INTRODUCTION

Why Light?

It is no coincidence that we interact with the world primarily through visual means—light has many ideal properties for communication of information. These same properties make light ideal for data storage. The wavelength of visible light is small (on the order of 0.5 microns or about 1/7000 the height of a letter written in 10-point font), which indicates it can be used to read or write a large amount of data per unit area.

Unlike direct-contact methods, such as printing or photography, light can access the storage medium from a distance, and so it does not risk damage to the data. This factor is particularly important when the medium is being moved very quickly, which is required both for rapid access and high read/write speed. Thus, optics has fundamentally attractive qualities to enable high density, fast access, rapid-transfer rate and long-term reliability. These aspects are the primary user requirements for any data-storage system.

These qualities were all exploited in microfiche, which was the first large-scale system for read-only optical data storage, distribution, and retrieval. The analogies with later optical disks are striking. Microfiche masters are first created by precision machines that photographically reduce the document to be distributed. These masters are then replicated by contact optical printing to make many inexpensive distribution copies. These distribution copies are shipped to end users such as libraries, where high information density and low cost make them much preferable to paper copies of the original text. Finally, individuals access the data rapidly by loading a fiche into a two-dimensional scan table and an optical projector. Many of these reading machines are equipped with copiers to print an image back to paper, which completes the distribution loop.

The need for a digital version of this system for distribution of audio and video data was financially motivated by the entertainment industry and made possible by the development of semiconductor electronics. The distribution needs are similar to the microfiche model—possibly high-cost mastering equipment to enable high-volume, inexpensive reproduction of distribution media, which will be read by low-cost machines to reproduce the original content. The primary difference is that microfiche provides manual access to two-dimensional images, whereas the entertainment application needs fully automatic access to primarily serial data, such as audio. This need suggests the use of a single “track” wrapped in a spiral onto a two-dimensional disk spun by a motor, which of course describes mechanically fabricated and mechanically accessed phonograph records.

An “optical phonograph” is not possible with conventional light sources, such as incandescent lamps, because of a version of the law of conservation of energy called the “conservation of radiant.” The radiance of a light source is given by its power in watts divided by its area and the solid angle into which it radiates. For example, a typical light bulb might radiate 10 watts from a mm² filament into a 2π hemisphere, which results in a radiance on the order of 10⁶ W m⁻² Sr⁻¹. The conservation of radiant does not allow any optical system such as a lens to increase this quantity because as the area of a focus is reduced, the angular divergence is increased proportionally. Thus, as one attempts to focus a light bulb onto a small spot, power is inevitably lost because the maximum solid angle the lens can capture is limited. For the example lamp, the maximum power that could be delivered into a μm² spot is roughly a microwatt, which is insufficient power to transfer data at audio rates.

The situation changed in 1960 with Maiman’s demonstration of light amplification through the stimulated emission of radiation (“laser”). It is often reported in the popular literature that laser light differs from conventional light sources by being of a single frequency. This is not necessarily true—the lasers used in optical disk players oscillate simultaneously at a set of frequencies. The primary interesting property of lasers as a light source is their radiance. The semiconductor lasers used in consumer read/write disk players today have roughly 10 mW of optical power, which is a factor of 1000 lower than our hypothetical light bulb, but they radiate this light from an aperture of a few square microns with a divergence angle of tens of degrees. Their radiance is thus on the order of 10¹³ W m⁻² Sr⁻¹, which is ten million times larger than the lamp. If expanded to mm² area, the divergence can be very low, which is why laser pointers can deliver a bright spot across a large conference room. Similarly, when focused, virtually all laser power can be delivered to a spot size limited only by the physics of diffraction to be roughly the size of the wavelength. Although the few milliwatts of total optical power is low, the intensity (formally irradiance) in watts per square meter is a million times greater than the surface of the sun. This high intensity is used during writing to heat the surface of the disk by hundreds of degrees Celsius in just a few nanoseconds.
Optical Audio Distribution: From Mouth to Ear

A breathtaking number of technologies are used in the complete process of delivering an accurate audio reproduction from a singer to a listener. Additional steps are required for capturing, coding, and delivering video or other data, but the audio example captures most of the fundamental steps, so this section will follow the process from beginning to end both as motivation for the specification of an optical drive and to provide an introduction to the following sections, which cover the broad field of optical data storage variants.

The process begins with an audible frequency in the range of 20 Hz to 20 kHz, which perhaps could be produced by a group of musicians. These musicians produce atmospheric pressure waves, which in turn displace the diaphragm of two microphones that generate two electrical voltages ideally proportional to air displacement. These stereo analog (continuous in amplitude and time) voltages could be directly recorded on an analog recording medium, such as a phonograph or magnetic audio tape. Such analog recordings have the potential to provide perfect fidelity but are also subject to degradation, which results in noise that cannot be separated from the sound signal.

Digital systems, in contrast, sample the analog audio signal to produce a pulse-code modulation (PCM) digital approximation. This analog-to-digital conversion, which is often referred to as simply “digitization,” is described by two parameters: quantization and sampling rate. The Shannon sampling theorem states that the sampling rate must be at least twice the rate of the highest audio frequency to be transmitted or 40 kHz samples/second. The compact disk (CD) standard of 44.1 kHz was chosen for compatibility with recording technology at the time the standard was established, and thus the standard can theoretically deliver audio frequencies up to 22.05 kHz. The quantization refers to the number of discrete sound levels that are recognized and is described by the number of binary bits required to represent all possible levels. Early in CD history, considerable debate surrounded the use of $2^{14} = 16,384$ or $2^{16} = 65,536$ quantization levels, with the higher fidelity 16-bit system eventually becoming the standard. These $2^{16}$ voltage levels give a maximum power signal to quantization noise ratio of $2^{10}$ or 96 dB, which is adequate for all but the most discerning listeners.

This digitized approximation of the stereo analog audio must be stored as two 16 bit $\times 44.1$ kHz $= 88.2$ kbytes/second data streams. The CD audio format does not use compression (such as the now-familiar MP3 algorithm), and thus it does not lose any information but does require storage of the complete signal. This signal is divided into 6-sample frames of 2 channels $\times 2$ bytes/sample $\times 6$ samples $= 24$ bytes, to which a subcode byte for control is added. In addition, eight bytes of cross-interleave Reed-Solomon parity data are added to each frame to correct errors caused by scratches or other noise sources, as shown in Fig. 1. Each of the total 33 bytes is encoded with the eight-to-fourteen modulation (EFM) and three bits of merging word yielding $33 \times (14 + 3) = 561$ bits. A 27-bit synchronization word completes the total of 588 bits in a frame, only 192 of which are audio data. The total data rate required is thus 588 bits/6 samples $\times 44.1$ kHz $= 540$ kbytes/second, which is three times the 176 kbytes/sec required for the digitized audio and about 10 times the bandwidth required for the original analog stereo signal. To store the standard 74 minutes of music on a CD thus requires a total capacity of 2.4 Gbytes of raw (total) capacity to store the 780 Mbytes of audio (user) data. Note that data formats such as CD ROM typically store less that this total, roughly 650 MByte, because of the addition of extra error correction and header data, which reduces the total available to the user.

