The Compact Disc Digital Audio System

By Thomas D. Rossing

One of the most exciting developments in data storage technology in recent years is the compact disc. Only 12 cm in diameter, a compact disc can store more than 6 billion bits of binary data to be read out by a laser. This is equivalent to 782 megabytes, or more than the capacity of 1500 half-megabyte floppy discs. Over 275 000 pages of text, each holding 2000 characters, could be stored on a compact disc for display on a television monitor.

Used for digital audio, a compact disc stores 74 minutes of digitally encoded music. This music can be reproduced with very high fidelity over the full audible range of 20 to 20 000 Hz. The dynamic range and the signal-to-noise ratios can both exceed 90 dB, and the sound is virtually unaffected by dust, scratches, and fingerprints on the disc. Unlike most other digital recording media, compact discs can be replicated in large quantities from a master disc.

Of importance to the physics teacher is that this new invention, which has captured the attention of the public and especially of young people, embodies a number of interesting applications of optical, acoustical, and mathematical principles. This article calls attention to some of these principles and suggests sources for further reading by interested teachers and students.

A Brief History of the Compact Disc

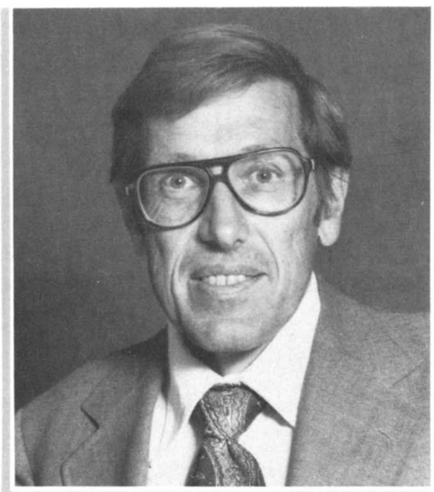
Although the compact disc incorporates technologies pioneered by a number of individuals and research organizations, successful development is largely due to a joint effort by Philips in the Netherlands and Sony in Japan. Optical technology, developed by Philips, merged with Sony's error-correction techniques to create the successful compact disc format, which has now been adopted by other manufacturers.

Philips first proposed storing audio material on an optically read disc around 1974. Analog modulation methods, used for video storage[1] were considered, but digital encoding was favored. Philips established a small disc diameter as a design prerequisite. Meanwhile, Sony had explored the possibility of a large-diameter audio disc with optical readout and had extensively studied error-processing requirements for such a system. In 1979, Philips and Sony reached an agreement to collaborate, and the Compact Disc Audio system was announced in 1980.

Major developments included a semiconductor laser for the optical pickup and large scale integrated (LSI) circuits for signal processing, control circuits, and D/A conversion. The Compact Disc Audio system was introduced in Europe and Japan in 1982 and in the United States in 1983. Over 900 000 players and 17 million discs were sold in 1984, making it the most successful electronic product ever introduced.[2] Players selling for less than $200 are available.

Information Storage

Recorded information is contained in pits impressed into the plastic surface which is then coated with a thin layer of aluminum to reflect the laser beam (Fig. 1). Pits



Tom Rossing *teaches courses in acoustics and directs the Acoustics Research Laboratory at Northern Illinois University. He has published resource letters and reprint books for the AAPT on* Musical Acoustics *and* Environmental Noise Control *as well as a textbook* The Science of Sound *(Addison-Wesely, 1982). A recent Compact Disc with* Auditory Demonstrations *by him and his Dutch colleagues Adrian Houtsma and Wil Wagenaars is available from the Acoustical Society of America. A biographical sketch appears in the Sept. 1986 issue of* TPT. *(Department of Physics, Northern Illinois University, DeKalb, IL 60115)*

are about 0.5 μm wide and 0.11 μm deep, arranged in a spiral track similar to the spiral groove in a phonograph record, but much narrower. The track spacing on a compact disc is about 1.6 μm, compared to about 0.1 mm (100 μm) for the groove of a long-play phonograph record.

The track on a compact disc, which spirals from the inside out, is about three miles in length. The track of pits is recorded and read at a constant 1.25 m/s, so the rotation rate of the disc must change from about 8 to 3.5 revolutions per second as the spiral diameter changes. Each pit edge represents a binary 1, whereas flat areas within or between the pits are read as binary 0's.

The laser beam, applied from below the compact disc as it lies on a turntable, passes through a transparent layer 1.2 mm thick and focuses on the aluminum coating, as shown in Fig. 2. The spot size of the laser on the transparent layer is 0.8 mm, but at the signal surface where the pits are recorded, its diameter is only 1.7 μm. Thus any dust or scratch smaller than 0.5 mm will not cause a readout error because it is out of focus. Larger blemishes are handled by error-correcting codes, which will be discussed in a later section.

The Optical Pickup

The optical pickup used to read a compact disc is shown in Fig. 3. A semiconductor laser emits a beam of red light (790-nm wavelength) that is eventually focused to a tiny spot 1.7 μm in diameter. The reflected beam is directed to a photodiode that generates an electrical signal to be amplified and decoded.

