1986-04

The compact disk ROM: Applications software: It optimizes access time, is compatible with various operating systems, and has the potential for multimedia use

Oren, Tim; Kildall, Gary A.

T. Oren and G. A. Kildall, "The compact disk ROM: Applications software: It optimizes access time, is compatible with various operating systems, and has the potential for multimedia use," in IEEE Spectrum, vol. 23, no. 4, pp. 49-55, April 1986, doi: 10.1109/MSPEC.1986.63708
http://hdl.handle.net/10945/68423

3M, and Sony are the major suppliers. A master disk costs between $3000 and $5000. Each CD-ROM disk made from the master costs about $10 in quantities of 1000 and $5 or less in quantities of 10 000.

Some suppliers are packaging databases on CD-ROMs. For example, Reference Technology bundles a library of almost 9000 software programs onto a single CD-ROM disk and offers it free with each purchase of its Clasix Series 500 CD-ROM DataDrive. The drive, including a PC-interface card and cables, is priced at $1595. Sold separately, the software CD-ROM disk costs $850. Reference Technology also offers a $149 audio card that lets its CD-ROM drive play compact audio as well as CD-ROM disks.

Once the retail price of CD-ROM drives drops to $200 or less, they are expected to do very well in the personal-computer market. As the need for more storage for personal computers increases, a 0.5-gigabyte drive for $200 would be an ideal solution, and the broad dissemination of these drives could change personal-computer use in ways that are as yet unfathomable.

To probe further

Since CD-ROM uses a technology similar to audio CD, readers may be interested in the special issue of the *Philips Technical Review* entitled "Compact Disk Digital Audio" (Vol. 40, no. 6, 1982), published by Philips Research Laboratories, Eindhoven, the Netherlands. The issue has four articles on the subject: M.G. Carasso et al, "The Compact Disc Digital Audio System"; J.P.J. Heemskerk and K.A. Schouhamer Immink, "Compact Disc: Systems Aspects and Modulation"; H. Hoeve et al, "Error Correction and Concealment in the Compact Disk System"; and D. Goedhart et al, "Digital-to-Analog Conversion in Playing a Compact Disc."

Readers interested in applications of optical disks or comparison of storage technologies may find the following useful: J. Rodriguez, "A Comparison of Optical and Magnetic Storage Devices," and A.E. Bell and V. Marrello, "Magnetic and Optical Data Storage: A Comparison of the Technical Limits," both in the digest of papers of the IEEE Computer Conference (Compcon), Spring 1984, Pub. No. CH2017-2/84; and P.P. Chen, "Applications of Video Disks to Office Automation," digest of 1981 National Office Automation Conference, Afips Press, Montvale, N.J.

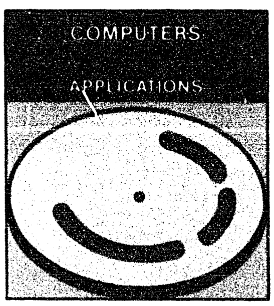
IEEE Spectrum articles on the subject include Miyaoka, S., "Digital Audio is compact and rugged," Vol. 21, no. 3, March 1984, p. 35.

About the author

Peter P. Chen (SM) is Foster Distinguished Professor of Computer Science and director of the Institute of Computer and Information System Research at Louisiana State University in Baton Rouge. His research interests include database systems, office automation, software design, and knowledge engineering. Internationally known for his contributions to the development of the entity-relationship data model, which is widely used in information systems analysis and database design, Chen taught at the Massachusetts Institute of Technology and the University of California, Los Angeles, before assuming his current post at LSU in 1983. He is editor in chief of the international journal *Data & Knowledge Engineering* and associate editor of *IEEE Computer, Information Sciences*, and *Future Generation Computing*. He was the founding chairman of the IEEE Computer Society Technical Committee on Office Automation, serving in the chair from 1981 to 1985. He has a Ph.D. in applied mathematics (computer science) from Harvard University (1973). ◆

The compact disk ROM: applications software

It optimizes access time, is compatible with various operating systems, and has the potential for multimedia use



Given the specialty of the compact-disk, read-only memory—putting massive databases on a user's desk inexpensively—many engineers will likely find themselves perfecting the technology and expanding its uses for years to come. A spinoff of compact-disk audio technology, the optically read CD-ROM presents various challenges to computer software and systems engineers:

- Overcoming a relatively slow rate of data access.
- Integrating audio, video, and graphics.
- Providing multiuser access to data.
- Accommodating a lack of standards for CD-ROM data structures.
- Updating the CD-ROM database.