These $2 \times 10^{10}$ bits are then written to the master, which will determine the physical structure of the replicated CDs. This process is performed by a laser recorder that has much the same structure as a consumer read/write disk drive (shown in Fig. 2) but with more precise mechanics and higher resolution via a smaller wavelength and a more powerful objective lens. The mastering disk is a polished glass plate that is wider and thicker than the eventual CD to help maintain flatness. This disk is rotated by the spindle motor while the writing optics are slowly moved outward. The focused laser is modulated with the binary data stream in order to inscribe the data as $\sim 100$-nm deep pits arranged in a continuous spiral over 5 kilometers in length, as shown in Fig. 3. The pits are formed in a spin-coated layer of photoresist, which is developed after optical writing, or a dye-polymer that is immediately vaporized by the writing laser. In the later case, a direct read after write (DRAW) laser can follow the writing spot to maintain and verify the disk quality. A $\sim 400$-nm layer of nickel is then...

![Figure 2](image-url)
vapor deposited onto the glass surface to provide the foundation for additional electroformed nickel, which is built up to a thickness of 300 µm. This nickel father is separated from the glass and used as a pattern to plate mirror-image mothers, which in turn are used to grow the sons, which are the same polarity as the fathers.

The replicated CDs are formed at a rate of one every few seconds using the sons as stampers for injection molding. Liquid polycarbonate is injected against the stamper, and a press closes to force the liquid into the contact with the pattern. Water cools the metal mold to solidify the polycarbonate; the mold opens and vacuum fixtures remove the green disk. After more cooling, the data side is sputtered with aluminum, which forms the reflective layer. A ~70-µm thick ultraviolet (UV)-cured lacquer is spun on to protect the aluminum and provide a printing area for the label. A cross-section of the resulting structure is shown in Fig. 4 (top). The 1.2-mm polycarbonate platter serves as the mechanical substrate; the replicated data layer (top) and the optical window (bottom) are placed at a sufficient distance from the data layer that dust and scratches are out of focus and do not cause read errors.

This replicated audio disk is then distributed to a listener who inserts it an audio CD player. The load mechanism clamps the spindle hole onto the motor spindle, and the disk is brought up to approximately 500 revolutions per minute (RPM) or a linear velocity of 1.3 m/second at the inner data radius of 25 mm. The infrared, semi-conductor diode laser is turned on and focused by the objective lens to a small spot that enters through the transparent face of the 1.2-mm polycarbonate disk, as shown in Fig. 4. This spot is very small—assuming the illumination uniformly fills the lens, the full width to half maximum (FWHM) of the optical intensity is given by 0.515 λ/NA, where NA is the numerical aperture of the lens, which is given by the sine of the lens radius over its effective focal length. CD players use lasers with λ = 780 nm and lenses with NA = 0.47, which yield a spot FWHM of 0.85 µm, as shown in Fig. 3, right.

To recover the data, this focused spot must be placed on a track to within 100 nm of the track center in the radial direction and 500 nm in depth. Because the placement of the disk on the hub and the flatness of the plastic disk cannot guarantee these tolerances, tracking and focus sensors are used to detect the offset of the focus from the disk in both dimensions, and the position of the objective lens is adjusted with voice-coil actuators driven by a closed-loop servo system. Figure 5 shows one means of detecting this radial offset. It is remarkable that these tolerances, which are measured in 100s of nm, can be maintained while the listener is jogging with the player, for example.

Once data is acquired, a clocking signal is recovered and compared with the desired raw data rate of ~540 kbytes/second to set the spindle rotation rate. This causes the spindle rotation to drop from roughly 500 RPM at the inner radius to approximately 200 RPM at the outer edge of the disk to maintain a constant linear velocity (CLV). This choice maximizes capacity since bits are always stored at the minimum size, but sacrifices data rate (which could be larger at the outer diameter) and access time (as the spindle

Figure 3. CD layout (left) and track structure (right). The optical intensity of the focus is shown in red and pit boundaries as white contours. Note that the infrared laser used for CD is beyond the typical visible spectrum.

Figure 4. Cross-section of a compact disk. Because of the replication sequence from master to son to disk, the pit formed in the master appears as a plateau to the laser. The pits diffract the light out of the lens aperture (see Fig. 5), which cause the read signal to decrease, as shown in the data signal, bottom. Micron-scale dust and scratches on the entrance face cause noise on the data but do not obscure the laser, which has mm-scale diameter at this surface.
speed changes). The alternative choice, constant angular velocity (CAV), has the opposite characteristics. An intermediate choice is zoned constant linear velocity (ZCLV), which maintains linear velocity in a set of radial regions somewhat like the transmission in a gasoline engine keeps the motor running at a small range of RPMs as the vehicle moves through a range of speeds.

Once tracking and clock rate are established, the focused laser light reflects from aluminum data layer, which is a replica of the original data pattern written by the mastering laser light reflects from aluminum data layer, which is somewhat like the transmission in a gasoline engine keeps which maintains linear velocity in a set of radial regions.

As shown in Fig. 3, this layer consists of a series of "pits" that actually protrude into the polycarbonate layer system. As shown in Fig. 3, this layer consists of a series of replica of the original data pattern written by the mastering laser light reflects from aluminum data layer, which is a somewhat like the transmission in a gasoline engine keeps which maintains linear velocity in a set of radial regions.

The lower plots show how radial runout can be detected—light is redistributed within the lens aperture and can be detected by the difference of signals from a two-segment photodiode.

Figure 5. Calculated reflected intensity distribution at the objective lens when the focus is on a smooth land (upper left) or a 100-nm deep, 500-nm wide pit (others). The lens aperture, which is shown in white, determines what light continues on to the detector. The lower plots show how radial runout can be detected—light is redistributed within the lens aperture and can be detected by the difference of signals from a two-segment photodiode.

