scan path for speed-up factor $Ps=2$, (b) Example of trick-play data location for phase-lock-based servo solution, (c) Example of trick-play data location for speed-lock-based servo solution.

The obtained trick-play quality is not only influenced by the applied servo system, but also by the applied video compression scheme. In the DV recording system, tape and compression related aspects have been combined, thereby facilitating trick play on the basis of re-using normal-play compressed video information. D-VHS is a recording system where storage part (bit engine) and compression have been separated, resulting in a trick-play solution that facilitates individual trick play (virtual) channels for various trick-play speed-up factors and their corresponding directions.

4.2 Trick play for digital video

The DV system (CEI/IEC61834, 1998) is a recording standard intended for 25-Hz (PAL) and 30-Hz (NTSC) television standards and was first announced in 1993 (Matsushita et al., 1993). The applied video compression was largely guided by system aspects such as editing on a picture basis, robustness for repetitive (de-)compression, number of tracks to store a compressed picture, very high forward and backward search on tape, overall robustness and high picture quality. In order to understand the DV trick-play mechanism, a brief system overview is presented, which introduces the essential aspects that enable trick play.

4.2.1 System architecture

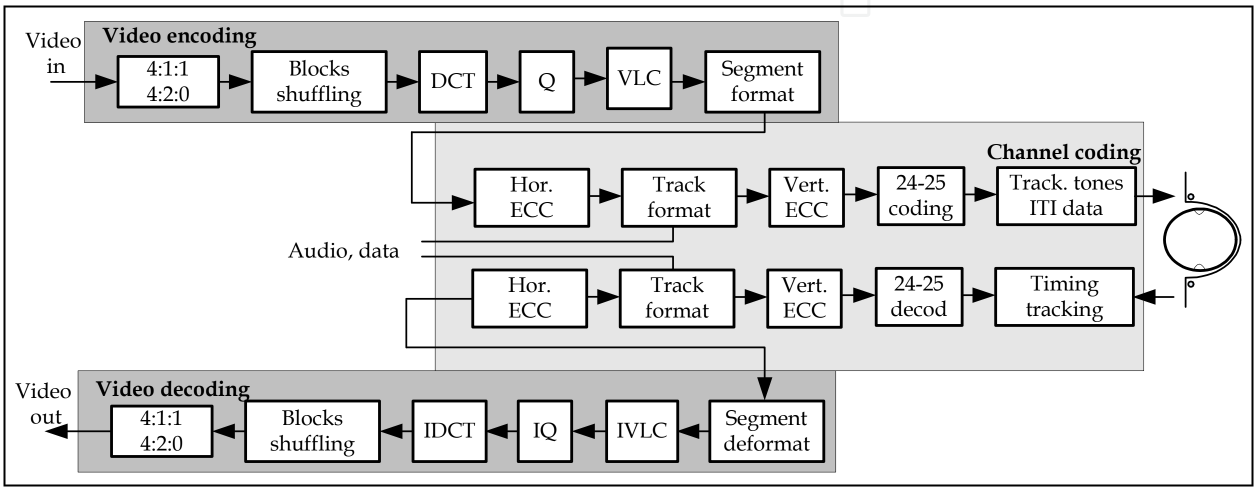


Fig. 5. Block diagram of complete processing of consumer DV recording system.

Figure 5 depicts the block diagram of the DV recording system. At the left-hand side, video enters the system and is compressed using a 4:2:0 sampling format for 25-Hz video or 4:1:1 sampling format in the case of 30-Hz video. Prior to compression, the video data is shuffled increasing the normal-play quality and enabling trick play. The compression is based on intraframe coding, using a Discrete Cosine Transform (DCT), and subsequent quantization and Variable-Length Coding (VLC) of the transformed video data. The compressed video or uncompressed audio data together with signalling information is packetized into Sync Blocks (SB), which are protected with parities from the Reed-Solomon-based horizontal forward error correction. To further improve the system robustness, a vertically-oriented error-correction layer is added, calculating parities over multiple SBs within a track, which are stored in dedicated parity SBs. Finally, the SBs are stored on tape using a highly efficient DC-free 24 -25 channel code with embedded tracking tones required for tracking purposes. For a 25-Hz frame rate, a picture is stored using 12 consecutive tracks or 10 tracks for a 30-Hz frame rate.

4.2.2 DV video compression and sync block mapping

DV video compression is based on macroblocks (MB) constructed from luminance and chrominance DCT blocks, see Fig. 6. Each macroblock contains a full-color area of 256 pixels, where the physical dimensions differ between 32×8 rectangular (30 Hz) and 16×16 square for (25 Hz).

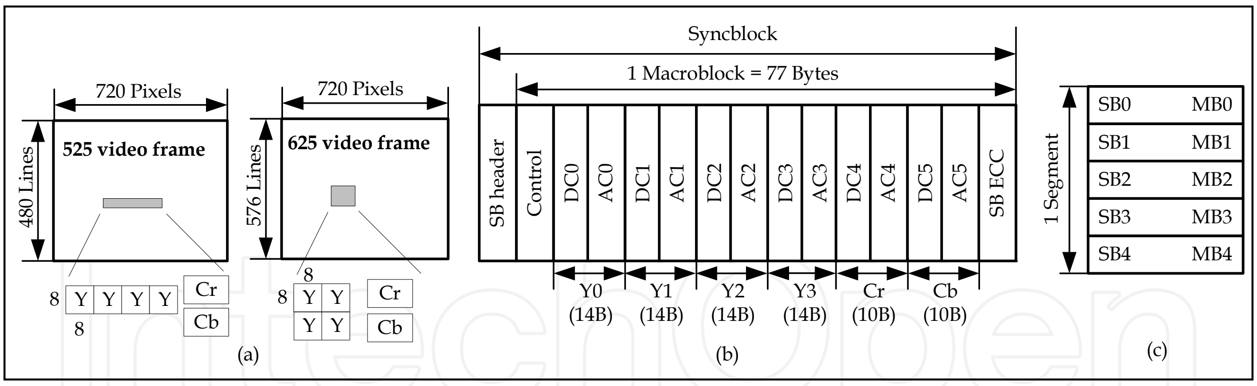


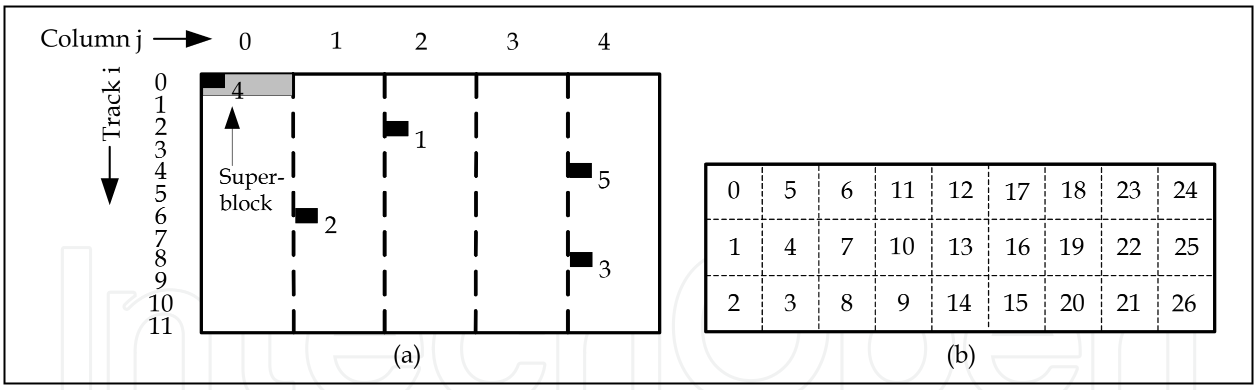
Fig. 6. DV Macroblocks. (a) Construction of macroblocks for 25-Hz and 30-Hz video, (b) SB data format showing fixed predetermined positions of low-frequent DCT coefficients, (c) Five pre-shuffled macroblocks forming a single segment.

The number of DCT blocks differ for a 25-Hz or 30-Hz system, but result in an equal number of 135 macroblocks per track (one MB per SB), which leads to a uniform recording concept that lowers the system costs. Segments are constructed from 5 pre-shuffled MBs, see Fig. 6 (c), to smoothen the coding performance and create a fixed bit cost per segment.

The compression of a segment is realized by applying a feedforward coding scheme, resulting, which high robustness and individual decodability of each segment. Although the segments are of fixed length, this does not mean that the macroblocks constructing the segment are also fixed-length coded. The optimal mapping of macroblocks on sync blocks is achieved when each macroblock is stored in a single sync block. This also enables the highest search speed (up to 100 times), due to individual decodability. By applying a fixed mapping, see Fig 6 (b) of the DCT low-frequency coefficients on a sync block, each sync block is individually decodable at the expense of a somewhat lower PSNR, due to the fact that some high-frequency coefficients may be carried by neigboring sync blocks belonging to the same segment.

The trick-play quality is determined by two factors: the trick play speed-up factor and the data shuffling. The system is designed such that the picture quality gradually deteriorates with increasing search speed. For very high search speeds (50-100 times), the reconstructed picture is based on individually decoded MBs, which are found in the individually retrieved SBs. In this mode, the vertical ECC cannot be deployed and therefore the decoder relies only on a very robust storage format where the low-frequency information of each DCT block is available on fixed positions inside the SB. For lower tape-speeds, larger consecutive portions of a track can be retrieved, so that full segments can be decoded in full quality and individual MB from partial segments. The shuffling is organized such that the quality is further increased in this playback mode. Figure 7 (b) depicts that consecutive segments contain MBs that are neighbors, which form a superblock resulting in a larger coherent spatial area from the same image. Hence, the lower the tape speed, the more neighboring MBs are successfully retrieved and the larger the coherent area that is constructed.