In addition, systems designers and managers must consider:

- Hardware requirements for workstations using the CD-ROM.
- Costs of converting data and producing ROMs.
- Rapid changes in a new technology.

Solutions to these concerns combine hardware, software, and systems approaches—which are themselves still evolving to a remarkable degree.

Tim Oren and Gary A. Kildall KnowledgeSet Corp.

All of the challenges, and some of the approaches to their solutions, arise naturally from the pathway that information follows on its way from publisher to user via the CD-ROM [Fig. 1]. First the data are scanned from printed pages or converted from magnetic media to a form suitable for the CD-ROM. At this data preparation stage, some indexing of information and optimization for CD-ROM storage is usually performed. The transformed data are stored on magnetic tape, which then goes to a premastering facility for error-correction coding and data interleaving and scrambling [see "The compact disk ROM: how it works," p. 44]. The premaster tape is used to create a glass master disk for verification. Mass production follows, with the pressing process embossing the data onto plastic disks for end users. The final step is playback in a CD-ROM drive, typically attached to a desktop computer.

It is here that the hardware concerns—primarily the performance of the drive unit and the capacity of the medium—become evident. The prime software goal is to create data structures and accessing strategies optimized for the drive's characteristics. Economic concerns arise at each stage of the process, with one-time costs predominating on the data preparation and mastering side, and unit costs in the production and retrieval steps.

CD-ROM drives: wait before hurrying

CD-ROM drives that provide access to up to 600 megabytes of data on each 120-millimeter disk have been available for under $1000 since March. Although this combination of cost and capacity is unprecedented, the drives suffer from slow access speed. The laser pickup head typically requires 0.5 second to access information on the innermost tracks, and 1.5 seconds to reach the outer tracks.

By contrast, hard-disk units can access data in tens of milliseconds. The CD-ROM is slowed by the mass of its pickup head, which contains a focusing system with several lenses and which must be positioned with extreme accuracy. A high-torque stepping motor would move the head more quickly but would increase the cost and mass of the drive beyond practicality. To help compensate for the long access time, most CD-ROM drives have a small tipping mirror that rapidly directs the laser beam to nearby tracks without any movement of the entire lens assembly. Once the head has moved to the desired spot on the disk, the sequential reading speed of the CD-ROM is quite good: 1.2 megabits per second.

The slow access rate is a severe drawback for the most common forms of information stored on CD-ROMs: sequential data files, conventional databases, and textual information bases. In any of these cases, data structures designed for magnetic disks will produce unacceptable results on the CD-ROM.

In a typical magnetic disk file and directory structure [Fig. 2A], each link from one block of a directory or file to another will cause a movement of the pickup head. The primary purpose of such links is the expansion of existing files and directories when information is added.

But on a CD-ROM there is no possibility of expanding the information once it is stored. Because the number and size of the data files are known when the data are prepared, a much simpler disk structure may be used [Fig. 2B]. Finding a file will require at most two head motions, the first to reach the directory and the second to move to the file. The data blocks are contiguous within each file and within the disk directory, taking advantage of the drive's sequential read rate and tipping mirror.

In a relational database, one of the most common types, stored information is considered to consist of tables. For instance, an auto insurance database might include a table with one column containing the names of policyholders, another column for policy numbers, and additional columns for renewal dates and descriptions of automobiles owned. Each row of the table would consist of the data for a single customer. The columns in another table in the same data base would designate the auto maker, model and year, estimated value, accident rate, and theft rate. Each row in this table would refer to a single type of car insured by the firm using the database.

Information in such a database is retrieved through indexes that store the order and location of each row as sorted by a key column. One such index might allow insurance policies to be accessed in order of renewal date. On a magnetic disk, the indexes are constructed as a "tree" of links to information. Each branching point, or node, in the tree contains links to further nodes, leading eventually to the data itself. Finding a particular record requires traversing several levels of the tree, and following each of the links requires a movement of the pickup head.

Such a relational database will perform very poorly on a CD-ROM unless it is modified. The data structure must be adapted to the characteristics of the drive. In particular, the number of levels in the tree, and thus the number of head seeks, can be reduced by making each node much larger—increasing the number of its links. This also takes advantage of the CD-ROM's high sequential read speed.

The large data capacity of the CD-ROM can also be used to advantage in the "join" operation that is commonly performed to merge two tables. For instance, there might be a need to selectively combine the two auto insurance tables to yield yet another one, consisting of data on policy number, auto value, and loss rate only. Because a join operation requires a great deal of access to the storage unit, it saves time to compute such operations in the data preparation stage and then store the redundant tables on the CD-ROM.