The photodetector converts the optical power into a current that is amplified by a transimpedance amplifier, which results in an analog voltage that is an approximation of the signal that was used to drive the laser in the mastering system. It is only an approximation because of both the timing jitter and the addition of noise from dust, scratches, laser fluctuations, and so on. A decoder examines this analog voltage at the clock period and digitizes it into a binary signal. If the noise and jitter are minimal, this signal is exactly the original 540 kbyte/second data that were recorded on the master. If not, the eight parity bytes per frame are now used to correct the errors, if possible. Errors that cannot be corrected are disguised in the final audio signal by interpolation between the known signals adjacent to the unrecoverable data or, if the gap is too lengthy, by blanking.

The encoding process is now reversed to extract the 192 bits equals 24 bytes of audio data from each frame. These data represent six samples of the left and right channel or about 140 microseconds of audio. These data are stored in a digital memory buffer that provides several seconds of uninterrupted audio to hide skips or tracking errors that would otherwise interrupt the playback. Digital samples are retrieved from this memory, which are converted to an analog voltage by a digital-to-analog converter, amplified, and sent to the speakers, which convert the voltages into sound waves for the left and right ear.

TECHNICAL DETAILS

A defining capability of optical data storage is read/write access from mm-scale distances and tolerance to surface wear and contamination. Equally important is the ability of a laser to be focused with high intensity to a submicron spot, which enables storage density measured in megabits/mm².

Data storage on disks is convenient because large area—8,600 mm² in the case of CD—can be moved past this spot at high linear velocity for fast transfer rate via simple
TABLE 1. PRIMARY PHYSICAL AND PERFORMANCE SPECIFICATIONS FOR 1×, ROM VERSIONS OF THREE OPTICAL DATA STORAGE GENERATIONS

<table>
<thead>
<tr>
<th>Target application</th>
<th>Audio</th>
<th>Video</th>
<th>HD Video</th>
<th>Units</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layers per side</td>
<td>Single</td>
<td>Single</td>
<td>Single</td>
<td>Double</td>
<td></td>
</tr>
<tr>
<td>Layers wavelength</td>
<td>780–790</td>
<td>635–6</td>
<td>405±15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lens NA</td>
<td>0.47</td>
<td>0.6</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser spot size</td>
<td>0.85</td>
<td>0.56</td>
<td>0.25</td>
<td>μm</td>
<td>Full-width to half max intensity</td>
</tr>
<tr>
<td>Cover thickness</td>
<td>1.2</td>
<td>0.6</td>
<td>0.1</td>
<td>mm</td>
<td>Surface to first data</td>
</tr>
<tr>
<td>Spacer thickness</td>
<td>None</td>
<td>55 ± 15</td>
<td>25 ± 5</td>
<td>mm</td>
<td>First data to second data</td>
</tr>
<tr>
<td>Working distance</td>
<td>1.9</td>
<td>1.7</td>
<td>0.5</td>
<td>mm</td>
<td>Lens to disk surface, typical</td>
</tr>
<tr>
<td>Data radii</td>
<td>25–58</td>
<td>24–58</td>
<td>24–58</td>
<td>mm</td>
<td></td>
</tr>
<tr>
<td>Land reflectance</td>
<td>70</td>
<td>45–85</td>
<td>35–70</td>
<td>% ROM</td>
<td></td>
</tr>
<tr>
<td>Pit width</td>
<td>0.5</td>
<td>0.3</td>
<td>0.13</td>
<td>μm</td>
<td></td>
</tr>
<tr>
<td>Pit depth</td>
<td>100</td>
<td>120</td>
<td>70</td>
<td>nm</td>
<td>Typical</td>
</tr>
<tr>
<td>Minimum pit length</td>
<td>0.83</td>
<td>0.4</td>
<td>0.15</td>
<td>μm</td>
<td>3T</td>
</tr>
<tr>
<td>Track pitch</td>
<td>1.6</td>
<td>0.74</td>
<td>0.32</td>
<td>μm</td>
<td></td>
</tr>
<tr>
<td>Capacity</td>
<td>0.65</td>
<td>4.7</td>
<td>25</td>
<td>GB ROM</td>
<td></td>
</tr>
<tr>
<td>Transfer rate</td>
<td>0.15</td>
<td>1.385</td>
<td>4.5</td>
<td>MB/s ROM, 1×</td>
<td></td>
</tr>
<tr>
<td>Linear velocity</td>
<td>1.3</td>
<td>3.49</td>
<td>4.9</td>
<td>m/s ROM, 1×</td>
<td></td>
</tr>
<tr>
<td>Rotation rate</td>
<td>200–500</td>
<td>580–1400</td>
<td>800–1950</td>
<td>RPM ROM, 1×, in data region, typ.</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Major Milestones in the History of the CD

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>VHS</td>
</tr>
<tr>
<td>1978</td>
<td>Laser disk</td>
</tr>
<tr>
<td>1980</td>
<td>CD audio standard “Red Book”</td>
</tr>
<tr>
<td>1982</td>
<td>CD audio drive, ~100 titles</td>
</tr>
<tr>
<td>1983</td>
<td>CD-ROM standard, 800k disks sold</td>
</tr>
<tr>
<td>1984</td>
<td>Automobile and portable CD players</td>
</tr>
<tr>
<td>1985</td>
<td>CD-ROM, 22M disks sold</td>
</tr>
<tr>
<td>1986</td>
<td>CD-I</td>
</tr>
<tr>
<td>1987</td>
<td>100M disks sold</td>
</tr>
<tr>
<td>1988</td>
<td>CD production exceeds LPs</td>
</tr>
<tr>
<td>1990</td>
<td>CD-R</td>
</tr>
<tr>
<td>1993</td>
<td>Electronic skip protection for portables</td>
</tr>
</tbody>
</table>

Nonetheless, the compact disk was built on the lessons learned from the laser disk with the addition of digital encoding made possible by the rapidly developing integrated circuit industry. Although battles occurred over the diameter of the disk, sampling frequency, and other features, the standard setting process generally happened without the contentious battles that took place in later generations, which resulted in the “Red Book” for CD-Audio from Philips and Sony who control the format. The market driver was to replace the venerable audio long playing (LP) vinyl record, which was first released in 1948, with a format that was smaller, offered more listening time and used digital sampling to deliver high quality that did not degrade with use.

The primary component of the CD drive is the semiconductor laser diode. Charge injected into a pn junction creates an inverted population of electrons in the conduction band that, when stimulated by a passing photon, emit an identical photon to amplify the optical signal. When placed in a resonant cavity, often formed via cleaved facets on each end of the laser, round-trip amplification can exceed the loss, and the device will spontaneously oscillate, which produces a high-radiance, coherent optical beam. This optical beam can be directly by modulated the injection current, which enables both feedback power control and simple switching for high-speed writing.

