Compact Disc: system aspects and modulation

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

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 sound from the studio 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'.

Fig. 2 shows the encoding system in more detail. The audio signal is first converted into a stream $B_i$ 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 $B_i$. 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 is 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 \( \eta \) 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 accommodated side by side on the disc. The information density \( n/A \) is thus given by:

\[
 n/A = \eta/d^2. \tag{1}
\]

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

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 \( \lambda \) 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; \( \lambda = 800 \text{ nm} \). 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

![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.](image)

The channel

The bit stream \( B_1 \) 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 \( B_1 \) 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.

tolerance for the disc thickness is proportional to \( NA^{-3} \), 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 \( \mu \text{m} \) 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 \( B_2 \); see fig. 3a. The signals originating from different pits and lands are superimposed 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. 3a), 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 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.

Fig. 3. 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.

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.

Table I. Manufacturing tolerances.

<table>
<thead>
<tr>
<th>Player</th>
<th>Thickness 1.2 ± 0.1 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flatness ± 0.6° (at the rim corresponding to a 0.5 mm)</td>
</tr>
<tr>
<td></td>
<td>Pit-edge positioning ± 50 nm</td>
</tr>
<tr>
<td></td>
<td>Pit depth 120 ± 10 nm</td>
</tr>
<tr>
<td><strong>Disc</strong></td>
<td><strong>BIBLIOTHEEK NAT. LAB, GLOEI/ FABRIKEN</strong></td>
</tr>
<tr>
<td>Thickness 1.2 ± 0.1 mm</td>
<td>POST S 80,000</td>
</tr>
<tr>
<td>Flatness ± 0.6° (at the rim corresponding to a 0.5 mm)</td>
<td>5600 JA EINDHOVEN</td>
</tr>
<tr>
<td>Pit-edge positioning ± 50 nm</td>
<td></td>
</tr>
<tr>
<td>Pit depth 120 ± 10 nm</td>
<td></td>
</tr>
</tbody>
</table>

**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 bit-stream 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. 4a). Uncoded recording (1: pit; 0: land, or vice versa) then gives the pattern of fig. 4b. This results in the rounded-off signal of fig. 4c at Q in fig. 1; fig. 4d gives the eye pattern. The signal in fig. 4c represents the highest frequency (f_m) for this mode of transmission, and we have f_m = ½ 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.

---

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 $f_c = (2NA/\lambda)v$.

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 $f_{m1} (= \frac{1}{2}f_d)$.

Fig. 6 gives the situation with modulation: an imaginary $8\rightarrow16$ modulation, which is very close to EFM, however. Each group of 8 incoming data bits (fig. 6a) is converted into 16 channel bits (fig. 6a'). 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. 6b). On the
time scale the minimum pit and land lengths ('the
minimum run length' \( T_{\text{min}} \)) have become \( \frac{1}{2} \) times as
long as in fig. 4, but a simple calculation shows that
about as much information can nevertheless be trans-
mitted 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. 6b and b'); the 'channel bit
length' \( T_c \) has decreased by a half.

With the modulation we have managed to reduce the
highest frequency \( (f_{\text{max}}) \) in the signal (see fig. 6c,
left; \( f_{\text{max}} = \frac{1}{2} f_a = \frac{1}{3} f_{\text{m}} \)). Therefore \( f_a \) and \( v \) can be
reduced by a factor of \( \frac{1}{2} \) for the case in which a very
small eye height is tolerable (see fig. 5); this represents
an increase of 50% in playing time.

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. 6d). 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 \( \frac{1}{2} \)
\( (a_2 > \frac{1}{3} \) 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 \( (T_e) \), 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' \( \sigma \) (the number
of incoming data bits per unit length of the track;
\( \sigma = f_a/v \)) for three systems: '8-24 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', \( T_{\text{min}} = 2T \), \( T_e = \frac{3}{4} T \)). 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 \( (2NA/\lambda)_{\text{v}} \), 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 perturba-
tions 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.

