Chapter 3

THE CD SYSTEM AS STANDARDIZED BY PHILIPS AND SONY

3.1 Introduction to publications of the Compact Disc digital audio system

J.B.H. Peek and J.P. Sinjou

Preface

It should be stressed that this introduction does not intend to mention and recognize all people within Philips and Sony who contributed to establish the CD digital audio system standard in 1980. This standard is based on the collaborative work of many persons, both from Philips and from Sony, and it would be impossible to properly acknowledge all these individuals in the space of only a few pages. More information on the persons involved can be found in the doctoral thesis (in German) by Jürgen Lang ("Das Compact Disc Digital Audio System", 1996, RWTH, Aachen, ISBN 3-00-001052-1). This introduction only aims at describing some important decisions that were made between the successful demonstration of the CD prototype on March 8, 1979, and the establishment of the Philips-Sony CD standard in June 1980.

The Philips-Sony partnership

Already at an early stage in the development of the CD prototype, the Philips Board of Management emphasized that the directors of the Philips Audio product division should aim at realizing a world standard for the CD. With this in mind the directors of Audio decided that, to achieve this goal, a first step would be to find a strong industrial partner that would be interested to cooperate with Philips in attaining a common CD system standard. Therefore, Philips went public with its digital audio disc innovation in a press conference on March 8, 1979. And as a follow-up, the directors of Audio decided to approach several Japanese companies and ask them if they would be interested to receive a delegation of Audio so that the Philips CD prototype could be shown and demonstrated. A positive response of these companies was received and from March 14 till March 23, 1979, the following companies and organizations were visited in succession: JVC, Sony, Pioneer, Hitachi, MEI (Matsushita) and the DAD (Digital Audio Disc Committee). The DAD had been installed by the Japanese Ministry of Industrial Trade and Industry with the task to evaluate various digital audio disc systems and to recommend a world standard. On the last day of the visit, J. van Tilburg, the general director of Audio, received a phone call from A. Morita, the president of Sony. Morita said that, after consulting the management of Sony, he had decided to cooperate with Philips. The vice-president of Sony, N. Ohga, would come to Eindhoven to discuss the contract.

The Philips-Sony collaboration

With Sony, Philips had an ideal partner. Sony not only had an excellent position in products related to digital recording of audio on magnetic tape, but they also had developed a prototype optical digital audio player and disc. The diameter of Sony's disc, however, was 30 cm, much larger than the 11.5 cm diameter of the Philips CD disc. To determine a common standard for the CD system, Philips and Sony agreed to a sequence of meetings to be held alternately in Eindhoven and Tokyo. During these meetings, the technical experts from Philips and Sony had to settle issues like the playing time of a disc, its diameter, the audio sampling frequency, the signal quantization (bits/sample), and the signal format to be used. Determining the signal format implied that Philips and Sony had to agree on the purpose and the interpretation of the successive bits in each block of data on the disc, including the modulation code and the error correcting code to be used.

The first meeting, of in total six meetings, was held in Eindhoven on August 27 and 28, 1979. The last meeting was in Tokyo on 17 and 18 June, 1980. The way in which the final error correcting code gradually emerged, illustrates the interchange of ideas between Philips and Sony engineers.

At their first meeting, Philips and Sony each proposed a different error correcting code. Philips proposed the rate 2/3 convolutional code that was used in the prototype CD system. This code was developed by L. Vries and is described in his paper [L.B. Vries, "The Error Control System of Philips Compact Disc", AES Preprint 1548, New York, November 1979], reprinted here in Sect. 2.4. Sony proposed a rate 2/3 code too, a b-adjacent code with

16-bit symbols (Sony opted for 16-bit quantization) and minimum distance 3, in combination with a simple parity-check code for error detection.

On February 5, 1980, after several in-depth and open discussions with Philips experts, Sony proposed a revised code that they called a cross b-adjacent code. This code, again with 16-bit symbols, was a combination of a couple of single-error or double-erasure correcting codes with a convolutional delay interleave in between [T. Doi, "Error Correction for Digital Audio Recordings", AES Premiere Conference, New York 1982, June 3-6, p.170]. L. Vries and L. Driessen, both engineers from Philips, subsequently analyzed this revised code. Their mathematical analysis of the performance of the revised Sony code appeared in an internal Philips Research Technical Note [L. Driessen, L. Vries, "The Performance of Sony's Cross-B-Adjacent Code on a Memoryless Channel", Technical Note Nr. 54, 1980]. This Technical Note was submitted as a discussion paper for the next Philips-Sony meeting. As a consequence of their analysis it became clear to Vries and Driessen that Sony's revised code had a better performance than the convolutional code as originally proposed by Philips.

While analyzing Sony's code, Driessen and Vries saw possibilities to enhance the correction and detection capabilities, both for errors and erasures, without changing the rate of the code. Instead of 16-bit symbols, Driessen (educated in algebraic coding theory) suggested to use 8-bit symbols corresponding to the Galois Field GF(2^8), which implied that a codeword (still having the same number of information and parity bits) doubled in length (counted in symbols) and that the minimum distance increased from 3 to 5. This modification offered a better protection against random errors and short burst errors, while keeping the same protection against long burst errors. A further attractive feature of the proposed improved code was that it allowed several decoding strategies, thereby increasing the freedom for each manufacturer to choose a distinguishing decoding strategy. Although implementing the improved code on a chip at the time was much more complicated than implementing Sony's revised code, the proven advantages were so convincing that Sony accepted the suggested improvements without any changes. Later, after having reached agreement on the lengths of the two interleaved codes and the interleave scheme itself, the improved rate 3/4 code was called CIRC (Cross Interleaved Reed Solomon Code) and it became the Philips-Sony error correcting coding standard for CD in June 1980.

As mentioned before, Philips and Sony also had to agree on a modulation code, which is needed to adapt the incoming bit stream to the characteristic of the CD storage channel. At their first meeting on August 27, 1979, Philips and

Sony each proposed a different modulation code. Philips presented the M3 code also used in their CD prototype, whereas Sony proposed a code called 3PM. Both codes were DC-free and run-length-limited. A DC-free code produces an encoded bit stream with very little spectral content at low frequencies, as required to prevent disturbance of the servo systems. A run-length-limited code produces runs of ones and zeros that are constrained to have a prescribed minimum and maximum length. The choice of the minimum run-length permits the power spectrum of the encoded data sequence to be adapted to the low-pass transfer characteristic of the CD-channel, thereby facilitating bit detection. A proper choice also helps to reduce the impact of various disc artifacts. The maximum run-length ensures that the encoded data stream contains enough timing information to permit reliable clock recovery.

First comparative experiments showed that the M3 code performed better if the disc was scratched or contaminated, while the 3PM code could achieve a higher data density in a clean, well-aligned environment. With this result in mind, the engineers from both sides proposed new modulation codes, supported by practical test results. Experimental data and test discs were exchanged. At some point in the discussions, the successive codes were called ASAP1, ASAP2, ASAP3, indicating the urgency of the project (ASAP=As Soon As Possible). The iterations were stopped as soon as further iterations did not bring significant further improvements, and the resulting code was later dubbed EFM (Eight to Fourteen Modulation). The EFM code has a minimum run-length of 3 bit intervals, a maximum run-length of 11 bit intervals, a code rate of 8/17, and state-independent low-frequency content. Both parties felt that the intensive period of cross testing and of mutual improvements had resulted in "the best of both worlds": a code which combined a high rate (or, equivalently, a high information density on the disc) with a high robustness against disc and player errors.

In June 1980, Philips and Sony decided to apply for two patents, one on CIRC and the other one on EFM. These patents were later granted by the U.S. patent office and are registered as:

1) K. Odaka, Y. Sako, I. Ikuo, T. Doi (all from Sony), L. Vries (Philips), "Error correctable data transmission method", U.S. Patent 4,413,340.
2) K. Immink, J. Nijboer (both from Philips), H. Ogawa, K. Odaka (both from Sony), "Method of coding binary data", U.S .Patent 4,501,000.

Together with the patent of P. Kramer ("Reflective optical record Carrier", U.S. Patent 5068846), mentioned also in Sect. 2.1, these patents are essential to the CD system.

Fig. 1. On the last day, August 18, 1980, at the end of the six meetings, a photograph was taken in Tokyo. It shows a happy smiling team. From left to right:
2nd row: Heemskerk, Harada, Miyaoka, Vries, Nijboer, Tsurushima, Doi, Ogawa, Naruse, Odaka.
Front row: Sinjou, Bögels, Nakajima, Mizushima.

In the so-called 'Red Book' the Philips-Sony standard is described in detail. This book mentions important parameters, such as the playing time of approximately 60 minutes, the 44.1 kHz sampling frequency, the 16-bit signal quantization, and the 12 cm diameter of a disc. The standard of the 'Audio recording-Compact disc digital audio system' is available at the International Electrotechnical Commission (IEC) in Geneva as document 60908 (second edition 1999).

The diameter of 11.5 cm of the CD prototype disc changed to 12 cm in the standard because of a personal wish of N. Ohga. The reason was that with a diameter of 12 cm a particular performance of Beethoven's ninth symphony with a length of 74 minutes could be recorded on a disc.

After the CD standard was established in 1980, many papers were published, not only by Philips and Sony authors separately but also by Philips and Sony authors jointly. The amazing success of the CD after 1982, when the CD player and disc came on the market, also resulted in many books and papers that explained various aspects of the CD system.

Collected papers in this chapter

On the pages following this introduction a number of papers on the CD, by or with Philips authors, are reprinted. In 1982, a special issue of the 'Philips Technical Review' (Vol. 40, No. 6, 1982) was completely dedicated to the CD system. The papers contained in this issue are all reprinted in this chapter. The special issue started with an introduction to the integral CD system with the title "The Compact Disc Digital Audio System". The next three papers explain various subsystems in a CD player. The first of these, "Compact Disc: system aspects and modulation", describes EFM. The second paper, "Error correction and concealment in the Compact Disc system", explains the error correction subsystem CIRC and also the method to conceal those errors that CIRC could not correct but only detect. In the third paper, "Digital-to-analog conversion in playing a Compact Disc", it is shown how the performance of a 14-bit D/A converter, in combination with digital signal processing, can be made equivalent to a 16-bit D/A converter.

A large part of the success of the CD system can be attributed to the attractive small shining disc of which by now about 220 billion have been sold. The disc-mastering process, a key step before mass production of CD discs, is outlined in the paper "Compact Disc (CD) Mastering - An Industrial Process" that is reprinted in Sect. 3.6.

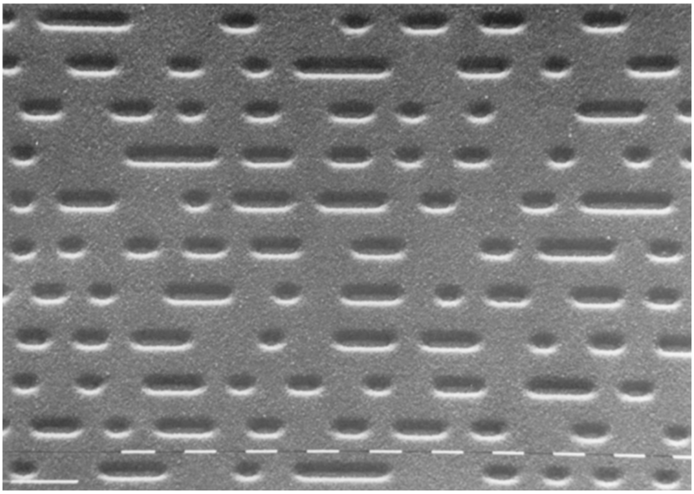
From a system point of view, the successive digital signal processing operations in a CD player are designed on the basis of communications concepts. These concepts encompass demodulation, error correction and detection, interpolation to conceal uncorrected but detected errors, and bandwidth expansion to ease D/A conversion. These ideas are explained in a paper "Communications Aspects of the Compact Disc Digital Audio System" that is reprinted in Sect. 3.7.

3.2 The Compact Disc Digital Audio system

M.G. Carasso, J.B.H. Peek, J.P. Sinjou

Drs M. G. Carasso and Dr Ir J. B. H. Peek are with Philips Research Laboratories, Eindhoven; J. P. Sinjou is with the Philips Audio Division, Eindhoven.

Abstract
Digital processing of the audio signal and optical scanning in the Compact Disc system yield significant advantages: insensitivity to surface damage of the disc, compactness of disc and player, excellent signal-to-noise ratio and channel separation (both 90 dB) and a flat response over a wide range of frequencies (up to 20 000 Hz). The Compact Disc, with a diameter of only 120 mm, gives a continuous playing time of an hour or more. The analog audio signal is converted into a digital signal suitable for transcription on the disc. After the digital signal has been read from the disc by an optical 'pick-up' the original audio signal is recreated in the player.



The information on the Compact Disc is recorded in digital form as a spiral track consisting of a succession of pits. The pitch of the track is 1.6 μm, the width 0.6 μm and the depth of the pits 0.12 μm. The length of a pit or the land between two pits has a minimum value of 0.9 and a maximum value of 3.3 μm. The scale at the bottom indicates intervals of 1 μm.

Reprinted with permission from Philips Tech. Rev. **40**, 151-155, 1982.

3.2.1 Introduction

During the many years of its development the gramophone has reached a certain maturity. The availability of long-play records of high quality has made it possible to achieve very much better sound reproduction in our homes than could be obtained with the machine that first reproduced the sound of the human voice in 1877. A serious drawback of these records is that they have to be very carefully handled if their quality is to be preserved. The mechanical tracking of the grooves in the record causes wear, and damage due to operating errors cannot always be avoided. Because of the analog recording and reproduction of the sound signal the signal-to-noise ratio may sometimes be poor (< 60 dB), and the separation between the stereo channels (< 30 dB) leaves something to be desired.

For these and other problems the Compact Disc system offers a solution. The digital processing of the signal has resulted in signal-to-noise ratios and a channel separation that are both better than 90 dB. Since the signal information on the disc is protected by a 1.2 mm transparent layer, dust and surface damage do not lie in the focal plane of the laser beam that scans the disc, and therefore have relatively little effect. Optical scanning as compared with mechanical tracking means that the disc is not susceptible to damage and wear. The digital signal processing makes it possible to correct the great majority of any errors that may nevertheless occur. This can be done because error-correction bits are added to the information present on the disc. If correction is not possible because there are too many defects, the errors can still be detected and 'masked' by means of a special procedure. When a Compact Disc is played there is virtually no chance of hearing the 'tick' so familiar from conventional records.

With its high information density and a playing time of an hour, the outside diameter of the disc is only 120 mm. Because the disc is so compact, the dimensions of the player can also be small. The way in which the digital information is derived from the analog music signal gives a frequency characteristic that is flat from 20 to 20 000 Hz. With this system the well-known wow and flutter of conventional players are a thing of the past.

Another special feature is that 'control and display' information is recorded, as 'C&D' bits. This includes first of all 'information for the listener', such as playing time, composer and title of the piece of music. The number of a piece of music on the disc is included as well. The C&D bits also contain information that indicates whether the audio signal has been recorded with pre-emphasis and should be reproduced with de-emphasis[1]. In the Compact Disc system a pre-emphasis characteristic has been adopted as standard with time constants of 15 and 50 μs. In some of the versions of the player the 'information for the listener' can be presented on a display and the different sections of the music on the disc can be played in the order selected by the user.

In the first article of a series of four on the Compact Disc system we shall deal with the complete system, without going into detail. We shall consider the disc, the processing of the audio signal, reading out the signal from the disc and the reconstitution of the audio signal. The articles that follow will examine the system aspects and modulation, error correction and the digital-to-analog conversion.

3.2.2 The disc

In the Laser Vision system[2], which records video information, the signal is recorded on the disc in the form of a spiral track that consists of a succession of pits. The intervals between the pits are known as 'lands'. The information is present in the track in analog form. Each transition from land to pit and vice versa marks a zero crossing of the modulated video signal. On the Compact Disc the signal is recorded in a similar manner, but the information is present in the track in digital form. Each pit and each land represents a series of bits called channel bits. After each land/pit or pit/land transition there is a '1', and all the channel bits in between are '0'; see Fig. 1.

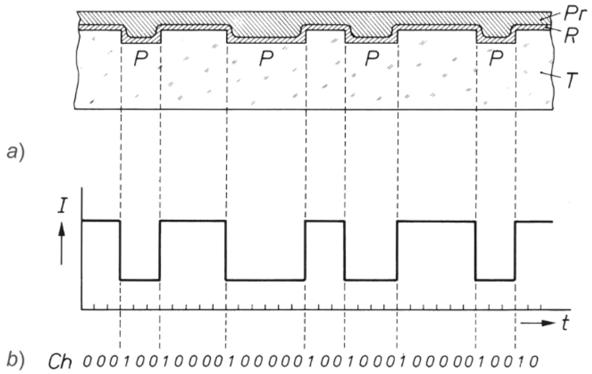


Fig. 1. *a)* Cross-section through a Compact Disc in the direction of the spiral track. T transparent substrate material, R reflecting layer, Pr protective layer. P the pits that form the track. *b*) I the intensity of the signal read by the optical pick-up (see Fig. 2), plotted as a function of time. The signal, shown in the form of rectangular pulses, is in reality rounded and has sloping sides[3]. The digital signal derived from this waveform is indicated as a series of channel bits *Ch*.

The density of the information on the Compact Disc is very high: the smallest unit of audio information (the audio bit) covers an area of 1 μm^2 on the disc, and the diameter of the scanning light-spot is only 1 μm. The pitch of the track is 1.6 μm, the width 0.6 μm and the depth 0.12 μm. The minimum length of a pit or the land between two pits is 0.9 μm, the maximum length is 3.3 μm. The side of the transparent carrier material T in which the pits P are impressed - the upper side during playback if the spindle is vertical - is covered with a reflecting layer R and a protective layer Pr. The track is optically scanned from below the disc at a constant velocity of 1.25 m/s. The speed of rotation of the disc therefore varies, from about 8 rev/s to about 3.5 rev/s.

3.2.3 Processing of the audio signal

For converting the analog signal from the microphone into a digital signal, pulse-code modulation (PCM) is used. In this system the signal is periodically sampled and each sample is translated into a binary number. From Nyquist's sampling theorem the frequency of sampling should be at least twice as high as the highest frequency to be accounted for in the analog signal. The number of bits per sample determines the signal-to-noise ratio in the subsequent reproduction.

In the Compact Disc system the analog signal is sampled at a rate of 44.1 kHz, which is sufficient for reproduction of the maximum frequency of 20 000 Hz. The signal is quantized by the method of uniform quantization; the sampled amplitude is divided into equal parts. The number of bits per sample (these are called audio bits) is 32, i.e. 16 for the left and 16 for the right audio channel. This corresponds to a signal-to-noise ratio of more than 90 dB. The net bit rate is thus $44.1 \times 10^3 \times 32 = 1.41 \times 10^6$ audio bits/s. The audio bits are grouped into 'frames', each containing six of the original samples.

Successive blocks of audio bits have blocks of parity bits added to them in accordance with a coding system called CIRC (Cross-Interleaved Reed-Solomon Code)[4]. This makes it possible to correct errors during the reproduction of the signal. The ratio of the number of bits before and after this operation is 3:4. Each frame then has C&D (Control and Display) bits, as mentioned earlier, added to it; one of the functions of the C&D bits is providing the 'information for the listener'. After the operation the bits are called data bits.

Next the bit stream is modulated, that is to say the data bits are translated into channel bits, which are suitable for storage on the disc; see fig. 1*b*. The EFM code (Eight-to-Fourteen Modulation) is used for this: in EFM code blocks of eight bits are translated into blocks of fourteen bits[5]. The blocks of fourteen bits are linked by three 'merging bits'. The ratio of the number of bits before and after modulation is thus 8:17.

For the synchronization of the bit stream an identical synchronization pattern consisting of 27 channel bits is added to each frame. The total bit rate after all these manipulations is 4.32×10^6 channel bits/s. Table I gives a survey of the successive operations with the associated bit rates, with their names.

From the magnitude of the channel bit rate and the scanning speed of 1.25 m/s it follows that the length of a channel bit on the disc is approximately 0.3 μm.

Name	Bit rate in 10^6 bits/s	Operations
Audio signal		PCM (44.1 kHz)
Audio bit stream	1.41	CIRC (+ parity bits)
		Addition of C&D bits
Data bit stream	1.94	EFM
		Addition of merging bits
		Addition of synchronization patterns
Channel bit stream	4.32	

Table I. Names of the successive signals, the associated bit rates and operations during the processing of the audio signal.

The signal produced in this way is used by the disc manufacturer to switch on and off the laser beam that illuminates the light-sensitive layer on a rotating glass disc (called the 'master'). A pattern of pits is produced on this disc by means of a photographic developing process. After the surface has been coated with a thin silver layer, an electroplating process is applied to produce a nickel impression, called the 'metal father'. From this 'father disc' impressions called 'mother discs' are produced in a similar manner. The impressions of the mother discs, called 'sons' or 'stampers', are used as tools with which the pits P are impressed into the thermoplastic transparent carrier material T of the disc; see Fig. 1.

3.2.4 Read-out from the disc

As we have seen, the disc is optically scanned in the player. This is done by the AlGaAs semiconductor laser described in an earlier article in this journal[6]. Fig. 2 shows the optical part of the 'pick-up'. The light from the laser *La* (wavelength 800 nm) is focused through the lenses L_2 and L_1 on to the reflecting layer of the disc. The diameter of the light spot *S* is about 1 μm. When the spot falls on an interval between two pits, the light is almost totally reflected and reaches the four photodiodes D_1-D_4 via the half-silvered mirror M. When the spot lands on a pit - the depth of a pit is about ¼ of the wavelength in the transparent substrate material - interference causes less light to be reflected and an appreciably smaller amount reaches the photodiodes. When the output signals from the four photodiodes are added together the result is a fairly rough approximation[3] to the rectangular pulse pattern present on the disc in the form of pits and intervals.

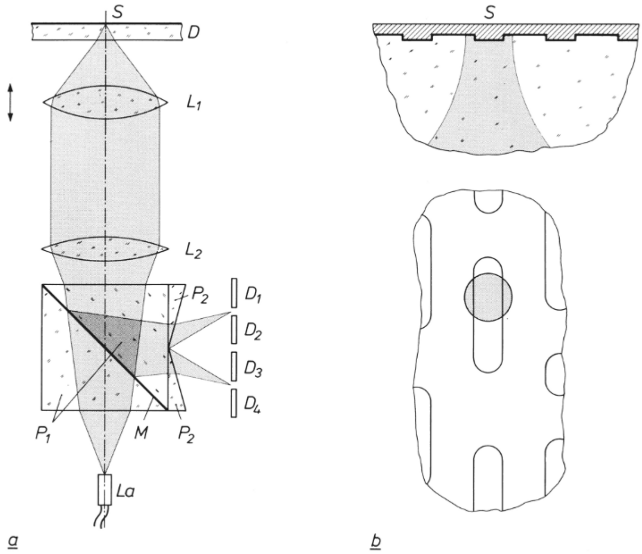


Fig. 2 *a)* Diagram of the optical pick-up. D radial section through the disc. S laser spot, the image on the disc of the light-emitting part of the semiconductor laser *La*. L_1 objective lens, adjustable for focussing. L_2 lens for making the divergent laser beam parallel. *M* half-silvered mirror formed by a film evaporated on the dividing surface of the prism combination P_1. P_2 beam-splitter prisms. D_1 to D_4 photodiodes whose output currents can be combined in various ways to provide the output signal from the pick-up and also the tracking-error signal and the focusing-error signal. (In practice the prisms P_2 and the photodiodes D_1 to D_4 are rotated by 90° and the reflection at the mirror *M* does not take place in a radial plane but in a tangential plane.) *b)* A magnified view of the light spot S and its immediate surroundings, with a plan view. It can clearly be seen that the diameter of the spot (about 1 μm) is larger than the width of the pit (0.6 μm).

