

CD ROM

VOLUME TWO

OPTICAL PUBLISHING

*A practical
approach to
developing CD ROM
applications.*

EDITED BY SUZANNE ROPIEQUET
WITH JOHN EINBERGER AND BILL ZOELICK

MICROSOFT
PRESS

Hyundai Ex. 1016

PUBLISHED BY

Microsoft Press

A Division of Microsoft Corporation

16011 N.E. 36th Way, Box 97017, Redmond, Washington 98073-9717

Copyright © 1987 by Microsoft Press

All rights reserved. No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.

Library of Congress Cataloging in Publication Data

CD ROM

Includes bibliographies and indexes.

Contents: v. 1. The new Papyrus v. 2. Optical publishing.

1. Compact discs. 2. Optical storage devices.

I. Lambert, Steve, 1945- II. Ropiequet, Suzanne, 1954-

TK7882.C56C3 1986 004.5'3 86-2369

ISBN 1-555615-000-8

Printed and bound in the United States of America.

1 2 3 4 5 6 7 8 9 FGFG 8 9 0 9 8 7

Distributed to the book trade in the

United States by Harper & Row.

Distributed to the book trade in

Canada by General Publishing Company, Ltd.

Distributed to the book trade outside the

United States and Canada by Penguin Books Ltd.

Penguin Books Ltd., Harmondsworth, Middlesex, England

Penguin Books Australia Ltd., Ringwood, Victoria, Australia

Penguin Books N.Z. Ltd., 182-190 Wairau Road, Auckland 10, New Zealand

British Cataloging in Publication Data available

C O N T E N T S

		1 Chapter One: Changing the Publishing Paradigm Bill Zoellick
SECTION I	THE CD ROM ENVIRONMENT	11 Chapter Two: Information Delivery Systems Leonard Laub
		31 Chapter Three: CD ROM Characteristics John Einberger
SECTION II	TEXT PREPARATION AND RETRIEVAL	45 Chapter Four: Preparing Text Raymond Bridge and James Morin
		63 Chapter Five: Selecting an Approach to Document Retrieval Bill Zoellick
		83 Chapter Six: Full-Text Retrieval and Indexing Jerry Fand
		103 Chapter Seven: Database Retrieval and Indexing Gregory Colvin
SECTION III	IMAGE AND SOUND PREPARATION	121 Chapter Eight: Displaying Images Gene Apperson and Rick Doherty
		169 Chapter Nine: Using Audio Bryan Brewer

NP
31

SECTION IV DISC PRODUCTION

185 Chapter Ten: Disc Origination
Bill Zoellick

195 Chapter Eleven: High Sierra Group Format Description
John Einberger and Bill Zoellick

217 Chapter Twelve: Premastering and Mastering
Anne Armstrong

SECTION V DATA PROTECTION
AND UPDATING

227 Chapter Thirteen: Data Protection
Joe Bremner

255 Chapter Fourteen: Updating CD ROM Databases
David Roux

SECTION VI CASE STUDIES

269 Chapter Fifteen: The First CD ROM Publication
Brian Martin

**283 Chapter Sixteen: A Clinical Medical Information System
on CD ROM**
Marilyn G. Winokur, Leonard S. Rann, and William Thornburg

SECTION VII APPENDICES

301 Appendix A: Glossary

337 Appendix B: Resources

349 Index

Information Delivery Systems

The small, shiny, rainbow-colored plastic discs called CD ROMs are so captivating that people can be excused for not realizing that these little wonders only *carry* information. Unlike books, the discs cannot deliver that information directly to people. As you might expect in this day of long but implicit names, that job falls to something commonly known as a *delivery system*.

In its simplest form, a delivery system contains:

- A drive, which reads the digital data stored on the CD ROM disc.
- A controller, which gives directions to the drive.
- A search engine, which consists of a microprocessor, memory, and software. This system interprets the user's requests and responses and generates instructions for the controller.
- An input device, which receives the user's requests and responses.
- An output device, which displays information taken from the disc for the user.