Included in the sophisticated optical pickup are a diffraction grating, a polarization beam splitter, a quarter-wavelength plate, and several lenses. A semiconductor laser employs an aluminum-gallium-arsenide (AlGaAs) pn junction, similar to that in a light-emitting diode (LED). The diffraction grating creates two secondary beams that are used for tracking the primary beam.

When the laser beam strikes a land area between two pits, it is almost totally reflected. When it strikes a pit (which appears like a bump from the reading side), whose 0.11-μm height is roughly one-quarter wavelength of the laser light, the part of the beam reflected from the pit cancels the part reflected from the land, so little or no light returns. The reflected light of varying intensity is directed on the photodiode; a change in intensity will

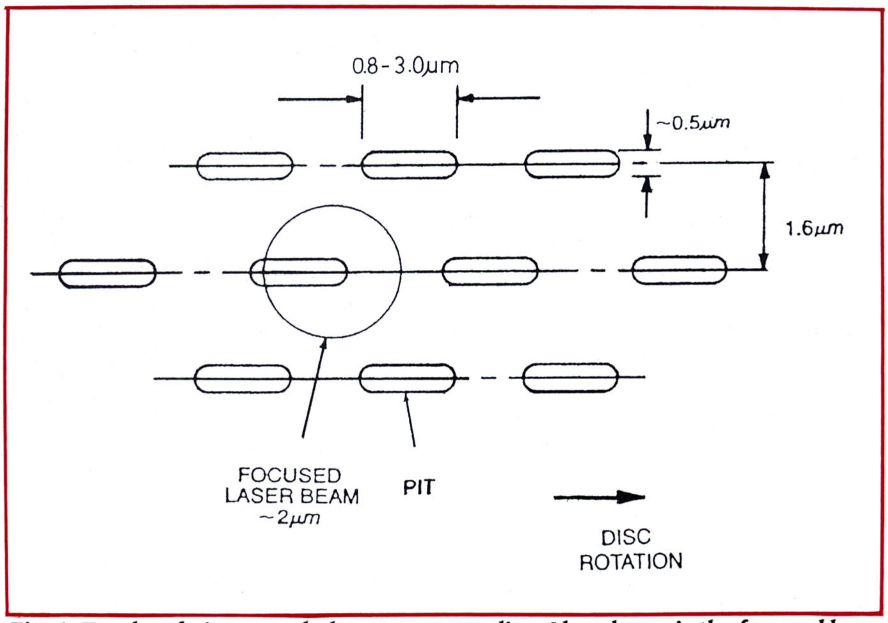


Fig. 1. Tracks of pits recorded on a compact disc. Also shown is the focused laser beam used to read the recorded information.

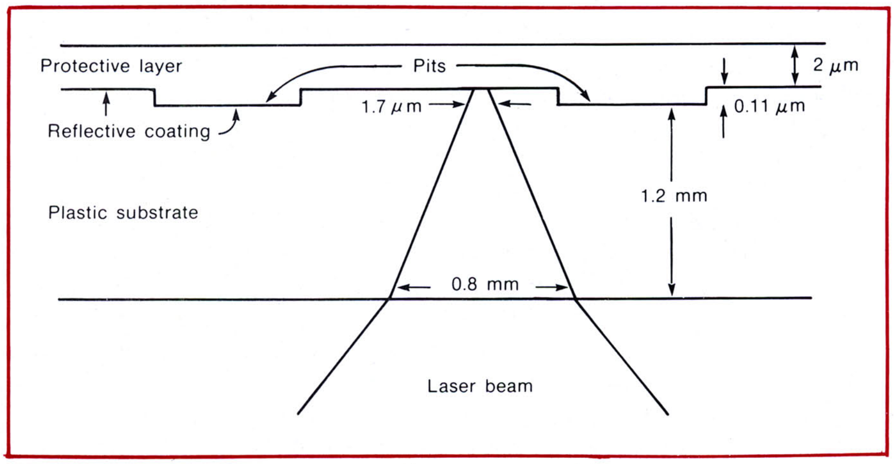


Fig. 2. Cross section of a compact disc (not to scale).

eventually be interpreted as a binary 1 and unchanged intensity as a 0.

Compact disc players incorporate rather sophisticated systems for automatically focusing the laser beam on the reflective surface and for centering the laser beam on the track being read. These will be described in a later section.

Semiconductor Laser

The laser light that is used to read out the recorded data is generated by a semiconductor laser of the *double heterojunction* type.[3-5] Developed at Bell Laboratories for optical communication systems, this compact, efficient light source is ideal for optical readout from audio compact discs and videodiscs. The laser consists of three layers having different energy gaps E_g, as shown in Fig. 4. Layer 2 is gallium arsenide GaAs, while layers 1 and 3 have the composition $Al_xGa_{1-x}As$ with x between 0 and 0.5 (typically 0.16). In layers 1 and 3, E_g is greater than in the "active" layer 2. Layer 1 is shown as n-type, layer 3 as p-type, and layer 2 as nearly intrinsic semiconductor.