Figure 7 (a) depicts how the individual MBs of one segment are chosen. It can be seen that within one segment, MB data is sampling the full image both horizontally and vertically. This ensures a smoothening of the data statistics, enabling a fixed bit cost per segment. Moreover, the shaded area in the upper-left corner indicates a superblock with that is constructed from the neighboring MBs, deploying a pattern as depicted in Fig. 7 (b).



0	5	6	11	12	17	18	23	24
1	4	7	10	13	16	19	22	25
2	3	8	9	14	15	20	21	26

Fig. 7. DV tracks. (a) Selection of MBs for segment construction and assignment of picture areas on tracks, (b) Ordering of macroblocks in repetitive clusters, called superblocks.

4.3 Trick play for D-VHS STD mode format

D-VHS (ISO60774-5, 2004) is a tape-based recording system, following the push-based architecture and developed in two stages. The first stage covers the record and playback of MPEG-2 compressed normal-play information (Fujita et al., 1996), whereas the second stage addresses trick-play aspects. The D-VHS standard describes three recording modes, each intended for a specific data rate, resulting in a format for STanDard (STD) mode, Low-Speed (LS) mode and High-Speed (HS) mode. The first two recording modes share the same trick-play speed-up factors ±4, ±6, ±12 and ±24, whereas the HS recording mode supports the speed-up factors ±3, ±6 and ±12. This section describes high-end and low-cost trick-play signal processing for the STD mode format supporting speed-up factors ±4, ±12 and ±24.

4.3.1 Trick play based on information carried by a virtual channel

Trick play for the D-VHS STD mode assumes that the recording system deploys a phase-lock tracking servo (track select). The standard describes areas within the tracks, which are allocated for a particual trick-play speed-up factor for forward as well as reverse playback direction. These trick-play regions form a pattern, which is repetitive modulo 48 tracks, see Fig. 8. For a particular playback direction, i.e. forward or reverse, the trick-play regions corresponding to a particular trick-play speed-up factor, form a virtual information channel. This channel is only present during fast search at a particular speed-up factor and corrsponding direction. During record, a track is filled with 336 sync blocks, see Fig. 9 (b), which are data units of 112 Bytes. Each Sync Block (SB) consists of synchronization information, main data area and inner parity based on a Reed-Solomon forward error correction, see Fig. 9 (a). Two main data areas store a time-stamped TS packet, which is constructed of a 4-Byte header containing the time stamp and a single TS packet of 188 Bytes. The physical storage position depends on the SB number and corresponding time stamp value.

Each SB, regardless whether it contains normal-play or trick-play data, is protected by means of inner parity. The corresponding inner parity is stored at the last Byte positions of a SB. To increase the playback robustness, a second Reed-Solomon forward error correction is applied. The corresponding parity data (outer parity) are stored in the last SBs of a track. The usage of outer parity information for trick play is optional. In a push-based architecture, the isochrone nature of the stored audiovisual information is reconstructed at playback with the aid of time stamps. Time stamping is a method that captures the position of an MPEG-2

Transprt Stream (TS) packet on the time axis. At playback, the time stamp is used to position the TS packet at the proper time location. This reconstructs the isochrone behavior of the TS sequence, which enables the usage of an external MPEG-2 decoder connected via a digital interface (IEC61883-4, 1994). Time stamping in D-VHS, is based on a more accurate 27-MHz clock (±20ppm versus ±40ppm in broadcast), which is locked to the incoming TS stream. D-VHS stores this time stamp together with the TS packet in either two consecutive SBs or in two separated SBs, with trick play SBs in between. The trick-play virtual information channel contain an MPEG-2 compliant TS stream. The isochronous playback of such a trick-play sequence requires the presence of time stamps. Moreover, these time stamps are also used for adequate placement of the TS packets in the virtual channel.

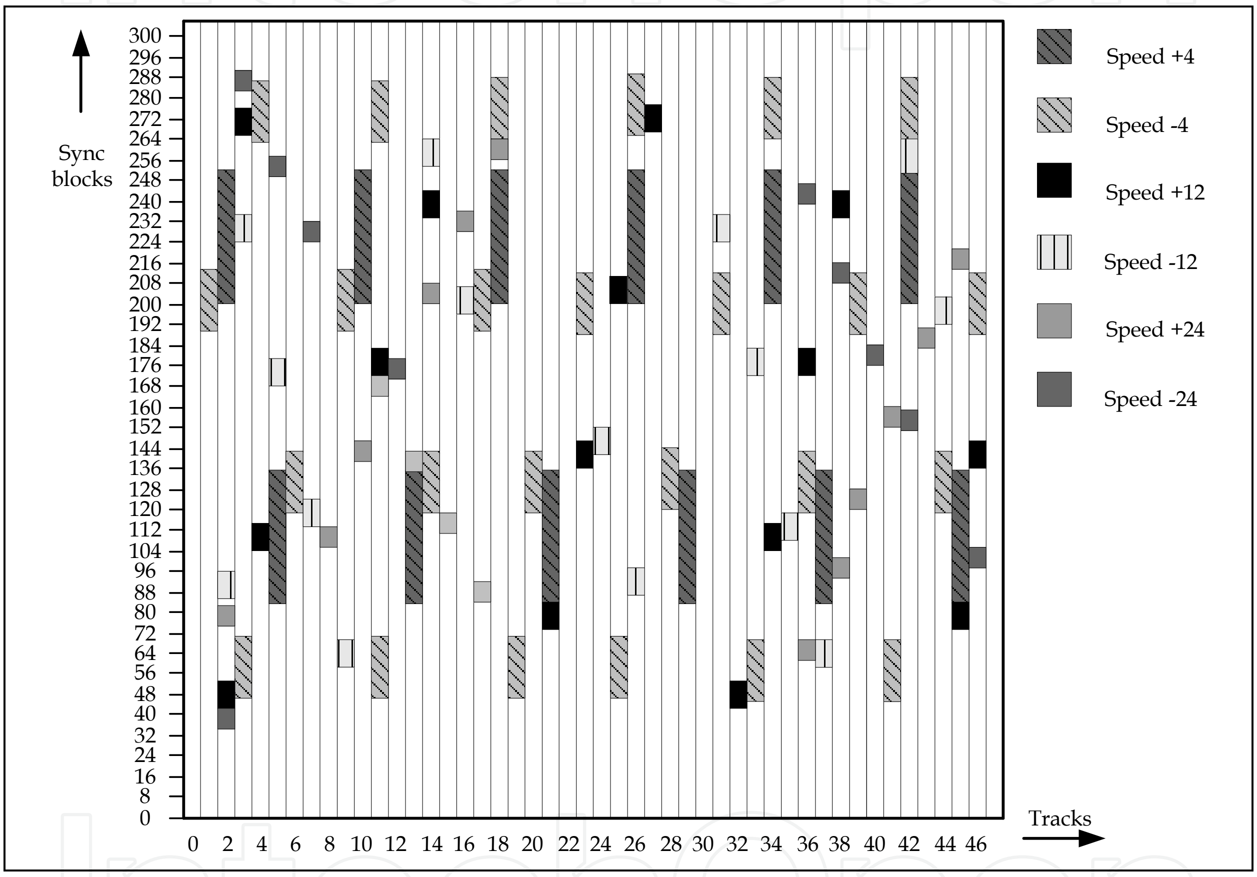


Fig. 8. D-VHS STD-mode trick play regions for speed-up factors: ± 4, ± 12 and ± 24.

Due to the tight relation between the time stamp and the physical storage location, the drum rotation phase needs to be synchronized during record and playback. Synchronization is essential because the local time stamp generator is reset modulo three revolutions of the scanner. Trick play in D-VHS is based on retrieving scattered fragments from various tracks. For the situation that the drum rotates at 1,800 rpm (30 Hz), the sync blocks allocated for each trick-play speed-up factor, result in a capacity of 2,301,120 bits/sec, see Table 1, regardless the playback direction. Trick-play data can be divided into two categories: high-end and low-cost. The former is characterized by high video quality and a refresh rate equal to the frame rate, regardless of the involved implementation cost. The latter is determined by minimizing the involved system costs. In any case, as trick play is in general a low-cost feature, the amount of signal processing should be bounded.

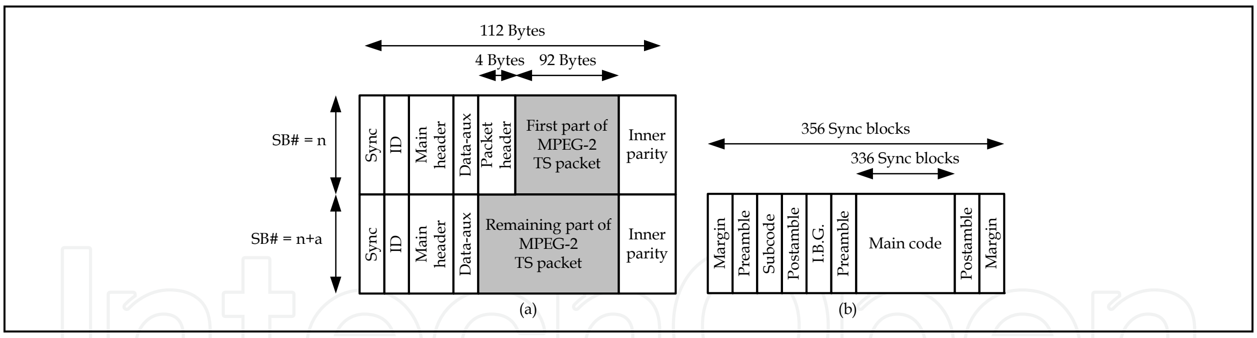


Fig. 9. D-VHS sync block. (a) MPEG-2 TS packet mapped at two SBs, (b) 336 SBs containing normal play, trick play, or padding data constructing one track.