Such delivery systems can be part of larger systems and include familiar computer applications such as word processors and spreadsheets. They also can be dedicated to special purposes, such as instruction or research.

This chapter introduces those interested in publishing or distributing information on CD ROM to the sorts of delivery systems people are building and describes what the systems can do.

The Concept of a Workstation

This chapter uses another new-age term, *workstation*. The term usually refers to a high-end personal computer tailored to perform a specific task such as data retrieval or engineering design. Here, workstation refers to a computer system designed to search for and find information stored on a

CD ROM disc. A CD ROM workstation can be a shared mainframe, a general-purpose microcomputer, or a dedicated delivery system.

A CD ROM workstation must contain, or have unrestricted access to, all elements of a delivery system. The concept of a CD ROM workstation goes hand in hand with the idea of individuals working on whatever interests them, at whatever time and rate, and with whatever emphasis makes sense to them.

Although large, centralized systems that process big batches of queries can use CD ROM discs, this discussion concerns applications that are basically "one person, one disc." Here, we don't presume that the user, although possibly skilled in other fields, is a computer specialist or a librarian. We presume the delivery system can respond correctly, appropriately, and clearly to casual, vernacular requests.

CD ROM is a low-cost medium, usually using low-cost hardware of moderate performance. The drives and the ways they attach to computers are appropriate for a low-cost computer environment. It is thus not surprising that a PC is today's most frequently used general-purpose host for CD ROM delivery systems.

Most PC-based delivery systems are designed for use by one person at a time. Such a system typically contains:

- A PC-XT, PC-AT, or compatible, depending on how much computing the search engine or graphic display requires.
- A floppy disk or hard disk drive to store either the system software (device driver) the PC needs to find and control the CD ROM drive or the application software needed if CD ROMs store only data and indexes. If the CD ROM is bootable and stores the application software, magnetic disks still are needed to temporarily store file directories and indexes downloaded from the CD ROM.
- A CD ROM drive.
- A CD ROM drive controller or SCSI (Small Computer System Interface) host adapter.
- Some type of input device.
- A video card and CRT display (usually).
- A printer.
- A hard disk to store indexes for the CD ROM application, data, system software, and other applications.

We now examine major components of such a CD ROM workstation. Multiuser and more specialized workstations are addressed later.

CD ROM Drives

In consumer electronics, machines that record information usually are called *recorders*. The machines that play CD audio discs are called *players*. In the computer world, these machines are called *drives*, partly because they both play back and record information and partly because that's just what they're called.

Freestanding drives. Many CD ROM drives come with their own power supplies and housings. Such a freestanding drive gets its power directly from a wall socket. It plugs into the rest of the delivery system (usually a microcomputer) through a standard external interface or through a special device interface card that plugs into an expansion slot of the host computer.

The initial popularity of freestanding drives comes from:

- Availability, since these drives usually are based on commercial CD players.
- Simple interfacing.
- Freedom from problems of limited power supplies and available slots.

Built-in drives. More recently, manufacturers began supplying CD ROM drives that can be connected inside microcomputers in the same manner as floppy disk and Winchester drives. The first generation of built-in CD ROM drives uses the standard, full-height, 5¼-inch design, but newer models fit into half-height packages for use with the newer, smaller desktop and portable computers.

As with their magnetic counterparts, built-in CD ROM drives take direct current (DC) from the host computer's power supply; eliminating a separate power supply and package saves money. At this writing, OEM prices of such drives already are less than \$500, with some quotations less than \$250 in large volume, and prices are expected to continue dropping as production increases.

Jukeboxes. The consumer CD market has become interested in multidisc machines to reduce the need to get up and change discs. Some machines hold six discs; others hold 100 or more. These all have been adapted for CD ROM use and provide direct, online access to tens of thousands of megabytes of data.

Audio CD jukeboxes take as long as 10 seconds to swap discs, which may seem like a long time to the user of a highly interactive computer. A powerful technique for reducing this swap time is to incorporate several CD ROM drives in the jukebox so that you can load discs containing data likely to be called soon in anticipation of the request. Such multidrive jukeboxes are not yet available.