The optical pick-up shown in Fig. 2 is very small (about 45 × 12 mm) and is mounted in a pivoting arm that enables the pick-up to describe a radial arc across the disc, so that it can scan the complete spiral track. Around the pivotal point of the arm is mounted a ‘linear’ motor that consists of a combination of a coil and a permanent magnet. When the coil is energized the pick-up can be directed to any required part of the track, the locational information being provided by the C&D bits added to each frame on the disc. The pick-up is thus able to find independently any particular passage of music indicated by the listener. When it has been found, the pick-up must then follow the track accurately - to within ± 0.1 μm - without being affected by the next or previous track. Since the track on the disc may have some slight eccentricity, and since also the suspension of the turntable is not perfect, the track may have a maximum side-to-side swing of 300 μm. A tracking servosystem is therefore necessary to ensure that the deviation between pick-up and track is smaller than the permitted value of ± 0.1 μm and in addition to absorb the consequences of small vibrations of the player.

The tracking-error signal is delivered by the four photodiodes D_1 to D_4. When the spot *S*, seen in the radial direction, is situated in the centre of the track, a symmetrical beam is reflected. If the spot lies slightly to one side of the track, however, interference effects cause asymmetry in the reflected beam. This asymmetry is detected by the prisms P_2, which split the beam into two components. Beyond the prisms one component has a higher mean intensity than the other. The signal obtained by coupling the photodiodes as (D_1+D_2) - (D_3+D_4) can therefore be used as a tracking-error signal.

As a result of ageing or soiling of the optical system, the reflected beam may acquire a slowly increasing, more or less constant asymmetry. Owing to a d.c. component in the tracking-error signal, the spot will then always be slightly off-centre of the track. To compensate for this effect a second tracking-error signal is generated. The coil that controls the pick-up arm is therefore supplied with an alternating voltage at 600 Hz, with an amplitude that corresponds to a radial displacement of the spot by ± 0.05 μm. The output sum signal from the four photodiodes - which is at a maximum when the spot is in the centre of the track - is thus modulated by an alternating voltage of 600 Hz. The amplitude of this 600 Hz signal increases as the spot moves off-centre. In addition the sign of the 600 Hz error signal changes if the spot moves to the other side of the track. This second tracking-error signal is therefore used to correct the error signal mentioned earlier with a direct voltage. The output sum signal from the photodiodes, which is processed in the player to become the audio signal, is thus returned to its maximum value.

The depth of focus of the optical pick-up at the position of *S* (see Fig. 2) is about 4 μm. The axial deviation of the disc, owing to various mechanical effects, can have a maximum of 1 mm. It is evident that a servosystem is also necessary to give correct focusing of the pick-up on the reflecting layer. The objective lens L_1 can therefore be displaced in the direction of its optical axis by a combination of a coil and a permanent magnet, in the same way as in a loudspeaker. The focusing-error signal is also provided by the row of photodiodes D_1 to D_4. If the spot is sharply focused on the disc, two sharp images are precisely located between D_1 and D_2 and between D_3 and D_4. If the spot is not sharply focused on the disc, the two images on the photodiodes are not sharp either, and have also moved closer together or further apart. The signal obtained by connecting the photodiodes as $(D_1+D_4)-(D_2+D_3)$ can therefore be used for controlling the focusing servosystem. The deviation in focusing then remains limited to ± 1 μm.

3.2.5 Reconstitution of the audio signal

The signal read from the disc by the optical pick-up has to be reconstituted to form the analog audio signal.

Fig. 3 shows the block diagram of the signal processing in the player. In *DEMOD* the demodulation follows the same rules that were applied to the EFM modulation, but now in the opposite sense. The information is then temporarily stored in a buffer memory and then reaches the error-detection and correction circuit *ERCO*. The parity bits can be used here to correct errors, or just to detect errors if correction is found to be impossible[4]. These errors may originate from defects in the manufacturing process, damage during use, or fingermarks or dust on the disc. Since the information with the CIRC code is 'interleaved' in time, errors that occur at the input of *ERCO* in one frame are spread over a large number of frames during decoding in *ERCO*. This increases the probability that the maximum number of correctable errors per frame will not be exceeded. A flaw such as a scratch can often produce a train of errors, called an error burst. The error-correction code used in *ERCO* can correct a burst of up to 4000 data bits, largely because the errors are spread out in this way.

If more errors than the permitted maximum occur, they can only be detected. In the CIM block (Concealment: Interpolation and Muting) the errors detected are then masked. If the value of a sample indicates an error, a new value is determined by linear interpolation between the preceding value and the next one. If two or more successive sample values indicate an error, they are made equal to zero (muting). At the same time a gradual transition is created to the values preceding and succeeding it by causing a number of values before the error and after it to decrease to zero in a particular pattern.

In the digital-to-analog converters *DAC*[7] the 16 bit samples first pass through interpolation filters F and are then translated and recombined to recreate the original analog audio signal A from the two audio channels L and R. Since samples must be recombined at exactly the same rate as they are taken from the analog audio signal, the DACs and also *CIM* and *ERCO* are synchronized by a clock generator C controlled by a quartz crystal.

Fig. 3 also illustrates the control of the disc speed n_D. The bit stream leaves the buffer memory at a rate synchronized by the clock generator. The bit stream enters the buffer memory, however, at a rate that depends on the speed of revolution of the disc. The extent to which n_D and the sampling rate are matched determines the 'filling degree' of the buffer memory. The control is so arranged as to ensure that the buffer memory is at all times filled to 50% of its capacity. The analog signal from the player is thus completely free from wow and flutter, yet with only moderate requirements for the speed control of the disc.

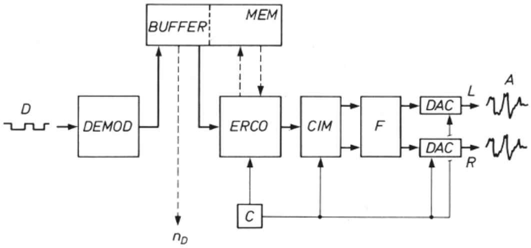


Fig. 3. Block diagram of the signal processing in the player. *D* input signal read by the optical pick-up; see Fig. 2. A the two output analog audio signals from the left (*L*) and the right (*R*) audio channels. *DEMOD* demodulation circuit. *ERCO* error-correction circuit. *BUFFER* buffer memory, forming part of the main memory *MEM* associated with *ERCO. CIM* (Concealment: Interpolation and Muting) circuit in which errors that are only detected since they cannot be corrected are masked or 'concealed'. *F* filters for interpolation. *DAC* digital-to-analog conversion circuits. Each of the blocks mentioned here are fabricated in VLSI technology. *C* clock generator controlled by a quartz crystal. The degree to which the buffer memory capacity is filled serves as a criterion in controlling the speed of the disc.

References

[1] See F. W. de Vrijer, Modulation, Philips tech. Rev. **36**, 305-362 (1976), in particular pages 323 and 324.

[2] See Philips tech. Rev. **33**, (Sect. 3.3).

[3] See Fig. 3 of the article by J. P. J. Heemskerk and K. A. Schouhamer Immink, Sect. 3.3

[4] See H. Hoeve, J. Timmermans and L. B. Vries, Error correction and concealment in the Compact Disc system, Sect. 3.4

[5] See J. P. J. Heemskerk and K. A. Schouhamer Immink, Compact Disc: system aspects and modulation, Sect. 3.3.

[6] J. C. J. Finck, H. J. M. van der Laak and J. T. Schrama, Philips tech. Rev. **39**, 37 (1980).

[7] See D. Goedhart, R. J. van de Plassche and E. F. Stikvoort, Digital-to-analog conversion in playing a Compact Disc, Sect. 3.5.

3.3 Compact Disc: system aspects and modulation

J.P.J. Heemskerk, K.A. Schouhamer Immink

Dr J. P. J. Heemskerk is with the Philips Audio Division, Eindhoven;
Ir K. A. Schouhamer Immink is with Philips Research Laboratories, Eindhoven.

Abstract
The Compact Disc system can be considered as a transmission system that brings sound from the studio into the living room. The sound encoded into data bits and modulated into channel bits is sent along the 'transmission channel' consisting of write laser — master disc — user disc — optical pick-up. The maximum information density on the disc is determined by the diameter *d* of the laser light spot on the disc and the 'number of data bits per light spot'. The effect of making *d* smaller is to greatly reduce the manufacturing tolerances for the player and the disc. The compromise adopted is $d \approx 1$ μm, giving very small tolerances for objective and disc tilt, disc thickness and defocusing. The basic idea of the modulation is that, while maintaining the minimum length for 'pit' and 'land' (the 'minimum run length') required for satisfactory transmission, the information density can be increased by increasing the number of possible positions per unit length for pit edges (the bit density). Because of clock regeneration there is also a maximum run length, and the low-frequency content of the transmission channel must be kept as low as possible. With the EFM modulation system used each 'symbol' of eight data bits is converted into 14 channel bits with a minimum run length of 3 and a maximum run length of 11 bits, plus three merging bits, chosen such that, when the symbols are merged together, the run-length conditions continue to be satisfied and the low-frequency content is kept to the minimum.

In this article we shall deal in more detail with the various factors that had to be weighed one against the other in the design of the Compact Disc system. In particular we shall discuss the EFM modulation system ('Eight-to-Fourteen Modulation'), which helps to produce the desired high information density on the disc.

Fig. 1 represents the complete Compact Disc system as a 'transmission system' that brings the sound of an orchestra into the living room. The orchestral sound is converted at the recording end into a bit stream B_i, which is recorded on the master disc. The master disc is used as the 'pattern' for making the discs for the user. The player in the living room derives the bit stream B_o - which in the ideal case should be identical to B_i - from the disc and reconverts it to the orchestral sound. The system between *COD* and *DECOD* is the actual *transmission channel*; B_i and B_o consist of 'channel bits'.

Reprinted with permission from Philips Tech. Rev. **40**, 157-164, 1982.

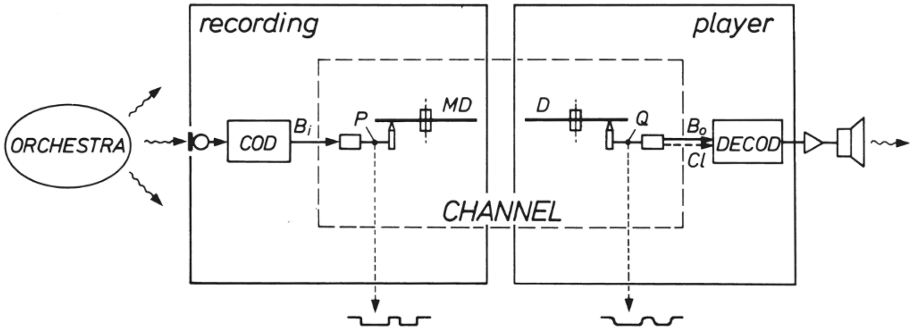


Fig. 1. The Compact Disc system, considered as a transmission system that brings sound from the studio into the living room. The transmission channel between the encoding system (*COD*) at the recording end and the decoding system (*DECOD*) in the player, 'transmits' the bit stream B_i to *DECOD* via the write laser, the master disc (*MD*), the disc manufacture, the disc (*D*) in the player and the optical pick-up; in the ideal case B_o is the same as B_i. The bits of B_o, as well as the clock signal (*Cl*) for further digital operations, have to be detected from the output signal of the pick-up unit at *Q*.

Fig. 2 shows the encoding system in more detail. The audio signal is first converted into a stream B_1 of 'audio bits' by means of pulse-code modulation. A number of bits for 'control and display' (C&D) and the parity bits for error correction are then added to the bit stream[1][2]. This results in the 'data bit stream' B_2. The modulator converts this into channel bits (B_3). The bit stream B_i is obtained by adding a synchronization signal. The number of data bits n that can be stored on the disc in given by:

$$n = \eta A/d^2,$$

where A is the useful area of the disc surface, d is the diameter of the laser light spot on the disc and η is the 'number of data bits per spot' (the number of data bits that can be resolved per length d of track). A/d^2 is the number of spots that can be accomodated side by side on the disc. The information density n/A is thus given by:

$$n/A = \eta/d^2.$$

The spot diameter d is one of the most important parameters of the channel. The modulation can give a higher value of η. We shall now briefly discuss some of the aspects of the channel that determine the specification for the modulation system.

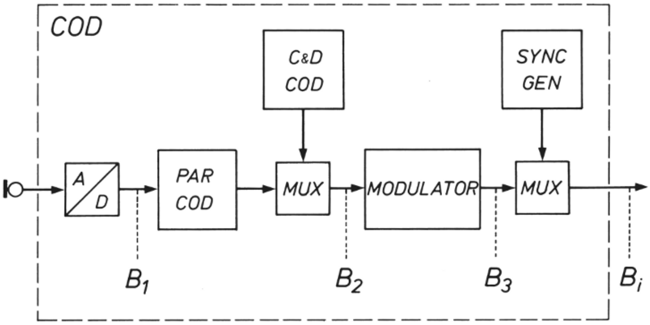


Fig. 2. The encoding system (*COD* in Fig. 1). The system is highly simplified here; in practice for example there are two audio channels for stereo recording at the input, which together supply the bit stream B_1 by means of PCM, and the various digital operations are controlled by a 'clock', which is not shown. The bit stream B_1 is supplemented by parity and C&D (control and display) bits (B_2), modulated (B_3), and provided with synchronization signals (B_i). *MUX*: multiplexers. Fig. 9 gives the various bit streams in more detail.

The channel

The bit stream B_i in Fig. 1 is converted into a signal at *P* that switches the light beam from the write laser on and off. The channel should be of high enough quality to allow the bit stream B_i to be reconstituted from the read signal at *Q*.

To achieve this quality all the stages in the transmission path must meet exacting requirements, from the recording on the master disc, through the disc manufacture, to the actual playing of the disc. The quality of the channel is determined by the player and the disc: these are mass-produced and the tolerances cannot be made unacceptably small.

We shall consider one example here to illustrate the way in which such tolerances affect the design: the choice of the 'spot diameter' *d*. We define *d* as the half-value diameter for the light intensity; we have

$$d = 0.6\ \lambda/NA,$$

where λ is the wavelength of the laser light and *NA* is the numerical aperture of the objective. To achieve a high information density (1) *d* must be as small as possible. The laser chosen for this system is the small CQL10[3], which is inexpensive and only requires a low voltage; the wavelength is thus fixed; $\lambda \approx$ 800 nm. This means that we must make the numerical aperture as large as possible. With increasing *NA*, however, the manufacturing tolerances of the player and the disc rapidly become smaller. For example, the tolerance in the local 'skew' of the disc (the 'disc tilt') relative to the objective-lens axis is proportional to NA^{-3}. The tolerance for the disc thickness is proportional

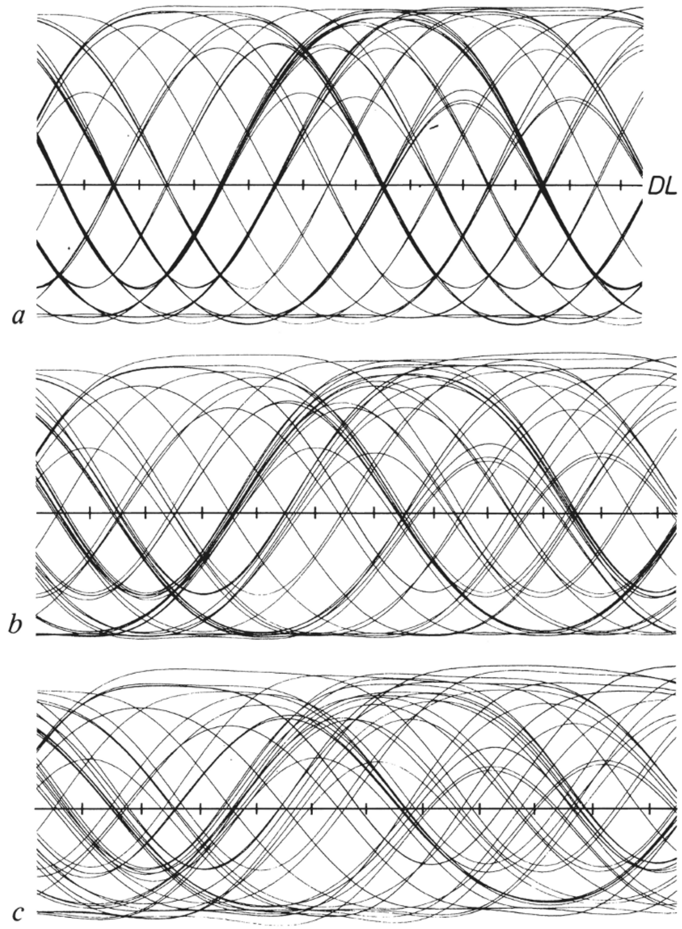


Fig. 3 *a)* Eye pattern. The figures give the read signal (at *Q* in Fig. 1) on an oscilloscope synchronized with the bit clock. At the decision times (marked by dashes) it must be possible to determine whether the signal is above or below the decision level (*DL*). The curves have been calculated for *a)* an ideal optical system, *b)* a defocusing of 2 μm, *c)* a defocusing of 2 μm and a disc tilt of 1.2°. The curves give a good picture of experimental results.

to NA^{-4}, and the depth of focus, which determines the focusing tolerance, is proportional to NA^{-2}. After considering all these factors in relation to one another, we arrived at a value of 0.45 for *NA*. We thus find a value of 1 μm for the spot diameter *d*.

The quality of the channel is evaluated by means of an 'eye pattern', which is obtained by connecting the point *Q* in Fig. 1 to an oscilloscope synchronized with the clock for the bit stream B_o; see Fig. 3*a*. The signals originating from different pits and lands are super-imposed on the screen; they are strongly rounded, mainly because the spot diameter is not zero and the pit walls are not vertical. If the transmission quality is adequate, however, it is always possible to determine whether the signal is positive or negative at the 'clock times' (the dashes in Fig. 3*a*), and hence to reconstitute the bit stream. The lozenge pattern around a dash in this case is called the 'eye'. Owing to channel imperfections the eye can become obscured; owing to 'phase jitter' of the signal relative to the clock an eye becomes narrower, and noise reduces its height. The signals in Fig. 3a were calculated for a perfect optical system. Fig. 3*b* shows the effect of defocusing by 2 μm and Fig. 3*c* shows the effect of a radial tilt of 1.2° in addition to the defocusing. In Fig. 3*b* a correct decision is still possible, but not in Fig. 3*c*.

This example also gives some idea of the exacting requirements that the equipment has to meet. A more general picture can be obtained from Table I, which gives the manufacturing tolerances of a number of important parameters, both for the player and for the disc. The list is far from complete, of course.

With properly manufactured players and discs the channel quality can still be impaired by dirt and scratches forming on the discs during use. By its nature the system is fairly insensitive to these[1], and any errors they may introduce can nearly always be corrected or masked[2]. In the following we shall see that the modulation system also helps to reduce the sensitivity to imperfections.

Player	Objective-lens tilt ± 0.2 °
	Tracking ±0.1 μm
	Focusing ± 0.5 μm
	R.M.S. wavefront noise of read laser beam 0.05 λ (40 nm)
Disc	Thickness 1.2 ± 0.1 mm
	Flatness ± 0.6° (at the rim corresponding to a sag of 0.5 mm)
	Pit-edge positioning ± 50 nm
	Pit depth 120 ± 10 nm

Table I: Manufacturing tolerances.

Bit-stream modulation

The playing time of a disc is equal to the track length divided by the track velocity *v*. For a given disc size the playing time therefore increases if we decrease the track velocity in the system (the track velocity of the master disc and of the user disc). However, if we do this the channel becomes 'worse': the eye height decreases and the system becomes more sensitive to perturbations. There is therefore a lower limit to the track velocity if a minimum value has been established for the eye height because of the expected level of noise and perturbation. We shall now show that we can decrease this lower limit by an appropriate bitstream modulation.

We first consider the situation without modulation. The incoming data bit stream is an arbitrary sequence of ones and zeros. We consider a group of 8 data bits in which the change of bit value is fastest (Fig. 4*a*). Uncoded recording (1: pit; 0: land, or vice versa) then gives the pattern of Fig. 4*b*. This results in the rounded-off signal of Fig. 4*c* at *Q* in Fig. 1; Fig. 4*d* gives the eye pattern. The signal in Fig. 4c represents the highest frequency (f_{m1}) for this mode of transmission, and we have $f_{m1} = \frac{1}{2} f_d$, where f_d is the data bit rate. The half eye height a_1 is equal to the amplitude A_1 of the highest-frequency signal.

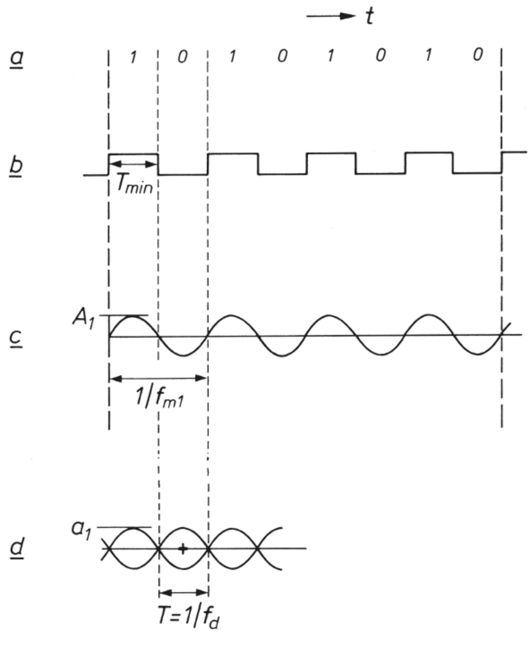


Fig. 4. Direct recording of the data bit stream on the disc. *a*) Data bit stream of the highest frequency that can occur. *b*) Direct translation of the bit stream into a pattern of pits. *c*) The corresponding output signal (at *Q* in Fig. 1); its amplitude A_1 is found with the aid of Fig. 5. *d*) The eye pattern that follows from (*c*). T_{min} minimum pit or land length; f_{m1} highest frequency; *T* data bit length; f_d data bit rate. We have $T_{min} = T$; $f_{m1} = \frac{1}{2} f_d$.

The relation between the eye height and the track velocity now follows indirectly from the 'amplitude-frequency characteristic' of the channel; see Fig. 5. In this diagram *A* is the amplitude of the sinusoidal signal at *Q* in Fig. 1 when a sinusoidal unit signal of frequency *f* is presented at *P*. With the aid of Fourier analysis and synthesis the output signal can be calculated from *A*(*f*) for any input signal. The line in the diagram represents a channel with a perfect optical system. In the first part of this section we shall take this for granted. The true situation will always be less favourable. The 'cut-off frequency' is determined by the spot diameter and the track velocity v; in the ideal case $f_c = (2NA/\lambda)v$.