The laser diodes available at the time the CD standard was written had a wavelength of 785 nm. The only other variable that impacts spot size (and thus capacity and transfer rate) is the NA of the objective lens. Unfortunately, cover thickness accuracy and sensitivity to disk tilt scale as the fourth and third power of NA, respectively, which causes rapid increases in disk and drive cost as NA is increased. The CD NA of 0.47 balanced disk capacity and manufacturing cost and results in a FWHM spot size of 0.85 microns.

This value may be immediately used to estimate the capacity of the CD and the required rotation rate to meet the audio bandwidth needs. As described in the introduction, the EFM modulation code and CIRC error correction parity data adds sufficient overhead to the sampled audio to triple the needed storage (588 channel bits/192 audio bits = 3.06). Also, recall that the RLL recording scheme compresses three channel bits into 0.83 microns along the track, which almost exactly equals the optical spot size. These two coding steps cancel such that the user capacity can be accurately estimated simply by dividing the disk data area by the spot size time the track pitch:

\[
\frac{\pi (582 - 252) \text{mm}^2}{(1.8 \mu m) (0.85 \mu m)} = 6.33 \text{ Gbit} = 0.79 \text{ Gbytes}
\]

which is close to the CD audio total of 0.78 Gbytes. Note that the value in Table 1 of 0.65 Gbytes is for the ROM (data) format, which adds additional overhead to decrease the error rate. Similarly, we may estimate the linear speed (and thus rotation rate vs radius) that must be maintained to meet the 176 Mbytes/sec digital audio bandwidth requirements:

\[
(176 \text{ Mbytes/sec}) \left( 8 \frac{\text{bits}}{\text{byte}} \right) \left( 0.85 \frac{\mu m}{\text{bit}} \right) = 1.2 \text{ m/sec}
\]

which is the minimum velocity in the CD audio specification.

These results provide a fundamental understanding of how optical storage specifications scale with the two primary design variables available, which are laser wavelength and objective lens NA. As will be shown in the following subsections, whereas gains can be made in more efficient codes or better servo for finer relative track spacing, the primary driver is spot size, which in turn is determined by wavelength and NA. One immediate conclusion is that capacity per layer will increase inverse to the square of the laser spot size and by additional factors of two and four for two-layer or/and double-sided disks. Transfer rate, however, increases only inverse to the laser spot size and does not depend on the number of layers. Transfer rate for optical data storage has thus grown much more slowly than capacity. Although this rate has been adequate for entertainment distribution applications, it is not appropriate for high-performance data storage. Magnetic disk and tape systems avoid this same scaling problem by using multiple read/write heads. This method has been applied with very limited success (Zen Multibeam) in optical storage and only for reading.

Variations within CD

The basic CD- digital audio (DA) format generated many technical and application variants. Many of these were released as official standards from the CD consortium, as shown in Table 3. Like the “Red book,” these standards are still proprietary and available for license under a nondisclosure agreement, which limits the availability of information in the public domain. Some formats were generated precisely to avoid these standards and the associated royalty fees. A complete discussion of every format announced by the large number of companies in this space would be a lengthy article on its own, so a selection of those with the largest impact is given in the discussion below.

Red Book (DA). The first specification created the CD-digital audio format that was described in the introduction.

Yellow Book (ROM). The acronym ROM for “read only memory” was not entirely appropriate for a digital distribution medium, but it was common from integrated circuits,
such as PROMs. It was built “on top of” the digital audio (Red book) standard and quickly became popular for the distribution of arbitrary digital data, particularly software. Unlike the audio application, in which bit errors could be concealed by interpolation or blanking, digital ROM requires nearly error-free performance. This is accomplished by adding another layer of Reed-Solomon error correction overhead in addition to sector identification and other control data required for the random-access data application, as shown in Fig. 6.

**Orange Book (Writable).** Given that the CD format was conceived as a replicated distribution system, the development of writable formats was a remarkable innovation. A complete understanding of the physics and engineering of these technologies is outside the scope of this article, which focuses primarily on the distribution format, but they are summarized here for completeness.

The magneto-optic (MO) format uses the very high intensity of the laser to rapidly heat a ferromagnetic layer beyond its Curie point, which enables an electromagnet on the opposite side of the disk to change the local magnetic polarization. Because data can only be detected by the Kerr rotation of the reflected polarization, not total intensity, the format was incompatible with existing drives and was never commercially released.

The –R or write once read many (WORM) format uses a thin layer of photo-sensitive dye between the polycarbonate substrate and the reflective layer, which is typically silver or gold to avoid corrosion by the dye. During writing, the laser is increased to roughly 5 mW (10-fold above the reading level) to “burn” the dye, which creates an amplitude and/or phase change at the focus, which can then be read similar to a CD-ROM, although at reduced contrast, which can make the disks unreadable in some players. Because stamped data that provided the radial tracking information for the original CD is not present, CD-R disks are stamped with grooves to guide the head. In addition, variations in the spindle motor speed, which in the CD-ROM format could be accommodated through buffering, must now be held to a tolerance similar to a laser mastering machine. The groove is therefore wobbled to provide an absolute time measurement (ATIP) as well as to store information about the disk type.

The –RW or rewritable format was known as –E during development, because it can be erased. It is based on a phase-change alloy, which will cool to an amorphous state when melted at ~400°C but can be returned to a crystalline state when annealed at a lower temperature. These two states can be created by the writing laser and read at low power by detecting their differing reflectivity. The lower total reflectivity makes these disks incompatible with some drives.

**White Through Green Books (Applications).** These formats are built off of the foundational physical specifications to create new applications.

**Purple Book (Double Density).** The original CD standard was somewhat conservative on manufacturing specifications, particularly track pitch, as can be observed in Fig. 3. Foreshadowing the DVD, this format doubled the capacity by halving the area per bit to a 1.1-μm track pitch and 0.62-μm minimum mark length. It was overcome by the greater performance of DVD.

**Scarlet Book (Super Audio).** See “Hybrid formats” in next section for this CD/DVD hybrid.

**Super Video CD and China Video Disk.** This format was a successor to the White book video-CD format sponsored by the Chinese government to avoid royalties on the DVD format. These formats have been combined into a single standard, the Chaoji Video CD as the official Chinese video format.