The jukebox is itself a data-storage device. Just as a disc drive is directed to a track or sector of data, a jukebox must be directed to a disc. So the jukebox requires a device interface and software to communicate with the operating system and application software in the host. Jukebox manufacturers can provide information about specific requirements.

Specialty drives. One of several refinements in CD ROM drive design recognizes the great similarity between CD ROM and CD audio, providing decoders for both data and audio. This hybrid is achieved by adding the Digital-to-Analog (D/A), low-pass filtering, and audio output stages of a standard CD player to a standard CD ROM drive. Thus, the microcomputer can control audio programs, and audio can accompany data presentations. OEM prices of the first generation of such drives are similar to those of standard CD ROM drives, in the range of \$500 to \$1000.

The focus and tracking servos of all optical disc drives provide a degree of shock and vibration protection, but CD ROM applications destined for moving vehicles require even more protection. These *severe environment* applications call for rugged drives, obtained either by repackaging standard drives (producing "ruggedized" drives) or by redesigning them from the ground up. Both methods were used in designing the CD players that were recently introduced commercially for automobiles. Because of this consumer product base, rugged CD ROM drives will also probably cost less than \$1000.

Consumer CD players usually take several seconds to move between non-adjacent tracks. This access time is too long for most computer applications, so CD ROM drives move faster. Most CD ROM drives have an average access time in the range of $\frac{1}{2}$ to 1 second, or 500 to 1000 milliseconds (msec). The heads of some newer drives are slightly lighter and their head positioners more powerful, enabling average access times of less than 300 msec. Future drives probably will be even faster and will not cost significantly more.

The Controller

Magnetic disk drives, particularly those made for microcomputers, contain little of the electronics that make them work. All dialogue between the drive and the computer about commands and status, and most of the error detection and correction of digital data recovered from the disk, are processed by a separate package of electronics called a *controller*.

CD ROM drives contain most of the functions associated with magnetic disk controllers, including data separation, error correction, and simple data buffering. But a data connection still must exist between the CD ROM drive and its host computer. This connection consists of a host adapter printed circuit card, device driver software in the host computer, and a cable that connects the CD ROM drive to the host adapter. The host adapter card converts the bus signals of the host computer to a series of signals known to the CD ROM drive; the cable is the electrical link between the two. Connectors on each end of the cable mate with connectors on the host adapter card and the drive. Electrical signals exchanged between the CD ROM drive controller and the host computer are sent through a *device interface*.

Many newer CD ROM drives contain their own controllers, which talk to standard high-level interfaces such as the IEEE 488 or SCSI (Small Computer System Interface—pronounced “Scuzzy”). Both SCSI and IEEE 488 are drive-independent interfaces, and both are configured as buses. This means, for example, that a single SCSI interface can deal with as many as eight devices, among which might be CD ROM drives, Winchester drives, writable optical disc drives, and, soon, printers and other peripherals. All the system developer needs is a software device driver that tells the operating system how to find and communicate with each device attached to the SCSI bus.

The SCSI bus is said to be *intelligent* because it can partially process the information coming from the CD ROM drive. This processing takes time. The time required to execute a command is known as *command latency*. Although command latency decreases the performance of SCSI interfaces, having the interface do this work removes a portion of the burden from the host computer.

SCSI is a high-performance interface and can move data at rates of a maximum of 32 megabits per second (Mb/s). This rate satisfies the system developer's desire to attach several drives, such as a Winchester disk (5 to 10 Mb/s), writable optical disc (1 to 5 Mb/s), or CD ROM (1.2 Mb/s), to the same interface.

The Search Engine

CD ROM discs store enormous amounts of data, but most people are concerned only with answers to their immediate questions. In addition, most people do not know (or really care) how the information is organized on the disc. They probably will pose questions in their own terms. The job of interpreting such questions and determining which information to take from the disc to gratify them belongs to the *search engine*.