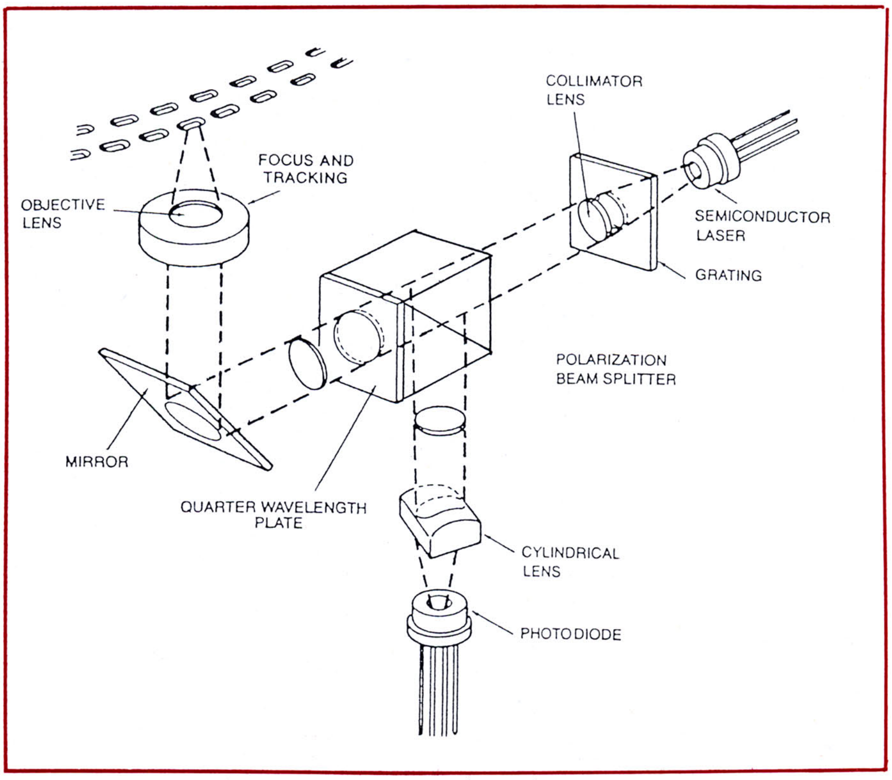


Fig. 3. Optical pickup in a compact disc player. Coherent light from a semiconductor laser is focused on one recorded track on the disc. The reflected light is directed to a photodiode.

Light is emitted when electrons and holes recombine, and this takes place mainly in layer 2. Because of the barriers ΔE_c and ΔE_v, the electrons and holes are trapped in layer 2, leading to what might be called a "population inversion"; that is, there are excess electrons and holes that can be stimulated to recombine (and emit light) by a photon having $h\nu = E_{g2}$.

If the semiconductor is in thermal equilibrium (forward current I = 0), there are few electrons in the conduction band and few holes in the valence band. More photons are absorbed (by creating electron-hole pairs) than are emitted by recombination. It is only when the forward current is fairly large that the population of holes and electrons is large enough for photon multiplication to take place. Early GaAs lasers required such large current densities that they could not operate continuously except at very low temperatures.[6] It was the invention of the double heterojunction structure that made operation at room temperature practical.

The difference in energy gap between layers not only creates the barriers ΔE_c and ΔE_v that trap holes and electrons, but it also creates a discontinuity Δn in the refractive index that traps photons in layer 2. Layers 1 and 3 have a lower refractive index than layer 2, so light traveling nearly in the z-direction in layer 2 (see Fig. 4a) is totally internally reflected.[7] The front and back mirrors necessary for laser action are cleavage planes of the crystal and consequently plane and parallel.

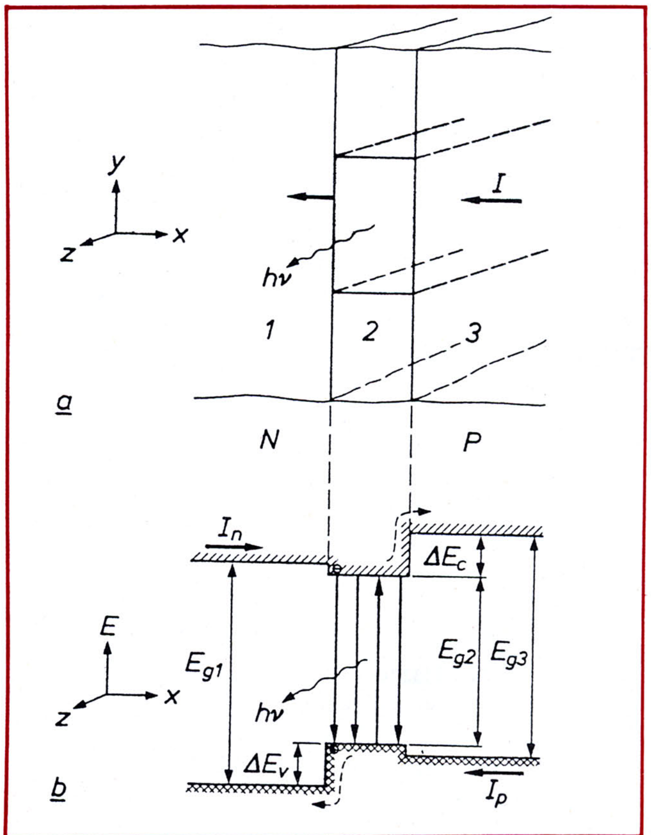


Fig. 4. a) Structure of a double heterojunction semiconductor laser. Layer 2 is GaAs; layers 1 and 3 are $Al_xGa_{1-x}As$ with n-type and p-type conductivity. b) Simplified energy-level diagram for a large forward current I. Electrons and holes recombine to emit light in layer 2 (from Ref. 5).