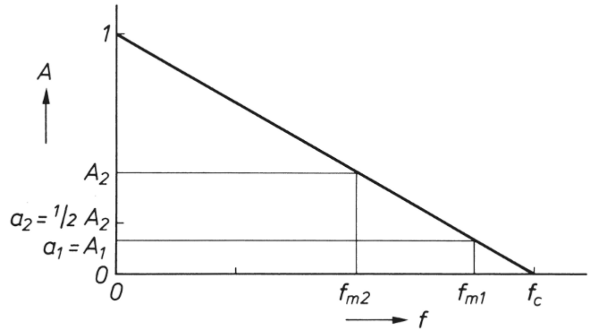


Fig. 5. Amplitude-frequency characteristic of the channel. The diagram gives the amplitude A of the sinusoidal signal at Q (Fig. 1) when a sinusoidal unit signal is presented at P as a function of the frequency f. The transfer is 'cut off' at the frequency f_c, which is given by $f_c = (2NA/\lambda)v$. The line shown applies to an ideal optical system; in reality A is always somewhat lower; the cut-off frequency is then effectively lower. The 'maximum frequencies' f_{m1}, f_{m2}, the amplitudes A_1, A_2 and the 'half eye heights' a_1, a_2 relate to the 'direct' and 'modulated' writing of the data bits on the disc; see Figs 4 and 6.

For a given track velocity we now obtain the half eye height a_1 in Fig. 4 directly from Fig. 5: it is equal to the amplitude A_1 at the frequency f_{m1}. If v, and hence f_c, is varied, the line in Fig. 5 rotates about the point 1 on the A-axis. For a given minimum value of a_1, the figure indicates how far f_c can be decreased; this establishes the lower limit for v. In particular, if the minimum value for a_1 is very small, f_c can be decreased to a value slightly above f_{m1} (= ½f_d).

Fig. 6 gives the situation **with** modulation: an imaginary 8→16 modulation, which is very close to EFM, however. Each group of 8 incoming data bits (Fig. 6*a*) is converted into 16 channel bits (Fig. 6*a*′). This is done by using a 'dictionary' that assigns unambiguously but otherwise arbitrarily to each word of 8 bits a word of 16 bits, but in such a way that the resultant channel bit stream only produces pits and lands that are at least three channel bits long (Fig. 6*b*). On the time scale the minimum pit and land lengths ('the minimum run length' T_{min}) have become 1½ times as long as in Fig. 4, but a simple calculation shows that about as much information can nevertheless be transmitted as in Fig. 4 (256 combinations for 8 data bits), because there is a greater choice of pit-edge positions per unit length (see Fig. 6*b* and *b*′); the 'channel bit length' T_c has decreased by a half.

With the modulation we have managed to reduce the highest frequency (f_{m2}) in the signal (see Fig. 6*c*, left; $f_{m2} = ⅓ f_d = ⅔ f_{m1}$). Therefore f_c and v can be reduced by a factor of 1½ for the case in which a very small eye height is tolerable (see Fig. 5); this represents an increase of 50% in playing time.

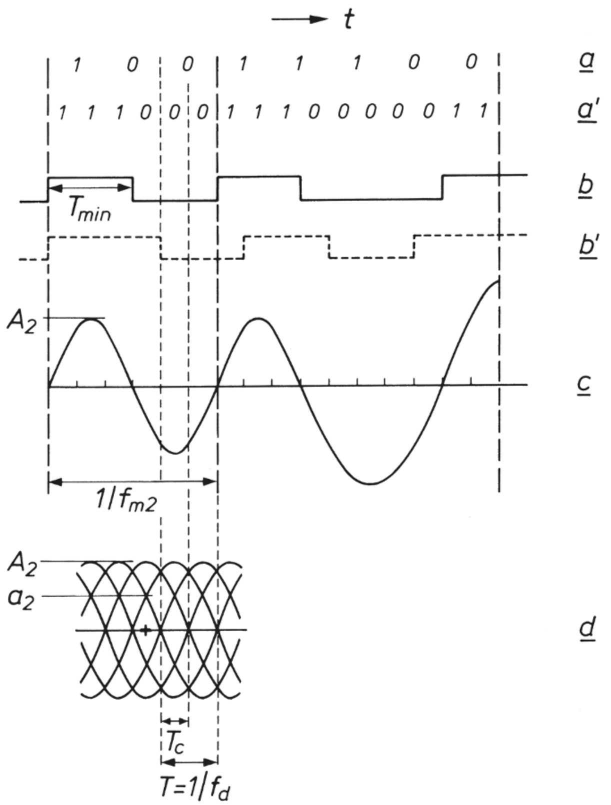


Fig. 6. Eight-to-sixteen modulation. Each group of 8 data bits (a) is translated with the aid of a dictionary into 16 channel bits (*a'*), in such a way that the run length is equal to at least three channel bits. *b*) Pattern of pits produced from the bit stream (*a'*). *b'*) pattern of pits obtained with a different input signal. *c*) The read signal corresponding to (*b*); its amplitude is again determined from Fig. 5.*d*) The resultant eye pattern. The half eye height (*a*2) here is only half the amplitude (A_2) of the approximately sinusoidal signal of maximum frequency (f_{m2}).

The modulation also has its disadvantages. In the first place the half eye height (a_2) in this case is only half of the amplitude (A_2) of the signal at the highest frequency (see Fig. 6*d*). This has consequences if the minimum eye height is not very small. For example, the modulation becomes completely unusable if the half eye height in Fig. 5 has to remain larger than ½ ($a_2 > ½$ implies $A_2 > 1$); uncoded recording is then still possible ($A_1 = a_1$). In the second place, the tolerance for time errors and for the positioning of pit edges, together with the eye width (T_c), has decreased by a half. In designing a system, the various factors have to be carefully weighed against one another.

To show qualitatively how a choice can be made, we have plotted the half eye height in Fig.7 as a function of the 'linear information density' σ (the number of incoming data bits per unit length of the track; $\sigma = f_d/v$) for three systems: '8→8 modulation' (i.e. uncoded recording), 8→16 modulation, and a system that also has about the same information capacity (256 combinations for 8 data bits) in which, however, the minimum run length has been increased still further, again at the expense of eye width of course ('8→24 modulation', $T_{min} = 2T$, $T_c = ⅓T$). The figure is a direct consequence of the reasoning above, with the assumption that the cut-off frequency is 20% lower than the ideal value $(2NA/\lambda)v$, as a first rough adjustment to what we find in practice for the function $A(f)$.

In qualitative terms, the 8→16 system has been chosen because the nature of the noise and perturbations is such that the eye can be smaller than at A in Fig. 7, but becomes too small at C. An improvement is therefore possible with 8→16 modulation, but not with 8→24 modulation.

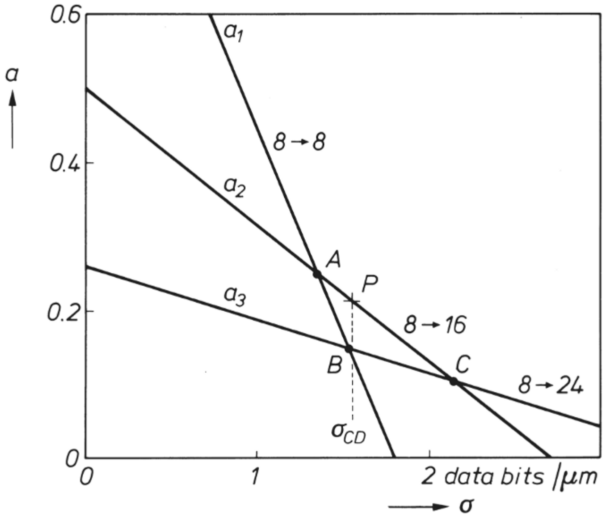


Fig. 7. Half eye height a as a function of the linear information density σ, for 8→8, 8→16 and 8→24 modulation. These systems are characterized by the following values for the channel bit length T_c and the minimum run length T_{min}:
8→8: $T_c = T$, $T_{min} = T$ (Fig. 4),
8→16: $T_c = ½\,T$, $T_{min} = T$ (Fig. 6),
8→24: $T_c = ⅓\,T$, $T_{min} = 2\,T$,
where T is the data bit length. The straight lines give the relations that follow from Fig. 5:
$a_1 = c_1(1 - f_{m1}/f_c) \rightarrow a_1 = 1 - \sigma/1.8$,
$a_2 = c_2(1 - f_{m2}/f_c) \rightarrow a_2 = 0.5(1 - \sigma/2.7)$,
$a_3 = c_3(1 - f_{m3}/f_c) \rightarrow a_3 = 0.26(1 - \sigma/3.6)$,
where σ is the numerical value of the linear information density, expressed in data bits per µm. The c's are the ratios of the half eye height to the amplitude, and the f_m's the maximum frequencies for the three systems ($c_1 = 1$, $c_2 = \sin 30° = 0.5$, $c_3 = \sin 15° = 0.26$, $f_{m1} = ½\,f_d$, $f_{m2} = ⅓\,f_d$, $f_{m3} = ¼f_d$; f_d is the data bit rate). The second set of equations follows from the first set by substituting $0.8 \times (2NA/\lambda)v$ for f_c, with $NA = 0.45$, $\lambda = 0.8$ µm, $v = f_d/\sigma$. The factor 0.8 is introduced as a rough first-order correction to the 'ideal' amplitude characteristic.

For our Compact Disc system we have $\sigma = 1.55$ data bits/µm ($f_d = 1.94$ Mb/s, $v = 1.25$ m/s[1]); the operating point would therefore be at *P* in Fig. 7. The model used is however rather crude and in better models *A*, *B* and *C* lie more to the left, so that *P* approaches *C*. But 8 → 16 modulation is still preferable to 8 → 24 modulation, even close to *C*, since the eye width is 1½ times as large as for 8 → 24 modulation.

EFM is a refinement of 8→16 modulation. It has been chosen on the basis of more detailed models and many experiments. At the eye height used, it gives a gain of 25% in information density, compared with uncoded recording.

Further requirements for the modulation system

In developing the modulation system further we still had two more requirements to take into account.

In the first place it must be possible to regenerate the *bit clock* in the player from the read-out signal (the signal at *Q* in Fig. 1). To permit this the number of pit edges per second must be sufficiently large, and in particular the 'maximum run length' T_{max} must be as small as possible.

The second requirement relates to the 'low-frequency content' of the read signal. This has to be as small as possible. There are two reasons for this. In the first place, the servosystems for track following and focusing[1] are controlled by low-frequency signals, so that low-frequency components of the information signal could interfere with the servo-systems. The second reason is illustrated in Fig. *8*, in which the read signal is shown for a clean disc (*a*) and for a disc that has been soiled, e.g. by fingermarks (*b*). This causes the amplitude and average level of the signal to fall. The fall in level causes a completely wrong read-out if the signal falls below the decision level. Errors of this type are avoided by eliminating the low-frequency components with a filter (*c*), but the use of such a filter is only permissible provided the information signal itself contains no low-frequency components. In the Compact Disc system the frequency range from 20 kHz to 1.5 MHz is used for information transmission; the servosystems operate on signals in the range 0-20 kHz.

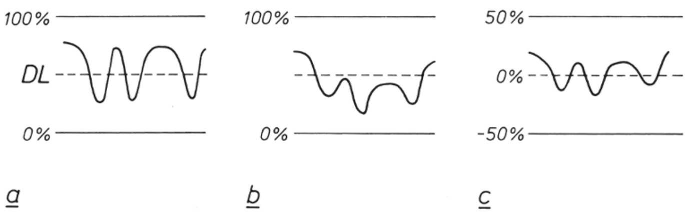


Fig. 8. The read-out signal for six pit edges on the disc, a) for a clean disc, b) for a soiled disc, c) for a soiled disc after the low frequencies have been filtered out. DL decision level. Because of the soiling, both the amplitude and the signal level decrease; the decision errors that this would cause are eliminated by the filter.

The EFM modulation system

Fig. 9 gives a schematic general picture of the bit streams in the encoding system. The information is divided into 'frames'. One frame contains 6 sampling periods, each of 32 audio bits (16 bits for each of the two audio channels). These are divided into symbols of 8 bits. The bit stream B_1 thus contains 24 symbols per frame. In B_2 eight parity symbols have been added and one C&D symbol, resulting in 33 'data symbols'.

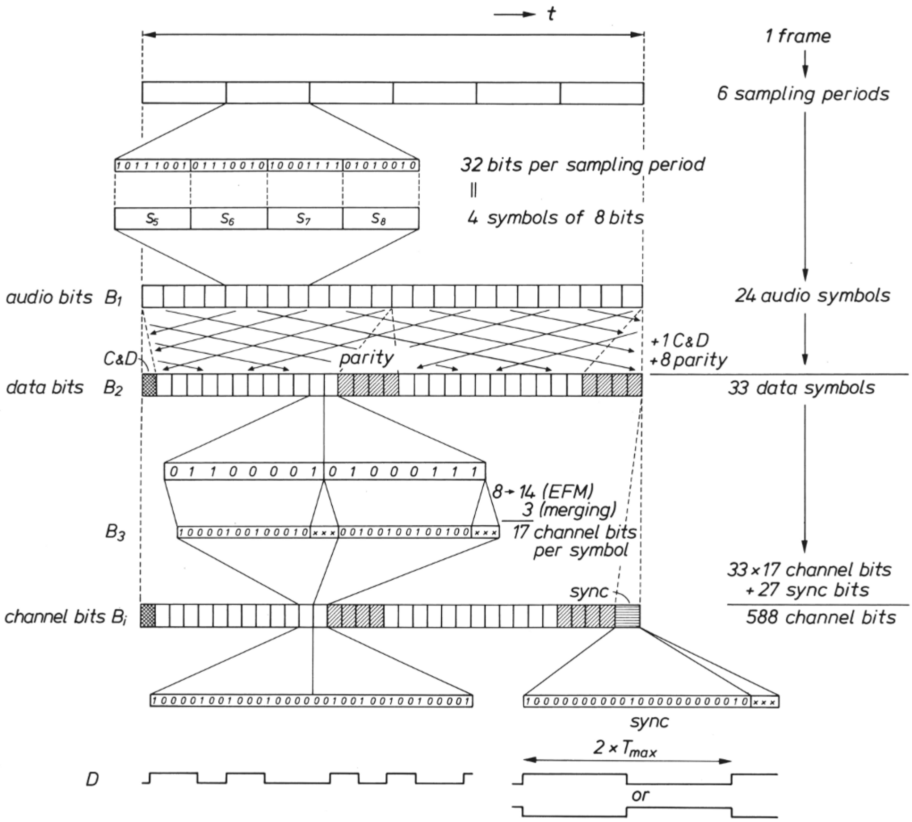


Fig. 9. Bit streams in the encoding system (Fig. 2). The information is divided into frames; the figure gives one frame of the successive bit streams. There are six sampling periods for one frame, each sampling period giving 32 bits (16 for each of the two audio channels). These 32 bits are divided to make four symbols in the 'audio bit stream' B_1. In the 'data bit stream' B_2 eight parity and one *C&D* symbols have been added to the 24 audio symbols. To scatter possible errors, the symbols of different frames in B_1 are interleaved, so that the audio signals in one frame of B_2 originate from different frames in B_1. The modulation translates the eight data bits of a symbol of B_2 into fourteen channel bits, to which three 'merging bits' are added (B_3). The frames are marked with a synchronization signal of the form illustrated (bottom right); the final result is the 'channel bit stream' (B_i) used for writing on the master disc, in such a way that each '1' indicates a pit edge (D).

The modulator translates each symbol into a new symbol of 14 bits. Added to these are three ‘merging bits’, for reasons that will appear shortly. After the addition of a synchronization symbol of 27 bits to the frame, the bit stream B_i is obtained. B_i therefore contains $33 \times 17 + 27 = 588$ channel bits per frame. Finally, B_i is converted into a control signal for the write laser. It should be noted that in B_i ‘1’ or ‘0’ does not mean ‘pit’ or ‘land’, as we assumed for simplicity in Fig. 6, but a ‘1’ indicates a pit edge. The information is thus completely recorded by the positions of the pit edges; it therefore makes no difference to the decoding system if ‘pit’ and ‘land’ are interchanged on the disc.

Opting for the translation of series of 8 bits following the division into symbols in the parity coding has the effect of avoiding error propagation. This is because in the error-correction system an entire symbol is always either ‘wrong’ or ‘not wrong’. One channel-bit error that occurs in the transmission spoils an entire symbol, but — because of the correspondence between modulation symbols and data symbols — never more than one symbol. If a different modulation system is used, in which the data bits are not translated in groups of 8, but in groups of 6 or 10, say, then the bit stream B_2 is in fact first divided up into 6 or 10 bit ‘modulation symbols’. Although one channel-bit error then spoils only one modulation symbol, it usually spoils two of the original 8 bit symbols.

In EFM the data bits are translated 8 at a time into 14 channel bits, with a T_{min} of 3 and a T_{max} of 11 channel bits (this means at least 2 and at the most 10 successive zeros in B_i). This choice came about more or less as follows. We have already seen that the choice of about 1½ data bits for T_{min}, with about 16 channel bits on 8 data bits, is about the optimum for the Compact Disc system[4]. A simple calculation shows that at least 14 channel bits are necessary for the reproduction of all the 256 possible symbols of 8 data bits under the conditions $T_{min} = 3$, $T_{max} = 11$ channel bits. The choice of T_{max} was dictated by the fact that a larger choice does not make things very much easier, whereas a smaller choice does create far more difficulties.

With 14 channel bits it is possible to make up 267 symbols that satisfy the run-length conditions. Since we only require 256, we omitted 10 that would have introduced difficulties with the ‘merging’ of symbols under these conditions, and one other chosen at random. The dictionary was compiled with the aid of computer optimization in such a way that the translation in the player can be carried out with the simplest possible circuit, i.e. a circuit that contains the minimum of logic gates.

The merging bits are primarily intended to ensure that the run-length conditions continue to be satisfied when the symbols are ‘merged’. If the run length is in danger of becoming too short we choose ‘0’s for the merging bits; if it is too long we choose a ‘1’ for one of them. If we do this we still retain a large measure of freedom in the choice of the merging bits, and we use this freedom to minimize the low-frequency content of the signal. In itself, two merging bits would be sufficient for continuing to satisfy the run-length conditions. A third is necessary, however, to give sufficient freedom for effective suppression of the low-frequency content, even though it means a loss of 6% of the information density on the disc. The merging bits contain no audio information, and they are removed from the bit stream in the demodulator.

Fig. 10 illustrates, finally, how the merging bits are determined. Our measure of the low-frequency content is the ‘digital sum value’ (DSV); this is the difference between the totals of pit and land lengths accumulated from the beginning of the disc. At the top are shown two data symbols of B_2 and their translation from the dictionary into channel symbols (B_3). From the T_{min} rule the first of the merging bits in this case must be a zero; this position is marked ‘*X*’. In the two following positions the choice is free; these are marked ‘*M*’. The three possible choices *XMM*=000, 010 and 001 would give rise to the patterns of pits as illustrated, and to the indicated waveform of the DSV, on the assumption that the DSV was equal to 0 at the beginning. The system now opts for the merging combination that makes the DSV at the end of the second symbol as small as possible, i.e. 000 in this case. If the initial value had been −3, the merging combination 001 would have been chosen.

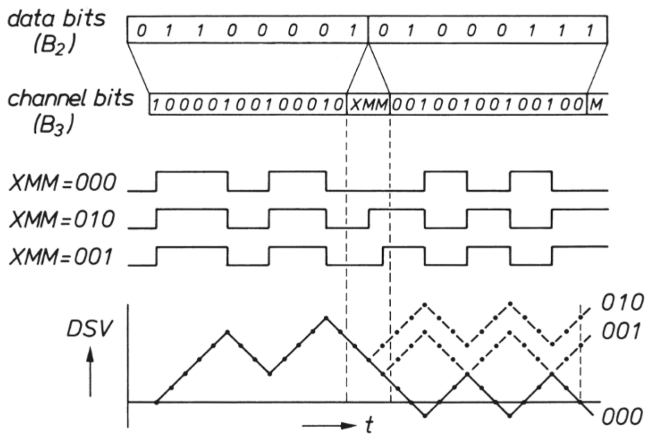


Fig. 10. Strategy for minimizing the digital sum value (DSV). After translation of the data bits into channel bits, the symbols are merged together by means of three extra bits in such a way that the run-length conditions continue to be satisfied and the DSV remains as small as possible. The first run-length rule (at least two zeros one after the other) requires a zero at the first position in the case illustrated here, while the choice remains free for the second and third positions. In this case there are thus three merging alternatives: 000, 010 and 001. These alternatives give the patterns of pits shown in the diagram and the illustrated DSV waveform. The system chooses the alternative that gives the lowest value of DSV at the end of the next symbol. The system looks ‘one symbol ahead’; strategies for looking further ahead are also possible in principle.

When this strategy is applied, the noise in the servo-band frequencies (< 20 kHz) is suppressed by about 10 dB. In principle better results can be obtained, within the agreed standard for the Compact Disc system, by looking more than one symbol ahead, since minimization of the DSV in the short term does not always contribute to longer-term minimization. This is not yet done in the present equipment.

References

[1] M. G. Carasso, J. B. H. Peek and J. P. Sinjou, The Compact Disc Digital Audio system, (Sect. 3.2).

[2] H. Hoeve, J. Timmermans and L. B. Vries, Error correction and concealment in the Compact Disc system (Sect. 3.4)..

[3] J. C. J. Finck, H. J. M. van der Laak and J. T. Schrama, A semiconductor laser for information read-out, Philips tech. Rev. **39**, 37-47 (1980).

[4] A more detailed discussion is given in K. A. Immink, Modulation systems for digital audio discs with optical readout, Proc. IEEE Int. Conf. on Acoustics, speech and signal processing, Atlanta 1981, pp. 587-589.

3.4 Error correction and concealment in the Compact Disc system

H. Hoeve, J. Timmermans, L.J. Vries

Ir H. Hoeve and J. Timmermans are with the Philips Audio Division, Eindhoven; Ir L. B. Vries is with Philips Research Laboratories, Eindhoven.

Abstract
After an example showing how errors in a digital signal can be corrected, the article deals with the theory of block codes. The treatment of random errors and error bursts is discussed. Error correction in the Compact Disc system uses a Cross-Interleaved Reed-Solomon Code (CIRC), which is a combination of a (32,28) and a (28,24) code. One of the two decoders in the CIRC decoding circuit corrects single errors, the other corrects double errors. The residual errors are interpolated linearly to a length of up to 12 000 bits, and longer errors are muted. The interpolation and the signal muting take place in a separate chip, whose configuration is briefly discussed.

3.4.1 Introduction

When analog signals such as audio signals are transmitted and recorded via an intervening system such as a gramophone record it is difficult to properly correct signal errors that have occurred in the path between the audio source and the receiving end. With suitably coded digital signals, however, a practical means of error correction does exist. We shall demonstrate this with the following example[1].