Scanner revolution (Hz)	Trick play channel bit rate (bits/s)
30	2,301,120
29.97	2,298,821.17

Table 1. Channel bit rate for 30-Hz and 29.97-Hz scanner frequency.

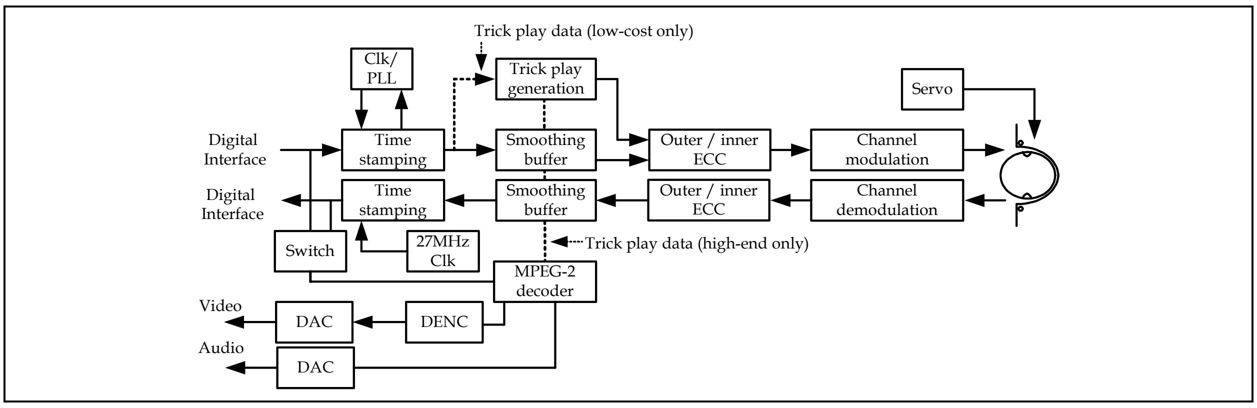


Fig. 10. D-VHS block diagram. The dashed connections indicate the information flow for the two trick-play flavors.

A basic functional block diagram that enables both high-end and low-cost trick-play implementation, is depicted in Fig. 10. The difference between the two concepts is only in the selection of normal-play pictures for the lowest forward trick-play speed, while the remaining processing steps are equal.

High-end trick play is implemented in the following way. During record of the normal play sequence, the normal play sequence is decoded and the decoded pictures are made available to the trick-play functional block, indicated by the dashed line, see Fig. 10. During trick play, the stored trick-play sequence is de-smoothed and reconstructed based on the time stamps. Low-cost trick play requires de-multiplexing of the normal-play sequence and video elementary stream parsing to locate the intraframe-compressed normal-play pictures. Due to the low trick-play bandwidth, see Table 1, high-end trick play is based on Common Intermediat Format (CIF) resolution. Almost 90,000 bits are available on a per picture basis for video with 25-Hz frame rate. Images of CIF-based resolution can be coded with sufficiently high quality using fixed bit-cost compression, resulting in a Peak Signal-to-Noise Ratio (PSNR) between 30 dB and 40 dB, depending on the spatial complexity, see Fig. 11 (b).

For low-cost trick play, the available bandwidth during trick play is not sufficient to enable a full refresh rate and preserve the picture quality when re-using normal-play intraframe-coded pictures. Hence, the concept for low-cost trick play is based on using transcoded normal-play intraframe-compressed pictures. Transcoding resulting in lower bandwidth can be obtained in two ways: removing picture energy (AC coefficients, see Table 3) or reduced refresh rate. Table 2 indicates the normal-play intraframe-compressed bit cost for various encoder settings. The minimum intraframe bit cost that remains when removing all AC energy is depicted in Table 3.

Normal play bit rate (Mbit/s)	**Number of pixels per line**	**GOP parameters**		**Average intraframe bit cost (bits)**	**Minimum intraframe bit cost (bits)**	**Maximum intraframe bit cost (bits)**
		N	**M**			
9.4	720	12	3	770,084	430,896	1,099,696
3.4	480	12	3	281,126	45,984	564,568
5.0	528	12	1	417,819	68,344	640,244
8.0	720	12	3	578,032	451,616	909,848

Table 2. Normal-play MPEG-2 intraframe bit cost for various bit rates, frequently used GOP structures, 576 lines per image, 25-Hz frame rate and 4:2:0 sampling format.

Normal play bit rate (Mbit/s)	**GOP parameters**		**Average intraframe bit cost (bits)**	**Minimum intraframe bit cost (bits)**	**Maximum intraframe bit cost (bits)**
	N	**M**			
9.4	12	3	108,326	83,600	122,488
3.4	12	3	77,329	53,944	97,200
5.0	12	1	55,012	48,032	60,176
8.0	12	3	78,915	75,336	81,840

Table 3. Transcoded DC-only bit costs for various bit rates, frequently used GOP structures, 576 lines per image, 25-Hz frame rate and 4:2:0 sampling format.

The normal-play intraframe-compressed pictures, as depicted in Table 2, result in refresh rates varying between 3 Hz and 8 Hz, which is insufficient from trick play perception (fast changing images) point of view. The DC-only intraframe-compressed pictures as depicted in Table 3, result in refresh rates up to 25 Hz, which is sufficient for trick-play perception, but the pictures lack detail due to the strong energy removal. An acceptable solution is obtained when the refresh rate is reduced to 8.3 Hz. In order to create an MPEG-2 compliant trick-play sequence, repetition pictures with or without "interlace kill" are applied, see also Section 3, resulting in a fixed bit cost of roughly 270,000 bits per trick-play GOP, see Fig. 11 (a). Various approaches have been reported to reduce AC energy from the intraframe-compressed normal-play pictures. The approaches vary between selecting the number of AC coefficients based on the amplitude of the AC coefficients (Ting & Hang, 1995) by means of prioritization (Boyce & Lane 1993), or on the basis of the differential DC value between successive DCT blocks (US6621979, 1998). Based on the concepts for low-cost trick play, the final trick-play quality depends on the normal-play quality and associated factors, due to the re-use of pictures. This results for luminance-only information, in a 25-dB PSNR for DC-only trick-play pictures, up to the original normal-play PSNR, when the spatial content of the pictures are of low complexity.

The trick-play information channel is filled with a multiplex of six individual transport streams, which are all derived from a single information stream, in this case the lowest fast-forward sequence. During trick-play, the higher and all supported reverse speeds are derived from fragments belonging to that particular trick-play sequence. Once the fixed bit-cost trick-play GOP for speed-up factor *Ps*=*+4*, is available (regardless whether high-end or low-cost), the trick-play video sequences for the other forward trick-play speed-up factors (*Ps*=*+12* and *Ps*=*+24*) are derived from this trick-play sequence by the "search speed data selection" block, see Fig. 11 (c), via sub-sampling in the compressed domain of the fixed bit-cost GOPs.

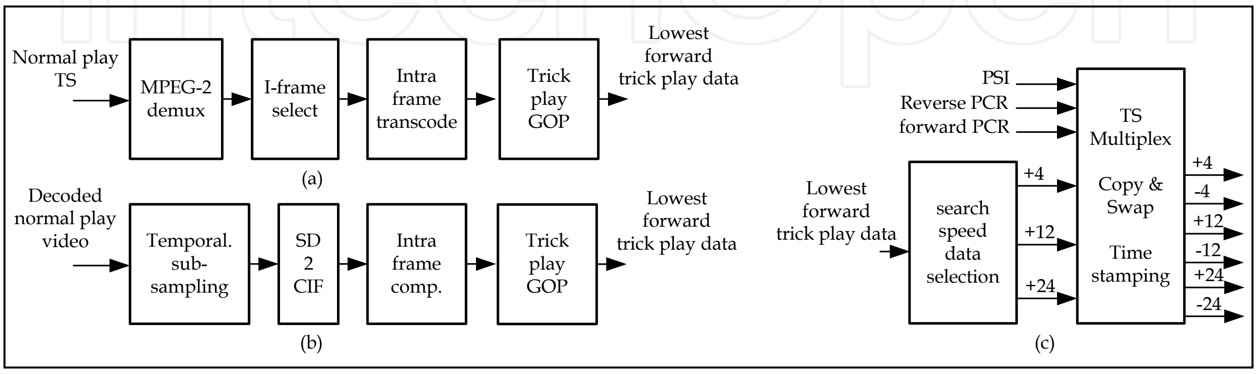


Fig. 11. D-VHS trick-play processing. (a) Low-cost trick-play video elementary stream generation, (b) High-end trick-play video elementary stream generation, (c) MPEG-2 TS multiplexing, time stamping and reverse trick-play generation.

The six trick-play streams are constructed as follows, assuming that the lowest search speed GOP is available, see Fig. 11 (a) and (b) . The functional block "TS multiplex Copy & Swap Time stamping", see Fig. 11 (c), performs three final operations on the forward trick-play elementary streams. The first operation is multiplexing the elementary streams into MPEG-2 Transport Streams (TS). The second operation is to copy and swap the TSs creating the trick-play sequences for the reverse search direction. Due to the fact that MPEG-2 decoding uses an increasing time axis, the reverse trick-play sequences are generated with a declining time axis. The swap operation on TS data is required to read the GOP in the correct order during trick play. Finally, the TS packets are equipped with a time stamp, such that the TS packets are properly mapped on the corresponding trick-play regions.