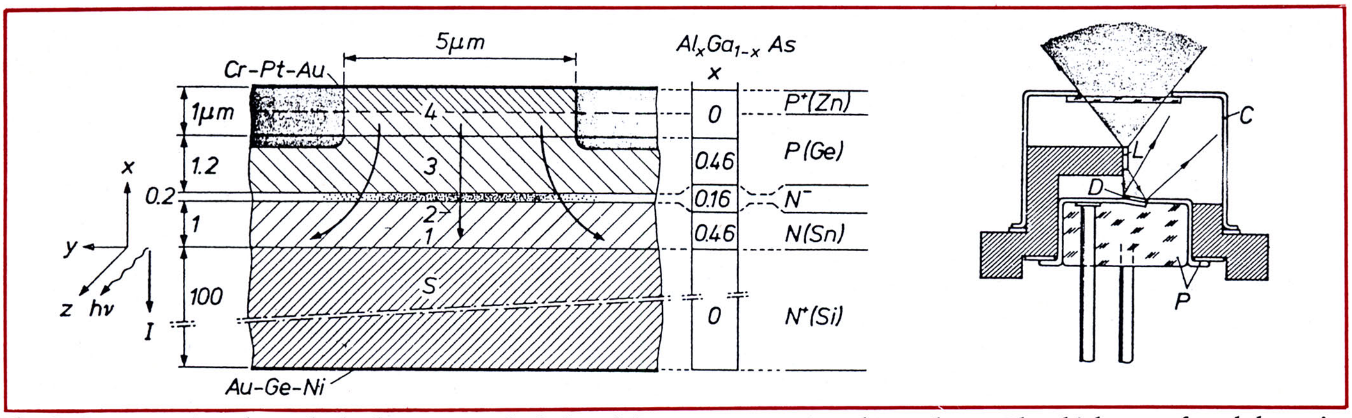


Fig. 5. Cross section through the laser crystal of the Philips CQL10 semiconductor laser. The thickness of each layer is given on the left and the Al content, conduction type, and dopant on the right. The upper part of layer 4 is doped to make it very strongly p-type. The encapsuled crystal is shown at the far right (from Ref. 5).

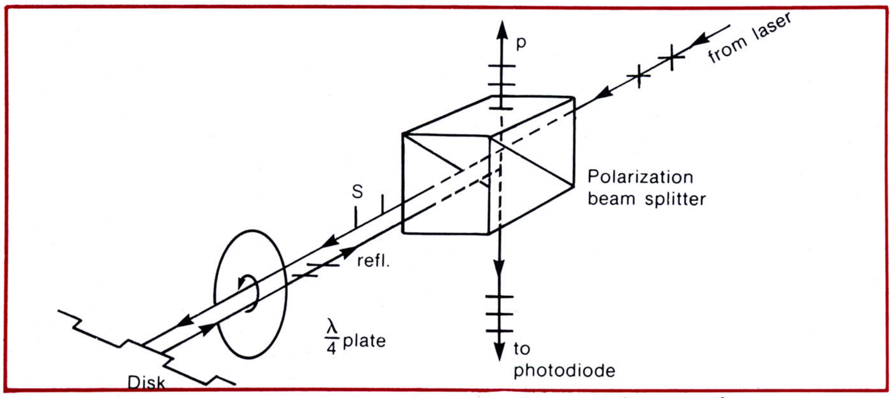


Fig. 6. Paths of light through the polarization beam splitter and quarter-wave plate.

The structure of the laser crystal in the Philips CQL10 light source for compact disc players is shown in Fig. 5. It consists of a substrate S with four layers (1–4) applied to it by means of liquid-phase epitaxy[8] and two metallic contact layers. The substrate S and layer 4 are pure GaAs. In the other layers some of the gallium has been replaced by aluminum; x = 0.16 in layer 2 and x = 0.46 in layers 1 and 3. The active layer 2 is about 0.2 μm thick.

The CQL10 emits red light with a wavelength λ = 780 nm into an elliptical cone with half-power angles of 34° and 60° parallel and perpendicular to the plane of the active layer, respectively. The optical efficiency, which ranges from 1 to 5%, depends strongly on temperature. To stabilize the threshold current and prevent thermal runaway, a photodiode to monitor the laser output is incorporated into the CQL10. A signal from this diode is used to control the supply current by means of a feedback circuit.