Suppose that a message of 12 binary units (bits) has to be transmitted (a stream of digital information can always be divided into groups of a particular size for transmission). The 12 bits x_{ij} are arranged as follows in a matrix, in which all x_{ij} can only have the value 0 or 1:

$$\begin{matrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{matrix}$$

To discover at the receiving end whether the message read there contains an error, and, if so, what the error is, one extra bit (called a ‘parity bit’) is added to

Reprinted with permission from Philips Tech. Rev. **40**, 166-172, 1982.

each row and column: x_{15}, x_{25}, x_{35} and x_{41}, x_{42}, x_{43}, x_{44} respectively. These parity bits provide a check on the correctness of the message received. The values assigned to them are such that x_{i5} (i = 1, 2, 3) makes the number of ones in row i even, for example, while x_{4j} (j = 1, 2, 3, 4) makes the number of ones in column j even. Next, a further parity bit (x_{45}) is added that has a value such that the number of ones in the block is made even. This results in the following matrix of four rows and five columns:

$$
\begin{matrix}
x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\
\mathrm{x}_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\
x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\
x_{41} & x_{42} & x_{43} & x_{44} & x_{45}
\end{matrix}
$$

It is easy to verify that the number of ones in the last row is also even, and so is the number of ones in the last column. If now a bit, say x_{23}, is incorrectly read at the receiving end, then the number of ones in the second row and the number of ones in the third column will no longer be even, and once this has been ascertained, a 0 at position x_{23} can be changed into a 1, or vice versa, thus correcting the error.

So as to be able in this way to correct one error in 12 information bits, it is necessary to send a total of 20 bits instead of 12: the 'code word' of $n=20$ bits consists of $k=12$ information bits and $n-k=8$ parity bits. The (n,k) code used here, a (20,12) code, makes it possible to correct single errors and also, as can easily be verified, to detect various multiple bit errors.

The 'rate' of an error-correcting code is taken to be the ratio of the number of information bits to the total number of bits per code word: k/n. The (20,12) code does not have a high rate, because it requires a relatively large number of parity bits. For the Compact Disc this would entail a considerable reduction in the playing time.

The theory of error-correcting codes[2] gives design methods that entail a minimal addition of parity bits when certain correction criteria are satisfied. An important concept in this theory is the 'distance' and in particular the 'minimum distance' d_m between two code words of n bits. Distance here is taken to be the number of places in which the bits of the two code words differ from each other. In the above example the minimum distance $\mathrm{d_m}$ is equal to 4: if one single bit of the k information bits changes, then the two parity bits of the associated row and column change at the same time, as does the one at the bottom right-hand corner, x_{45}, so that the entire code word has changed at four places. Theory tells us that to correct all the combinations of t errors occurring within one word, the minimum distance must be at least $2t+1$. To correct single errors, therefore, the minimum distance need be no greater than three. Examples of this are the single-error-correcting Hamming codes[4].

The statement that a code word x, which is received as a different word z because of terrors, can be restored to its original form if the minimum distance is $2t + 1$, can be seen from Fig. 1. A decoder provided with a list in which all the code words are stored can compare z with each of these code words and thus recover the correct code word unambiguously.

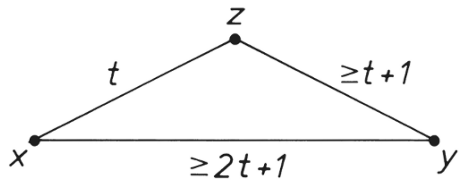


Fig. 1. The original transmitted code word x is received as z owing to t bit errors. Any code word y differing from x lies at a distance $\geq 2t + 1$ from x. To cause z to change into y it is necessary to change at least $t + 1$ bits. It follows that x is the only code word that has a distance t from z.

3.4.2 On the theory of block codes

In the foregoing we have shown with a simple example that it is possible to correct errors. Error-correcting systems do have their limitations, of course. To make this clear we shall consider how error-correcting codes should be designed to guarantee a specific measure of correction, with as few extra bits as possible added to the digital information to be transmitted. It will help if we first say something about the theory of block codes.

So that known and efficient error-correcting codes can be applied, groups of bits are formed by adding together a fixed number s of consecutive bits; these groups are called symbols. With these symbols we now set to work in the same way as with the bits in the foregoing: the information symbols are grouped together to form blocks with a length of k symbols. For error-correction we now add parity symbols to expand each block of k information symbols into a code word of n symbols. The $n - k$ parity symbols to be added are calculated from the k information symbols, and this is done in such a way as to make the error correction as effective as possible. Thus, of the very large number of possibly different words of n symbols only a small fraction, i.e. $2^{(k-n)s}$, become code words (see Fig. 2).

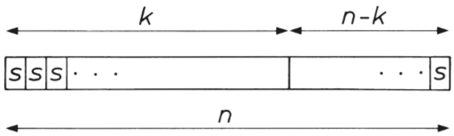


Fig. 2. A code word of length *n* consists of an information block of *k* symbols and a parity block of *n-k* symbols; each symbol comprises *s* bits. The number of possible words of *n* symbols is 2^{ns}. The parity bits are fixed for each combination of the *ks* information bits in accordance with established encoding rules . The number of code words is thus 2^{ks}. It follows that the fraction $2^{(k-n)s}$ of the number of possible words consists of code words.

For a given encoding system both *n* and *k* are fixed.

As already mentioned in the article on modulation in the Compact Disc system[3], the start of each word is marked by a synchronization symbol. (A word marked by a synchronization symbol is called a 'frame'.) The error-correcting system therefore knows when a new word begins, and the only errors it has to deal with are errors that occur in the transmission of data.

There are two kinds of errors: those that are distributed at random among the individual bits, the random errors, and errors that occur in groups that may cover a whole symbol or a number of adjacent symbols; these are called 'bursts' of errors. They can occur on a disc as a result of dirt or scratches, which interfere with the read-out of a number of adjacent pits and lands.

The best code for correcting random errors is the one that, for given values of *n* and *k*, is able to correct the largest number of independent errors within one code word. In the detection and correction of errors the symbols have to undergo a wide variety of operations. Large *k*-values (as with the Compact Disc) require extremely complex computing hardware. Practice has shown that the only acceptable solution to this problem is to choose a convenient code. And the only usable codes that enter into consideration, so far as we know at present, are the 'linear codes'.

A code is linear if it obeys the following rule:

If $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are code words, then their sum $x + y = (x_1 + y_1, ..., x_n + y_n)$ is also a code word.

In this sum the symbol $x_i + y_i$ is produced - irrespective of the number of bits *s* per symbol - by a modulo-2 bit addition. The special feature of the linear code is thus that each sum of code words yields another code word, i.e. a word of *n* symbols, which also belongs to the small fraction of symbol combinations permitted in the code.

It is this linearity feature that makes it possible to cut down considerably on the extent of the decoding equipment. The *Reed-Solomon codes*[2] are examples of such a linear code. They are also extremely efficient, since for every $s > 1$

and $n \leq 2^S - 1$ there exists a Reed-Solomon code with

$$d_m = n - k + 1.$$

Together with the general condition $d_m \geq 2t + 1$ mentioned earlier, which the minimum distance must satisfy for the correction of t errors, this yields $n - k \geq 2t$. Put in another way: to correct t symbol errors it is sufficient to add $2t$ parity symbols. (By 'distance' between two words we mean here the number of positions in which there are different symbols in the two words; it does not matter how many corresponding bits differ from each other within the corresponding symbols.)

In practice a less cumbersome algorithm will generally be used for error correction than the comparison with the aid of a list of all the code words, as described at the end of Sect. 3.4.1. We shall not consider the details of the algorithm here. We shall, however, try to give some idea of the manner in which error bursts are tackled with block codes. To do this we must introduce the concept of 'erasure'.

The position (i) of a particular symbol (x_i) in a transmitted code word (x) is called an erasure position if a decoder-independent device signals that the value of x_i is not reliable. This value is then erased, and in the decoding procedure the correct value has to be calculated. The decoding is now simpler and quicker because the positions at which errors can occur are known. (We assume for the moment that no errors occur outside the erasure positions.) The advantage of correcting by means of the erasures is expressed quantitatively by the following proposition:

If a code has a minimum distance d_m, then $d_m - 1$ erasures can be reconstituted.

Since the number of errors that can be corrected without erasure information is $\frac{1}{2}(d_m - 1)$ at most, the advantage of correcting by means of erasures is clear. In the Compact Disc system the value of the analog signal to be reproduced is converted at every sampling instant into a binary number of 16 bits per audio channel. For error correction the digital information to be transmitted is divided into groups of eight bits, so that in each sampling operation four information symbols (consisting of audio bits) are generated. In fact, eight parity symbols are added to each block of 24 audio symbols[4]. The calculation of the parity symbols will not be dealt with here.

3.4.3 Cross-Interleaved Reed-Solomon Code

The error-correcting code used in the Compact Disc system employs not one but two Reed-Solomon codes (C_1,C_2), which are interleaved 'crosswise' (Cross-Interleaved Reed-Solomon Code, CIRC). For code C_1 we have: n_1 = 32, k_1 = 28, s = 8, and for C_2: n_2 = 28, k_2 = 24, s = 8. The rate of the CIRC we use is $(k_1/n_1)(k_2/n_2) = 3/4$.

For both C_1 and C_2 we have $2t = n - k = 4$, so that for each the minimum distance d_m is equal to $2t + 1 = 5$. This makes it possible to directly correct a maximum of two (= t) errors in one code word or to make a maximum of four (= $d_m - 1$) erasure corrections. A combination of both correction methods can also be used.

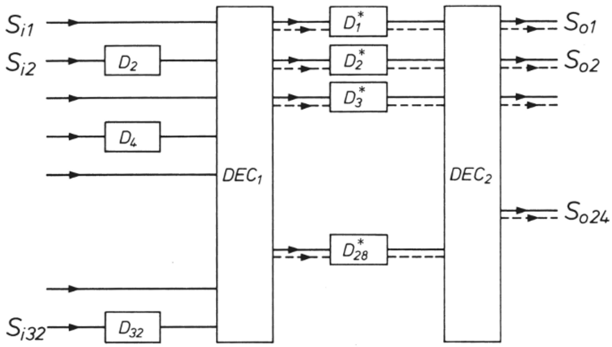


Fig. 3. Schematic representation of the decoding circuit for CIRC. The 32 symbols (S_{i1}, …, S_{i32}) of a frame (24 audio symbols and 8 parity symbols) are applied in parallel to the 32 inputs. The delay lines D_{2i} (i = 1, …, 16) have a delay equal to the duration of one symbol, so that the information of the 'even' symbols of a frame is cross-interleaved with that of the 'odd' symbols of the next frame. The decoder DEC_1 is designed in accordance with the encoding rules for a Reed-Solomon code with n_1 = 32, k_1 = 28, s = 8. It corrects one error, and if multiple errors occur passes them on unchanged, attaching to all 28 symbols an erasure flag, sent via the dashed lines. Owing to the different lengths of the delay lines D_j^* (j = 1, …, 28), errors that occur in one word at the output of DEC_1 are 'spread' over a number of words at the input of DEC_2. This has the effect of reducing the number of errors per DEC_2 word. The decoder DEC_2 is designed in accordance with the encoding rules for a Reed-Solomon code with n_2 = 28, k_2 = 24, s = 8. It can correct a maximum of four errors by means of the erasure-positions method. If there are more than four errors per word, 24 symbol values are passed on unchanged, and the associated positions are given an erasure flag via the dashed lines. S_{o1}, …, S_{o24} outgoing symbols.

Decoding circuit

The error-correction circuit[5] is shown schematically in Fig. 3; Fig. 4 is a photograph of the actual IC. The circuit consists of two decoders, *DEC*, and a number of delay lines, *D* and *D**. The input signal is a sequence of frames[6].

The 32 symbols of a frame are applied in parallel to the 32 inputs. In passing through the delay lines D_2,D_4, …., D_{32}, each of length equal to the duration of one symbol, the even symbols of a frame with the odd symbols of the next frame form the words that are fed to the decoder DEC_1. (The symbols of the frames are 'cross-interleaved'. In fact they are 'deinterleaved', because the 'interleaving'[4] has already taken place, before the information was recorded on the disc.) If there are no errors in the transmission path, the decoder DEC_1 will receive code words that correspond to the encoding rules for C_1, and it will pass on 28 symbols unchanged. DEC_1 is designed for correcting one error. If it receives a word with a double or triple error, that event is detected with certainty; all the symbols of the received word are passed on unchanged, and all 28 positions are provided with an erasure flag. The same happens in principle for events from 4 to 32 errors, but here there is a small probability ($\approx 2^{-19}$) that this detection will fail. We shall return to this probability later.

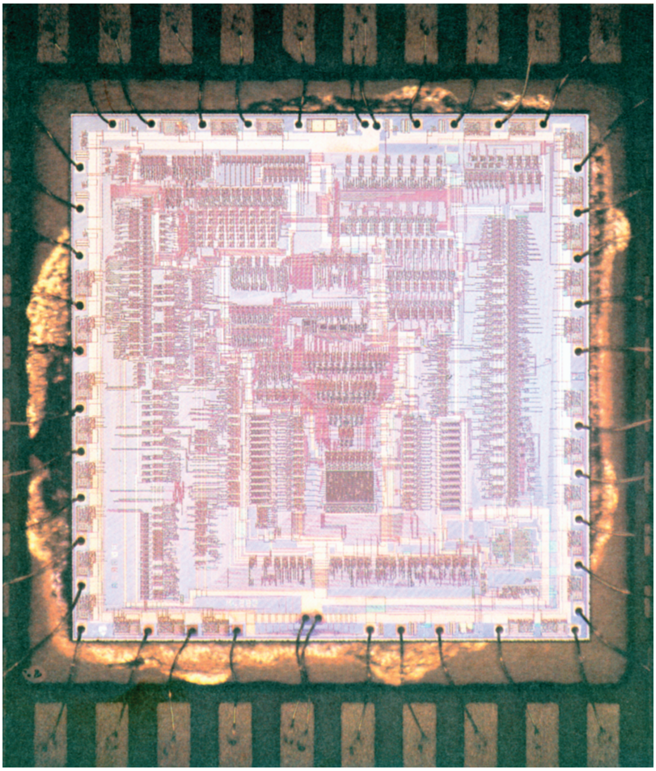


Fig. 4. The integrated circuit for error detection and correction is fabricated in n-channel MOS silicon-gate technology. It has an area of 45 mm^2 and contains about 12 000 gates.

The symbols arrive via the delay lines D_1*,..., D_{28}*, which differ from each other in length, at the input of DEC_2 in different words. If there are no errors present, DEC_2 will receive words that correspond to the encoding rules for C_2, and it will pass on 24 audio symbols unchanged. DEC_2 can correct up to four errors, by means of erasure decoding. (In the current Compact Disc system full use is not made of this facility: DEC_2 is arranged in such a way that only two errors are corrected.) If DEC_2 receives a word containing five or more errors with given erasure positions, it will pass on 24 symbols unchanged, but provided with an erasure flag at the appropriate positions; this flag has in fact already been assigned by DEC_1. A value for the erroneous samples can still be calculated with the aid of a linear interpolation.

As already mentioned, DEC_1 has been designed to allow the correction of single errors, and the detection of double and triple errors. The probability that DEC_1 will not detect quadruple or higher multiple errors is only about 2^{-19}. It may seem strange that the possibility of correcting two random errors is not utilized: in fact it would considerably increase the chance of DEC_1 failing to detect quadruple or higher multiple errors.

The probability P of quadruple or higher multiple errors passing DEC_1 without being detected can be approximated by the expression

$$P = \frac{1 + n_1(2^s - 1)}{2^{s(n_1 - k_1)}} \approx 2^{-19} .$$

The numerator contains the number of error patterns with one error or none. (The factor $(2^s - 1)$ is the number of possibilities for one bit error per symbol; such a symbol can occur at n_1 positions. The value 1 is added because zero errors can be achieved in exactly one way.) This complete expression is to be related to the number of possibilities for filling in the parity: $2^{s(n_1-k_1)}$). For proof of this equation the reader is referred to the literature[7].

When a disc is used for the recording and read-out of digital signals there are few random errors; most errors then occur as bursts. This is because the dimensions of a pit are small in relation to the most common mechanical imperfections such as dirt and scratches. It is therefore very important that multiple errors of this type cannot pass DEC_1 without being indicated with a high degree of certainty.

Since the bursts are 'spread out' over several words at the input of DEC_2, the number of errors per word hardly ever exceeds the limit value $d_m - 1 = 4$. In this way most error bursts are fully corrected.

3.4.4. Specifications of CIRC

In assessing the quality of our CIRC decoder for Compact Disc applications its ability to correct both error bursts and random errors is of great importance.

The quality characteristics for the correction of bursts are the maximum fully correctable burst length and the maximum interpolation length. The first is determined by the design of the CIRC decoder and in our case amounts to about 4000 data bits, corresponding to a track length on the disc of 2.5 mm. The maximum interpolation length is the maximum burst length at which all erroneous symbols that leave the decoder uncorrected can still be corrected by linear interpolation between adjacent sample values. This 'length' is about 12000 data bits; see the next section.

Random errors can also introduce multiple errors within one code word now and again; we shall return to this presently. The greater the relative number of errors ('bit error rate', BER) at the receiving end, the greater is the probability of uncorrectable errors. A measure for the performance of this system is the number of sample values that have to be reconstituted by interpolation for a given BER value per unit time. This number of sample values per unit time is called the sample interpolation rate. The lower this rate is at a given BER value, the better the quality of the system for random-error correction.

Aspect	Specification
Maximum *completely* correctable burst length	≈ 4000 data bits (i.e. ≈ 2.5 mm track length on the disc)
Maximum interpolatable burst length in the *worst* case	≈ 12 300 data bits (i.e. ≈ 7.7 mm track length)
Sample interpolation rate	One sample every 10 hours at BER = 10^{-4}; 1000 samples per minute at BER = 10^{-3}
Undetected error samples (clicks)	Less than one every 750 hours at BER = 10^{-3}; negligible at BER $\leqslant 10^{-4}$
Code rate	3/4
Structure of decoder	One special LSI chip plus one random-access memory (RAM) for 2048 words of 8 bits
Usefulness for future developments	Decoding circuit can also be used for a four-channel version (quadraphonic reproduction)

Table I. Specifications of CIRC.

An objective assessment of the quality of the error-correcting system also requires an indication of the number of errors that pass through unsignalled and are therefore not corrected by the system. These unsignalled and uncorrected errors may produce a clearly audible 'click' in the reproduction.

The main features of the CIRC system are summarized in Table I. Details of the calculation relating to the quality can be found in the literature[7].

3.4.5 Concealment of residual errors

The purpose of error concealment is to make the errors that have been detected but not corrected by the CIRC decoder virtually inaudible. Depending on the magnitude of the error to be concealed, this is done by interpolation or by muting the audio signal[8].

Two consecutive 8 bit symbols delivered by the decoder together form a 16 bit sample value. Since a sample value in the case of a detected error carries an erasure flag, the concealment mechanism 'knows' whether a particular value is reliable or not. A reliable sample value undergoes no further processing, but an unreliable one is replaced by a new value obtained by a linear interpolation between the (reliable) immediate neighbours. Sharp 'clicks' are thus avoided; all that happens is a short-lived slight increase in the distortion of the audio signal. With alternate correct and wrong sample values, the bandwidth of the audio signal is halved during the interpolation (10 kHz).

If the decoder delivers a sequence of wrong sample values, a linear interpolation does not help. In that case the concealment mechanism deduces from the configuration of the erasure flags that the signal has to be muted. This is done by rapidly turning the gain down and up again electronically, a procedure that starts 32 sampling intervals before the next erroneous sample values arrive. To achieve this the reliable values are first sent through a delay line with a length of 32 sampling intervals, while the unreliable values are processed immediately. The gain is kept at zero for the duration of the error and then turned up again in 32 sampling intervals. The gain variation follows a cosine curve (from 0 to 180° and from 180 to 360°) to avoid the occurrence of higher-frequency components. This also means that there are no clicks when the audio signal is muted, as in switching the player on and off, during an interval in playing or during the search procedure.

Maximum burst-interpolation length

Two associated 16 bit sample values, one from the left and one from the right audio channel, together form a sample set. If these sets were fed to the concealment circuit in the correct sequence, it would not be possible to interpolate more than one set from their reliable neighbouring sets. This would mean that in the case of an error longer than the maximum correctable burst length signal muting would very soon have to be applied.

By interleaving the sample sets it becomes possible to interpolate new sets for a given length of consecutive erroneous sets. This is done by alternating groups of 'even' sample sets with groups of 'odd' sets. Such a group, odd or even, can be interpolated from its neighbouring group or groups. The maximum burst-interpolation length is thus equal to the length of such a group. In our system we have grouped the twelve 16 bit sample values of a frame in the way shown in Fig. 5. The odd and even groups are separated by the parity values *Q*. Since these are not necessary for the reconstitution of the original signal and may therefore permissibly be unreliable, they increase the interpolation length. The maximum length with this grouping is certainly seven or even eight sample values, for some error patterns.

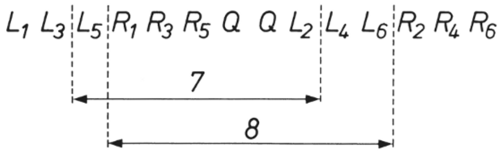


Fig. 5. Grouping of the sample values within a frame; L_i values for the left channel, R_i values for the right channel. For each sequence of seven unreliable values, new values can be calculated with certainty from reliable neighbours (e.g. if L_5 to L_2 are unreliable, the new values are interpolated from R_6 of the preceding frame and from the reliable values of the above frame). Given a favourable situation, new values can in fact be derived for eight consecutive values (e.g. values for R_1 up to L_6 from R_6 of the preceding frame, the reliable values of the above frame and L_1 of the succeeding frame).

The delay lines corresponding to D_i* (see Fig. 3) in the encoder[4] have placed eight frames between two successive sample values, after interleaving. The maximum burst length that can always be interpolated is therefore 56 frames. This presupposes, of course, that we are working with sample values consisting of two immediately consecutive symbols; the distance between all successive symbols is four frames, however. This is also the work of the delay lines D_i*.

The delay lines corresponding to D_i (again Fig. 3) in the encoder[4] now ensure, however, that this distance is alternatively three and five frames, after interleaving. The distance of five frames is responsible for a decrease in the maximum interpolation length from 56 to 51 frames. We have tacitly assumed here that the burst also comes within a block of eight frames. If we discount this assumption, there is still a reduction of a length of 1 frame - 2 symbols. The maximum burst length that can be interpolated with certainty has now become 50 frames + 2 symbols.

So far we have taken no account of random errors that can be interpolated; this is the subject of the next and final section. At this point we shall simply mention the effect of the interpolation of such errors on the maximum interpolation length.

To achieve good results in the treatment of random errors, the symbols are finally sent through a further set of delay lines Δ_i with a length of two frames. These delay lines, which serve purely and simply for 'restoring' uncorrected random errors, cause in their turn a reduction of the interpolation length by two frames. The final maximum burst length that is guaranteed capable of interpolation is thus 48 frames + 2 symbols, which corresponds to 12 304 bits.

Interpolation of random errors

If the symbols S_{oi} (Fig. 3) after the decoder DEC_2 were already in the correct sequence, a pattern of errors might arise that would rule out any possibility of interpolation, even though there were no long error bursts. This would happen if DEC_1 failed to detect an error but DEC_2 had detected it, resulting in the rejection of the entire frame at the output of DEC_2. As described in Sect. 3.4.3, however, the chance of DEC_1 failing is very small.