5. Trick play for optical storage systems

Optical storage systems differ from the traditional tape-based storage system in the sense that they offer random access to the stored audiovisual information, enabling trick-play generation at playback using the stored normal-play audiovisual information. Although no separate trick-play sequences are stored on the storage medium, additional information is stored facilitating trick play. In this section, trick play is described for two optical standards: SuperVCD and DVD.

5.1 Super Video Compact Disc (SuperVCD or SVCD)

SuperVCD is the successor of Video Compact Disc (VCD) and is also known as SVCD, offering up to 800 MByte storage capacity, using a 780 nm laser. This standard appeared on the market in the early 1990's (IEC62107, 2000). SuperVCD utilizes better audio and video quality compared to VCD, which used MPEG-1 video compression at Common Intermediat Format (CIF) resolution with a fixed bit rate of 1.1458 Mbit/s and MPEG-1 layer 2 audio at 256 kBit/s. The improved video quality of SuperVCD was later obtained by going to higher spatial resolution 480×576 (H×V) at 25 Hz for PAL TV system and 480×480 (H×V) at 29.97 Hz for the NTSC TV system. To code the interlaced video more efficiently, MPEG-2 video coding is applied in combination with variable bit rate. Multi-channel audio has been added to the system to increase the audio quality. Figure 12 depicts the SuperVCD reference model. From Fig. 12, it can be seen that the amount of data delivered by the CD module differs for MPEG sectors and data sectors, which is caused due to unequal error protection resulting in a higher (14%) storage capacity for an MPEG sector and less protection of the MPEG-compressed data. The layout of an SVCD consists of a lead-in, program area and a lead-out area, as depicted in Fig. 13 (a).

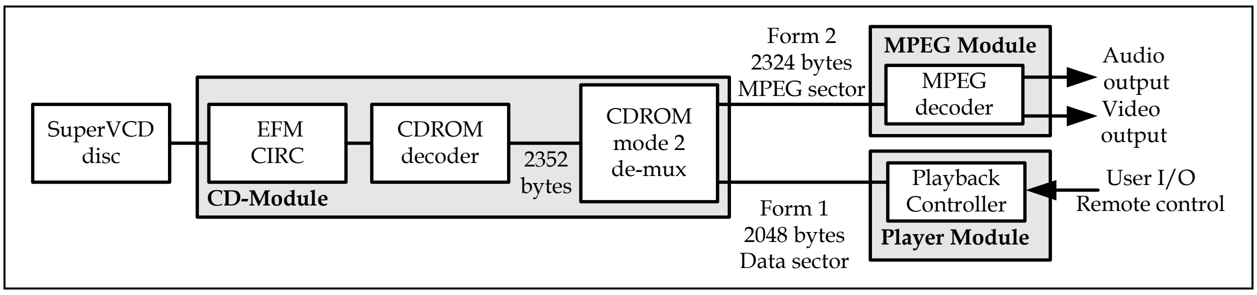


Fig. 12. SuperVCD system reference Model.

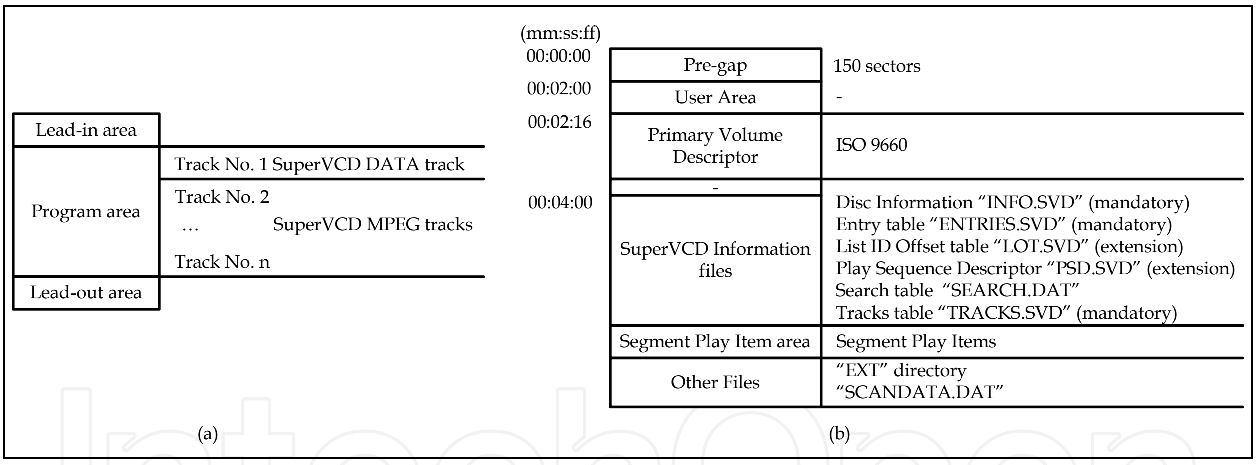


Fig. 13. SuperVCD disc and data track. (a) SuperVCD disc layout, (b) Data track layout example.

The program area is split into a SuperVCD data track and SuperVCD MPEG tracks. The former track stores additional information required by the SuperVCD player. The latter contains the MPEG-2 Program Stream containing the audiovisual information, which can be sub-divided into parts or chapters and are stored on the SuperVCD directory "MPEG2". SuperVCD is based on a pull-based architecture. Traditional trick play for such a system, as discussed in Section 3, can be on the basis of using only normal-play intraframe pictures or intraframe pictures in combination with uni-directional coded pictures. In SuperVCD, only intraframe-compressed pictures are used to realize traditional trick play. Traditional trick play on a SuperVCD stored program is established by means of locator information indicating the start position of an intra-coded picture. Furthermore, due to the random access of the stored information, a new navigation feature has been added to the system enabling to jump forward or backward in a stored program. To support this new form of navigation, another list with locator information is available.

Traditional trick play on a SuperVCD stored program is established by means of the "SEARCH.DAT" information or via the scan information data, which is multiplexed as user data together with the video elementary stream. The presence of the "SEARCH.DAT" file depends on whether the system profile tag is set to 0x00. Due to the Variable Bit Rate (VBR) coding and buffering applied by MPEG-2, there is no longer a direct relation between playing time and the position on disc. To solve this relation problem for trick play, the "SEARCH.DAT" file contains a list with access point sector addresses, indicating the nearest intraframe compressed picture in the MPEG track on a regular incremental time grid of 0.5 seconds. The total number of entries is minimally 1 and maximally 32,767 using a storage format of (*mm:ss:ff*), where *mm={0 ..99 }* indicates the minutes part of the sector offset value, *ss={0..59 }* the seconds part of the sector offset value and *ff={0..74}* represents the fraction part of the sector offset value. This information is useful for features such as time search. It should be noted that, as indicated in Fig. 13 (b), the MPEG-2 coded information and associate data are embedded into the regular CD-Audio sectors, which form the fundamental storage framework on a CD disc. Each sector corresponds to a certain time instance indicates in minutes, seconds and fragments.

If the "SEARCH.DAT" is not present, a second information mechanism, called scan information data, is provided by the SuperVCD standard to enable trick play. Scan information data is information indicating the location of the next and previous intraframe coded picture. This information, which is transmitted using the MPEG-2 user data structure, precedes each intra-coded picture and consists of two field pairs, one pair for a video stream and one pair for a still image stream. The fields are coded as a sector offset value referencing to the start sector of the MPEG track, as indicated by the table of contents. The scan information fields share the same storage format, which is *mm:ss:ff*. The scan information data refers to an access point sector.

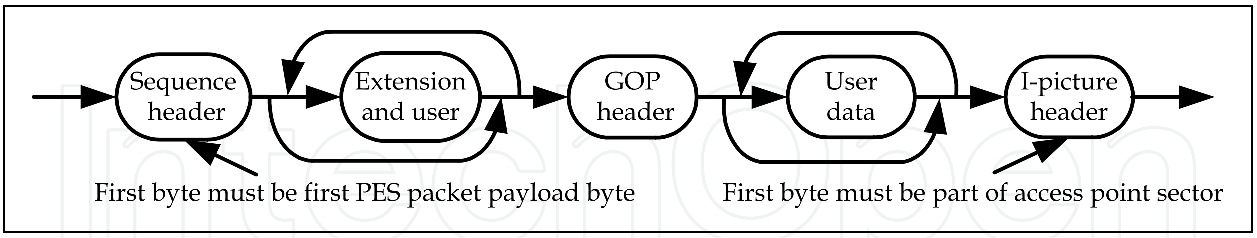


Fig. 14. SuperVCD access-point sector structure.

An access point sector is an MPEG video sector, in which the first byte of a Packetized Elementary Stream (PES) packet belongs to a sequence header, see Fig. 14. MPEG-2 allows the usage of different quantization tables. MPEG-2 video encoders that use this feature will transmit the different quantization matrices in the sequence header. It is for this reason that in order to have proper trick play, these quantization matrices have to arrive at the MPEG-2 decoder, which is guaranteed by the access-point structure as depicted in Fig. 14.

Time-based jumping is another navigation method supported by SuperVCD and involves jumping forward or backward in time to the next chapter. This feature is enabled by means of the "ENTRIES.SVD" file, which stores the entry point addresses of the MPEG-2 tracks on disc. Due to the limited size of 2,048 Bytes, which corresponds to one sector, the number of entries of the "ENTRIES.SVD" file is limited to a maximum of 500.