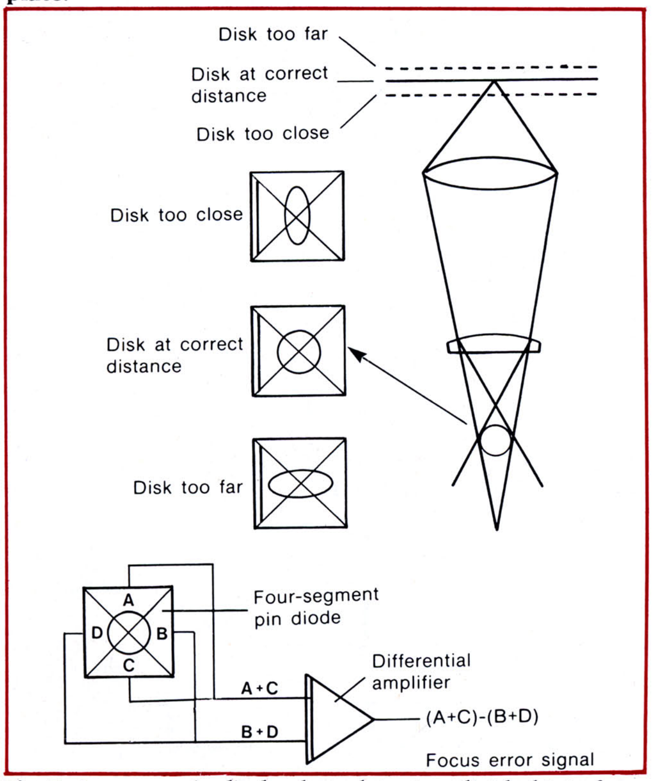


Fig. 7. A square pin diode photodetector, divided into four identical triangular segments, and a difference amplifier control a focus servo that moves a cylindrical lens up and down to correct for disc surface irregularities. (from Ref. 12).

Polarization Beam Splitter

The polarizing beam splitter consists of two 90° isosceles prisms cemented together at the hypotenuse. The hypotenuse faces are coated with a polarization-sensitive multilayer dielectric film. Light from the laser is split into two beams; the p-beam polarized parallel to the surface is internally reflected, while the s-beam polarized perpendicular proceeds without deviation toward the disc, as shown in Fig. 6.

Before reaching the disc, the transmitted s-beam passes through a quarter-wave plate which changes its polarization to circular. On the return journey the reflected light, returning through the quarter-wave plate, is polarized parallel to the prism interface and is reflected 90° toward the photodiode.

Auto Focus and Tracking

To distinguish between pits and land areas, the laser beam must stay focused within about 0.5 μm of the reflecting surface. The flat surface of the disc, however, may have deviations as large as 0.5 mm, one thousand times greater. Thus the objective lens must refocus rapidly as the surface deviates during the rotation. A servo-driven auto-focus system that uses a four-quadrant photodiode and a differential amplifier to control the servo disc makes the rapid refocus possible.

The cylindrical lens, shown just above the photodiode in Fig. 3, projects a circular laser spot on the photodiode if the reflecting surface of the disc is exactly at the focus of the objective lens, but an elliptical spot if it is above or below the focus, as shown in Fig. 7.

The four-quadrant diode provides electrical signals proportional to the light intensity on each of its four quadrants, and these are combined to form the error focus signal, as shown in Fig. 7. The objective lens is attached to a movable coil in a magnetic field, similar to that which drives a loudspeaker cone.

The spiral track of a compact disc is only 1/60 as wide as the grooves of a long-play phonograph disc. A very sensitive auto-tracking system is required to keep the laser beam centered on the track. One widely used auto-tracking system makes use of two secondary beams created by the diffraction grating in the optical pickup (see Fig. 3), which are separated from the main beam by about 20 μm. When the main beam is centered on a signal track, the secondary beams are partly on that track and partly on adjacent land areas. As the beam moves to either side, one of the secondary beams encounters more pit area, while the other encounters more land area. The difference in the signals from the two tracking photodiodes tells the servo drive which way to move the beam.

Other methods for tracking the pit spiral use the main beam itself. They are described as "push-pull," "differential phase detection," and "high-frequency wobble" methods, and each has its own unique advantages.[9]

Coding and Decoding

The data actually recorded on the disc are the result of applying a rather elaborate coding scheme. First the audio signals are sampled at a 44.1 kHz rate and converted into binary data. Then they undergo error-correction encoding and EFM modulation (which I explain shortly), and subcode and synchronization information is added.

Rarely is digital information recorded on tape or disc in its raw state. The packing density and the reliability can be improved markedly by using a clever scheme for *channel coding* or *modulation*. A number of these schemes have been developed for computer peripherals; some work best when recording on magnetic tape, some on magnetic discs, and some are particularly well suited for optical discs.

One of the earliest channel codes was the NRZ (nonreturn to zero) code, in which ones and zeros were represented directly by two opposite states of the recording medium, as shown in Fig. 8a. (On magnetic tape these would be two different remnant states of magnetization; on an optical disc these are represented by a pit and a land area). This system, which is used in video tape recorders, requires external synchronization to decode strings of ones or zeroes; therefore, for use in computer peripherals the NRZI (nonreturn to zero inverse) code, shown in Fig. 8b, is preferred. In this code a one is represented by a transition, a zero by no transition (transitions usually occur in the middle of a bit cell).[10]

Other modulation or channel codes used in digital recording include PE (phase encoding), FM (frequency modulation), MFM (modified FM), Miller code, HDM (High-density modulation),and EFM (eight-to-fourteen modulation).[11] I will describe only the EFM code, shown in Fig. 8c, which is the one used in compact discs.