Defining terms

Block: a unit of data on a disk.
Directory: a data structure that associates names with the data files carrying the names.
File: a collection of data, organized logically in blocks.
Full-text search: the ability to find any word in a text.
Link: the address of a location in a file where related information may be found.

Searching the whole text on a desktop

Full-text searching, which examines an entire database for the occurrence of a single word or a complex combination of terms, is becoming possible for the personal computer for the first time, thanks to CD-ROM technology. With a complete full-text search package, a user will be able to specify the desired proximity of multiple-search terms; to use Boolean AND, OR and NOT operators in the specification; and to specify a nonunique search term with "wild-card" characters.

The data structures needed for full-text search are generated by the inversion method. The entire database is scanned to find and count the occurrences of each word. At this point, words that occur so frequently as to be meaningless, like "of" and "to," are discarded. In material about data processing, for example, terms such as "computer" might also fall into this category.

The database is reread once to record the position of every occurrence of each word. The resulting large table is the word index, or full-text inversion, of the database. A magazine article on semiconductor technology, for example, might use the word "micron" four times. The inversion of the article would contain an entry for "micron," along with the four locations where the word occurred.

On a magnetic disk system such an index is usually stored in a multilevel data structure similar to that of a relational database. The strategy for conversion to CD-ROM also follows the database pattern: the levels of the data structure are reduced in number, so the index for each letter of the alphabet can be stored in one undivided area, and read in a single operation. The pointers to text locations would appear in another table nearby.

A search for each paragraph containing the two words "micron" and "leakage" would require four seek operations on the CD-ROM. One motion is required to read the index for the letter "m" and verify the existence of "micron." A second motion reads the locations of the word. The same two operations are performed for "leakage." The various text locations, obtained from the index, are then compared to see if any paragraph contains both terms.

Of course there is a price to pay for such extensive indexing. A full-text inversion can occupy as much storage space on the CD-ROM as the text itself does. If complete indexing exceeds the storage capacity of the ROM, it may be possible to discard more terms during the first phase of inversion. On the other hand, storing only key words or an index for abstracts of papers would occupy only a few percent of the total capacity.

When these expedients are unacceptable, compressing the data and indexes may be possible. Optimal encoding of text information can double the effective storage, but at the cost of retrieval efficiency. Some compression of the inversion can be achieved by storing the distance between successive locations of a word, rather than storing the actual location of every occurrence. Again, the penalty is a slower search.

Because the CD-ROM pickup movement is slow, it is a poor device for timesharing by several users or tasks. If several simultaneous searches are competing, the pickup head will be moved frequently as each task takes its turn. This "thrashing" destroys carefully optimized sequences of head motion, and performance quickly degrades.

The reproducible nature of the CD-ROM disk offers some solutions on a systems level. For example, identical CD-ROMs may be clustered in multiple drives under an intelligent controller to serve several users simultaneously. The controller can switch read requests to the available drive that has its pickup head closest to the requested location on the disk. Systems designers must carefully assess the load factor per user to determine the number of identical drives needed to prevent thrashing and give all users adequate access times.

If multiple drives are too costly, other users may be locked out when a search is in progress. The penalties in response time and bottlenecking of tasks are obvious, but they may not be excessive when there are only a few users.