5.2 Digital Video Disc (DVD)

The DVD is the optical media successor of the SuperVCD featuring higher recording density of up to 4.7 GByte for the basic format, using a 650 nm laser and deploying a pull-based architecture. The DVD-Video format (Taylor, 2001) specification is a video playback application of the DVD-ROM standard, applying MPEG-2 video compression and allowing MPEG-1 and run-length coding for still images. Various audio formats are supported such as Pulse Code Modulation (PCM), MPEG-1 or MPEG-2 compressed audio, or Dolby Digital multi-channel, which are packetized into a Packetized Elementary Stream (PES) and multiplexed into an MPEG-2 Program Stream. Also included in the MPEG-2 Program Stream multiplex are the real-time private stream, called Presentation Control Information (PCI), and Data Search Information (DSI), resulting in a maximum multiplexed bit rate of 10.08 Mbit/s. A player based on the DVD format, is typically connected to an standard TV display, either via baseband or modulated composite coax, or via analog component YCrCb/RGB video. Moreover, the signal can be of either interlaced or progressive nature. Figure 15 indicates the block diagram of a DVD player. A DVD player is based on a presentation engine and a navigation manager, as shown in Fig. 16. The presentation engine uses the information in the presentation data stream, see Fig. 15, to control what is presented to the viewer. For example, the navigation manager creates menus, provides user interfaces and controls branching, all on the bases of retrieved information from the DVD disc. This allows content providers to disable search-mode functions on particular DVD content, e.g. a trailer indicating that the audiovisual content is only to be used in the domestic environment. Per video title set, a particular movie can be made available as a single MPEG-2 Program Stream, or can be split up in maximally 9 parts, as indicated in Fig. 17 (a). This figure shows an example of a DVD volume layout containing a DVD-Video zone (DVD-Audio zone have been left out for simplicity). A Video Object Block (VOB) file, see Fig. 17 (b), is constructed of one or more cells containing a group of pictures or audio blocks; the cell forms the smallest addressable entity for random access. A cell, also known as scene, can be as short as 1 second, or cover the whole sequence and is uniquely identified with its cell ID and corresponding video object ID. A cell is further sub-divided into video object units, an entity containing zero or more Group-Of-Pictures (GOP). For the situation that a video object unit contains a single GOP, this GOP should start with a sequence header followed by a GOP header, which by default is followed by an intra-coded picture. Video object units are further split into packs and packets, which are compliant with the MPEG-2 Program Stream standard (ISO/IEC13818-1, 2000). Navigation data is a collection of information that determines how the physical data is accessed and controls interactive playback. The information is grouped into four categories: control, search, user interface and navigation commands, resulting in navigation information being split over five levels. The levels "program chain" and "data search" contain information required to perform navigation as discussed previously. For example, the "program chain" information enables the remote control navigation buttons "previous" and "next". The "data search" information is situated in the "navigation pack", see Fig. 17 (b) and thereby an integral part of the Program Stream.

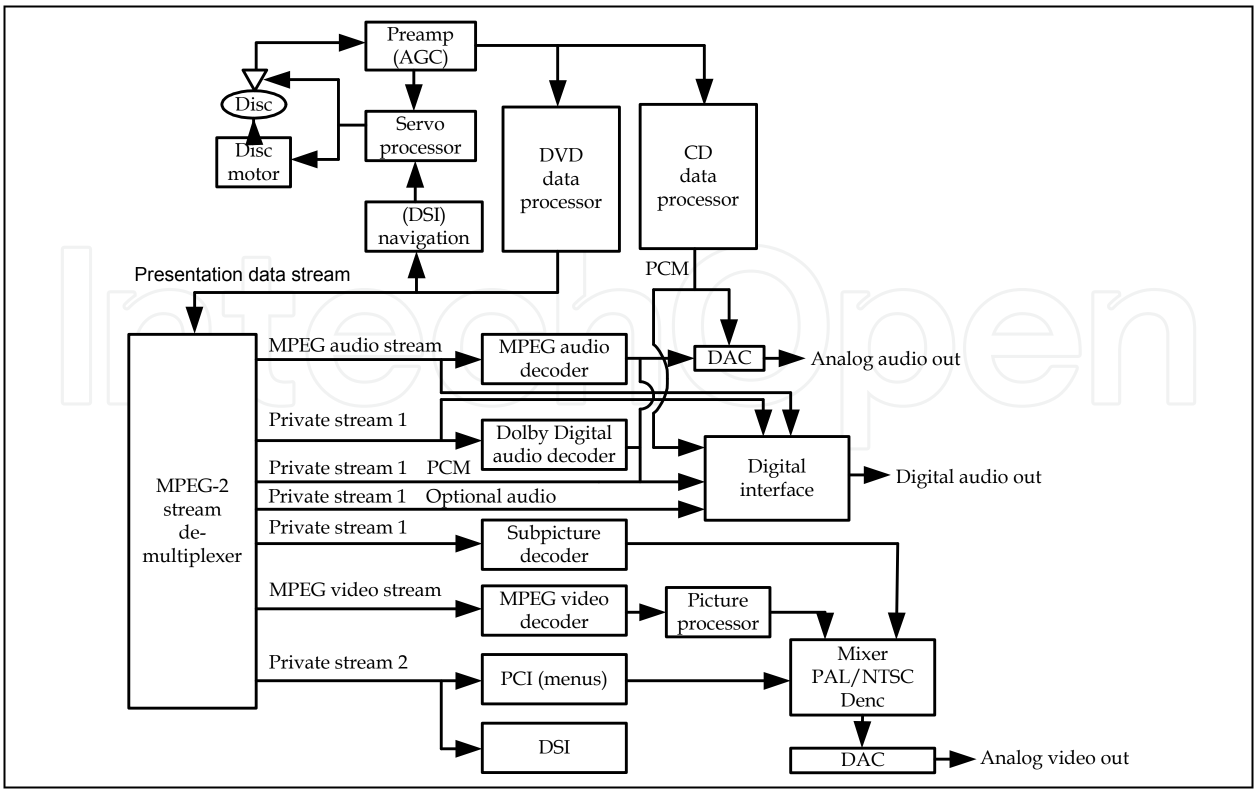


Fig. 15. DVD-Video player block diagram.

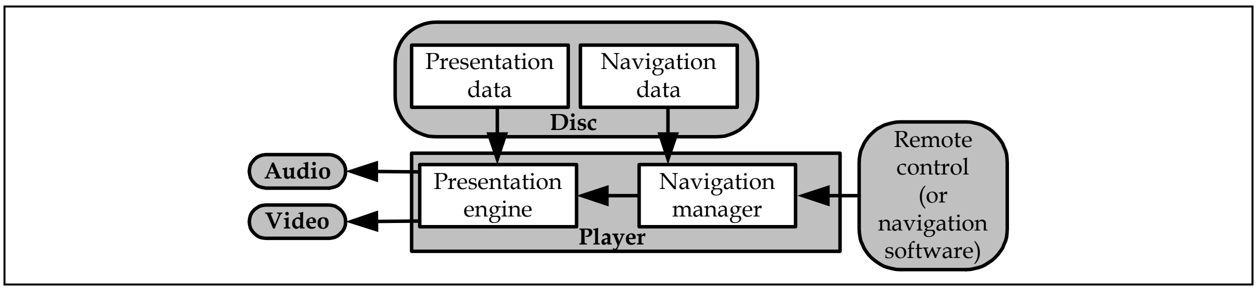


Fig. 16. DVD navigation and presentation model.

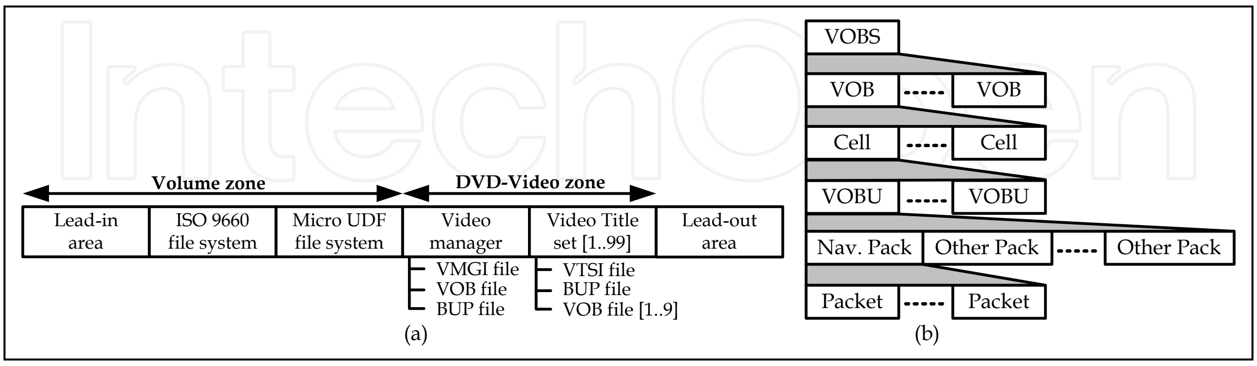


Fig. 17. DVD-Video data construction. (a) DVD volume layout containing a DVD-Video zone, (b) DVD abstraction layers on top of an MPEG-2 Program Stream.

The "navigation pack" holds for each video object unit a pointer for forward / reverse scanning, either referring to I-type or P-type pictures, which is an improvement with respect to SuperVCD that is only supporting I-type picture search.