The EFM is a kind of NRZI code in which eight bits of data are represented by fourteen channel bits. This may seem wasteful, but by requiring that every one in channel bits (corresponding to a transition) be separated by two to ten channel zeros, EFM allows bit sizes to be less than the diameter of the laser beam that reads them. Thus EFM actually results in a greater bit density. In the compact disc system, a 1.7 μm diameter laser spot reads pits whose length varies incrementally from 0.833 to 3.054 μm.

It turns out that of the 16 384 (2^{14}) possible 14-bit

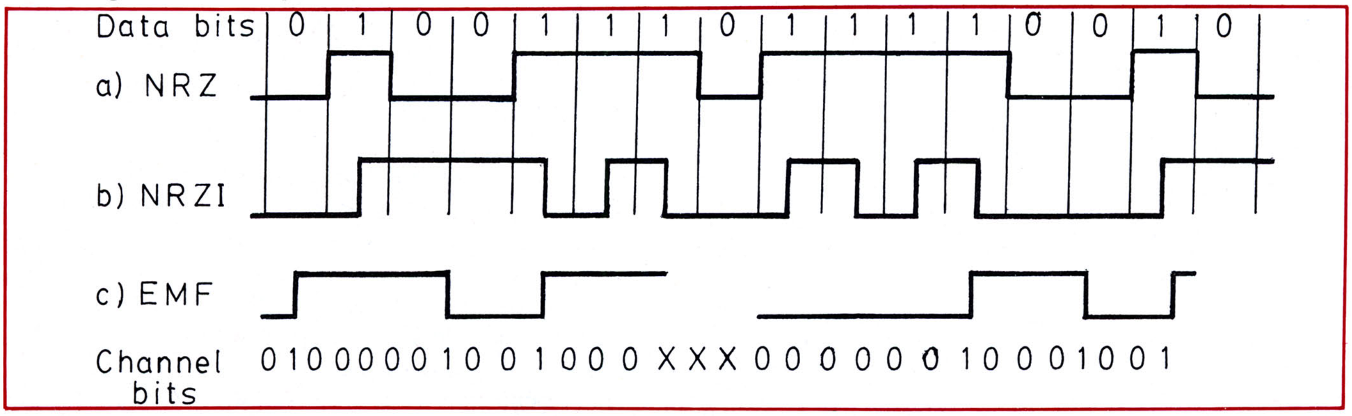


Fig. 8. Examples of modulation or channel codes used in digital recording (after Ref. 11).

patterns, only 267 of them satisfy the specified requirement of having 2 to 10 binary 0's in succession, but this is more than enough to code the 256 (2^8) possible 8-bit blocks. Blocks of 8 bits are translated into blocks of 14 bits using a conversion table.[12]

An effective error-correction system is essential in a digital disc system. Two types of errors must be corrected. The first includes random errors on the disc caused by inaccurate cutting, etc. The second are burst errors caused by dirt and scratches on the disc. The CIRC (cross interleave Reed-Solomon code), used in compact discs, deals with both types of errors.

Details of the CIRC are beyond the scope of this discussion. The code is a powerful one, capable of correcting bursts of over 8000 consecutive recorded bits. Errors up to 28 200 bits can be partially corrected by interpolation and estimation of erroneous bits.[13]

Data on a compact disc are formatted in *frames*. A frame contains 408 bits of audio data along with parity bits, sync bits, and a subcode, making a total of 588 bits. To scatter possible errors, the symbols recorded in each frame originate from different frames.

Sampling and Oversampling

Recording of an analog signal (such as music) in a digital format always begins with sampling and digital conversion. To represent a continuous waveform as a sequence of numbers, one must sample the waveform at regular intervals (44 100 times per second in the compact disc system) and express these samples as numbers. The conversion of the sample voltages to digital numbers is done in an analog-to-digital converter (ADC).[14]

According to Nyquist's sampling theorem, 2R samples per second are needed to represent a waveform with a maximum frequency R. If the waveform has components of frequency greater than R, a rather serious condition called *aliasing* or *foldover* can occur. Frequencies f_i above R will lead to foldover frequencies $f_F = 2R - f_i$ in the sampled output. Thus, to prevent aliasing, it is important to remove frequencies above R by means of a low-pass anti-aliasing filter.