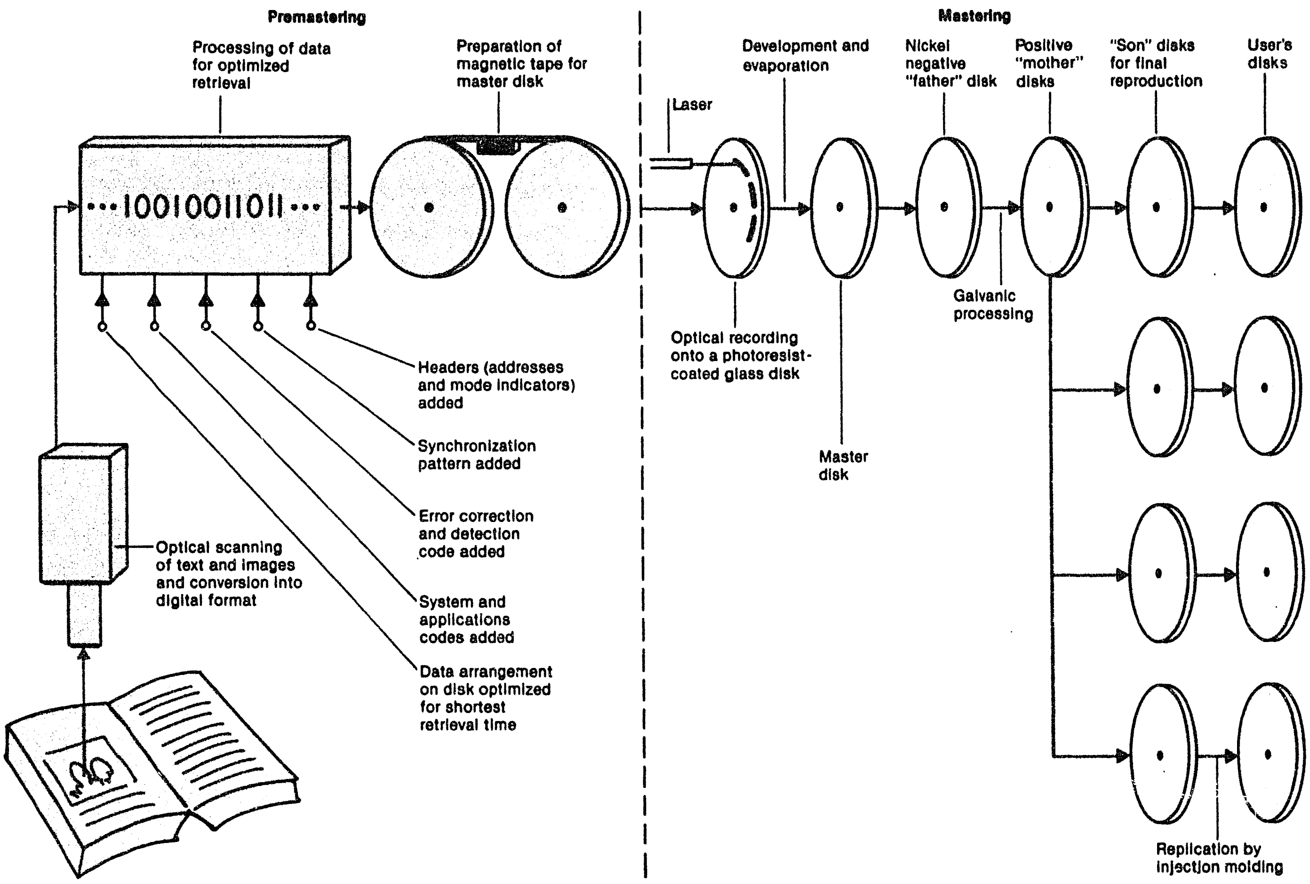
Multiple media are coming

Currently CD-ROM is a closed system with no capability for audio or video information, but this may change soon. On Feb. 24 Sony and Philips announced their intent to create a specification for interactive audio and video applications on CD-ROM—the "CD Interactive Media" (CD-I). The specification would be a complete format for interactive use of CD-ROM, including speech; natural still and animated pictures; and computer graphics, files, and programs as well as audio and video. The CD-I standard will tentatively include specifications for a low-cost player based on the Motorola 68000 microprocessor, with the first player expected to be introduced sometime next year.

In the interim, multimedia presentations based on CD-ROM will require an additional peripheral player. In such cases, the CD-ROM would hold search-key information for a remotely controlled audio CD or a videodisk player.

If cost prohibits the use of two drives, inexpensive stopgap measures are available. Sound sequences may be digitized and recorded on the CD-ROM. Similarly, printed images may be captured with scanners, or video images with frame grabbers, and the results placed on the disk. However, both the audio and the video techniques may yield inferior reproduction, and they are not portable from one "player" to another. For example, the playback of digitized audio may differ on computers with different hardware clock rates and may not be feasible at all on timeshared systems. Scanned images may not match the resolution, aspect ratio, or color capability of a display.

One technique for overcoming graphics constraints on the CD-ROM is to include standard graphics metafiles, such as the one based on the Graphical Kernel System (GKS). A metafile captures the sequence of output operations, such as lines, pie segments, or text, that make up a picture. Given the availability of GKS driver



[1] The production of a compact-disk read-only memory (CD-ROM)—a peripheral multimegabyte optical storage device for computers—is an elaborate process that includes data preparation, premastering, mastering, and replication. Data acquired through optical scanning or playback of machine-readable magnetic tape is first processed for optimized retrieval time on the disk. Premastering includes the addition of application and error-correction codes as well as data synchronization patterns and headers that identify addresses of data. All of these data are recorded on magnetic tape and are then transferred to a master disk in two stages—optical recording by laser and photolithography. Two more production stages lead to a master negative disk. Several positive disks follow, leading to negatives from which CD-ROMs are mass-produced by injection molding.

software for a display, a metafile may be played back with no loss of resolution. However, picture libraries in this form are not generally available, so the technique is limited to applications where graphics can be generated from scratch.