5.3 Blu-ray Disc (BD)

Blu-ray Disc (BD, 2006) is the optical media successor of DVD, featuring higher recording density of up to 25 GByte for the basic format enabled by a 405 nm laser. BD is an optical storage system suitable for recording as well as playback-only of audiovisual information applying a pull-based architecture, see Fig. 18 (b). Video can be compressed using either: MPEG-2, MPEG-4 AVC, also known as H.264, or SMPTE VC-1 video compression. Multi channel audio is supported up to a 7.1 audio system and encoding is based on either, using Dolby Digital AC-3, DTS, or uncompressed using PCM. Besides the previous audio coding standards, three other standards are optionally supported: Dolby Digital Plus and two lossless coding methods, called Dolby TrueHD and DTS HD. The BD-ROM standard defines two platforms: a High Definition Movie (HDMV) and a Java™ platform, also known as BD-J, and these platforms categorize the BD-ROM features. In the BD-ROM system, exactly one mode, either HDMV mode or BD-J mode, is active at any given point of time during playback. HDMV supports features such as seamless multi-angle and multi-story, Language Credits, Directors Cuts, etc., see Fig. 18 (b).

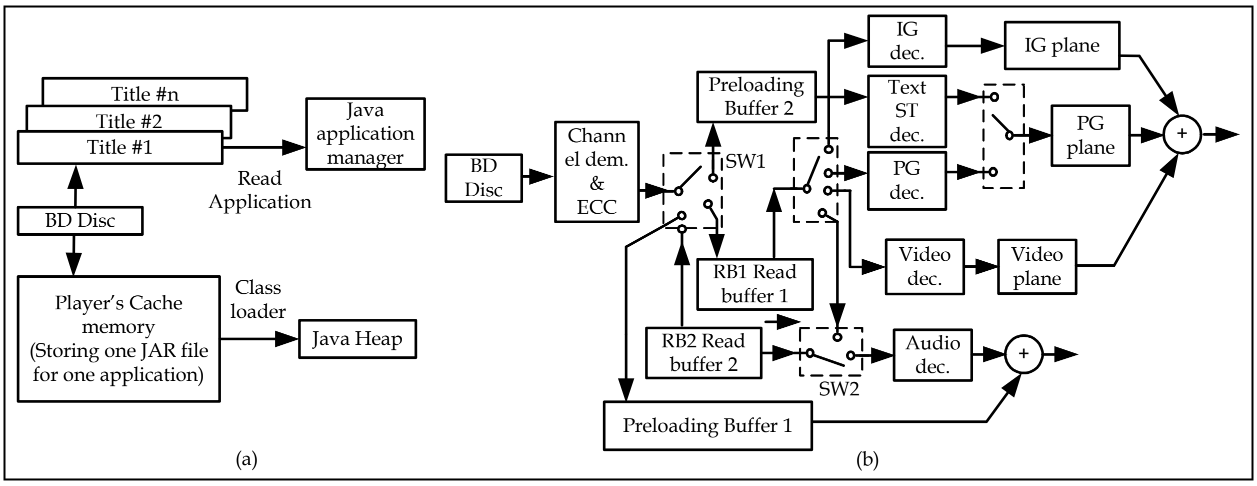


Fig. 18. BD-ROM system standard. (a) Java application tables in BD-ROM, (b) HDMV decoder model.

BD-J enables a fully programmable application environment with network connectivity, thereby enabling the content provider to create highly interactive, updateable BD-ROM titles, see Fig. 18 (a), and is based on the Java-2 Micro-Edition (J2ME) Personal Basis Profile (PBP), a Java profile that was developed for consumer electronics devices. It is assumed that the BD-ROM disc will be the primary source for media files, but it alternatives are the studio's web server and local storage. The unit of playback in BD-J is the PlayList, just as in HDMV. All features of HDMV, except Interactive Graphics (IG), which is replaced by BD-J graphics, can be used by a BD-J application. Supported features include Video, Audio, Presentation Graphics (PG), Text Subtitle component selection, media-time and playback-rate (trick-mode) control. The BD-J video device is a combination of the HDMV video and presentation graphics planes. Both video and presentation graphics will playback in the video device. BD supports traditional trick play via additional locator information, situated in the Clip Information file. Moreover, BD optionally supports Title Scene Search (TSS), a metadata-based advanced navigation method.

An AV stream file, together with its associated database attributes, is considered to be one object. The AV stream file is called a "Clip AV stream file", and the associated database attribute file is called a “Clip Information file". An object consisting of a “Clip AV stream file” and its corresponding “Clip information file” is called a Clip. A “Clip AV stream file” stores data, which is an MPEG-2 Transport Stream defined in a structure called the BDAV MPEG-2 Transport Stream. In general, a file is a sequence of data bytes. But the contents of the “Clip AV stream file” are developed on a time axis, therefore the access points into a “Clip AV stream file” are specified with time stamps. The “Clip Information file” stores the time stamps of the access point into the corresponding AV stream file. The Presentation Engine reads the clip information, to determine the position where it should begin to read the data from the AV stream file. There is a one-to-one relationship between a “Clip AV stream file” and a “Clip Information file”, i.e., for every “Clip AV stream file”, there is one corresponding “Clip Information file”.

The Playback Control Engine in the BD-ROM Player Model uses PlayList structures, see Fig. 19 (b). A “Movie PlayList” is a collection of playing intervals in the Clips, see Fig. 20 (a). A single playing interval is called a PlayItem and consists of an IN-point and an OUT-point, each of which refers to positions on a time axis of the Clip, see Fig. 20 (a). Hence, a PlayList is a collection of PlayItems. Here, the IN-point denotes a start point of a playing interval, and the OUT-point indicates an end point of the playing interval.

In Section 3, it was discussed that traditional trick play on compressed video data depends on the GOP structure, see also Fig. 1 (a). To enable random access, BD stores an EP_map (Entry Point), which is part of the Clip Information file, for each entry point of an AV stream. Unlike the previous optical storage standards, not only information regarding the position of intraframe-compressed pictures is stored, but optionally, also information regarding the length of each GOP and the coding type of the pictures constructing a particular GOP. This information is stored in the so-called *GOP_structure_map* located in the Supplemental Enhancement Information (SEI), which is stored in the user data container of the firstly decoded Access Unit (AU) of a GOP. The meaning of trick play is defined by the manufacturer of the BD-ROM Player, e.g. the manufacturer may define that $Ps=1.2$ forward play is not trick play, and instead define that $Ps>2$ forward play is considered trick play.

EP_map is a part of the “Clip Information file”, and this information is mainly used for finding addressing information of data positions, where the BD-ROM player should start to read the data in the AV stream file. The corresponding access point is given to the Clip, see Fig. 19 (b), in the form of time indications, referring ultimately to a particular data byte in the AV stream. EP_map has a list of Entry Point data (EP-data) that is extracted from the AV stream, where decoding can start. Each EP-data is a pair of a PTS (Presentation Time Stamp) for an access unit and a data address of the access unit in the AV stream file. EP_map is used for two main purposes. First, it enables to find the data address of the access unit in the AV stream file that is pointed to by the PTS in PlayList. Second, it is used for facilitating fast forward and reverse trick play. The EP_map gives the relationships between Presentation Time Stamp (PTS) values and addresses in the AV stream file. The PTS entry point is called PTS_EP_start, where the actual AV stream entry point address is called SPN_EP_start, see Fig. 20 (b). In order to reduce the size of the table and to improve the searching performance, the EP_map for one stream is split in two sub-tables: EP_coarse and EP_fine. The “Clip Information file”, which has a maximum file size of 2 MBytes is stored in “Clipinf”, a sub-directory of “BDMV”. Note that it is allowed that the CLIPINF directory contains no Clip Information file.

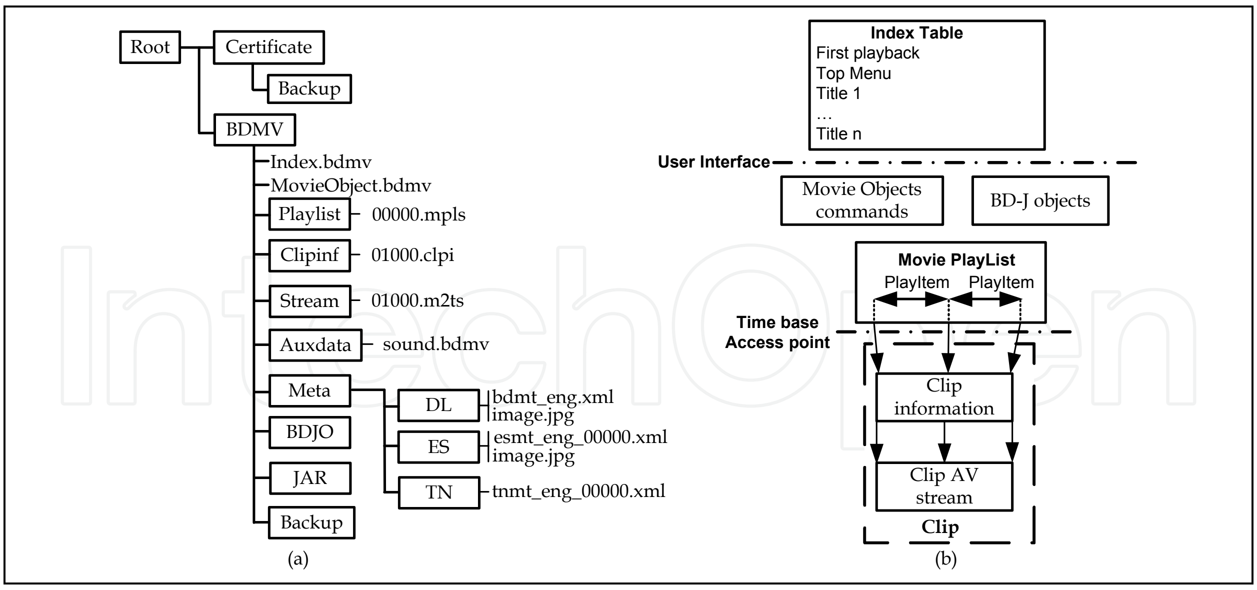


Fig. 19. BD-ROM data. (a) Example BD directory, (b) Simplified structure of BD-ROM.