For our Compact Disc system we have \( \sigma = 1.55 
\) data bits/\( \mu \)m \( (f_a = 1.94 \) 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 \( P \) ap-
proaches \( C \). But 8-16 modulation is still preferable
to 8-24 modulation, even close to \( C \), since the eye
width is \( \frac{1}{2} \) 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 par-
cular the 'maximum run length' \( T_{\text{max}} \) must be as
small as possible.

The second requirement relates to the 'low-fre-
quency content' of the read signal. This has to be as

\[ \begin{align*}
\sigma &= \frac{f_a}{v} \\
8-24: T_e &= \frac{3}{4} T, T_{\text{min}} = 2T \quad \text{(fig. 4)}, \\
8-16: T_e &= \frac{1}{2} T, T_{\text{min}} = T, T_c = T \quad \text{(fig. 6)}, \\
8-8: T_e &= T, T_{\text{min}} = T \quad \text{(fig. 4)}, \\
\end{align*} \]

where \( T \) is the data bit length. The straight lines give the relations
that follow from fig. 5:

\[ \begin{align*}
a_1 &= c_1(1 - f_{\text{m}1}/f_a) \rightarrow a_1 = 1 - \sigma/1.8, \\
a_2 &= c_2(1 - f_{\text{m}2}/f_a) \rightarrow a_2 = 0.5(1 - \sigma/2.7), \\
a_3 &= c_3(1 - f_{\text{m}3}/f_a) \rightarrow a_3 = 0.26(1 - \sigma/3.6),
\end{align*} \]

where \( \sigma \) is the numerical value of the linear information density,
expressed in data bits per \( \mu \)m. The \( c \)'s are the ratios of the half
eye height to the amplitude, and the \( f_{\text{m}} \)'s the maximum frequencies
for the three systems \( (c_1 = 1, c_2 = \sin 30° = 0.5, c_3 = \sin 15° = 0.26, \\
f_{\text{m}1} = \frac{1}{2} f_a, f_{\text{m}2} = \frac{1}{3} f_a, f_{\text{m}3} = \frac{1}{4} f_a, f_a \) is the data bit rate). The
second set of equations follows from the first set by substituting
\( 0.8 \times (2NA/\lambda)_{\text{v}} \) for \( f_a \), with \( NA = 0.45, A = 0.8 \) \( \mu \)m, \( v = f_a/\sigma \). The
factor 0.8 is introduced as a rough first-order correction to the
'ideal' amplitude characteristic.
small as possible. There are two reasons for this. In the first place, the servosystems for track following and focusing are controlled by low-frequency signals, so that low-frequency components of the information signal could interfere with the servosystems. 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.

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 B1 thus contains 24 symbols per frame. In B2 eight parity symbols have been added and one C&D symbol, resulting in 33 'data symbols'. 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 B3 is obtained. B1 therefore contains $33 \times 17 + 27 = 588$ channel bits per frame. Finally, B1 is converted into a control signal for the write laser. It should be noted that in B1 '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 B3 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 $T_{\text{min}}$ of 3 and a $T_{\text{max}}$ of 11 channel bits (this means at least 2 and at the most 10 successive zeros in $B_3$). This choice came about more or less as follows. We have already seen that the choice of about $\frac{1}{2}$ data bits for $T_{\text{min}}$, with about 16 channel bits on 8 data bits, is about the optimum for the Compact Disc system. 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 $T_{\text{min}} = 3$, $T_{\text{max}} = 11$ channel bits. The choice of $T_{\text{max}}$ 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 con-
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 are shown two data symbols of \( B_2 \) and their translation from the dictionary into channel symbols \( (B_3) \). From the \( T_{\text{min}} \) rule the first of the merging bits in this case must be a zero; this position is marked \('X'\).

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 are shown two data symbols of \( B_2 \) and their translation from the dictionary into channel symbols \( (B_3) \). From the \( T_{\text{min}} \) rule the first of the merging bits in this case must be a zero; this position is marked \('X'\). In

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 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

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

If the initial value had been -3, the merging combination 001 would have been chosen.

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.

Summary. 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 \(d = 1\) \(\mu\)m, 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.
This prototype player, which will be put on the market later, will display 'information for the listener' such as title, composer, 'track number' and playing time of the piece of music. The different sections of the music on the disc can also be played in the order selected by the user — the numbers on the far right.