Looking for software standards

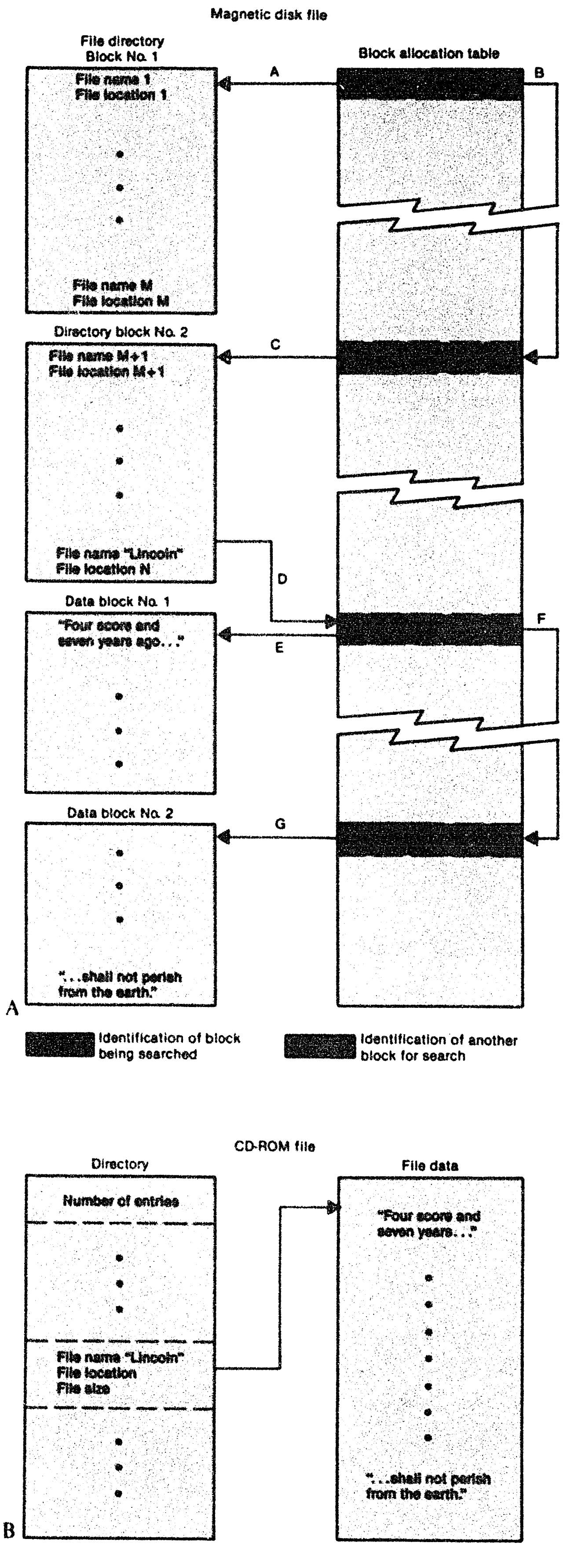
Like other peripherals, the CD-ROM must contend with a cluttered computer marketplace. While designers of dedicated CD-ROM workstations may to some degree choose their own environment, the developer of general-purpose CD-ROM hardware or software faces a variety of competing processor and bus architectures, operating systems, and user interfaces.

Compounding this difficulty is one of the most controversial issues now facing CD-ROM technology: the lack of standards for formatting data on the disk. While the CD-ROM hardware standard guarantees that a given disk may be read in any drive, the lack of standards for file or index structures on the disk means that a disk prepared by one CD-ROM company cannot be read by the software of another vendor. Thus, the purchaser of several different CD-ROM databases would have to use different software to access each one, with no convenient way to integrate the data extracted.