To avoid the use of sharp cutoff filters and to improve the signal-to-noise ratio, some manufacturers have introduced digital filters with oversampling, as shown in Fig. 9. Rather than suppress high-frequency components after the signal has been converted back to analog form by the digital-to-analog converter (DAC), a digital oversampling filter is added before the DAC. In this filter, each 16-bit word is multiplied four times with different coefficients. The product of each multiplier is 28 bits; these products are summed, and a weighted average of a larger number of samples is obtained. When a signal is sampled four times, the noise density is reduced by four, resulting in a 6-dB improvement in the signal-to-noise ratio.[15]

Compact Disc Players

A compact disc player is considerably more complicated than a phonograph disc player. A variable-speed servo motor rotates the disc, and an optical pickup reads the recorded data. Servo drives, controlled by electronic circuits, position the optical pickup and also the lens that focuses the laser beam to a tiny spot on the reflective surface. Demodulation, demultiplexing, and error correction circuits assemble the digital data, and digital-to-analog converters translate the digital data to an analog signal suitable for audio amplifiers. A block diagram of a compact disc player is shown in Fig. 10.

A considerable amount of sophisticated electronic circuitry is incorporated into a compact disc player. Several microprocessors are included. Most players allow the user to select a song anywhere on the disc or to program the selections to be played in any desired sequence. This is possible because the optical pickup is positioned by a servo motor rather than by following a fixed groove. Since the basic mechanism is fairly standard, much of the difference between various compact disc players is in the sophistication of their control circuitry.

Sound Reproduction

Finally, we come to the "bottom line." How much better does a compact disc reproduce music? That is a difficult question to answer, because there is so much more to high-fidelity sound reproduction than merely the choice of recording medium. In most home systems, the recording medium is probably not the main factor that limits fidelity. I tried to address this question in a series of five articles that have appeared in this journal.[16]

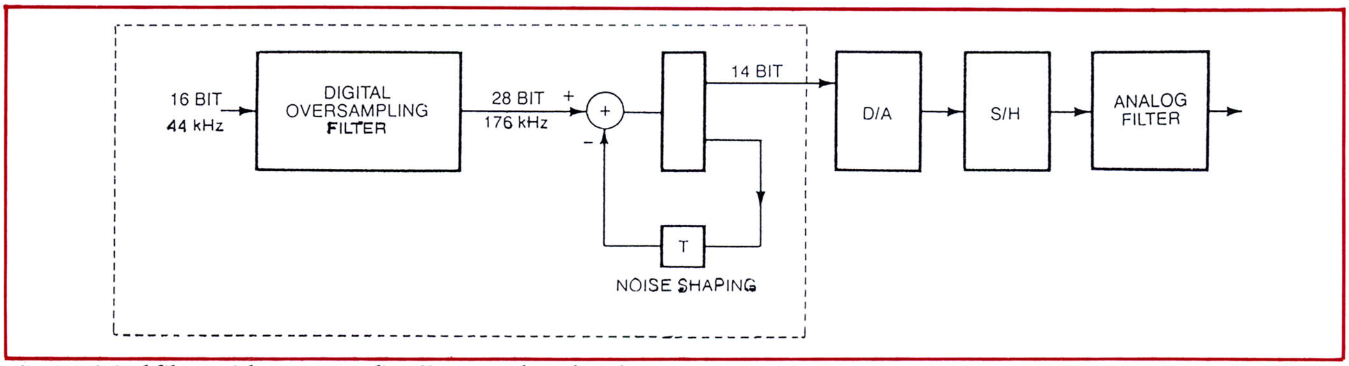


Fig. 9. Digital filter with oversampling.[20] (Reproduced with permission of the publisher, Howard W. Sams & Co., Indianapolis, *Principles of Digital Audio* by Kenneth C. Pohlmann, ©1985.