Since we prefer not to have to mute the audio signal, the concealment network contains a set of delay lines Δ_i, with a length of two frames, which ensure that the symbols of a single or double completely rejected frame from DEC_2 can still be interpolated from the reliable adjacent frames (see *fig. 6*). The probability that three completely rejected frames will occur within the interpolatable length determined by Δ_i is negligible.

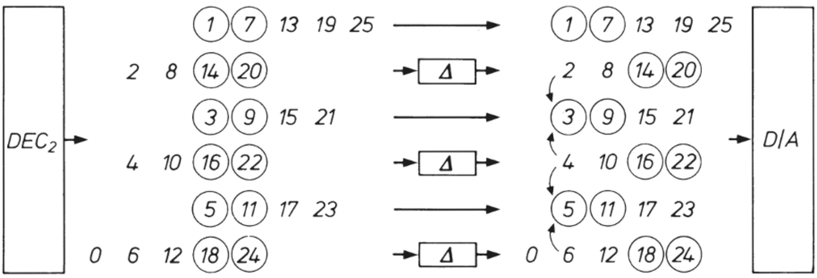


Fig. 6. The effect of the delay lines Δ_i with a length equal to the duration of two frames on the signal from the decoder DEC_2. Each number represents a sample set, and a circle around a number is an erasure flag. A frame, consisting of 24 symbols or 6 sample sets, is represented by a complete column. The succession of frames on the left in the figure (sample sets that are irrelevant in the present context have been omitted) comes direct from DEC_2 and comprises a pattern of random errors, causing the total rejection of two consecutive frames (1, 14, 3, … 11, 24). It can be seen, however, that the chosen grouping enables a new value from reliable neighbours to be interpolated for each unreliable sample set, e.g. a value for 5 from 4 and 6. After passing through the delay lines Δ_i with a length equal to the duration of two frames, the sample sets are applied in the correct sequence to the D/A converter. If a frame in the succession of frames on the right in the figure were to be completely rejected, no interpolation would be possible.

After the symbols have passed through the delay lines Δ_i, they are in the correct sequence. Most of the errors have been corrected and the signal is ready for the digital-to-analog conversion[9].

References

[1] This example is taken from S. Lin, An introduction to error-correcting codes, Prentice-Hall, Englewood Cliffs 1970.

[2] See for example F. J. MacWilliams and N. J. A. Sloane, The theory of error-correcting codes, North-Holland, Amsterdam 1978.

[3] See J. P. J. Heemskerk and K. A. Schouhamer Immink, Compact Disc: system aspects and modulation Philips Tech. Rev. **40**, 157-164 (1982), (Sect. 3.3).

[4] The calculation and addition of the parity symbols take place in the encoding circuit *PAR COD* in fig. 2 of the article of note [3]. Delay lines are used for interleaving the audio and parity symbols.

[5] This circuit corresponds to the *ERCO* chip in fig. 3 of the article by M. G. Carasso, J. B. H. Peek and J. P. Sinjou, Philips Tech. Rev. **40**, 151-155 (1982), (Sect. 3.2).

[6] In fig. 9 of [3] a frame of this kind is represented by the bit stream B_2, from which the C&D symbol has already been removed.

[7] L. M. H. E. Driessen and L. B. Vries, Performance calculations of the Compact Disc error correcting code on a memoryless channel, in: 4th Int. Conf. on Video and data recording, Southampton 1982 (IERE Conf. Proc. No. 54), pp. 385–395.

[8] Error concealment takes place in the *CIM* chip in Fig. 3 of the article of note [5].

[9] See D. Goedhart, R. J. van de Plassche and E. F. Stikvoort, Digital-to-analog conversion in playing a Compact Disc, Philips Tech. Rev. **40**, 174-179 (1982), (Sect. 3.5).

3.5 Digital-to-analog conversion in playing a Compact Disc

D. Goedhart, R.J. van de Plassche, E.F. Stikvoort

Ing. D. Goedhart is with the Philips Audio Division, Eindhoven; Ir R. J. van de Plassche and Ir E. F. Stikvoort are with Philips Research Laboratories, Eindhoven.

Abstract
The 16 bit words from the error-correcting circuit are converted into an analog signal by a 16 bit conversion system. This system consists of a digital transversal filter, in which the signal is oversampled 4 times (sampling rate 176.4 kHz) and then filtered in such a way that signals at frequencies above 20 kHz are attenuated by 50 dB after digital-to-analog conversion. The filter is followed by a noise shaper, which rounds off to 14 bits with negative feedback of the rounding-off error of the preceding sample. Next there is a 14 bit digital-to-analog converter, which is followed by a low-pass third-order Bessel filter. The signal-to-noise ratio of the complete system is about 97 dB. Even though the lowpass filter has a sharp cut-off the system is phase linear. The entire system, except for a few operational amplifiers, is contained in three integrated circuits; one for the digital filter (for both of the stereo channels) and two for the two digital-to-analog converters.

3.5.1 Introduction

The last stage in the series of operations on the signal in the Compact Disc system is the return from the digital code to the analog signal, which has the same shape as the acoustic vibration that was picked up by the microphone.

After decoding and error correction the digital signal has the form of a series of 16 bit words. Each word represents the instantaneous numerical value of the measured sound pressure in binary form, and is therefore a sample of the acoustic signal. There are 44 100 of these words per second.

The digital-to-analog converter in the Compact Disc player generates an electric current of the appropriate magnitude for each word and keeps it constant until the next word arrives. The electric current thus describes a 'staircase' curve that approximates to the shape of the analog signal (Fig. 1a). In terms of frequency, the steps in the staircase represent high frequencies, which extend beyond the band of the analog audio signal (20 Hz - 20 kHz). These high frequencies have to be suppressed by a lowpass filter; in the Compact

Reprinted with permission from Philips Tech. Rev. **40**, 174-179, 1982.

Disc player their level should be reduced to at least 50 dB below that of the maximum audio signal.

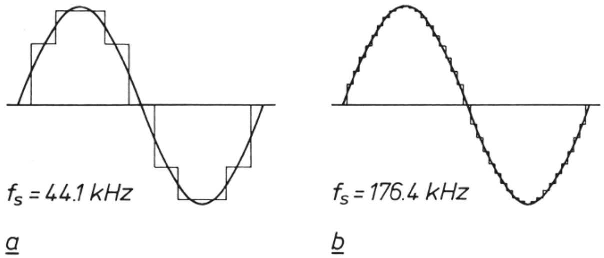


Fig. 1. A sinusoidal signal at 4.41 kHz sampled with a sampling rate f_s of 44.1 kHz (*a*) and with a frequency four times higher (*b*). In (*b*) the 'staircase' curve approximates more closely to the analog waveform, and the high frequencies present in the staircase signal are more easily filtered out.

If this high attenuation of the frequencies above the audio band is to be achieved solely with an analog lowpass filter, the filter must meet a very tight specification. It was decided to avoid this problem in the Philips Compact Disc player by introducing a filter operation, earlier in the digital stages. This was done by 'oversampling' by a factor of four: a digital filter, operating at four times the sampling rate (4 × 44.1 kHz = 176.4 kHz) delivers signal values at this increased frequency, thus refining the staircase curve (Fig. 1b) and making it easier to filter out the high frequencies. As a result it is possible to make do with a relatively simple lowpass filter of the third order after the digital-to-analog conversion.

The conversion of the 16 bit words into an analog signal is performed in the Philips Compact Disc player by a 14 bit digital-to-analog converter available as an integrated circuit and capable of operating at the high sampling rate of 176.4 kHz. Partly because of the fourfold oversampling and partly because of the feedback of the rounding-off errors in antiphase, rounding off to 14 bits does not result in a higher noise contribution in the audio band. This remains at the magnitude corresponding to a 16 bit quantization (signal-to-noise ratio about 96 dB), so that even though there is a 14 bit digital-to-analog converter it is still possible to think in terms of a 16 bit conversion system.

In comparison with direct 16 bit digital-to-analog conversion, which must be followed by a lowpass filter with a sharp cut-off to give sufficient suppression of signals at frequencies above 20 kHz, our conversion system has a number of advantages. The first is the linear phase characteristic, which can be obtained with a digital filter, but not with an analog filter; the second is a filter characteristic that varies with the clock rate and is therefore insensitive to variation in the speed of rotation of the disc. Finally, because the quantization steps are smaller, the maximum 'slew rate' that these circuits must be able to process is lower (the slew rate is the rate of variation of output voltage). There is therefore less chance of intermodulation distortion because the permitted slew rate has been exceeded.

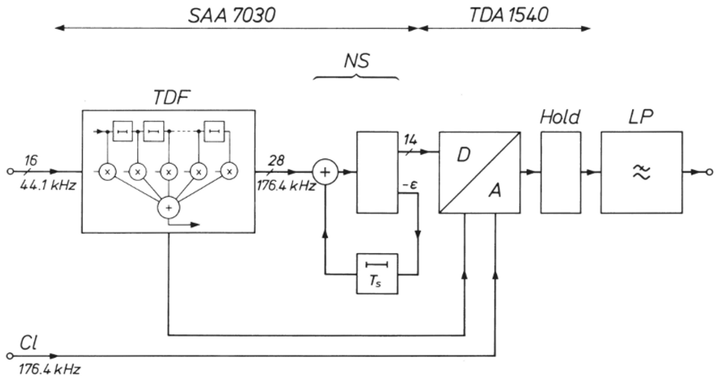


Fig. 2. Block diagram of the digital-to-analog conversion. TDF digital transversal filter which brings the sampling rate of 44.1 kHz to 176.4 kHz and attenuates signals in the bands around 44.1 kHz, 88.2 kHz and 132.3 kHz. NS noise shaper in which the rounding-off error is delayed by one period Ts after rounding-off to 14 bits and then fed back in the opposite sense. D/A 14 bit digital-to-analog converter. Hold hold circuit. Cl clock signal. LP lowpass 3rd-order Bessel filter.

The entire series of operations in the digital-to-analog conversion is shown as a block diagram in Fig. 2. The oversampling takes place in the digital filter TDF to which the input signal is fed. The filter output signal is then rounded off to 14 bits, and the rounding error is fed back in the opposite sense in the noise shaper NS. The digital filter and noise shaper are located in a single integrated circuit in NMOS technology (type SAA 7030). This IC processes both stereo channels. Then follow the digital-to-analog converter D/A and a hold circuit, combined in a single IC type (TDA 1540) in bipolar technology; for each stereo channel there is a separate IC. The analog signal finally passes through a lowpass filter.

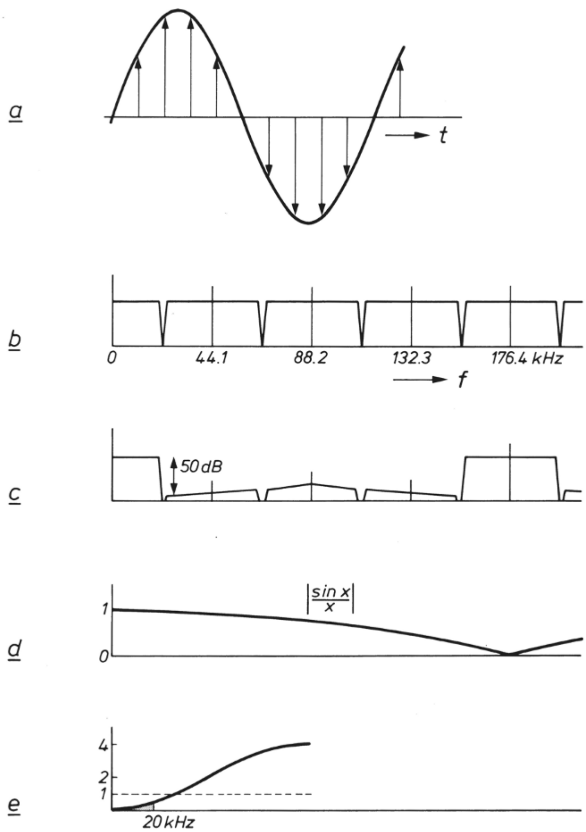


Fig. 3. a) A train of periodic pulses that sample an analog signal waveform. b) Frequency spectrum of such a pulse train. The pulse repetition frequency is 44.1 kHz, the sampled signal occupies the audio frequency band (0-20 kHz). c) Frequency spectrum for over-sampling and filtering of the same signal at 176.4 kHz. It is now much easier to filter out the frequencies above the audio band. d) A hold circuit after the digital-to-analog converter keeps a signal sample at the same value until the arrival of the next sample. The frequency spectrum in c is thus multiplied by the function |(sin x)/x| with a first zero at 176.4 kHz. e) Noise spectrum after the noise shaper. In the audio range of interest the noise is considerably attenuated compared with the flat noise spectrum (dashed line) that would be obtained without noise shaping.

3.5.2 Suppression of frequencies above the audio band

Direct digital-to-analog conversion of the presented signal provides a series of analog signal samples (Fig. 3a). These have the form of pulses that - in theory - are infinitely short, but have a content (duration times amplitude) corresponding to the sampled signal value. The repetition frequency is 44.1

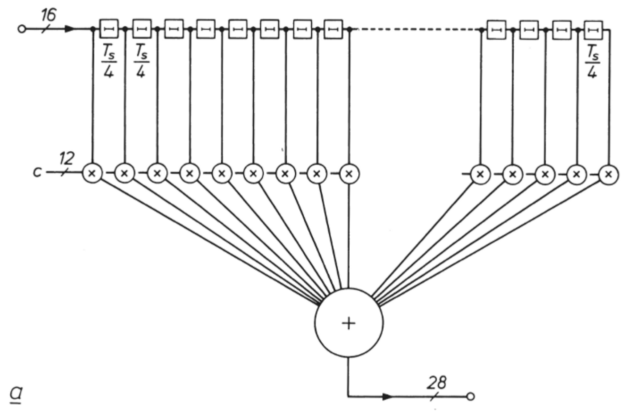
kHz. The frequency spectrum of such a series is illustrated in Fig. 3b[1]. In theory it is infinite; above the baseband 0-20 kHz can be seen integral multiples of the sampling frequency with their left-hand and right-hand sidebands. Between these bands there are transition regions, the first for example being between 20 kHz and 24.1 kHz.

This entire spectrum must not be passed on to the player amplifier and loudspeaker. Even though the frequencies above 20 kHz are inaudible, they would overload the player amplifier and set up intermodulation products with the baseband frequencies or possibly with the high-frequency bias current of a tape recorder. Therefore all signals at frequencies above the baseband should be attenuated by at least 50 dB.

To produce such an attenuation, an analog filter after the digital-to-analog converter will inevitably have to contain a large number of elements and require trimming. In addition a linear phase characteristic is required in the passband so that the waveform of pulsed sound effects will not be impaired. In the Philips Compact Disc player these requirements are met in a different way, by means of:

- fourfold oversampling of the signal in the digital phase,
- a digital filter operation,
- a hold function after the digital-to-analog conversion,
- a third-order Bessel filter in the analog-signal path.

A digital transversal filter is used for the filtering after oversampling. To understand the operation of the filter, we can think of it as consisting of 96 elements (Fig. 4a), while the delay in each element is $(176.4 \times 10^3)^{-1}$ s, i.e. a quarter of the sampling period or ¼Ts. Four times in each period the filter takes up new data. At three of these four times the content of this data is zero, since the oversampling is done by the introduction of intermediate samples of value zero. This means that only 24 of the 96 elements are filled at any one time. The contents of each element are multiplied by a coefficient c. The filter provides data at a rate of 176.4 kHz; each number is the sum of 24 non-zero multiplications. In this way the filter always calculates three new sample values at the locations of the zero samples.



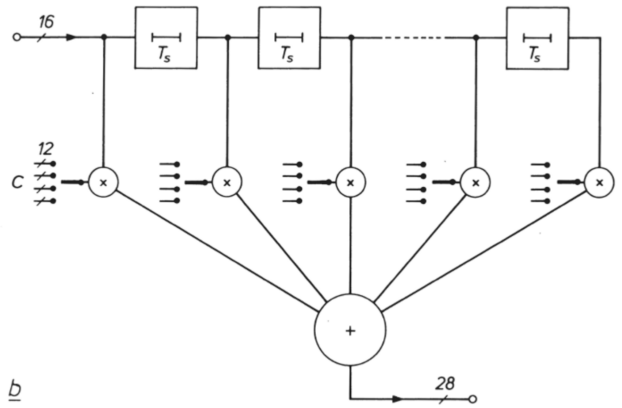


Fig. 4. Digital transversal filter. a) Filter consisting of 96 elements. A 16 bit word remains in each element for a quarter of the sampling period Ts. Since a new 16 bit word is only offered once per Ts, three-quarters of the elements are filled by the value zero. During the period Ts there are four multiplications by the 96 coefficients c; only 24 multiplications produce a product different from zero. These products are summed; in this way an output is provided four times in each sampling period, i.e. at a frequency of 4 × 44.1 kHz = 176.4 kHz. This means that there is a fourfold oversampling. b) An equivalent circuit that has been used in practice instead of (a) because it has 24 delay lines and multipliers instead of 96.

The practical version of the filter is in fact some-what different from the version referred to in the above explanation. In practice the filter consists of only 24 delay elements and a 16 bit word remains in each element for a time Ts (Fig. 4b). During this time Ts the word is multiplied four times by a coefficient c, which is different for each multiplication. The products are also summed four times during the time Ts and passed to the output. The frequency at which these summated values appear at the output is therefore 4/Ts = 176.4 kHz) again.

The coefficients are numbers with 12 bits. Each product has a length of 16 + 12 = 28 bits. The numbers have been chosen in such a way that the summation of 24 products does not introduce extra bits, so that the filter output consists of 28 bits with no rounding off.

The frequency spectrum of the oversampled and filtered signal is shown in fig. 3c. It can be seen that the bands in this spectrum around 1 ×, 2 × and 3 × 44.1 kHz are suppressed.

The digital-to-analog converter generates a current whose magnitude is proportional to the last digital word presented. This current is kept constant in a hold circuit until the next sample value is delivered, producing the staircase curve mentioned above. The signal samples have thus in theory changed from infinitely short pulses to pulses with the duration of a sampling period. This also has consequences for the frequency spectrum; the spectrum in Fig. 3c is multiplied by a curve of the form |(sin x)/x| that has a first zero at 176.4 kHz (see Fig. 3d). This gives an attenuation of signals in the 20 kHz sidebands on either side of 176.4 kHz by more than 18 dB. The hold effect causes no phase distortion.

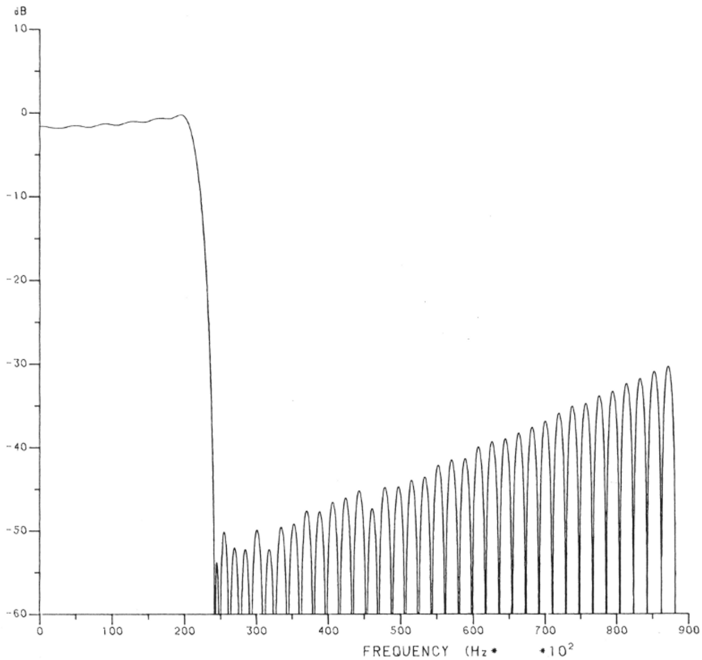


Fig. 5. Computer calculation of the detailed passband characteristic of the digital transversal filter. This has a small overshoot at the highest audio frequencies, which is used to compensate for the slight attenuation produced here by the curve in Fig. 3d and the analog Bessel filter. A very sharp lowpass cut-off of 50 dB is obtained. The irregularity in the suppressed band is caused by rounding-off the filter coefficients to 12 bits.

The attenuation is still not sufficient, however. As a supplement, a lowpass Bessel filter of the third order is used, which has its −3 dB point at 30 kHz. The Bessel type of filter has been selected because of its linear phase characteristic in the passband. This filter is simple and requires no highly accurate elements.

The hold function and the Bessel filter introduce some slight attenuation at the top of the passband. The digital filter is designed to correct this with a small overshoot (Fig. 5).

3.5.3 Suppression of the quantization noise

The presented signal, quantized to 16 bits, will contain some noise on conversion into an analog signal. This reproduces the errors due to the quantization in fixed steps. The root-mean-square value of the noise voltage in the sampled frequency band is $q/\sqrt{12}$, where q represents the magnitude of the quantization step. We see that when the quantization step is doubled, i.e. coding with one bit less, the noise voltage is also doubled, or, in other words, the noise level rises by 6 dB.

The samples that leave the filter at a repetition frequency of 176.4 kHz describe a signal with a band-width of 88.2 kHz. The quantization noise added due to the subsequent rounding off to 14 bits is spread over this band. With a signal of sufficient amplitude and a sufficiently broad frequency spectrum this distribution is uniform, since the quantization errors for successive samples are in principle uncorrelated; the quantization noise is ‘white’ noise. Only the band from 0 to 20 kHz is relevant; this is only about a fourth part of the sampled band, and the noise power in the band from 0 to 20 kHz is therefore only a fourth part of the total noise power. This means that because of the fourfold oversampling the signal-to-noise ratio in the relevant frequency band is 6 dB better than would be expected with 14 bit quantization. It is thus about 90 dB, which is what would have been obtained with a 15 bit system without over-sampling.

In rounding off from 28 to 14 bits it is useful to compare successive rounding-off errors. If the analog signal is a direct voltage, successive samples will have the same rounding-off error. The audio signal will not contain any direct current; it will however contain slowly varying signals that will resemble a direct current in a short time interval. If the error produced in the rounding-off from 28 to 14 bits is now changed in sign and added to the next sample to arrive (see Fig. 2), the average quantization error for slowly varying signals - i.e. low frequencies - can be reduced. This appears in the shape of the frequency spectrum of the quantization noise (see Fig. 3e); at low frequencies the noise level is lower, at high frequencies it becomes higher. With a sampling rate of 176.4 kHz, it follows that a 7 dB gain in signal-to-noise ratio is obtained in the audio band (0-20 kHz). The ratio of the maximum signal to the noise contributed by the entire digital-to-analog conversion system described above is thus brought to about 97 dB, i.e. the value corresponding to a 16 bit quantization.

3.5.4 The digital-to-analog converter

The 14 bit digital-to-analog converter has been dealt with in detail elsewhere[2]. Here we shall only indicate how it differs from other digital-to-analog converters.