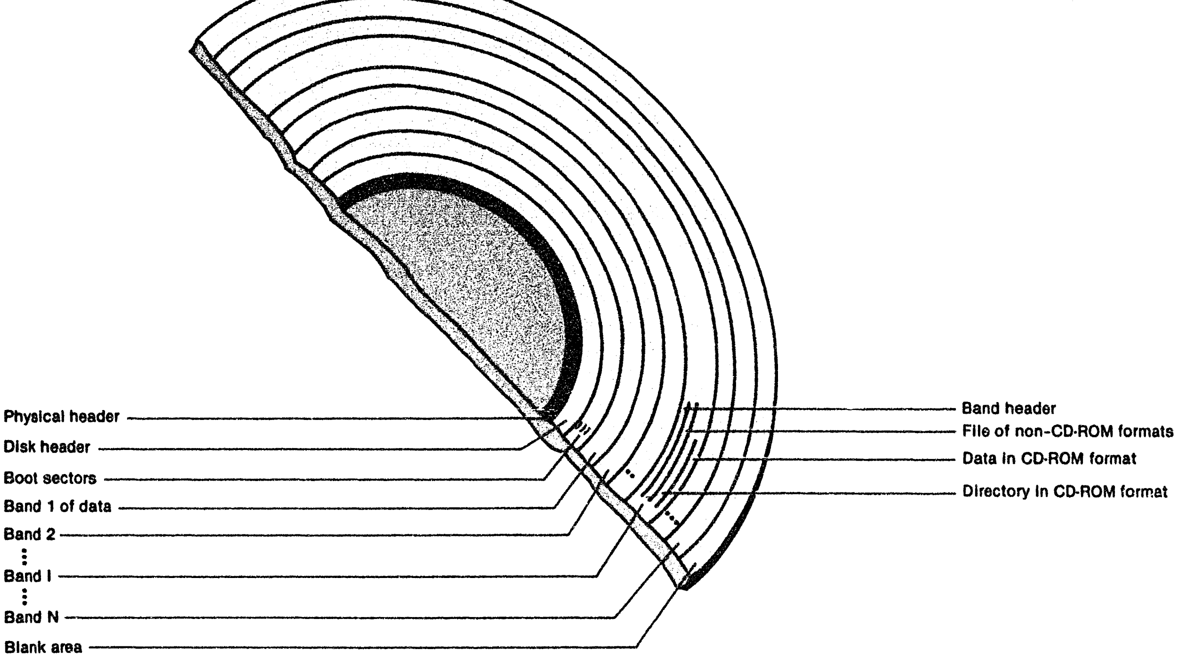
A standard called Unifile was proposed last year by Digital Equipment Corp. This proposal and others, ranging from descriptions of bands on the disk to detailed file directory and index standards, are under consideration by an informal industry association called the High Sierra Group.

Choosing single standards for directories, file organization, and indexing structures may be premature at this time. Mimicking magnetic-disk data structures on current CD-ROM drives has severe performance penalties. Furthermore, it would be equally shortsighted to adopt these expedients as standards without further experience with current systems, not to mention the improvements that may occur in CD-ROM drives.

Two other factors argue against a standard now. First, information providers such as news organizations, book publishers, and database proprietors want their CD-ROM products to be protected from piracy. They have requested that the developers of CD-ROM systems scramble the indexes and encrypt the data to protect their investments. Second, CD-ROM technology and the techniques for efficient handling of very large masses of data are too immature to judge what types of information may be found on



*[2] A generalized data file system on magnetic disk creates a block allocation table at a fixed place on the disk [A]. Each entry in the table is linked with one disk block (*A,C,E,G*) and it also contains an indication that that block is in use. A file directory that originates from a fixed location in the table contains names and starting block numbers for logical groups of information. The file system is designed so that the directory and files can be easily extended to blocks on the disk that are noncontiguous with previous ones. An additional field in the table links each file or directory block to its successor. Here a search for the file named Lincoln would examine the first directory block and follow link* B *to the second block to find the name. After following directory link* D *to the first data block of the file, link* F *is used to reach the text. The process may be repeated indefinitely. The design for expandability creates a complex linked data structure that may require many movements of the pickup head to reach the desired data. Since a CD-ROM [B] maintains its data structure throughout its life, it does not require an allocation table. A file search proceeds immediately to the directory, all of which is located at a fixed spot on the disk. Directory entries contain the predetermined file size as well as its name and location. All file data are written contiguously. If the frequency of use for major portions of the data is known, the files in the directory and their layout on the CD-ROM can be arranged to minimize access time for them. The overall design minimizes the movement of the pickup head.*



[3] One proposed CD-ROM arrangement includes a disk header table to identify the location and type of each data band on the disk. It is followed by an area for the startup (boot) codes of various computers, along with tag fields that allow each machine to find its own code. Each data band contains its own headers, partitioning the band into a CD "native" directory and file structure, with the option of including "foreign" files. Native files obey a CD-ROM optimized standard format, while foreign files use a magnetic disk standard, such as Unix or PC-DOS.

CD-ROMs within, say, five years, or to predict what accessing structures may be appropriate for this information.