**Proprietary Formats for Gaming.** The purpose of cross-industry standards is to enable disks produced by one manufacturer to be read by another’s player. In the case of game software for a proprietary console, the situation is reversed; by switching to a custom format, the game manufacturer makes duplication more difficult. In particular, the lack of writing hardware and blank media are significant obstacles to the average user. The primary CD version of this type is the GigaDisk ROM (GD-ROM), which was developed for the Sega DreamCast. Track pitch is decreased to obtain approximately 1.2 GB total capacity in a CAV mode.

**Mini CD.** The mini CD is an 80-mm diameter disk readable by most players using a well in the disk loading tray. The format is used for some digital cameras or other

---

**Table 3. CD “Rainbow Book” Standards**

<table>
<thead>
<tr>
<th>Book</th>
<th>Formats</th>
<th>Content or Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>CD-DA</td>
<td>Digital audio</td>
</tr>
<tr>
<td>Yellow</td>
<td>CD-ROM</td>
<td>Digital data storage</td>
</tr>
<tr>
<td>Orange</td>
<td>CD-MO, -R, -RW</td>
<td>Various writable</td>
</tr>
<tr>
<td>White</td>
<td>VCD</td>
<td>Video</td>
</tr>
<tr>
<td>Blue</td>
<td>E-CD, CD+, CD+G</td>
<td>Audio plus data, video</td>
</tr>
<tr>
<td>Beige</td>
<td>PCD</td>
<td>Photo</td>
</tr>
<tr>
<td>Green</td>
<td>CD-i</td>
<td>Interactive</td>
</tr>
<tr>
<td>Purple</td>
<td>DDCD</td>
<td>Double density</td>
</tr>
<tr>
<td>Scarlet</td>
<td>SACD</td>
<td>Super audio</td>
</tr>
</tbody>
</table>

---

1 Sector = 2352 bytes = 98 frames of 24 user bytes per frame

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12 sync</td>
<td>4 sector</td>
<td>2048 user data</td>
<td>4 error detect</td>
<td>8 blank</td>
</tr>
</tbody>
</table>

---

Figure 6. CD-ROM sector layout. In all, 98 digital audio frames as shown in Fig. 1 are grouped into a sector that contains 2352 bytes of user data stored as 98 × 33 = 3234 total bytes. The additional overhead for ROM reduces the available capacity per sector to 2048 user bytes and the total disk capacity to 780 Mbyte × 2048 / 2352 = 680 MByte.
portable applications or for small-content distribution, such as software applications or music singles. The format should not be confused with the MiniDisc (MD), which is a magneto-optic format not in the CD family.

**Shaped CD.** These CDs are similar to a mini CD but formed in noncircular shapes to be eye-catching as marketing novelties. The data region is restricted to the largest continuous circle that fits within the disk outline.

**Generation 2 – DVD**

Although CD formats for video were available, in the early 1990s major motion-picture studios began to push for a format that could hold an entire movie at high quality without the need to replace or flip the disk. As shown in Table 4, two formats, the MultiMedia CD and the Super Density CD, were in development by rival industrial consortia. The Technical Working Group, which was lead by IBM, demanded that both sides agree to a common format, leading to the virtually complete adoption of the Super Density CD proposal under the name DVD. DVD has variously stood for “digital video disk” “digital versatile disk” and (officially) nothing at all.

The primary improvements from DVD to CD are reduced laser spot size and the capability for multiple layers. Laser development during the 1980s had pushed the feasible wavelength down to 640 nm (red), and improved manufacturing allowed the performance versus cost tradeoff to be shifted to a higher lens NA of 0.6. Together, these improvement decreased the spot size to 0.56 microns FWHM giving a potential 2.3 \times capacity increase over CD, as illustrated in Fig. 7. Improved servo systems and data channels were also designed so that the track pitch and minimum pit sizes could be reduced even more relative to the spot size to increase density by another factor of 1.9 as shown in Table 1. Using the simple estimation method, these specs result in single-layer capacity and transfer rate predictions of 3.7 Gbytes and 1.1 Mbytes/sec. These estimations are somewhat lower than the actual because of improved modulation and error-correction codes that reduced the overhead ratio.

The other major innovation in the DVD is the use of multiple layers and potentially multiple sides. As shown in Fig. 8, the optical cover layer has been reduced to 0.6 mm, which is one half the CD standard. This decreases the tilt sensitivity caused by the larger NA while also allowing for the possibility of double-sided disks in the standard 1.2-mm total thickness. The upper layer is partially reflective so that, when reading or writing the deeper layer, the laser can penetrate and reflections can return to the detector. The laser spot diameter on the out-of-focus layer is over 40 μm, which covers roughly 60 tracks. This keeps the intensity on the out-of-focus layer low and averages the signal from the other layer to a small noise term. The track spirals out from the hub on one layer, then it typically spirals back in on the second to minimize the seek time from one layer to the next.

**Variations within DVD**

Similar to the Rainbow books (Table 3) released to document the variety of CD formats, the DVD Forum has released a series of standards that specify the versions of the DVD specification, which are summarized in Table 5. Although the technology is improved in numerous ways, each is similar to its CD antecedent. Also like the CD standards, these are proprietary documents made available under license fee and a nondisclosure agreement; thus, not all the information is publicly available.

**Book A (ROM).** This book was described above and in Table 1. The B and C books define the disk layout and interface features for video (by far the largest application) and audio applications.

**Books D and E (Writeable).** The DVD-R is similar technologically to CD-R, which uses a dye to permanently inscribe marks in a pregrooved disk, although the dye is different because of the shorter wavelength. DVD-RW is analogous to CD-RW in that it uses a phase-change alloy to enable ~1000 write/erase cycles. Both single (4.7 GB) and dual (8.5 GB) layer formats are available. Sector and

<table>
<thead>
<tr>
<th>Year</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>MPEG-1 video format for CD-ROM</td>
</tr>
<tr>
<td>1994</td>
<td>MultiMedia CD standard announced</td>
</tr>
<tr>
<td>1994</td>
<td>MPEG-2 video format</td>
</tr>
<tr>
<td>1995</td>
<td>SuperDisk standard announced</td>
</tr>
<tr>
<td>1995/6</td>
<td>DVD Consortium and standard</td>
</tr>
<tr>
<td>1996</td>
<td>DVD Japanese release</td>
</tr>
<tr>
<td>1997</td>
<td>US release</td>
</tr>
<tr>
<td>1997</td>
<td>~500 titles available</td>
</tr>
<tr>
<td>1998</td>
<td>European release</td>
</tr>
<tr>
<td>1998</td>
<td>~2000 titles available</td>
</tr>
<tr>
<td>2003</td>
<td>US DVD sales greater than VHS</td>
</tr>
</tbody>
</table>

**Figure 7.** Optical spot (red) and track layout (white contours) for single-layer DVD that covers the same physical area as shown in Fig. 3 for CD.
control information is pre-stamped like a ROM disk in a system called land pre-pit (LPP), and the tracks are wobbled with timing information for spindle speed control. Similar to CD, the lower contrast of the writable formats can make them unreadable in drives designed for ROM. DVD Multidrives are certified to read and write all DVD Forum standards. Neither format was optimized for storage of many small files, which are typical of computer or camcorder use, that spawned several new formats, described next.