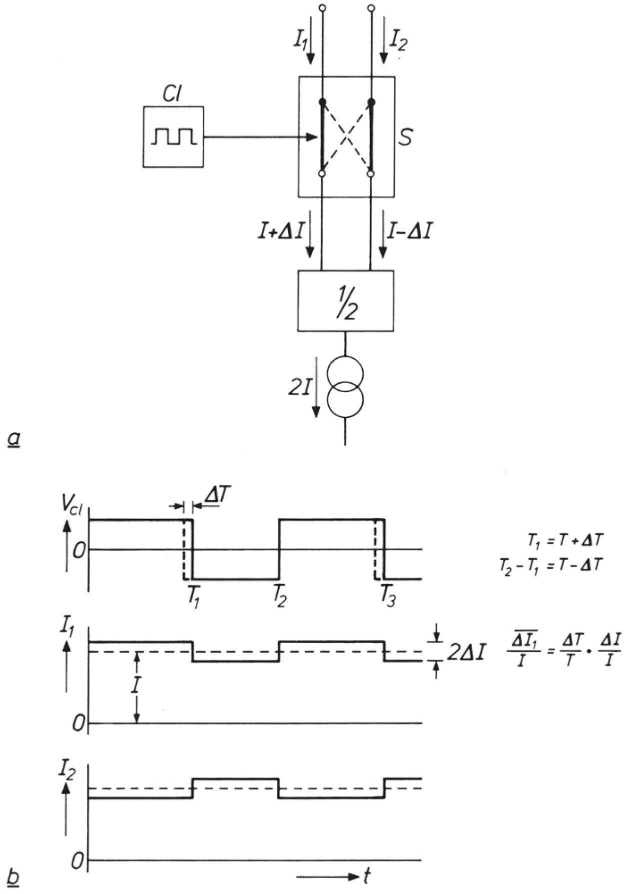


Fig. 6. a) Division of a current 2*I*. *Cl* clock generator. *S* switches for periodically interchanging the two half-currents. *b*) The output currents I_1 and I_2 as a function of time *t*. Their mean value is the same. A difference between the mean output currents can be caused by an asymmetry ΔT of the clock signal V_{cl}. This difference is however an order of magnitude smaller than ΔI.

A characteristic feature is the way in which currents are generated that are accurately related by a factor of 2; a digital-to-analog converter requires a set of such currents. The exact ratio is obtained by periodically interchanging the currents that are derived by dividing down by two from a constant reference current (see Fig. 6), so that small differences are averaged out. This system is known as 'dynamic element matching'. Accurate division by four can be carried out with a slightly more complicated circuit, also based on periodic interchange. The full series of current dividers is shown in fig. 7. Here Cl is the clock signal that controls the periodic switching; only for the four least-significant bits are the currents obtained from a passive division by means of differences in emitter area.

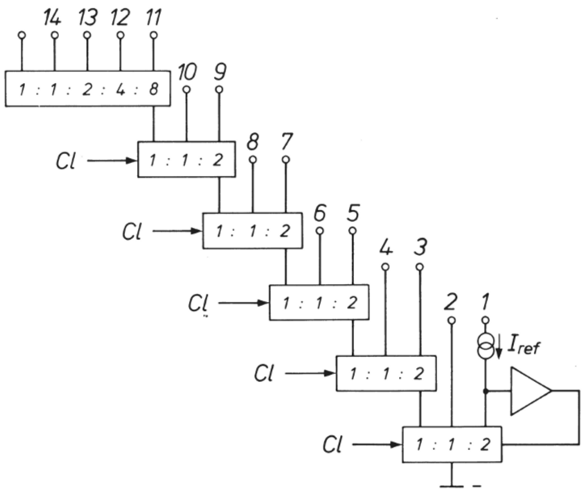


Fig. 7. Cascade of current dividers in the 14 bit digital-to-analog converter TDA 1540. The starting point is the reference current Iref. Currents that are accurately equal to a half and a quarter of the input current are obtained in the divider stages by periodic interchanges; the clock signal Cl controls these interchanges. Only the four least-significant bits 11 … 14 are obtained by passive division.

Fig. 8 shows the complete switching diagram of the 14 bit digital-to-analog converter. The cascade of divider stages can be seen in the figure. The ripple caused by the periodic switching is smoothed at the seven most significant bits by an RC filter; the seven capacitors (above in Fig. 8) are externally connected.

The nonlinearity of the digital-to-analog converter is extremely low: between −20°C and +70°C it is less than 3×10^{-5}, or half the least-significant bit. The TDA 1540 integrated circuit is followed by the low-pass Bessel filter of the third order, and the analog signal appears at the output.

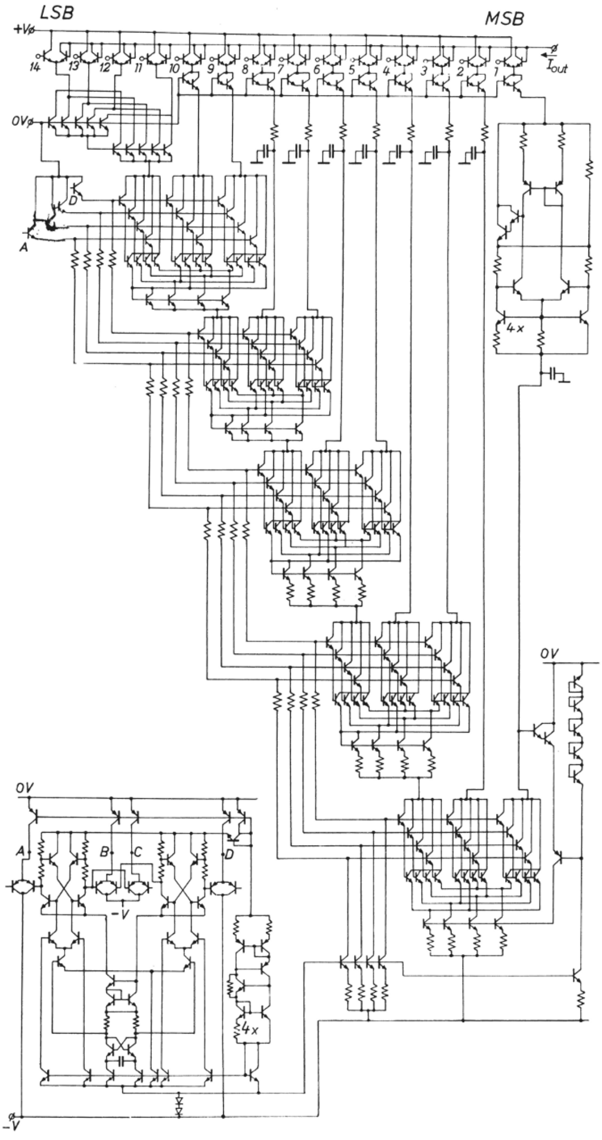


Fig. 8. Complete circuit diagram of the 14 bit digital-to-analog converter. The cascade of current dividers in Fig. 7 can be identified here. The capacitors (above), which smooth out the ripple on the divider-output currents, are external. Bottom left: The clock generator.

References

[1] The principles of the digital processing of audio signals are explained in a very readable account by B. A. Blesser in Digitization of audio: a comprehensive examination of theory, implementation, and current practice, J. Audio Engng Soc. **26**, 739-771 (1978).

[2] R. J. van de Plassche and D. Goedhart, A monolithic 14-bit D/A converter, IEEE J. SC-14, 552-556 (1979).

3.6 Compact Disc (CD) Mastering - An Industrial Process

W. Verkaik

Philips Electro-Acoustics Division, Optical Disc Mastering, Eindhoven, The Netherlands

Abstract
Compact Disc (CD) mastering is a process in which digital audio and subcode information is encoded into the standard CD format and recorded on a disk surface. The information is contained in pits of discretely varying lengths arranged in a spiral.
The disk-mastering process lies between tape mastering and replication. It involves the application of thin photoresistant layers onto glass substrates, encoding and recording the audio and subcode information, and developing and testing to generate the required pit dimensions (pit geometry).
The parameters influencing the pit geometry and other quality parameters of masters are many, and the process requires a specific philosophy and discipline to be performed industrially. This philosophy and the resulting equipment, operating requirements, quality control, and test methods are described.

3.6.1 Introduction

The introduction of Compact Disc (CD) digital audio signifies a new era in sound technology. The CD sets new standards in reproduction quality, impossible to achieve with traditional sound reproduction techniques. These standards, combined with the disk's compact size-both sides of a full LP on one side of a 120-mm diameter disk-make it a vital contribution to the future of commercial audio.

With CD digital audio, program origination and replication techniques show certain similarities to those for normal LPs. However, the mastering process is completely different. It is a process that Philips has developed against a considerable background of experience, gained with the LaserVision optical disk[1]. The LaserVision mastering technology has now led to the introduction of second-generation mastering equipment, specifically dedicated to the Philips Compact Disc.

Courtesy Audio Engineering Society (www.aes.org). Reprinted with permission from:
Digital Audio: AES Premiere Conference, Rye, New York, 1982, 189-195.

3.6.2 The production of CDs

The production chain from original sound recording to finished disks can be divided into the following stages (Fig. 1):

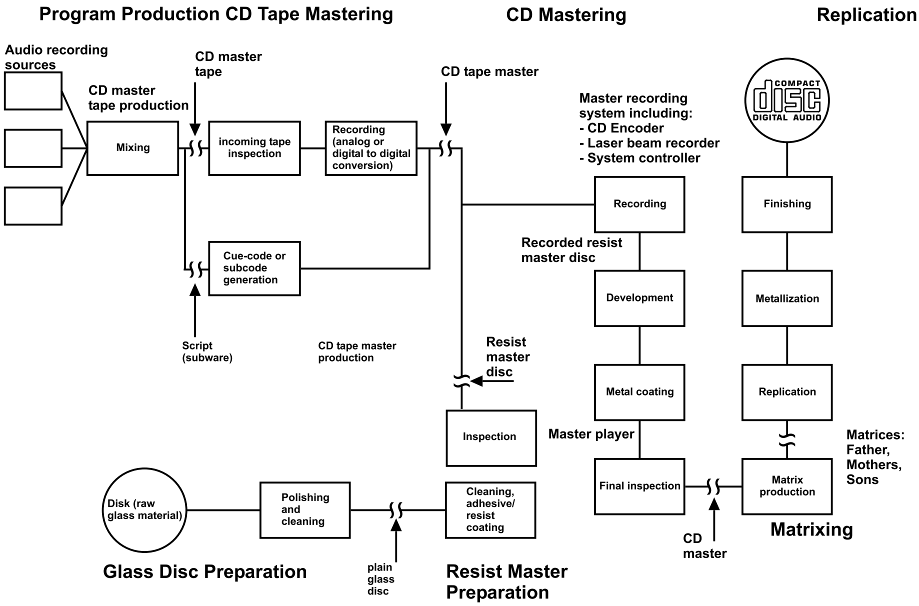


Fig. 1. Block diagram of CD production

1) *Program Production.* Here the original sound is recorded, mixed, and transcribed to generate the CD master tape, either analog or (preferably) digital, carrying the desired two stereo audio channels.
2) *CD Tape Mastering.* The Master tape at this stage is converted (analog to digital, or, if necessary, digital to digital), and the CD subcode information is generated and recorded (possibly in the form of cue codes) on the CD tape master. This tape master fulfills the requirements as specified in [2] and is the standard carrier of the CD digital audio information.
3) *CD Disk Mastering.* In this process the information from the CD tape master is encoded into the CD standard format and recorded (cut) on the surface of a photoresist-coated glass disk, the CD resist master disk. The result of this process is the CD disk master, the first disk-shaped carrier of the CD standard information contained in a vast number of pits arranged in a continuous spiral. This surface structure determines to a large extent the basic parameters of CDs and is optimized toward subsequent mass replication.

4) *Matrixing and Replication.* By galvanic processing the CD disk master surface is transferred onto a nickel shell (father) which, by the same process, can generate a number of positives (mothers). Each mother can generate a number of negatives (sons or stampers), which, after adequate processing, are used in replication. By compression or injection molding the stamper surface, information is pressed into a transparent plastic carrier, which after aluminum mirror coating (for reflection), protective lacquer coating, and label printing forms the final CD.

3.6.3 Disk mastering

The process steps involved in the disk-mastering process are the following:

1) *Glass Disk Preparation*. The glass substrate required to enter the mastering process is made by grinding, polishing, and cleaning. This substrate, which is standardized with regard to dimensions, clamping possibilities, and surface quality, is called “plain glass disk” and is effectively manufactured and distributed by a glass factory.
2) *Resist Master Preparation.* The plain glass disk enters the process area of the mastering facility. This is a clean room, climatically controlled, with a dustfiltering class of 10 000. In certain areas, where necessary, the equipment has dust filtering class 100 and facilities for the exhaust of chemical vapors. The disk is first visually checked for the minutest imperfection. It is then introduced into the resist master preparation system. Passing through the system, the disk is first thoroughly cleaned. It then receives an adhesive layer, followed by a coat of photoresist, after which careful inspection is carried out. The inspected disk is then placed in a special cassette, cured in an oven, and held in the store. The CD resist master disk, which has a shelf life of several weeks, is now ready for disk mastering.
3) *Recording and Developing.* Recording takes place in the recording area, which is very moderately controlled (class 100 000). With the CD tape master prepared, a CD resist master disk is taken from the store and passed to the CD master recording system. The system comprises a laser beam recorder with its own dust filtering of class 100, a system controller, encoder, and digital tape recorder. The signals from the CD tape master are recorded by the laser beam recorder, which exposes the CD resist master disk according to the CD tape master’s content. A well planned facility will be designed for expansion, to include a second CD master recording system. This permits a doubling of output, without the need for extra facilities in the resist master preparation system. After recording, the exposed CD resist master disk is returned in its special cassette to the process area. There it passes through a developing

and evaporating stage. The latter imparts a silver coating, which permits inspection and subsequent galvanic processing prior to the replication process. The CD disk master is now ready for inspection and testing.

4) *Quality Control.* At every stage in the process, quality inspection on samples is undertaken with equipment installed in clean sections, which have dust class 100. Final testing of the CD disk master is carried out by the master player system, which permits playing the CD disk master. The readout signals are relayed to a silent room for audio assessment. The system also permits the testing of signals which determine other quality aspects of the recording and the status of the mastering process. Prior to passing on to the matrixing department for further processing, there is a visual and microscopic final inspection.

3.6.4 The readout mechanism

To understand the effects of the basic dimensions of the pits formed in the mastering process on the CD system quality; it is worthwhile to give an elementary model of the readout mechanism.

The information contained in the discretely varying length of the pits is read out by a focused laser beam in the CD player. The size and the energy distribution of the laser spot hitting a pit in the information surface are illustrated in Fig. 2.

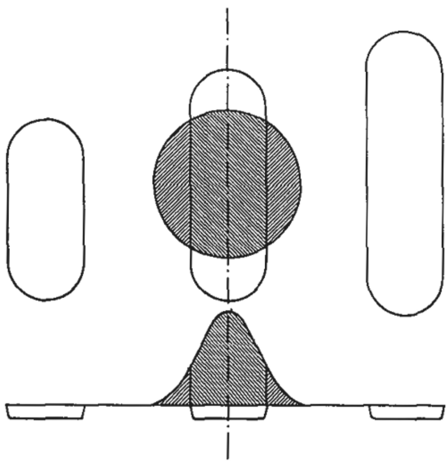


Fig. 2. Optical readout.

The light reflected from this surface is influenced by the presence of a pit and measured. This signal has to yield both the high-frequency signal containing all audio and subcode information (Fig. 3), and the radial tracking signal forming a servo signal for track following (Fig. 4).

The optimum high-frequency signal is achieved if the presence of a pit results in a total loss of reflected intensity. This situation occurs when the pit depth equals one-quarter of the apparent wavelength of the light, while the pit width is such that the intensity of the light reflected from the bottom of the pit equals the intensity of the light reflected from the surface (shaded areas in Fig. 2). In that case destructive interference will take place. Since the size and shape of the readout spot are standard in the CD system, there will be only one pit depth, and also one pit width, fulfilling the former requirement.

The optimum radial tracking signal unfortunately is not achieved at the same depth and width. On the contrary, an optimum signal is achieved when the pit depth equals one-eighth of the wavelength of the light.

Therefore a very carefully chosen compromise concerning the basic pit dimensions governs the mastering process, in which an optimum situation is specified yielding:

1) An acceptable high-frequency signal
2) An acceptable radial tracking signal
3) Mass-replicable structures
4) Minimum sensitivity to unwanted process parameters

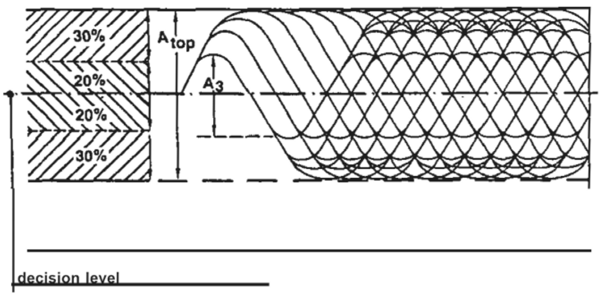


Fig. 3. High-frequency signal.

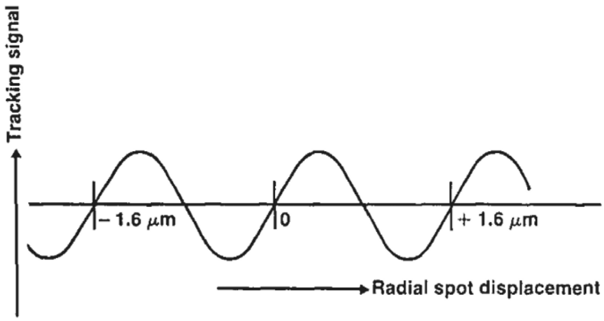


Fig. 4. Characteristic of radial differential signal.

3.6.5 Generation of pits

There are three steps involved in the generation of pits: 1) encoding, 2) recording, and 3) developing.

3.6.5.1 Encoding

The digital audio information read from the tape master and the subcode information generated by the subcode processor are fed into the professional CD encoder, the principle of which is shown in Fig. 5.

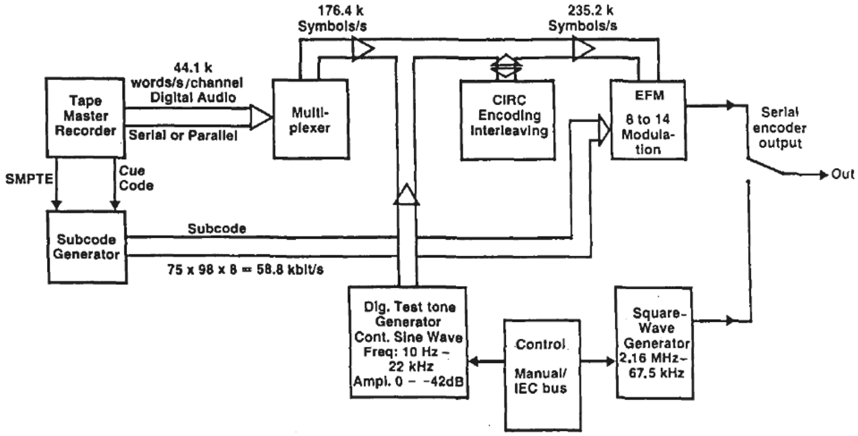


Fig. 5. CD encoder principle. Features: self-testing, decoding to digital audio optional. Serial encoder output: Audio-235.2 ksymb/s, 17 bits; Synchr-7.35 ksymb/s, 27 bits; Subcode-7.35 ksymb/s, 17 bits.

Apart from the multiplexing, coding according to the cross interleave Read-Solomon code (CIRC), and modulation according to the eight-to-fourteen modulation (EFM) principle, this encoder has facilities for the generation of test signals (sine waves and square waves adjustable in frequency and amplitude).

3.6.5.2 Recording

The serial encoder output (high-frequency signal) is connected to the driver of the acousto-optical (AO) modulator in the lightpath of the CD laser beam recorder. The optical configuration of this recorder is shown in Fig. 6. The light beam of the argon-ion laser is modulated by the AO modulator under control of the high frequency signal. The modulated laser beam, after passing various optical elements, is projected onto the objective lens, which focuses the laser beam on the surface of the resist master disk. By very accurately rotating the resist master disk and simultaneously translating the objective lens assembly, the focused recording spot will intermittently illuminate the photosensitive layer in a spiral fashion. The focusing of the objective lens on the moving resist master disk surface requires an active focusing servo system, comprising a primary focusing system using a separate diode laser beam and a secondary focusing system using part of the reflected light of the recording spot for fine tuning. The spot can be constantly monitored on a TV monitor. Alignment of the optical configuration after laser replacement, or maintenance, is greatly facilitated by the microcomputer-controlled beam-positioning facilities.

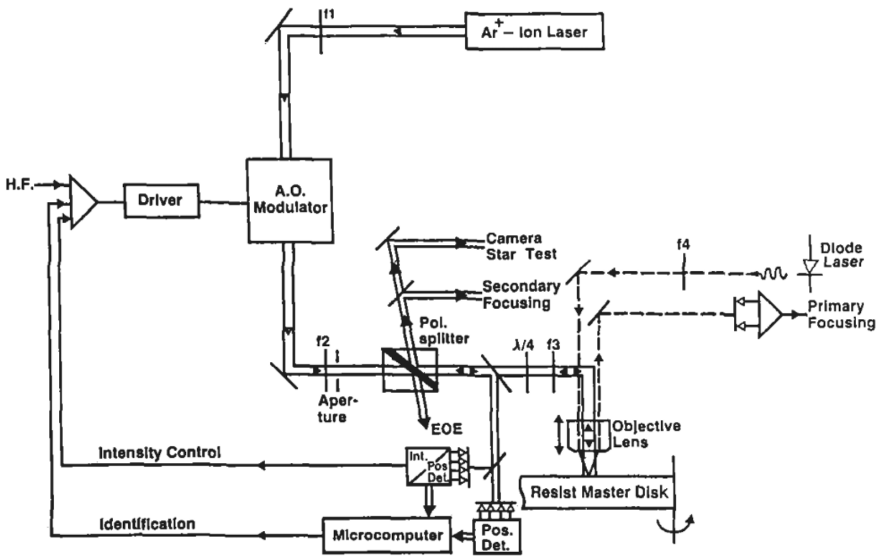


Fig. 6. Optical configuration of CD laser beam recorder.

3.6.5.3 Developing

The exposed resist master disk is developed in a developer system, where the rotating master is subjected to a flow of developing fluid, which is selectively etching away the illuminated portions of the photoresist. This etching process continues until the glass surface is reached and is terminated when the desired pit geometry is achieved. The progress of the pit formation is constantly monitored by measuring the zero-order and first-order diffracted intensities of the laser beam projected through the master.

3.6.6 Pit geometry

The following dimensions are related to the pit geometry: pit length, pit depth, pit width, slopes, and track pitch.

The pit length is dictated by the digital high-frequency signal to the optical modulator. Fig. 7 shows the correspondence between such a typical digital pattern and the resulting pit lengths. At a recording speed of 1.2 m/s this means that the pit length can have values between 0.833 and 3.054 µm, with minimum increments of 0.278 µm. The accuracy with which the pit length must be controlled must be one order of magnitude smaller than the smallest increment, such as ±30 nm. The pit depth depends on the thickness of the photoresist layer. During developing the exposed photoresist is etched away until the glass substrate is reached. The thickness and the homogeneity of the resist layer depend on the performance of the resist master preparation system and are crucial parameters in the mastering process.