How, then, is bedlam to be averted? The solution may lie in discarding the false analogy of the CD-ROM retrieval system and a standardized record player. Unless the CD-I player comes to dominate the business environment, there will be an excess of architectures for the foreseeable future. Further, the record player analogy ignores the adaptability of a general-purpose computer. The capacity of a CD-ROM is ample for storage of unique versions of the retrieval software for dozens of different information delivery systems. A CD-ROM standard need include only a disk header that allows the delivery system to identify and retrieve software, making a standardized data architecture less urgent.

A simple CD-ROM disk design should include program storage in the format of each intended delivery system. The programs would include retrieval software, video-display drivers when necessary, and any other desirable software. A single band on the disk [Fig. 3] would be allocated to each combination of processor and operating system. For instance, one band might be allocated to Intel microprocessors operating under the MS-DOS standard, a second band to VMS running on VAX machines, and a third to Motorola 68000-based Unix systems.

The database itself would reside in one or more bands occupying most of the disk, and would include the accessing structures appropriate to the application. This approach is flexible enough to reserve disk areas for future audio and video storage.

Direct updates of a CD-ROM database are, of course, impossible, so the entire disk must be replaced if changes are made in the data. Given current lead times for mastering and production, a quarterly update cycle is the most frequent that can be reliably maintained. If more frequent changes are necessary, updates must be distributed onto magnetic media or transferred from online sources to magnetic disk by the user.

However, since even a 1 percent change in a full CD-ROM represents over 5 megabytes of material, a few such changes can easily fill in a hard disk. Another consideration is the method of integrating the new and the old data, which must be done smoothly to avoid undue annoyance to the user—say an engineer searching an updated database for articles on a new type of MOSFET not found in the original CD-ROM database. The search software must direct the query to the new material on the magnetic disk without explicit instructions from the user.

Toward this end, the software that controls access to the data must be able to recognize queries to updated portions of the database, and to redirect them transparently to the magnetic disk files. The various indexes on the CD-ROM would also have to be checked against the updated information, disabling entries for obsolete data and finding those for new data. New indexing information could be distributed on the magnetic disk as part of the update, or it could be created by the user's computer, based on a version of the indexing software used in the original data preparation.

The mastering and production costs and the production schedules for CD-ROMs depend heavily on the availability of manufacturing facilities, which are essentially the same as those for producing compact audio disks. Most production facilities are dedicated to the audio CD market, which leads CD-ROMs by orders of magnitude.

Since March the production phases from premastering to the first mass-produced disks have typically required four to six weeks. This rate would permit quarterly updates of CD-ROMs for legal, medical, and financial data, as well as other types of databases in which timely information is crucial. New production capacity is coming on-stream rapidly, however, and a monthly update cycle could be practical by the end of this year.

Costs for premastering and mastering are now around $3000 for each master disk, and $4 to $5 per CD in lots of 10 000. These rates will fall as the production capacity grows.

Justified for multimegabyte storage

Not all databases belong on CD-ROMs, of course. If the information does not exceed 10 megabytes, there is little reason to put it on a CD-ROM, unless it is merely part of a broader, general program. At present, a standard CD-ROM can store about 550 megabytes, which may include anywhere from 300 to 450 megabytes of raw information, depending on the density of indexing.

Databases that exceed this limit may be accompanied with multiple drives, or "jukebox" changers—but at a greater cost, naturally. It may be possible to break up some large databases into chunks, each of them residing on a single disk. The chunks ought to be sufficiently autonomous, however, to allow a search to continue without disruption through a change of one or more disks.

The markets with the greatest immediate potential for CD-ROMs include lawyers, doctors, and engineers and other manufacturing professionals who regularly retrieve specifications. Many of these markets are already served by dedicated workstations and online data utilities. For workstations, CD-ROM offers an immediate cost savings over magnetic disks. A workstation that is based on CD-ROMs may help a user recover its cost in as little as a few months by substantially curtailing the bill for online data services.

A vendor who wants to deliver CD-ROM technology for the office or home personal computer, however, faces a problem. With no currently installed base of CD-ROM drives, publishers are reluctant to convert information from print to the CD-ROM format. And without many disk titles available, the consumer is hesitant to buy a CD-ROM drive. This impasse may be broken in the next year as the price of drives continues to fall, and two or three CD-ROM titles of general interest—like encyclopedias, dictionaries, movies, sports features, and reference sets of classical and religious literature—appear on the market. As the first set of consumers enter the marketplace, publishers, in turn, will be encouraged to jump into the business.

The CD-ROM is likely to have the greatest impact when it penetrates schools and municipal libraries. To save space and money, such institutions might choose the CD-ROM format for storing references that are not frequently accessed.