The “+” Formats. No removable storage technology seems to be complete without a format war—this is DVDs. DVD+R and DVD+RW are similar to their “dash” counterparts but generally with increased reliability because of more robust error-recovery and session-management features. The main difference observable by the consumer is increased flexibility in recording small sessions. A clear winner is not yet obvious between the “dash” and “plus” formats, and many players now advertise they are compatible with both using the notation DVD±R or DVD±RW.

DVD±VR. This format specifies how any writable disks should record video such that the resulting disk will play in a standard player.

DVD-RAM. The final contender in the writable DVD format war, the -RAM format is optimized to operate much like any computer-storage peripheral. Data are stored in concentric tracks whose sector headers can be observed by eye as small rectangular regions around the disk surface. Error management at the hardware level and, typically, a cartridge extend the phase-change writing material to as many as 100,000 write/erase cycles.

Table 5. DVD Specification Books

<table>
<thead>
<tr>
<th>Book</th>
<th>Formats</th>
<th>Content or Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DVD-ROM</td>
<td>Basic foundation</td>
</tr>
<tr>
<td>B</td>
<td>DVD video</td>
<td>Video</td>
</tr>
<tr>
<td>C</td>
<td>DVD audio</td>
<td>Multichannel audio</td>
</tr>
<tr>
<td>D</td>
<td>DVD-R</td>
<td>Write once</td>
</tr>
<tr>
<td>E</td>
<td>DVD-RW</td>
<td>Rewritable</td>
</tr>
</tbody>
</table>

Figure 8. Dual-layer, single-sided DVD structure. The upper 0.6-mm polycarbonate layer is only a mechanical structure in a single-sided disk, which provides stiffness and protects the fragile data layer from label-side damage. In a double-sided disk, both the top and bottom polycarbonate covers provide access to their own single or double layers. Note that no label is available in this case.

MiniDVD. An 80-cm format developed for camcorders, which is also known as three inch DVDs. Unfortunately, most 120 mm players will not play this format. They hold 1.5 GB SL / 2.66 GB DL.

National Formats. These formats, which are analogous to the Chinese SVCD, are designed as an alternative to the DVD format and its licensing royalties. The enhanced versatile disk (EVD) is the Chinese format, while Taiwan designed the forward versatile disk (FVD). It uses a smaller track pitch to attain 5.4 GB per layer and allows for up to three layers to attain a total capacity of 15 GB.

Limited Play. Although this article focuses on hardware and has thus not mentioned issues of copy protection or digital rights management, several formats were designed for just this purpose.

- Self-destructing formats. DVD-D and FlexPlay were conceived to be single-use disks that would become unusable in a short amount of time. The latter technology uses a sealed package that, when opened, allows atmospheric oxygen to attack a dye layer in the disk. After 48 hours, the disks darken and become unreadable.
- Remotely managed access. An alternative to a physical change in the disk is to remotely control access to content. DIVX was a short-lived format in which a special player used a phone line to register charges to a customer account when a disk was viewed. DataPlay disks are a custom format of 32-mm diameter and 250 MB / side using a cartridge and no cover material. They were designed both for portable storage and as a distribution medium in which music tracks on ROM disks could be unlocked for a fee.

Hybrid CD/DVD Formats. The large installed base and customer acceptance of CD for music has motivated several hybrid CD/DVD formats that attempt to use the greater capacity of DVD for increased fidelity or capacity while also presenting a CD layer for legacy players.

- DualDisk (which is similar to a European format DVD Plus) presents a CD on one side and a DVD on the
The obvious problem with this concept is that a 1.2-mm CD plus 0.6-mm DVD exceed the specified thickness of 1.2 mm for either. The compromise is to reduce the CD layer to 0.9 mm for a total 1.5 mm thickness. This violation of the standard can prevent some CD players from reading the disks because of spherical aberration and/or players of either type to be unable to load the extra-thick disks.

- Super Audio CD approaches the same problem but uses two layers rather than two sides. Using both CD and DVD layers in various combinations, this format is an attempt to displace the still-popular CD-DA by offering increased fidelity. The large installed base of standard CD players has so far prevented a large-scale shift to this format or its later competitor, DVD-Audio.

Proprietary Formats for Gaming. These formats are analogous to the CD proprietary formats but are updated to red lasers and generally higher NA.

- Universal Media Disk is a Sony format for the PlayStation that holds up to 0.9 GB (SL) / 1.8 GB (DL) on a ~65-mm diameter disk in cartridge.
- Nintendo optical disks for GameCube and Wii are an 80 mm, 1.5 GB mini-DVD for GameCube, which are extended to 120 mm and 8.5 GB for Wii. The format does not conform to the DVD Forum spec, which addresses both issues of piracy and license fees.

Versatile Multilayer Disk (VMD). This format is listed last because, while using generation two technology, chiefly a red laser, it is intended to compete with generation 3 products, such as BD. The format specifies 5 GB per layer, but it proposes 4-, 8-, and 10-layer disks to reach 50 GB total capacity.

Generation 3–BD

Virtually as DVD was being released, new high-definition (HD) TV sets were becoming available, and a limited number of stations began to broadcast HD. By analogy with digital audio (CD) and digital video (DVD), the market application for generation 3 is HD video. HD data streams require both more capacity and transfer rate than DVD can easily provide. However, adoption of HDTV has been slow, partly because of format uncertainty for both HD itself and the optical storage format.

As shown in Table 6, the history of the third generation of consumer optical storage has been dominated by a format war between the HD-DVD, which was promoted by the DVD Forum, and the BluRay Disk (BD) from a rival industrial group. Both use new violet lasers at 405 nm to reduce spot size. The HD-DVD format keeps the 0.6 NA and 0.6-mm cover layer to match existing disk fabrication machinery, but this introduces the need to control lens tilt actively to avoid coma. BD, in contrast, increases NA to 0.85 and reduces the cover layer to only 0.1 mm, as shown in Fig. 9. This thin cover required the development of new hard-coating technology to reduce the impact of scratches. The HD-DVD format was abandoned in favor of BD in early 2008, so it will not be discussed here.