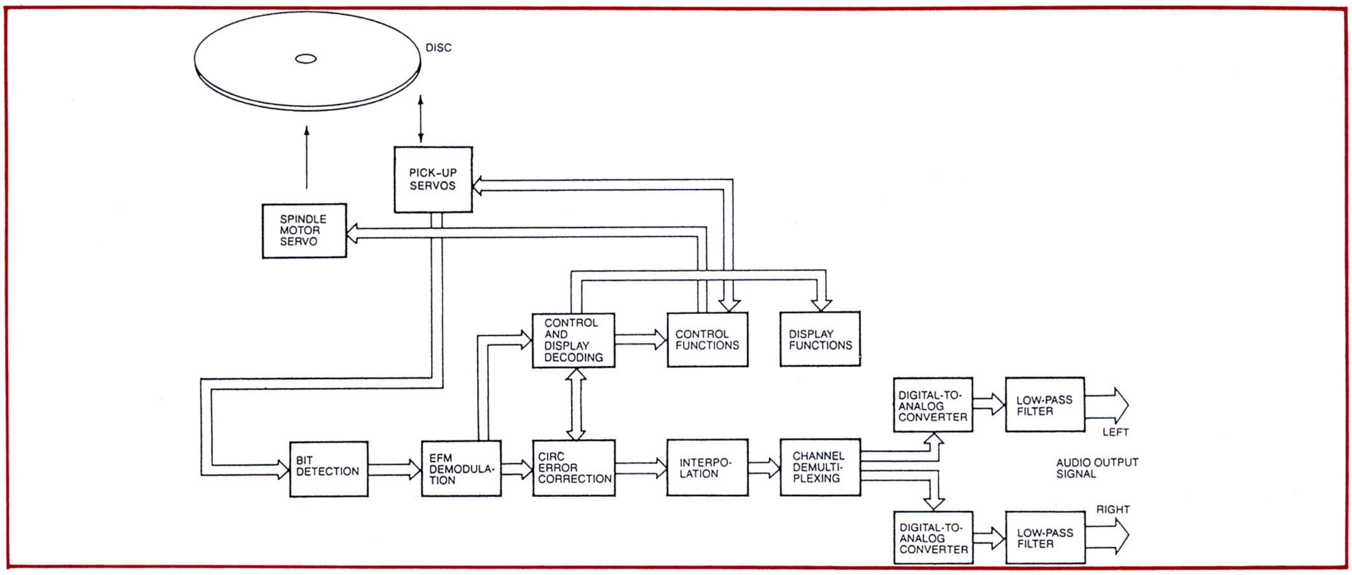


Fig. 10. Block diagram of a compact disc player.[21] (Reproduced with permission of the publisher, Howard W. Sams & Co., Indianapolis, *Principles of Digital Audio* by Kenneth C. Pohlmann, ©1985.

Nevertheless, compact discs offer great potential for high-fidelity sound reproduction, as I pointed out earlier in this article. It is a little amusing to me that manufacturers are now putting more than a little effort into methods for dynamically compressing the impressive 90-dB dynamic range of compact discs to a more manageable figure for apartment dwellers and especially for automobiles![17]

Erasable Compact Discs

Although present compact disc systems are intended to play prerecorded mass-produced discs, considerable research has been done on erasable discs. Most such research has focused on magneto-optical thin films such as GdFe, GdCo, and GdFeTb. The recording of information is done by locally heating the amorphous magnetic layer with a laser while applying a small external magnetic field.[18] The information density is about 40% of that obtained from a compact disc.

Another method for recording directly on a compact disc is to have the laser actually burn a tiny hole in the coating by ablation. Thin films of tellurium appear to be a promising material; a hole 1 μm in diameter could probably be opened in 50 ns with a 20-mW laser.[19] This type of writing is suitable for data storage discs of the WORM ("write once, read many") type.

Although compact disc home recorders may be a few years off, compact disc players are with us now, and they should be understood as well as enjoyed!

Acknowledgment

Thanks are due to my students who ask me so many good questions about compact discs that I could not avoid writing this article.□

References

1. The Philips VLP (video long-play) system is described in four articles in *Philips Tech. Rev.* **33**, 178–193 (1973). An interesting sidelight is that the VLP was first suggested by members of the Education Group as a way of disseminating visual information.
2. K.C. Pohlmann, *Principles of Digital Audio* (H.W. Sams, Indianapolis, 1985), Chapter 7.
3. I. Hayashi and M.B. Panish, "GaAs-Ga$_x$Al$_{1-x}$As Heterostructure Injection Lasers Which Exhibit Low Thresholds at Room Temperature," *J. Appl. Phys.* **41**, 150–163 (1970).
4. M.B. Panish and I. Hayashi, "A New Class of Diode Lasers," *Sci. Am.* **225** (1), 32–40 (1971).
5. 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).
6. Panish and Hayashi, p. 32.
7. Fink, Van der Laak, and Schrama, p. 30–39.
8. T.G.J. van Oirschat, W.J. Leswin, P.J.A. Thijs, and W. Nijman, "LPE Growth of DH Structures with the Double Source Method," *J. Crystal Growth* **45**, 262 (1978).
9. K. Pohlmann, "Aural Arguments," *Audio* **70** (7), 18–19 (1986).
10. Pohlmann, Chapter 3.
11. T. Doi, "Channel Codings for Digital Audio Recordings," *J. Audio Eng. Soc.* **31**, 224–238 (1983).
12. S. Miyaoka, "Digital audio is compact and rugged," *IEEE Spectrum* **21** (3), 35–39 (1984).
13. Miyaoka, p. 37–38.
14. T.D. Rossing, *The Science of Sound* (Addison-Wesley, Reading, MA, 1982), Chapter 28.
15. Pohlmann, Chapter 4.
16. T.D. Rossing, "Physics and psychophysics of high-fidelity sound," Part I, *Phys. Teach.* **17**, 563 (1979); Part II, *Phys. Teach.* **18**, 278 (1980); Part III, *Phys. Teach.* **18**, 426 (1980); Part IV, *Phys. Teach.* **19**, 293 (1981); Part V, *Phys. Teach.* **22**, 84 (1984).
17. E.F. Stilvoort, "Digital Dynamic Range for Audio," *J. Audio. Eng. Soc.* **34**, 3–9 (1986); see also W.M. Waagenaars, A.J.M. Houtsma, and R.A.J.M. van Lieshout, "Subjective Evaluation of Dynamic Compression in Music," *J.Audio Eng. Soc.* **34**, 10–18 (1986).
18. K.A. Schouhamer Immink and J.J.M. Braat, "Experiments Toward an Erasable Compact Disc Digital Audio System," *J. Audio Eng. Soc.* **32**, 531–538 (1984).
19. R.M. White, "Disk-Storage Technology," *Sci. Am.* **243** (2), 138–148 (1980).
20. Pohlmann, Chapter 4.
21. Pohlmann, Chapter 7.