Refinements are imminent

Current technical developments indicate that solutions are imminent for the most severe applications problems of the CD-ROM. Lighter pickups and erasable optical media are now under development and could reach the mass market within two years.

Even when these major problems are overcome, however, that will still leave the far-reaching question of how to deal effectively with massive amounts of information. Users such as educators and businessmen unaccustomed to complex information retrieval systems may be overwhelmed by a deluge of unstructured data, two orders of magnitude greater than that seen up to now.

A CD-ROM system that can deliver the equivalent of 400 volumes of text must include powerful but transparent accessing methods if the information is to be useful. Current retrieval systems, limited to full-text search, must give way to workstations for research and writing with optical databases at the core.

CD-ROM databases will continue to offer document title indexing and full-text search, but in the future they will also link related documents in what is becoming known as a "hypertext" system.

Such a system provides direct links to cited articles at the point of reference and, through a full indexing system, it can show all the citations of a document under study. With this capability, the researcher can move quickly through networks of related information, saving the browse path as a personalized index.

Database word indexes are likely to be augmented with a thesaurus of synonymous terms, so the search system can suggest extensions and refinements to the searches made on it. The use of hypertext links as a data flow graph will enable searches to be restricted to "nearby" information—information about closely related or similar topics.

When indexing information becomes available to the user during a search, the retrieval software can be reconfigured to the research task at hand. An offshoot will be such secondary products as encyclopedia study guides, case law studies, and medical tutorials, to allow the reader to examine the underlying source documents immediately.

Text and indexing information may be extracted from the database or entered by the user to form new documents that will be added transparently to the existing data bases. A single document might be viewed as contiguous text, as a skeletal outline, or as a collection of notecards. Document editing tools are likely to include notecard and outline processors for organizing ideas, and text and image editors to add character and picture information. Database references in a newly formed document will remain linked to the underlying information, and may also be linked to other documents.

Progress in CD-ROM technology could make such workstations a reality soon. With these tools, existing records can then be converted into structures of related knowledge. When mixed-mode drives become available, video and audio information can be incorporated directly into these archives, and animations and simulations will be added to encyclopedias as personal computers grow ever more powerful.

To probe further

The future of CD-ROM workstations is addressed in "Reading and writing the electronic book," by Nicole Yankelovich, Norman Meyrowitz, and Andries Van Dam, *IEEE Computer*, October 1985, pp. 15-30.

Techniques of textual search are described in "An overview of information retrieval subjects," by Martha Bartschi, *IEEE Computer*, May 1985, pp. 67-84.

The lack of hardware standardization is discussed in "Optical Disks Draw Attention as They Ship Out," *PC Week*, August 13, 1985, pp. 98-99.

The evolution of CD-ROM technology, standards, and applications is regularly discussed in *CD Data Report*, Langley Publications, Suite 115-324, 1350 Beverly Rd., McLean, Va. 22101 (telephone 703-241-2131), and in *Videodisc and Optical Disk Update*, available from Eckler Publishing, 11 Ferry Lane West, Westbury, Conn. 06880.

"A Prototype Electronic Encyclopedia," by Stephen A. Weyer and Alan H. Borning, appears in the ACM *Transactions on Office Information Systems*, January 1985, pp. 63-88.

Database techniques are discussed in the books *Principles of Data Base Systems*, by Jeffrey D. Ullman, Computer Science Press, Rockville, Md., 1980, and *The Art of Computer Programming, Vol. 3: Sorting and Searching*, by Donald E. Knuth, Addison Wesley, Reading, Mass., 1973.

A discussion of computer graphics standards is included in *Introduction to the Graphical Kernel System (G.K.S.)*, by F.R.A. Hopgood et al, Academic Press, London, 1983.

About the authors

Tim Oren is the user-interface designer for KnowledgeSet Corp., which develops CD-ROM software. He was previously with Digital Research, where he developed the interface editor for the GEM windowing system and worked on the Dr. Logo project. A native of Kokomo, Ind., Oren devotes his spare time to personal computing and to writing one of the first electronically published personal-computer columns to appear on the Compuserve information service. A member of several technical groups of the ACM, Oren holds an M.S. in systems science from Michigan State University (1982).

Gary A. Kildall (M) is founder and chairman of the board of Digital Research and founder of KnowledgeSet Corp. (formerly Activenture Corp.). He created Digital Research's first product, the CP/M operating system. He also designed the Dr. Logo programming language for the IBM PC and developed PL/I, one of the first high-level languages for microcomputers. A native of Seattle, Wash., Kildall taught computer science at the Naval Postgraduate School, Monterey, Calif. He holds a Ph.D. in computer science from the University of Washington (1972). ◆