If Title Scene Search is present, the corresponding information is separately stored from the content in "ES", a sub-directory of the "META" directory, which also stores other metadata, see Fig. 19 (a). The value part of the metadata filename links the file to the corresponding playlist. Other sub-directories of the "META" directory are "DL" and "TN". The directory "DL" (Disc Library) contains the disc-library metadata, which is mainly composed of two types of information: disc information and title information. Disc information contains the metadata of a disc itself, and title information includes the metadata of a title. The directory "TN" (Track/chapter Name) stores files containing metadata information regarding the names of tracks and chapters, which are sequentially stored. The track/chapter name metadata file exists per PlayList of a movie title. The corresponding metadata file is linked to the PlayList file via the value part of the file name.

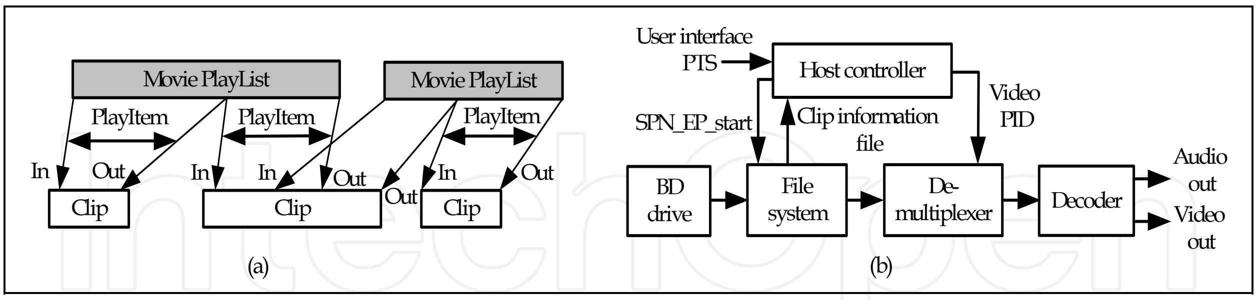


Fig. 20. Playback BD movie. (a) Movie playback using information from the PlayList, (b) Player model for I-picture search using EP_map.

6. Trick play for hard or solid-state disk based storage systems

With the advances in storage capacity and an associated firm price erosion, Hard Disk Drive (HDD) based storage systems have replaced the traditional tape-based storage media in the consumer arena. Recently, the Solid-State Drive (SSD) has made its appearance in the consumer arena, working its way into systems dominated by the HDD. Due to the enormous storage capacity of this type of media, a massive amount of audiovisual material can be stored. Efficient navigation through this data requires fast random access, enabling high-speed search and new forms of audiovisual navigation. HDD and SSD storage solutions have an advantage regarding the access time, see Fig. 21, when used for audiovisual storage applications. For the situation that the access time of the storage medium is less than the reciprocal of the frame rate, the storage system responsiveness enables traditional trick play at full refresh rate and paves the way for advanced navigation methods.

The disk-based storage system architecture can be either push- or pull-based and is mainly determined by the context in which the storage system is deployed. A good solution is to offer both architectures, and make a choice depending on where the audiovisual decoding is performed, so that the best of both architectures is deployed. When audiovisual decoding is performed locally, the pull-based solutions give the best trick-play quality. When operated in a networked manner, trick-play quality cannot be guaranteed, as this depends on the normal-play encoding settings. It should be kept in mind, that the involved network communication is minimal for push-based architectures, in which only commands are required when changing the mode of operation, i.e. play, stop, fast search forward, etc. Moreover, when supporting both architectures, also advanced navigation features become possible, where the availability of the navigation mode depends on the location of decoding.

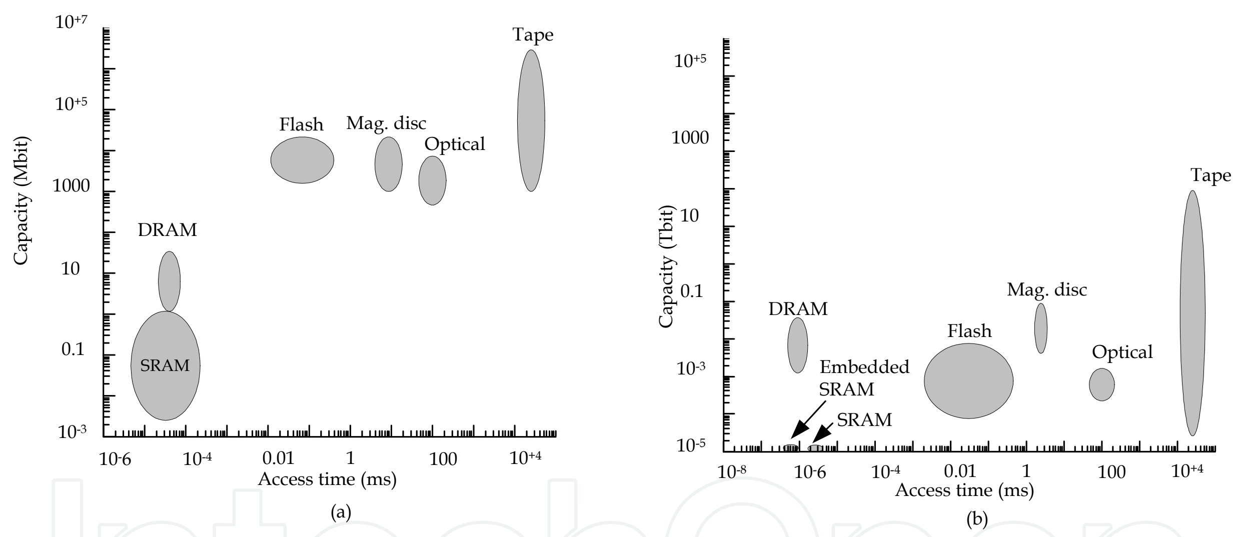


Fig. 21. Storage solutions and their access time. (a) Storage technology mid 1990s, (b) Storage technology 2009.

As a result, some advanced navigation modes may only be partly available when decoding is performed on remotely located decoders. Figure 22 depicts a functional block diagram of an HDD or SSD-based PVR, supporting both pull- and push-based playback. The pull mode requires the decoder to request for information, indicated via the dashed line. Let us elaborate on the two advanced navigation concepts introduced in Section 2. The pull-based architecture is suitable for traditional trick play and supports both advanced navigation concepts, whereas the push-based mode supports traditional trick play, hierarchical navigation and partly audiovisual trick play as discussed in Section 3.

During record, an MPEG-2 Transport Stream enters the system at the left-hand side, see Fig. 22. The CPI block adds time stamps to the incoming TS packets, which are stored together with the corresponding TS packet on disk. Furthermore, the MPEG-2 TS is de-multiplexed and the video signal is parsed.

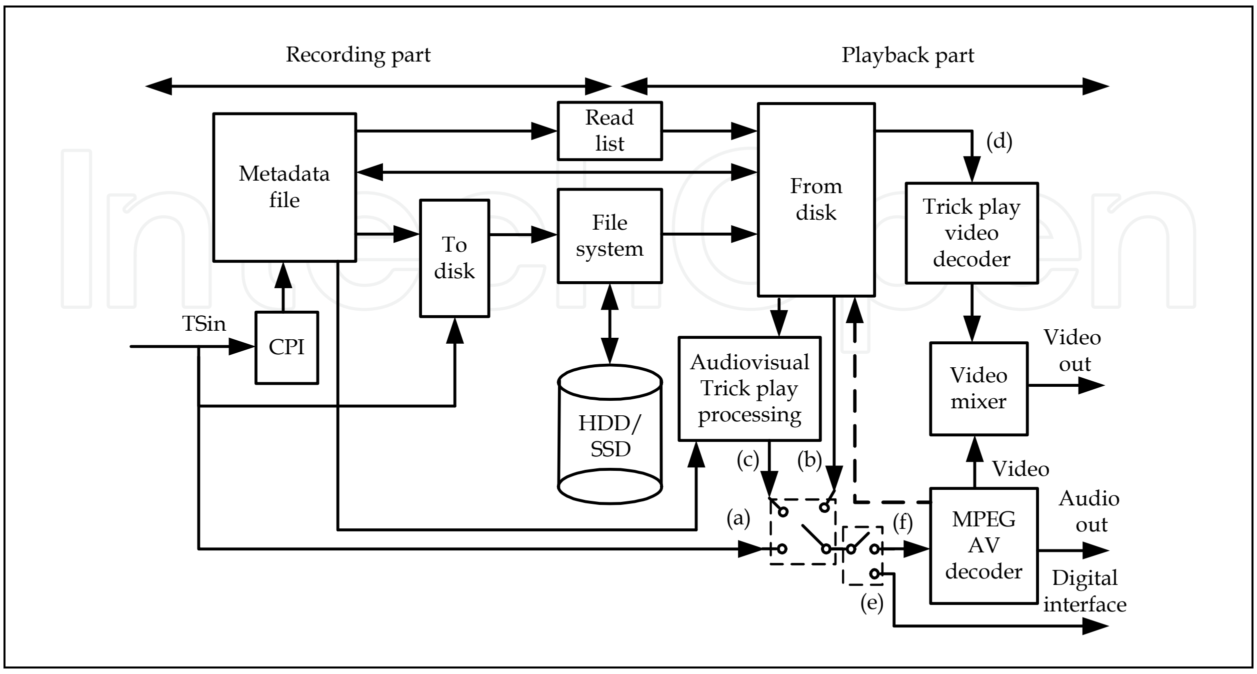


Fig. 22. Functional block diagram of HDD or SSD-based PVR. Operation modes: (a) View real-time, (b) View time-shift, (c) Traditional trick play or trick play based on normal-play AV fragments, (d) Trick play PiP to be used in combination with normal-play AV fragments, (e) AV playback according to push mode, (f) Pull-mode playback for local decoding.