Figure 10 shows the BD track layout on the same scale as the previous two figures. Although the track pitch and minimum mark length relative to the spot size have been slightly decreased from their DVD values, the change is small in comparison with the CD to DVD change. Most of the capacity increase is therefore driven by the spot size, which is less than half that of DVD. Note that the simple capacity and transfer rate estimation based on the Equations (1) and (2) using data from Table 1 yields 23 GB and 4.1 MB/s, which are relatively close to the actual values of 25 GB and 4.5 MB/s.

Variations within BD

Because of the relative youth of the third generation formats, fewer variants to date are available:

Writable. BD-R is a write-once format based on dye, and BD-RE is a rewritable format based on phase-change, analogous to their CD and DVD cousins. Capacity currently matches BD-ROM at 25 GB SL / 50 GB DL. A variant of BD-R called BD-R Low To High (LTH) has been developed that can use existing DVD-R manufacturing equipment;

![Figure 9. Dual-layer, single-sided BD structure. The upper 1.1-mm polycarbonate layer is only a mechanical structure in a single-sided disk, which provides stiffness and protects the fragile data layer from label-side damage. The cover layer has been reduced to 0.1 mm to allow a reasonable tilt tolerance at the high (0.85) NA. To avoid damage by scratches, this cover must be a special, hard material.](image-url)
however, players released prior to the format require a firmware upgrade to read the disks.

Mini BD. The 8-mm diameter format, which follows Mini CD and MiniDVD. Single-layer capacity is 7.5 GB.

Increased Layer Count. Although no new standards have yet been formalized, industrial research labs have demonstrated four-and six-layer versions of BD that reach up to 200 GB total capacity. Clearly, room to increase capacity is available, however, disk manufacturing cost and player tolerances will influence the commercial introduction of these advances.

Hybrid DVD/BD Formats.

- BD5 and BD9 use DVD hardware but store BD-compliant audio and video data. The transfer rate must be matched to the BD spec, so these DVD drives are required to operate at 3× transfer rate or more.
- AVCREC is a similar system targeted at low-cost applications that do not need the full capacity of a feature-length movie.

Ultra Density Optical (UDO). This format, using phase-change media, is very similar to a dual-sided BD except that it is aimed at the professional market. The 130-mm disk stores 60 GB on two sides and uses a cartridge for both disk protection and potential automatic handling.

POTENTIAL TECHNOLOGIES FOR A NEXT GENERATION

Unlike CD (audio), DVD (video), and BD (HD video), no single consumer market need has yet emerged to motivate this fourth generation. The increasing speed of home inter-net and convergence of personal computer and entertainment technology may provide an alternate path for content distribution, which has been the prime driver of previous generations. Alternative markets for optical storage include portable electronics or large-scale backup.

Academic, government, and industrial researchers are actively developing technologies for these potential applications. The literature is broad, so this section presents a summary of active research that covers the major approaches. The fundamental scaling laws introduced previously motivate the efforts. Generally, research falls into one of two approaches: reduced spot size or increased number of layers. The one exception is MultiLevel Recording, which uses variable pit depth to encode more that binary information at each surface location. One can view this as a straightforward application of information theory in which excess signal-to-noise ratio is traded for increased capacity. Despite successful industrial demonstration, there is little active development in this area.

In the area of reduced spot size, three general approaches are as follows: reduced laser wavelength, increased lens NA, and nonlinear interactions with the disk.

- Reduced wavelength. One can turn to very short wavelength lasers, X-rays or even electron beams to demonstrate very high capacity. Without a compact light source such as a laser diode, however, these efforts are largely confined to the laboratory. Beyond these academic demonstrations, it is currently assumed that laser wavelength has reached its practical limit. In addition to the huge effort to develop near UV diode sources, many materials begin to be opaque at this band, and no change as large as the recent shift to 405 nm from 640 nm is likely.
- Increased NA. Although tolerance scaling would seem to indicate that 0.85 (BD) is a reasonable upper limit on numerical aperture, it is possible to remove sensitivity to tilt by bringing the lens into near contact with a coverless disk. The gap can be maintained either by an air bearing similar to Winchester hard drives or by accurate servo actuation. When this gap is much less than a wavelength in size, it is possible to tunnel an optical spot from the lens to the disk surface via frustrated total internal reflection. In this case, NA is equal to the index of refraction of the solid immersion lens times the sine of the largest ray angle inside the lens and can be as large as ~2.0, which leads to spots sizes less than 100 nm. This near-field recording can only access one layer, and this layer must have no covering, which potentially exposes it to damage. An alternative is to tunnel the converging beam across the gap and into a thin cover with potentially multiple data layers, but this limits the NA to the index of the cover.
- Nonlinear disk interaction. In conventional writable storage products such as dye or phase-change, single photons are absorbed and initiate the desired thermal or chemical change, which results in a spot profile similar to the intensity. It is possible to design materials with low one-photon absorption such that at

Figure 10. Optical spot (blue) and track layout (white contours) for single-layer BD covering the same physical area as shown in Fig. 3 for CD and Fig. 7 for DVD.
sufficiently high intensity, two photons must be simultaneously or nearly simultaneously absorbed to initiate the writing mechanism. Generally, multiphoton cross-sections are sufficiently small that diode-lasers have insufficient power to write into these materials at commercially interesting transfer rates. However, new materials, particularly fluorescent, are in development, which may solve this problem. Most such techniques are combined with homogeneous multilayer storage, which is discussed in the following section.

An alternative nonlinear interaction that can be efficient is the heating of the medium itself. Super-resolution optical storage uses thermal layers that change their transmittance at high temperature. This transmittance change masks a portion of the spot, which effectively reduces its size via a local traveling mask. 100-GB-dual layer ROM disks have been demonstrated with this method.

The other primary approach is to increase the number of layers in the disk. Traditional methods of multilayer fabrication have been discussed above. Assuming cost and tolerance issues can be overcome, at some point the partially absorbing and reflecting properties of each layer reduce the laser power sufficiently that one cannot write to or read from the bottom layer. To go much beyond 10 layers therefore typically involves finding ways to break this limit. One approach is to electrochromically alter the layers therefore typically involves finding ways to break this limit. One approach is to electrochromically alter the absorption or reflection of the layer being addressed, leaving the others nearly transparent. This layer selection recordable (LSR) disk requires electrical addressing of a complex multilayer disk but breaks the absorption versus depth limitation. Another approach called the stacked volumetric optical disk (SVOD) is an optical floppy disk with 10 or more <0.1-mm layers held in a single cartridge. Single layers are pulled from the stack for reading and writing and reinserted when finished. The latter method avoids the need for optical access to variable depths and the associated difficulty with variable spherical aberration, which is common to all multilayer methods.