Serial Input to Optical Modulator:

Audio:	235.2	ksymb/sec	17 bits
Subcode:	7.35	ksymb/sec	17 bits
Synch.:	7.35	ksymb/sec	27 bits

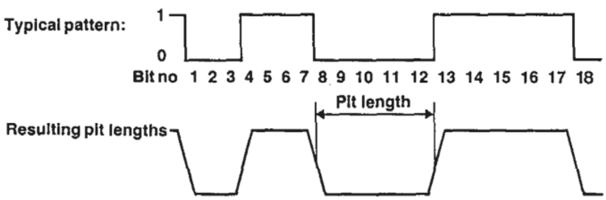


Fig. 7. CD pit length generation. Pit length increments at V_{lin} = 1.2 m/s-0.278 µm; minimum pit length (3 increments)-0.833 µm; maximum pit length (11 increments)-3.054 µm; tolerance-±30 nm. Serial input to optical modulator: Audio-235.2 ksymb/s, 17 bits; Subcode-7.35 ksymb/s, 17 bits; Synch-7.35 ksymb/s, 27 bits.

The pit width and slopes depend on the focused recording laser spot size and intensity distribution, together with the developing process. Depth, width, and slopes are carefully chosen to achieve the optimum as discussed in Sect. 3.6.4.

The track pitch is the distance between successive tracks on the disk and has the specified value of 1.6 μm. During mastering this track pitch depends on the rotational velocity of the resist master disk and the translational speed of the sledge carrying the focusing assembly. The specified track-pitch accuracy demands very stable and sophisticated control systems in the laser beam recorder.

3.6.7 Test parameters and methods

As has been shown in the previous paragraphs, the pit geometry is of basic importance for the quality of masters. Direct measurement of these dimensions is possible by means of electron microscopy, but this method is destructive for the test item, is very time consuming, and gives only a local indication. For routine measurements to ascertain master quality, the pit geometry is measured by playing the master on a master player and deriving test signals from the readout high-frequency signal. Track pitch and track-form stability are measured by monitoring the radial tracking signal during playback of the master.

Also a scan is made of information layer defects by counting appropriate indications and flags derived from the demodulating circuitry.

Finally (the "proof of the pudding is in the eating"), a master is released only after assessment of the total audio program quality and subcode integrity.

All these measurements are performed during the same real-time playback test session, using the specially designed CD master player system, which can also be used to perform similar measurements on stampers and replicas.

An additional aspect of the master quality is processability, which means the master's fitness to be processed in the subsequent matrixing department.

All test parameters and methods are summarized in Tables 1 and 2.

Table 1. Test parameters

Pit geometry
Carrier-to-noise ratio (CNR)
Surface noise
Symmetry
Phase depth
Track pitch
Track-form stability
Information layer defects
Block error rate (BLER)
C1,2 flags
Interpolations
Mutes
Overall program assessment
Audio signal quality
Ticks and clicks
Audio channel phase relation
Subcode
Processability
Metal coating
Scratches
Stains
Dust
Fibers

Table 2. Test method.

Master player with
Spectrum analyzer
Oscilloscope
Audio amplifier
Headphones
Loudspeakers in silent room
Subcode reader
Counters and chart recorders
Microscopes
Film viewer
Naked eye

3.6.8 Quality characteristic sourcing

Figure 8 shows how the important CD system performance parameters are influenced by the successive processes of disk making. The first column indicates the specified system performance characteristics. These characteristics are determined by the qualities of both the CD player and the disk. Since in this context our attention is focused on the disk production chain, in column 2 the corresponding disk parameters are indicated assuming an “ideal” readout spot.

Disk parameters can be generated entirely by the matrixing and replication process (indicated as source in column 3), or will be influenced by this process. For example, information layer defects of disks can stem from defects generated in the mastering process and magnified by matrixing and replication, but can also be generated in the latter process itself.

The pit geometry of the disk obviously stems from the pit geometry of the master, but will be influenced by the matrixing and replication process.

In general, if the effects on pit geometry in matrixing and replication are

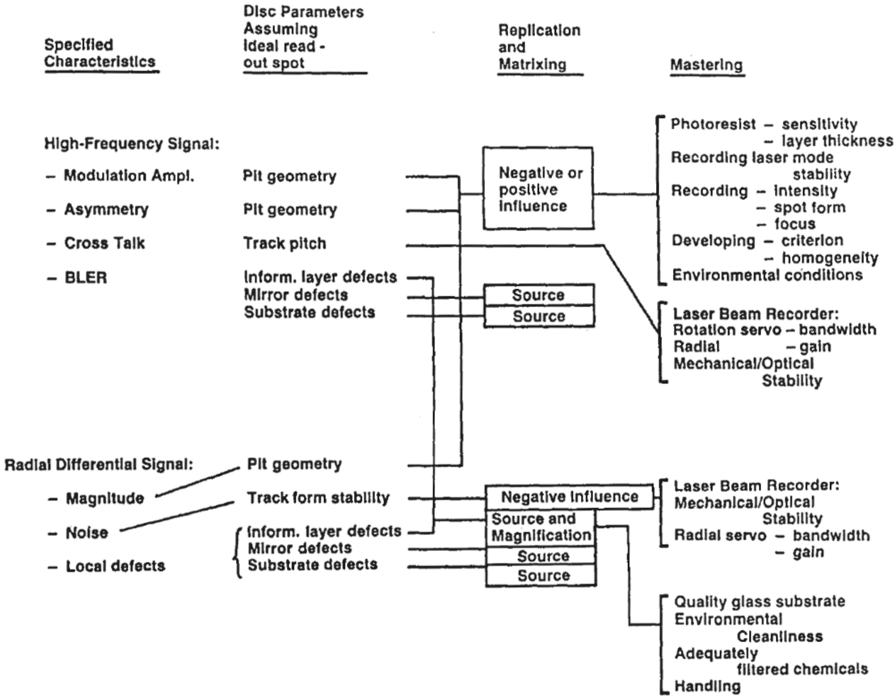


Fig. 8. Quality characteristics sourcing.

consistent, compensation of these effects during mastering can be executed. In those circumstances the replication process will have a "positive" influence on the pit geometry.

In general the mastering parameters in column 4 can be classified as quality of incoming materials, performance of mastering equipment, and environmental conditions and handling.

To arrive at an industrially acceptable situation concerning incoming materials, the Philips CD mastering process requires only commercially available chemicals and materials from several suppliers and standard items such as the tape master and the plain glass disk.

The CD mastering equipment is second-generation equipment specifically designed for long, trouble-free operation, with the help of well-defined quality control and maintenance procedures.

In order to make the mastering process less dependent on the environmental conditions and the skills of the operators, the CD mastering equipment was designed with built-in dust filtering (requiring much less investment in clean-room and air-conditioning facilities) and vastly automated handling. This not only has a direct positive effect on the costs of mastering, but also improves quality and yield.

3.6.9 Conclusions

From the previous paragraphs it can be concluded that the Philips CD mastering process

1) Is a process in which the basic parameters determining the CD system performance, as far as the disk is concerned, are well understood and under control
2) Makes use of equipment specifically designed for routine production
3) Is supported by a vast amount of basic and operational know-how
4) Is designed toward optimum quality and minimum cost of disk replication

   May well be called “an industrial process.”

References

[1] F. Olijkhoek, T. H. Peek, and C. A. Wesdorp, “Mastering Technology for the Philips Optical Disc Systems,” Video Disc Technology Overview 25/2. Electro/81.

[2] “Specification of the 3/4-Inch Cassette Type CD Master Tape,” Sony/Philips Publ.

3.7 Communications aspects of the Compact Disc digital audio system

Sophisticated coding and signal processing principles applied to a mass-marketed consumer product

J.B.H. Peek

3.7.1. Introduction

The compact disc digital audio system has already been introduced in a large number of countries. After an agreement between Philips and Sony in 1979, a common system standard was defined. This standard gradually became the world standard for this completely new system of storage and reproduction of audio signals. An extensive catalog of discs with various labels, and several brands of Compact Disc (CD) players are now available. Most people who have had the opportunity to listen to this new sound medium, not least performing artists, acknowledge that a more intense musical experience is achieved. The improvement in sound quality is in essence obtained by accurate waveform coding and decoding of the audio signals, and, in addition, the coded audio information is protected against disc errors.

From a systems point of view, the CD system was designed on the basis of communications concepts. The communications ideas that have been used will be described in this paper. The concepts applied in the CD player encompass demodulation, error correction and detection, interpolation, and bandwidth expansion to ease the D/A conversion. The paper describes an application of sophisticated communications coding and signal processing principles to a mass-marketed consumer product, and is therefore of general interest. It can be concluded that communications engineers can make valuable contributions in areas not traditionally part of the communications industry.

This restriction to communications concepts implies, however, that important and interesting aspects of the CD player such as the laser optical system, the tracking and focusing principles and control, and the integrated circuits designed for the player-will not be considered [1,2,3].

3.7.2 General System Description

As is usual in a communications system, some of the signal operations at the receiving end of the CD digital audio system are the inverse of those at the transmitting end. A block diagram showing the various signal operations is given in Fig. 1. Before a more detailed description of the successive signal operations in the CD player, we shall briefly describe the signal path from the studio to the optical readout in the CD player.

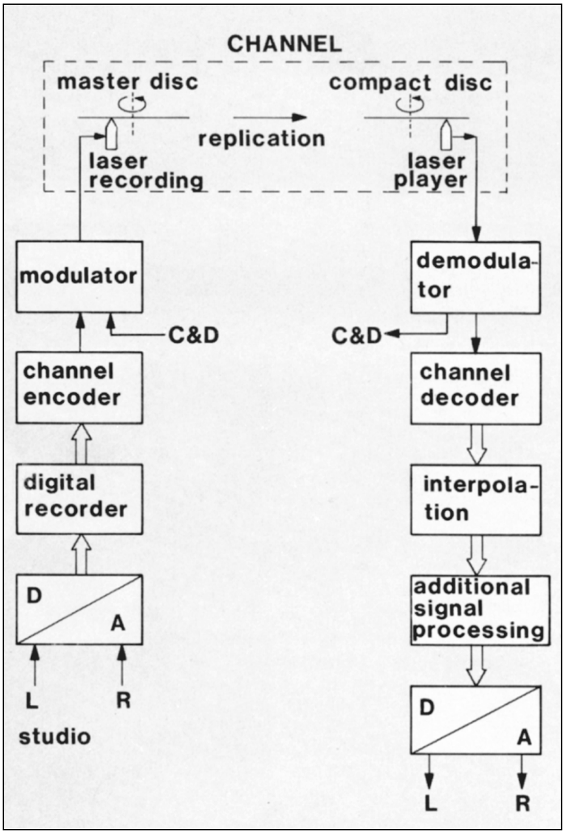


Fig. 1. The compact disc digital audio system, considered as a transmission system

Analog to Digital Conversion

Leaving aside any sound mixing, the two audio signals (left and right) that originate from the studio or concert hall are converted from analog to digital (A/D). The sampling frequency of the signals is quartz-crystal-controlled and is equal to 44.1 kHz. This sampling frequency of 44.1 kHz allows a recorded audio bandwidth of 20 kHz. The samples of both signals are uniformly quantized to 16 bits [4]. As a consequence of the quantization into 16-bit words and the crystal-controlled sampling, the conversion noise level is suppressed by more than 90 dB with respect to the peak signal level, and a total harmonic distortion of less then 0.005% can be achieved. The channel separation is more than 90 dB.

Recording

A video recorder is often used in combination with a PCM interface unit for digital recording of the audio signals on magnetic tape. It is because this video recorder uses the PAL television standard that the sampling frequency has been set at 44.1 kHz, which is $\frac{625-37}{625} \times 3 \times 15\,625 = 44.1$ kHz, where 625 is the number of lines in a PAL picture, 37 is the number of unused lines, 3 the number of audio samples recorded per line, and 15625 Hz the line frequency[5].

Channel encoding

Together with a subsequent modulation, channel encoding is part of the so-calIed disc mastering process. In this process, the information from the video tape recorder system is encoded into the standardized CD format.

In the channel encoding step, the digital information is protected against channel errors by adding parity bytes derived in two Reed-Solomon [6] error-correction encoders. Because the channel mainly has a burstlike error behavior, the wellknown communications technique of interleaving is used to spread the errors out over a longer time[6]. The data streams entering the first encoder, between the two encoders and leaving the second encoder, are scrambled by means of sets of delay lines. As a result of this the burst byte errors will, after deinterleaving, be spread over a longer time so that they can be more easily corrected. Those errors which cannot be corrected but are still detected, which would give corresponding unreliable samples, are restored by interpolation. This will be described in more detail later.

After the channel encoder, digital control and display (C&D) information is added to the encoded data. This information contains music-related data and a table of contents of the disc. With this table of contents, a CD player can be programmed so that only desired musical sections will be reproduced.

Modulation

Before the output data of the channel encoder can be conveyed to the master disc, a modulation operation, achieved by bit mapping, is necessary[7,8]. The reasons for modulation are the following:

- The frequency spectrum of the signal read from the Compact Disc should have low power at the lower frequencies such that the tracking control system is minima1ly disturbed. This requirement is similar to that encountered in digital magnetic recording.
- The binary signal transferred to the master disc must be such that the.bit clock frequency can be regenerated from the signal detected in the CD player. This requirement can be met by suitably mapping a block of n bits onto m (m > n) bits and by imposing an upper limit (say eleven) on the allowable length of a sequence of all ones or all zeros[8].
- Since the light spot with which the CD is scanned in the CD player has finite dimensions, intersymbol interference results which is compensated by processing a sequence of symbols. This imposes a lower limit on the length of a sequence of ones or zeros. A minimum run length of three turns out to be a good choice in practice. Thus, assuming for example m = 11, a sequence like 01010011010 is forbidden.

In the CD digital audio system, a modulation scheme calIed EFM (eight-to-fourteen modulation) is used which meets these requirements satisfactorily[8]. In EFM, a group of 8 bits (also called a byte or symbol) is mapped into 14 channel bits. It can be shown that there are 267 distinct 14-bit sequences that meet the run-length constraints. For a unique mapping of 8 bits, only 256 sequences are needed, so that 11 sequences can be discarded. At the receiver end, that is, the player end, the inverse operation can be obtained by a table look-up. The 14 bit sequences cannot, however, be run after the other without violating the constraints of at least 3 and at most 11 consecutive ones and zeros. By inserting 3 properly chosen merging bits between 14-bit blocks, the run-length requirements can again be satisfied while at the same time suppressing the lower signal frequencies.

In the section describing the error correction and detection systems, we use the concept of frame. A frame consists of 12 audio samples of 16 bits each. (This is equivalent to 24 bytes.) To such a frame, parity bytes and C&D bits are added and EFM is applied. After the addition of merging bits and a synchronization pattern, a final frame consisting of 588 channel bits results. Finally, a sequence of these frames is transferred to the master disc at a channel data rate of 4.32 Mb/ s.

The Channel

Next, the CD standardized format is optically recorded on the surface of a glass disc which is coated with photoresist[9]. Following development and evaporation, the result is the socalled master disc. By galvanic processing, the master disc surface is “transferred” into a nickel shell (or “father”). From this "father," "sons" or stampers are made, which are suitable for replication. By compression or injection molding, the information contained on the surface of the stamper is transferred in the form of about a billion minute pits to a transparent plastic disc. This CD has a diameter of 120 mm, a thickness of 1.2 mm, and a track pitch of 1.6 μm. Finally, after receiving a reflective aluminum coating, over which a protective lacquer is applied, the "Compact Disc" is ready for playing. In the CD player, the track on the disc is optically scanned by an AlGaAs laser (wavelength $\approx$ 0.8 μm) at a constant velocity of about 1.25 m/s. The speed of rotation of the disc therefore varies from 8 r/s when scanning the inner side of the disc to about 3.5 r/s when scanning the outer side. The maximum playing time is about 67 minutes (stereo, of course).

There are several sources of channel errors. First, small unwanted particles or air bubbles in the plastic material, or pit inaccuracies due to stamping and stamper errors, may be present in the replication process. This can cause errors when the information is optically read out. Second, fingerprints or scratches on the disc may occur when it is handled. Together with surface roughness, these disturbances cause additional channel errors. The channel mainly has a burstlike error behavior. As a consequence, a scratch or fingerprint will cause several 14-to-8 demodulated blocks to be in error, which in turn will result in several consecutive byte errors.

3.7.3 Some Error-Correcting Coding Principles

Before describing the error correction and detection that is used in the CD decoder (the channel decoder in Fig. 1), it might be useful to review some principles of error-correcting coding [6, 10].

Without any protective measures, channel errors would result in erroneous audio samples which in turn could cause considerable audible disturbances. It is the purpose of the channel code to reduce the errors at the output of the decoder to a sufficiently low level. In data communications systems, it is common practice, when retransmission is not practical, to use error-correcting codes to achieve such a goal. Since error-correcting block codes are used in the CD system, we will focus our attention solely on these codes. In a block code, a block of *k* information bits is encoded into *n* bits (Fig. 2).

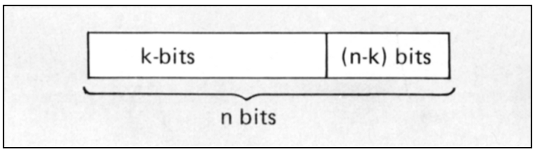


Fig. 2. A block code.

The $(n-k)$ bits which are computed from the k bits according to the mathematical structure of the code are called the parity bits. A block code is often specified by its (n,k) value.

Table I shows an illustrative example of a single-error correcting block code that is obtained by repeating the bit to be transmitted three times. The last two bits can be regarded as parity bits. If we assume that at the most one channel error can occur in a block of three bits, then it can be seen that if a zero were transmitted the number of zeros in a received block of three bits would still be in the majority. The same holds if a one were transmitted. This observation offers a simple single-error correction method based on a majority decision rule. If, however, at the most two errors can occur in a block of three bits, error correction is not always possible. Nevertheless, error detection is still possible, since any received code word other than 000 or 111 is detected as an error. In this simple example, correction and detection cannot be done simultaneously.

data bit	single error correcting code	channel outputs (max. one error)
0	0 0 0	0 0 0 0 0 1 0 1 0 1 0 0
1	1 1 1	1 1 1 1 1 0 1 0 1 0 1 1

Table I. Example of Single Error Correcting Code

At this point, it is useful to introduce the concept of "Hamming distance" between two code words. If two code words, each n bits long, differ in d $(d \leq n)$ positions, then the Hamming distance between these code words is d. Hence, if d errors occur in a transmitted code word the distance between this word and the original code word becomes d.

The effect of applying our example of a triple-repeating, single-error-correcting code can now be clarified with the aid of the Hamming distance concept. Originally, the Hamming distance between the two data bits 0 and 1 is $d = 1$, which is too small to give protection against channel errors. Using the triple repeating code, the distance between the two code words 000 and 111 is increased to $d = 3$. A maximum of one channel error (in a block of three bits) will result in a distance of one (at most) between the received and transmitted code word. This distance is small enough to enable one to decide without doubt which word was transmitted. The principle of looking for the nearest neighbor is called maximum-likelihood decoding.

In general, if up to t errors in an arbitrary code word have to be corrected, then the minimum distance d_{min} min must satisfy the condition

$$d_{min} \geq 2t + 1.$$

This fact is visualized in Fig. 3, which is a two-dimensional representation of a multidimensional codeword space. The point z represents a received word, while x and y are code words. Furthermore, it can be seen that up to $2t$ errors can be detected in this case, provided correction is not attempted simultaneously.

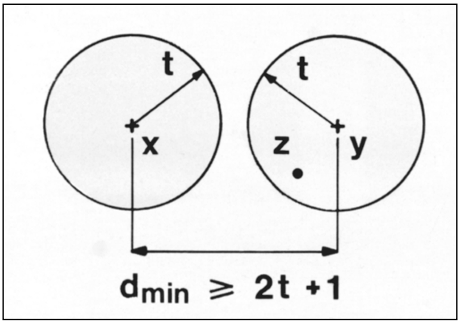


Fig. 3. Relation between minimum distance d_{min} and the maximum number of correctable errors t. The point z shows a received word, while x and y are code words.

In the theory of error-correcting codes, the concept of erasure decoding is of importance. The ith position in a block code, as given in Fig. 2, is called an erasure position if the bit value at that position is unreliable. How such an indication of unreliability can be obtained will become clear in the next section. It is the purpose of erasure correction to determine the correct bit values at a given number of erasure positions. Since, in the case of erasure correction, the positions of the unreliable bits are known, one can imagine that more bits can be corrected than when the positions are unknown.

This can be illustrated with the aid of the triple repeated code described previously. If the received word is unreliable at two arbitrarily chosen but known erased positions (and no further errors are present), then error correction is possible. By deciding on the nonerased bit as being the transmitted bit, simple error correction is obtained. In summary, the code given in Table I has only single-error-correction capability but double erasure correction capability. In general, for a code with minimum distance d_m, $(d_m - 1)$ erasures can be corrected at $(d_m - 1)$ given positions.

The principles of error-correcting block codes as described on a bit level can be extended to the symbol or byte level. Thus, from a block of k information symbols, (*n*-*k*) parity symbols can be calculated and added so that a block of *n* symbols results. With symbols of *s* bits, only a small number, that is, 2^{ks} of the large number 2^{ns} of possible different words of n symbols become code words so that a large d_{min} can be created.

Reed-Solomon codes are particularly efficient since only $2t$ parity symbols have to be used to correct *t* symbol errors. In other words,

$$d_{min} = n\text{-}k + 1.$$

The decoding algorithm will not be described here. In the next section, attention will be given to the decoding strategy which is used by the CD decoder, and the way in which burst errors are treated.

3.7.4 The Compact Disc Decoder

Both audio channels (left and right) are sampled with a frequency of 44.1 kHz. Each sample is represented in 16 bits using uniform quantization. The audio samples are gathered in frames of 12 audio samples each, 6 samples from the left audio channel (L_n) and 6 samples from the right channel (R_n), as shown in Fig. 4. Now each sample of 16 bits consists of 2 bytes or symbols, so that each frame can also be viewed as consisting of 24 audio bytes.

$$L_{6n}\ ,R_{6n}\ ,L_{6n+1}\ ,\ R_{6n+1}\ ,\ \ldots\ldots,L_{6n+5}\ ,R_{6n+5}$$

Fig. 4. The *n*th frame. A frame contains 12 audio samples, 6 samples from the left audio channel (L_n) and 6 samples from the right channel (R_n).

In the CD encoder, the bytes of a number of consecutive frames are scrambled and parity bytes are added such that disc errors can be corrected (or detected if correction fails). The entire process of scrambling and adding parity bytes can best be explained with the help of the CD decoder scheme (Fig. 5) which is, of course, the inverse of the encoder scheme.

Roughly speaking, the CD decoder consists of two decoders (called C_1 and C_2) in series[11-14]. These two decoders have the same structure and are capable of correcting and detecting byte errors. Both codes are Reed-Solomon codes with (*n,k*) values (32,28) and (28,24) so that each uses four parity bytes. Thus, the minimum distance $d_{min} = n\text{-}k + 1 = 5$ and, since $2t+1 \leq d_{min}$, we have $2t \leq 4$ for each code. On the other hand, we have seen that a code with $d_{min}=5$ can correct $e=d_{min}-1=4$ erasures. Hence, it is plausible that each code can correct any number of errors (t) and erasures (e) simultaneously, provided

$$2t + e \leq 4 \quad (e,t \text{ in bytes}).$$

As has been mentioned earlier, an erasure is a byte in a known position of which the byte value is uncertain (It might be erroneous).