During the parsing process, various characteristics of the MPEG-2 Video Elementary Stream are stored, such as relative positions of I- and P-type pictures, GOP length and transmission time of the intraframe-encoded pictures. This data is stored in a metadata file, which is stored as an extra descriptor file on the disk. Besides the information for traditional trick play, sub-pictures are derived from the normal-play video sequence and stored in the metadata file. The sub-pictures are generated in such a way that they can be used to construct mosaic screens via signal processing in the MPEG-2 domain, resulting in MPEG-2 compliant mosaic screen pictures (Eerenberg & de With, 2003). The information in the metadata file offers navigation features such as traditional trick play, but also the more advanced navigation methods for local (pull-mode) as well as remote users (push-mode).

For traditional trick play, the read-list block (using data from the metadata file) controls the disk retrieval process, resulting in fetching MPEG-2 TS fragments. These fragments are either processed by the audiovisual trick-play processing, which creates an MPEG-2 compliant signal, see Fig. 22 (c) and (e), or send the fragments directly to the local MPEG-2 decoder, see Fig. 22 (c) and (f).

For hierarchical navigation, the audiovisual trick-play block generates the mosaic screens for the pull- and push-based playback. For video navigation reinforced with audible sound, the audiovisual fragments are retrieved from disk and made available to the local decoder. A fast-search trick-play sequence is generated by means of a second video decoder, which could also be the control processor. The CIF-based trick-play sequence is merged with the normal-play video fragments by means of a video mixer.

7. Conclusions

Trick play is a feature that needs to be embedded within the design of a storage system. The system implementation should have low costs and be of suitable quality for quickly searching through the data. However, implementation of trick play differs for the various storage systems. Based on the discussed systems in this chapter, the following techniques have become apparent.

Trick play can be realized by re-use of normal-play compressed video data, or via additional dedicated trick-play information streams. Digital helical-scan recorders apply both mechanisms. For example, DV shuffles normal-play data to facilitate trick-play for a broad range of speed-up factors and D-VHS deploys dedicated virtual channel for each supported speed-up factor, which is filled with special trick-play information. The disadvantage of this solution is the reduction of the normal-play recording bandwidth, e.g. in D-VHS the reduction is about 5 Mbit/s.

Common optical-disc storage systems all deploy the re-usage of normal-play information and selectively address individual decodable pictures in the normal-play video stream. This is achieved via so-called locator information stored within the recorded stream. The optical pick-up unit is positioned, using this locator information, such that the retrieved information contains individually decodable pictures. In this way, the proper placement of the optical-pickup unit becomes a dominant factor in the random access time. As a consequence, the final trick-play quality depends on the normal-play GOP structure. The BD-based optical storage solution also optionally deploy normal-play GOP information, enabling the usage of P-type and B-type pictures for trick play. Moreover, metadata-based retrieval is facilitated by an optional descriptor file. With the current advances in optical drive technology, random access time can be avoided due the readout of the optical disc at a significant higher rate, causing the video source decoder to be the dominant factor for trick-play quality.

With the introduction of HDD-based PVRs, trick play can be deployed in full temporal quality due to the short access time, when operated in pull-mode. Both HDD and SSD-based systems deploy locator information, enabling basic and advanced forms of navigation. When operated in a push-mode, the trick-play temporal quality most probably decreases somewhat, but the decoder always receives a compliant stream for decoding, thereby making this approach suitable for networked-based storage systems. Drawback of a push-based architecture is that it limits the advanced navigation modes, as the final signal must comply with the deployed standards, when system cost should be kept low.

Future work in the field of trick play should focus on low-cost algorithms which are capable of finding useful metadata to create the metadata search files. This will facilitate searching techniques, which are based on semantic understanding of the data content, so that high-level searching becomes possible.

8. References

BD (2006). System Description Blu-ray Disc Read-Only Format, Part 3: Audio Visual Basic Specifications (3-1 Core Specifications) Version 2.01 DRAFT2

Boyce, J. & Lane, F. (1993). Fast scan technology for digital video tape recorders. *IEEE Trans. Consumer Electron.*, Vol. 39, No. 3, August 1993, pp. 186-191, ISSN 0098-3063

CEI/IEC61834 (1993). Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems

Eerenberg, O. & de With, P.H.N. (2003). System requirements and considerations for visual table of contents in PVR, *Proceedings of ICCE 2003*, pp. 24-25, ISBN 0-7803-7721-4, USA, June 2003, Los Angeles, CA

Eerenberg, O. & Aarts, R.M. & de With, P.H.N. (2008). System Design of Advanced Video Navigation Reinforced with Audible Sound in Personal Video Recording, *Proceedings of ICCE 2008*, pp. 1-2, ISBN 978-1-4244-1458-1, USA, Jan. 2008, Las Vegas, NV

ETSI_TS_101154 (2009). Technical Specification Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG 2 Transport Stream, Version 1.9.1, March 2009

Fujita, M.; Higurashi, S. & Hirano, S. & Ohishi, T. & Zenno, Y. (1996). Newly developed D-VHS digital tape recording system for the multimedia era. *IEEE Trans. Consumer Electron.*, Vol. 42, No. 3, (August 1996), pp. 617-622, ISSN 0098-3063

van Gassel, J.P. & Kelly, D.P. & Eerenberg, O. & de With, P.H.N. (2002). MPEG-2 compliant trick play over a digital interface, *Proceedings of ICCE 2002*, pp. 170-171, ISBN 0-7803-7300-6, USA, June 2002, Los Angeles, CA

IEC62107 (2000). Super Video Compact Disc – Disc-interchange system-specification International Standard

IEC61883-4 (2004). Digital Interface For Consumer Audio/Video Equipment Part 4: MPEG-TS data transmission

ISO/IEC11172-2 (1993). Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 2: Video

ISO/IEC 14496-3 (1999). PICOLA Speed change algorithm, Annex 5.D.

ISO/IEC13818-1 (2000). International Standard Information technology — Generic coding of moving pictures and associated audio information: Systems, Dec. 2000

ISO/IEC13818-2 (2000). International Standard Information technology — Generic coding of moving pictures and associated audio information: Video, Dec. 2000

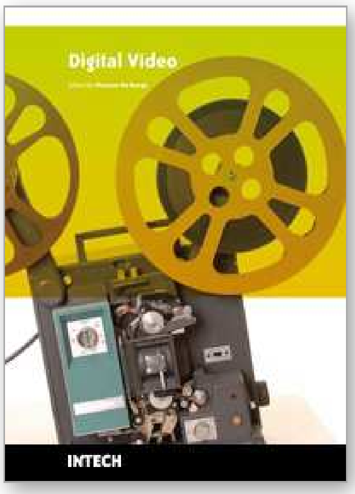
ISO60774-5 (2004). Helical-scan video tape cassette system using 12,65 mm (0,5 in) magnetic tape on type VHS –Part 5: D-VHS

Matsushita, Philips, Sony & Thomson. (1993). Outline of Basic Specification for Consumer-use Digital VCR

Taylor, J. (2001). *DVD Demystified,* McGraw-Hill, ISBN 0-07-135026-8

Ting, H. & Hang, H. (1995). Trick play schemes for advanced television recording on digital VCR. *IEEE Trans. Consumer Electron.*, Vol. 41 No. 4, Nov. 1995, pp. 1159-1168, ISSN 0098-3063

US6621979 (1998). Trick play signal generation for a digital video recorder using retrieved intra-encoded pictures and generated inter-encoded pictures, Patent



Digital Video

Edited by Floriano De Rango

ISBN 978-953-7619-70-1
Hard cover, 500 pages
Publisher InTech

This book tries to address different aspects and issues related to video and multimedia distribution over the heterogeneous environment considering broadband satellite networks and general wireless systems where wireless communications and conditions can pose serious problems to the efficient and reliable delivery of content. Specific chapters of the book relate to different research topics covering the architectural aspects of the most famous DVB standard (DVB-T, DVB-S/S2, DVB-H etc.), the protocol aspects and the transmission techniques making use of MIMO, hierarchical modulation and lossy compression. In addition, research issues related to the application layer and to the content semantic, organization and research on the web have also been addressed in order to give a complete view of the problems. The network technologies used in the book are mainly broadband wireless and satellite networks. The book can be read by intermediate students, researchers, engineers or people with some knowledge or specialization in network topics.

O. Eerenberg and P.H.N. de With (2010). Trick play on Audiovisual Information for Tape, Disk and Solid-State based Digital Recording Systems, Digital Video, Floriano De Rango (Ed.), ISBN: 978-953-7619-70-1, InTech, Available from: http://www.intechopen.com/books/digital-video/trick-play-on-audiovisual-information-for-tape-disk-and-solid-state-based-digital-recording-systems

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821