The other subcategory of multilayer disks is to give up fabricated layers and instead write into a homogeneous volume of approximately mm thickness. Servo tracking becomes an immediate issue in these approaches because stamped pregrooves and timing information are not available. The most common solution is to stamp such information into one cover layer and read it with a second laser, often of a second color to avoid data/servo crosstalk.

Most homogeneous multilayer optical storage systems are writable because limited options are available for rapid replication of data into the 3-D volume. Most are WORM (R) both to simplify the physics and because the value of hundreds of GB of data is generally larger than that of the disk itself. Virtually any mechanism of optical alteration of matter can be used for writing and any microscopy technique known to detect these alterations can be used for reading, which leads to a huge body of literature. Rather than try to summarize all the different examples, instead let us examine each sequential step in the process and enumerate the possible choices and their impact. Specific cases that are currently under active investigation will be highlighted.

1. **Absorption of writing light.** Light must be absorbed to initiate a change in the storage medium. Typically only total intensity is important, but in some cases (e.g., photoisomerization) the polarization state is also used to address a particular subset of the constituents. The absorption can be:
   a. **One-photon, linear.** The absorption length must be comparable with the thickness for light to reach the far side of the disk efficiently. Thus, only a small fraction of the writing light will be absorbed in the region of the focus, which demands that the material have very high sensitivity if it is to support an interesting write transfer rate. This in turn makes it likely that the material cannot be exposed to ambient light, which require a light-tight cartridge. Another consequence is that writing even a single layer slowly exposes the entire volume of the disk, which consumes a portion of the dynamic range.
   b. **Two-photon photochromic.** Photochromic molecules can be photo-excited into a thermally unstable state, which then has a strong change of absorption at a second wavelength. A second photon at that temporary absorption band can then be absorbed to initiate the storage mechanism. Because this event can only take place where a photon of both colors has been absorbed, most limitations of the one-photon absorption case are avoided. The intermediate state has a typical thermal lifetime of seconds, and thus a typical architecture uses orthogonal beams to restrict writing to a particular layer.
   c. **Resonant two-photon.** An important subclass of photochromic absorption is when both photons are of the same wavelength. In this case, a single focused beam excites the storage mechanism only near the focus (avoiding loss of dynamic range and writing beam attenuation) and can theoretically experience large absorption at the focus (for efficient use of laser power or high write rate). The absorption cross-section is increased with the lifetime of the resonant metastable state, but it cannot increase beyond the writing bit period, or many benefits are lost.
   d. **Nonresonant two-photon.** In the limit that the intermediate state is virtual, two photons must be absorbed simultaneously to initiate storage. This method has all the benefits listed for resonant two-photon absorption, but the cross-sections are typically too low to be excited by diode lasers at commercially interesting transfer rates.

2. **Modification of the medium.** The absorption of writing light must make some change to the medium that then leads to a modified optical property in preparation for later reading. The list below is not exclusive, for example, bleaching via isomerization or bleaching followed by polymerization followed by diffusion are possible combinations.
   a. **Isomerization.** Isomerization is a reversible (typically) change in the structure of a molecule. This
3. Fixing or development. Some changes are not visible to the read head until a development step, such as heating, is applied. Additionally, some media (e.g., photopolymers) require a fixing step to remove remaining sensitivity to allow nondestructive read-out. This step is not always permanent (e.g., thermally fixed data in lithium niobate can be erased by out-diffusion of the oligimeric products) or, at high powers, voids may be formed because of vaporization of the media.

d. Diffusion. Species may move into or out of exposed regions in response to consumption or generation by the writing event or may move because of osmotic pressure.

e. Decomposition. Polymers can be cleaved (possibly leading to out-diffusion of the oligimeric products) or, at high powers, voids may be formed because of vaporization of the media.

f. Thermal annealing/crystallization. Local heating by the laser focus can change the local molecular arrangement via annealing, thermal expansion, or crystallization.

g. Reorientation/poling. A portion of the material reorients along the polarization or is enabled, typically by laser heating, to orient according to an external electric or magnetic field.

3. Fixing or development. Some changes are not visible to the read head until a development step, such as heating, is applied. Additionally, some media (e.g., photopolymers) require a fixing step to remove remaining sensitivity to allow nondestructive read-out. This step is not always permanent (e.g., thermally fixed data in lithium niobate can be erased by high-temperature annealing).

4. Reading. The local modification caused by writing must be detected by the read head. To suppress response from out-of-focus layers, many architectures employ a confocal filter.

a. Reflection. The closest system to existing drives is reflection. Not all write mechanisms are compatible with reflection reading, however. For example, smooth index of refraction changes written by a focused beam do not reflect efficiently. This fact motivates microholograms in which counter-propagating coherent foci write micron-scale reflection gratings, which can then be efficiently read in reflection.

b. Refraction. A simplified version of any phase-contrast microscope can be used to detect phase in transmission, which requires a detection head on the opposite side of the medium. For example, a simple split detector can sense the presence of an index mark via deflection of the focus on the edge of the mark.

c. Absorption. Similar to (b), a single photodiode on the far side of the medium can detect absorption near the focus. Absorption does not scale well to many layers, so it is uncommon as a mechanism.

d. Retardance or other polarization effects. Similar to MO, changes in the optical polarization can be used to retrieve data.

e. Fluorescence. This area is currently undergoing intense research. Here data are detected by fluorescence at the read focus and stimulated by one- or two-photon absorption. A particular advantage is that a dichroic filter can reject all of the excitation light and associated surface and volume scatter. One primary challenge is the long-term stability of the fluorophores, which as a class have the tendency to exhibit a finite number of read cycles. Companies using this approach include D-Data (based on IP from Constellation 3-D), Mempile, and Call/Recall.

Finally, holographic optical storage is a potential 4th-generation technology that will not be discussed in this article, because it is described in another entry.

FURTHER READING


Technical Digest of the International Symposium on Optical Memory.

Technical Digest of the Optical Data Storage Topical Meeting.


ROBERT R. McLEOD
University of Colorado
Boulder, Colorado