Error-detecting capabilities are dependent on the number of errors and erasures that simultaneously have to be corrected. In general, the larger the correcting capability used, the smaller the detecting capability. Hence there is a trade-off between error correction and detection. An undetected erroneous sample can give an annoying audible click, while for detected erroneous samples, interpolated sample values can be computed such that the result is inaudible. The decoders (C_1 and C_2) are separated from each other and from the demodulator by deinterleaving delay lines which are intended to scatter a burst of disc errors among many code words such that the number of errors per code word is minimized, which in turn maximizes the correction and detection probabilities. The first deinterleaving delay lines and the first decoder (C_1) are intended for the correction of most of the small random single byte errors and the detection of the larger burst errors. The second set of deinterleaving delay lines and the second decoder (C_2) are intended for the correction of burst errors and other error patterns which the C_1decoder could not correct. As will be described in more detail, the delay lines Δ after the C_2 decoder scramble uncorrectable but detected byte errors (which become unreliable samples) in such a way that these can often be interpolated between reliable neighbor samples.

The various parts of the CD decoder scheme (Fig. 5) will now be described in more detail.

The deinterleaving delay lines (D) before the C_1 decoder consist of one-symbol (byte) delays used in every even-numbered byte of the 32 byte codewords. The term “code word” will be used only for the full length n. By this procedure, two consecutive bytes on the disc will always end up in two different C_1 code words, thus ensuring that a relatively small disc error lying on the boundary of two bytes will not cause two byte errors in a single C_1 word.

In the currently available Philips CD players, the following strategy (Table II) is used in the C_1 decoder: First, try to correct at most one byte error; if this fails, detect a multiple byte error pattern (put erasure flags on all bytes of the outgoing C_1 word which is derived from a 24-byte frame, as explained earlier). From the mathematical properties of the code it can be proved that the C_1 decoder (given the strategy) will detect all double and triple byte errors with certainty, while error events leading from 4 up to a maximum of 32 error bytes per code word have a probability of not being detected equal to:

Pr (undetected error pattern in code word/ ≥ 4 erroneous bytes) $\approx 1.9 \times 10^{-6}$, where the symbol / denotes a conditional probability.

After the C_1 decoder, the 28 remaining bytes (the 4 parity bytes used in the C_1 decoder are no longer used) and the possible erasure flags are deinterleaved by a triangular shaped network of delay lines (Fig. 5).

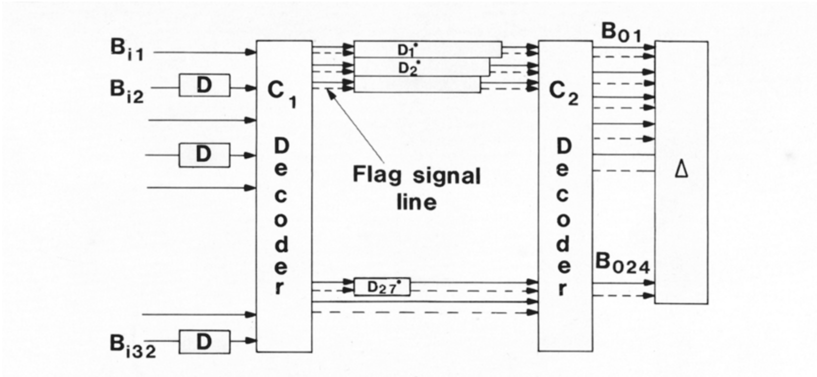


Fig. 5. Scheme of the CD decoder. The 32 bytes (B_{i1}, ... , B_{i32}) of a frame (24 audio samples and 8 parity bytes) are applied in parallel to the 32 inputs. The delay lines *D* have a delay equel to the duration of one byte, so that the information of the "even" bytes of a frame is cross-interleaved with that of the "odd" bytes of the next frame. The C_1 decoder is designed in accordance with the rules for a Reed-Solomon code with (n=32, k=28). It corrects one error, and if multiple errors occur passes them on unchanged, attaching to all 28 bytes an erasure flag, sent via the dashed lines. Due to the different lengths of the delay liner Di^* (i=1, ..., 27), errors that occur in one word at the output of the C_1 decoder are "spread" over a number of words at the input of the C_2 decoder. This result in reducing the number of errors per input word of the C_2 decoder. The second decoder C_2 is also designed to decode a Reed-Solomon code with (n=28, k=24). If the errors cannot be corrected, 24 bytes are passed on unchanged and the associated positions are given an erasure flag via the dashed output lines, B_{01}, ..., B_{024}. In most cases, the unreliable output samples (corresponding with the unreliable bytes) can still be restored by interpolation.

This network of delay lines (D_i*) is such that the length of the delay lines from bottom to top changes with increments of four bytes. Because of this network, the 28 symbols belonging to a single C_1 word and the possible attached erasure flags will be allocated to 28 different C_2 words which are equidistantly spaced. Fig. 6 illustrates how all the symbols of a single C_1 word which are given a flag (indicated by circles) arrive at the output of a triangular shaped delay network in distinct C_2 words, assuming for simplicity delay increments of one byte instead of the actual four bytes. This configuration of C_1 and C_2 code words explains the abbreviation CIRC (Cross Interleaved Reed-Solomon Code).

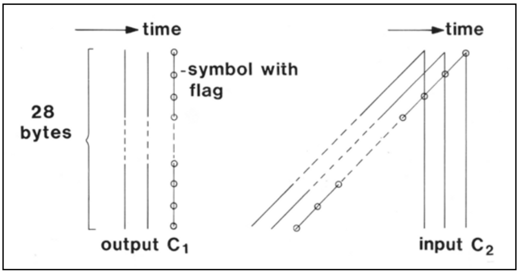


Fig. 6. Effect of deinterleaving: 28 bytes, with detected error flags, in a code word emerging from the C_1 decoder are distributed to 28 consecutive codewords which are then input to the C_2 decoder.

Suppose the increments in the delay lengths of the triangular network were indeed one byte. It would then be possible to correct a burst error encompassing four consecutive C_1 code words if four-erasure correction at the C_2 decoder was used (Fig. 7). In the actual CD system, the increment equals 4 bytes, thus offering a maximum burst-error-correcting capability of 16 consecutive uncorrectable C_1 words.

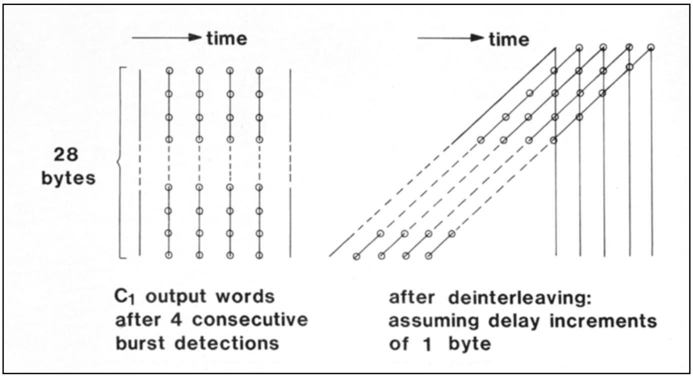


Fig. 7. Example: showing 4-erasure capability.

In current CD players, the possibility of correcting up to four erasures is not used since this would cause too high a probability of an undetected error (a "click"). The strategy adopted in these players allows up to two-erasure correction for the C_2 decoder (Table II). The translation of the C_2 correction strategy to the maximum correctable burst length on the disc is somewhat complicated because of the deinterleaving delay lines before the C_1 decoder. This is the reason for the unusual values of correction and interpolation length given in Table III. It must also be mentioned that the numbers given in Table 111" do not take error propagation (due, perhaps, to synch loss) into account.

For random symbol errors only, the probability of an interpolation or a click (see next section) can be calculated as a function of the disc random byte error probability. From measurements it appears that the average symbol error rate lies around 10^{-4} to 2×10^{-4}.

Starting from the currently used strategy as given in Table II, it can be calculated that the click probability for this case is negligible. The probability of an unreliable sample is 8.3×10^{-10} (once every 3-3/4 hours) if the random disc byte error is about 10^{-3} . For a random disc byte error rate of 10^{-4}, a realistic figure, the sample interpolation rate is about 10^{-15}.

C_1 decoder	C_2 decoder
if single- or zero-error is detected **then** modify at most one symbol accordingly **else** assign erasure flags to all symbols of the received word	**if** single- or zero-error is detected **then** modify at most one symbol accordingly **else** if more than 2 flags **then** copy C_2 erasure flags from C_1 erasure flags **else** if two flags **then** try 2 erasure decoding; **if** less than two flags or **if** 2-erasure decoding fails **then** assign erasure flags to all symbols of the received word

Table II. Currently Used Error-Correction and Detection Strategy.

Up until now, decoder performance has been expressed in terms of the maximum correctable burst length and the interpolation and click rates for the case of random byte errors. The question, however, is: Do these quantities reflect the actual performance of the decoders? In practice it turns out that the interpolations can be attributed, in most cases, to clusters of small error bursts such as can be caused by fingerprints or scratches on the surface of the disc. In spite of the interleaving, such relatively small bursts can lead to errors which will meet at the input of the C_2 decoder (Fig. 8) if they fall within the constraint length (≈ 1.8 cm on the disc). Hence there will be C_2 code words which cannot be corrected and will thus cause interpolations.

From the above, it can be seen that it is worthwhile to increase the correction capabilities of the decoders. In order, however, not to increase the click probability at the same time, it is necessary to introduce multiple-level reliability information (that is, distinction in flag qualities such as certainly in error, and less probable in error) at the entrance of both decoders. Current IC technology offers the possibility to implement these more-complex decoders, and they may be provided in future generations of CD players.

A final observation on the subject of error correction and detection in CD players is that all error control procedures are in vain if track loss occurs either through an improper design of the optical tracking servo system or because of excessive disc damage.

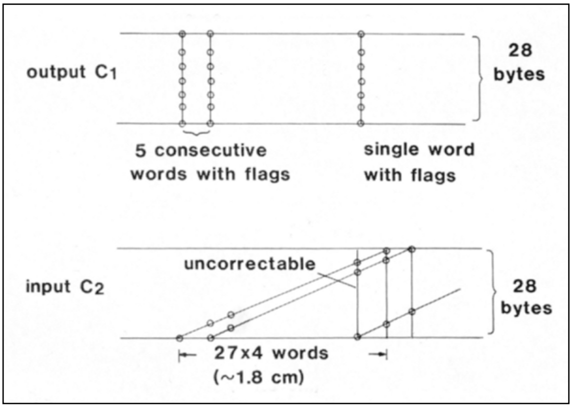


Fig. 8. Uncorrectable situation due to two smaller bursts. If, at the end of the C_1 decoder, 5 consecutive words are attached with flags and if, in addition, a single word attached with flags follows within a distance of 27x4 words (constraint length), an uncorrectable situation can occur. In that case, the input of the C_2 decoder can consist of three erroneous bytes which the present decoder cannot correct.

3.7.5 Interpolation and Muting

As has been mentioned earlier, those byte errors which cannot be corrected by thc C_2 decoder can still be detected. Without any further signal processing these unreliable samples could cause large audible disturbances. It is the purpose of interpolation to insert new samples instead of the unreliable ones [13]. Of course, the interpolated samples should be such that the final result gives no audible disturbance.

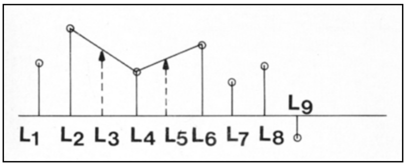


Fig. 9. First-order linear interpolation.

If two reliable neighbor samples are present, an interpolated sample can be obtained from a linear (straight line) interpolation (Fig. 9). Listening tests indicate that the result of this interpolation method in CD systems gives inaudible effects. If an entire C_2 word is detected as unreliable, this would, without taking precautions, make it impossible to apply the suggested interpolation method since both the even and odd numbered samples are declared unreliable. This situation arises if the C_1 decoder fails to detect an error but the C_2 decoder detects it. It is the purpose of the deinterleaving delay lines (Δ) in Figs. 5 and 10 to obtain a pattern, in such a situation, where the unreliable even-numbered samples can be interpolated from the reliable odd-numbered samples or vice versa. Two successive unreliable words consisting of 12 sample pairs are indicated in Fig. 10. A sample pair consists of a sample from the right and a sample from the left audio channel. After the delay lines Δ (length= two frames) the pattern is suitable for interpolation.

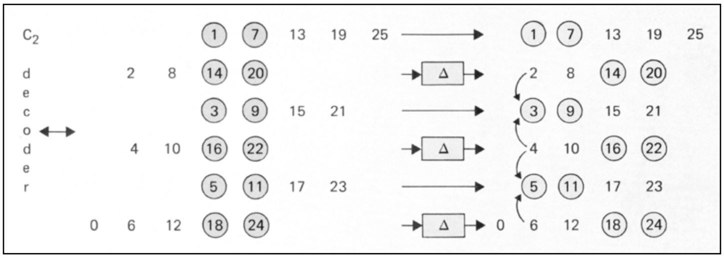


Fig. 10. The effect of delay lines Δ (2 frame times) on sets of samples. The numbers indicate the ordering of the sets of samples. An encircled sample set denotes an erasure flag. After the delay lines, the unreliable samples shown in the figure can be estimated by a first-order linear interpolation.

Because of the various deinterleaving operations and the shuffle of the samples at the input of the deinterleaving delay lines Δ, it is again somewhat complicated to determine the maximum burst length (on the disc) that can be dealt with using first-order linear interpolation. This maximum burst length turns out to be 48 frames (Table III).

C_2 decoder	correction length	interpolation length
1-symbol correction	4 frames 0.68 mm (track length) on disc	48 frames 8.16 mm
2-symbol correction	8 frames 1.36 mm	48 frames 8.16 mm
4-symbol erasure correction	15 frames 2.55 mm	48 frames 8.16 mm

Table III. Maximum Burst Correction and Interpolation Length.

In current CD players, a last remedy is provided in case a burst length of 48 frames is exceeded and two or more consecutive unreliable samples result. In this case, a gradually increasing attenuation of the reliable samples before the burst, then an insertion of zero-valued samples instead of the unreliable samples, and finally a decreasing attenuation of the reliable samples after the burst is applied. This muting of the signal is inaudible provided the muting time does not exceed a few milliseconds and the muting is only incidental.

Digital audio signals can be processed with a digital computer and listened to in a specially designed listening room. Various interpolation methods for the case of two or more consecutive unreliable samples have been tested using such a digital audio computer facility. Since the Compact Disc turns about 10 times a second when the inner side of the disc is read out, error patterns that occur every 0.1 seconds were used in the computer simulations. From these tests it can be concluded that simple straight-line interpolation performs satisfactorily if the number of consecutive unreliable samples is less than eight.

Further research revealed that if 16 consecutive samples are unreliable, restoration is always possible by using adaptive interpolation[15]; the word adaptive indicates an interpolation that uses the statistical properties of the music before and after the burst. Although adaptive interpolating is not used in current players, it is a future possibility.

3.7.6 Additional Signal Processing and D/ A Conversion

As has been mentioned earlier, the two audio signals (left and right) are uniformly quantized in 16 bits at a sampling rate of 44.1 kHz. After the interpolation or muting, the digital signal is in principle ready for conversion to the analog domain. The implementation of a 16-bit D/A converter at an acceptable price level is not, however, an easy task. Besides, as will be explained later, the analog filter following the D / A converter would be complex and expensive if a direct conversion were used.

It will be shown that a 16-bit D/ A performance is obtained from a 14-bit D/ A converter together with additional signal processing. A 14-bit D/A converter is easier to realize, but the 16-bit accuracy would be lost and a dynamic range of more than 90 dB would no longer be maintained. The operations on a 16-bit signal $s(n)$ that is first rounded off to 14 bits and D / A converted are depicted in Fig. 11. Rounding introduces an error $e(n)$ that can be considered as an independent (white) noise sample added to the signal $s(n)$. We have

$$-\frac{q}{2} < e(n) \leq +\frac{q}{2}$$

where q is the step size of the least significant bit (LSB), in our case the 14th bit. The mean square error $\overline{e^2(n)}$ is approximately

$$\overline{e^2(n)} \approx \frac{1}{12}\, q^2.$$

A rounded digital signal, with roundoff noise, can be modeled as a signal which has passed through a noisy communications channel. From communications theory it is known that a signal can be protected against noise by introducing redundancy, which requires bandwidth expansion at the transmitter end. This bandwidth expansion idea can be used to ease the D/ A conversion[16,17].

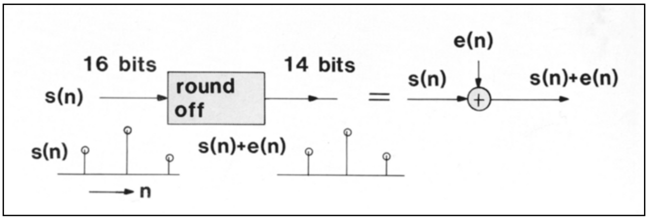


Fig. 11. Rounding a digital signal *s(n)* can be regarded as if the signal has passed through a communications channel.

In the D/ A conversion system, a bandwidth expansion by a factor of four is realized by what is called interpolation in the area of digital signal processing. The interpolated samples are obtained in a way that differs from that described in the previous section. Here the interpolated signal values are obtained by first a fourfold increase of the sample frequency through insertion of three zero signal values between every two input samples, and next by lowpass filtering this signal with a finite impulse response (FIR) digital filter. This digital filter has 96 taps, and the 96 coefficients are each represented in 12 bits, so that an attenuation in the stop band (above 24 kHz) of about 50 dB results (Fig. 12). In a digital system, only the signal frequencies in the band from zero to half the sampling frequency are relevant and consequently only these frequency bands are indicated in Fig. 12.

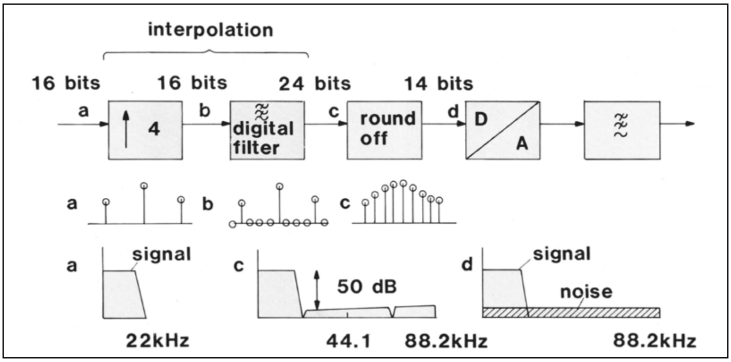


Fig. 12. D/A conversion based on interpolation.

Because of the digital lowpass filter, the signal at the filter output has acquired a word length of 16 + 12 = 28 bits. Reducing this word length by rounding to 14 bits gives a mean square error of $\frac{1}{12}q^2$ (where q is the step size of the LSB of a 14-bit D/A converter). This noise power, however, is now evenly distributed over a four-times-larger interval (Fig. 12). For reasons of simplicity, the attenuated signal components around 44.1 kHz and near 88.2 kHz are no longer indicated in the picture that shows the round-off noise spectrum. The noise power in the 0-22 kHz bandwidth, however, is four times less than in the case of a direct round-off from 16 to 14 bits. A factor of four in noise power (6 dB) corresponds with a factor of two in amplitude, and thus with one bit.

In the Philips D/A conversion system, a noise-shaping filter is used after the digital filter in Fig. 12 which redistributes the noise power in such a way that the noise power in the audio bandwidth 0-20 kHz is reduced at the expense of an increase in noise power outside this bandwidth. Since the ear responds only to frequencies up to 20 kHz, the 7-dB gain in signal-to-noise ratio obtained with the noise-shaping filter in this bandwidth can directly be translated as an extra bit gained. Thus the combination of interpolation (factor of four), the noise-shaping filter, and a 14-bit D/A gives about the same performance as a straight 16-bit D/A converter.

As mentioned earlier, the digital lowpass filter attenuates the frequencies above 24 kHz by about 50 dB. Consequently, the analog filter following the D/A converter is rather simple. Such an analog filter is necessary in order to prevent signals around multiples of the sampling frequency from overloading the power amplifier or from mixing with other signals (such as the bias signal from a tape recorder) and thus causing audible distortion. If, however, a direct 16-bit D/A is used, the analog filter after this converter would have to be quite complex. For this filter the transition band would have to be only a few kHz without affecting the flatness of the passband, while giving an attenuation above 24 kHz of at least 50 dB.

Acknowledgment
I would like to thank S. Baggen for valuable discussions and comments.

References

[1] M. G. Carasso, J. B. H. Peek, and J. P. Sinjou, “The compact disc digital audio’ system,” Philips Tech. Rev., (Special Issue), **40**, 151-156 (1982).

[2] S. Miyaoka, “Digital audio is compact and rugged,” IEEE Spectrum, pp. 35-39, March 1984.

[3] “Draft Compact Disc Digital Audio System,” now under discussion in the International Electrotechnical Commission as document 60A (Central Office) 82. For text, reference is made to doc. 60A (Secretariat 96).

[4] B. A. Blesser, “Digitization of audio,” J. Audio Eng. Soc., **26**, 739-771 (1978).

[5] T. Doi, Y. Tsuchiga, and A. Iga, “On several standards for converting PCM signals into video signals,” J. Audio Eng. Soc., **26**, 641-649 (1978).

[6] G. C. Clark and J. B. Cain, Error-Correction Coding for Digital Communications, Plenum Press, 1981.

[7] K. A. Immink, “Modulation systems for digital audio discs with optical readout,” IEEE Int. Conf. on Acoust., Speech, and Sig. Process., pp. 587-589, Atlanta, GA (1981).

[8] J. P. J. Heemskerk and K. A. Schouhamer Immink, “Compact disc: system aspects and modulation,” Philips Tech. Rev., (Special Issue), **40**, 157-164 (1982).

[9] W. Verkaik, “Compact Disc (CD) mastering-an industrial process,” The AES Premier Conference- The New World of Digital Audio, New York, (1982).

[10] A. B. Carlson, Communication Systems, McGraw-Hill, 1975.

[11] L. B. Vries and K. Odaka, “CIRC-the error correcting code for the Compact Disc,” The AES Premier conference-The New World of Digital Audio, New York (1982).

[12] L. M. H. E. Driessen and L. B. Vries, “Performance calculations of the compact disc error correcting code on a memoryless channel,” Int. Conf. Video and Data Recording, University of Southampton (1982).

[13] H. Hoeve, J. Timmermans, and L. B. Vries, “Error correction and concealment in the compact disc system,” Philips Tech. Rev., (Special Issue), **40**, 166-172 (1982).

[14] T. Doi, “Error correction for digital audio recordings,” The AES Premier Conference- The New World of Digital Audio, New York (1982).

[15] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, “Adaptive interpolation of time-discrete signals that can be modelled as autoregressive processes,” submitted for publication to IEEE Trans. Acoust., Speech, and Sig. Process.

[16] D. Goedhart, R. J. van de Plassche, and E. F. Stikvoort, “Digital-to- analog conversion in playing a compact disc,” Philips Tech. Rev., (Special Issue), **40**, 174-179 (1982).

[17] T. A. C. M. Claasen, W. F. G. Mecklenbräuker, J. B. H. Peek, and N. van Hurck, “Signal processing method for improving the dynamic range of A/D and D/A converters,” IEEE Trans. Acoust., Speech, and Sig. Process., ASSP-28, no. 5, Oct. 1980.