Search engine, yet another new-age term, refers to the combination of:

- A scheme for indexing and organizing information on the CD ROM disc, carried out during mastering by (or for) the publisher.
- Software that interprets queries and generates citations (listings of where relevant information can be found on the CD ROM disc).
- Processor hardware, which runs the software, receives queries from the input device, sends requests to the CD ROM drive device interface, processes the information as needed, and sends the information to the output device.

Many types of search engines are available, but, as with other technologies, none can be called "best" out of the context of an application. The most prevalent search engine designs in CD ROM applications are described elsewhere in this book.

Input Devices

Most CD ROM delivery systems are designed to retrieve information. The user enters a request for information by means of an input device. The most common input devices are:

Keyboards. To the delight of touch-typists everywhere, keyboards have become the most common input device for general-purpose computers. In CD ROM delivery systems, keyboards make sense when the user's queries involve large amounts of text.

Touch screens. In simple delivery applications such as point-of-sale kiosks in retail stores, the user indicates choices by touching menu selections. When a finger, pencil, or other rod-shaped object touches the surface of the screen at a certain location, the program compares this location with those of the names or symbols for the various current choices. It then selects the one corresponding to the touched location.

Touch screens can be constructed by means of several technologies, including:

- A yoke of LEDs and photocells that surround the screen. Touching the screen breaks some of the beams, and special circuitry calculates the horizontal and vertical locations of the touch.
- Ultrasonic transducers that launch sound waves across the glass faceplate. The touch causes sound waves to be reflected. Location is calculated from the delay times that occur when the reflected sound waves return to the source.
- A pair of transparent plastic membranes with conductive coatings that are stretched over the screen. Where touched, the coatings contact each other. Location is calculated from the resulting electrical resistance.

These methods require that you add components either to a standard monitor or to a TV screen or that you use a special monitor with a circuit card added to the host. All three options cost several hundred dollars. They also lead to "messy faceplate" as fingerprints build up. And all three suffer from poor resolution (pointing accuracy), because fingertips typically are large and people point them in a casual way.

A related method, the use of a *light pen* (a wand with a photocell at the tip that picks up the light from the phosphor as the scanning spot moves by), doesn't mess up the faceplate and can provide better resolution, but the user must play with a valuable and fragile object. Light pens also require special circuit cards to talk to the host, and they may cost several hundred dollars. However, light pen hardware and software drivers generally are cheaper than touch screen systems.

Mice, joysticks, and trackballs. In environments in which users can be trusted with something valuable and fragile, any of several cursor positioning devices can perform the function of a touch screen with much greater reliability and resolution.

The *mouse* is a palm-sized box that slides around on a desktop. As the user moves it, the cursor moves in the same way on the screen. When the cursor points to the desired item, the user presses a button on the mouse, selecting that item. After years of obscurity, the mouse now is in common use with general-purpose microcomputers, largely because it is bundled with the Macintosh.

Joysticks are familiar to many people as the control devices for video games. They are wands that you tip to direct the cursor. For delivery systems use, they are inferior to the mouse in resolution, but exceed the

mouse in durability and economy. Joysticks are quite inexpensive—sometimes costing as little as \$5 for parts.

Hard-core video game people know that serious players use *trackballs*. These are large plastic balls in a housing. The user rolls the ball to position the cursor. A trackball provides resolution at least equal to that of the mouse, doesn't require a lot of clear desk space, and usually is more durable than a mouse.

Most of these pointing devices talk to the host through a serial (RS 232C) port, of which most host computers have only two. Alternatively, they can talk to the PC's bus through a special circuit card. This approach is preferable except when no slots are available and power-supply capacity is spoken for already.

Voice recognizers. For text-based applications with vocabularies that, while small, are too large to fit on the screen as a menu, voice recognition is an attractive alternative for input or selection. These systems currently can handle only a few hundred words, which must be spoken clearly and distinctly (no continuous speech) into a recognition device that has been trained for a specific speaker. Voice recognizers are expensive and may increase the hardware budget by hundreds of dollars.

The user of a voice recognition system need not touch a keyboard or screen (very good for repair or manufacturing personnel with both hands on a machine), but must carry or wear a microphone. Practical systems show the word they think they've heard and ask for confirmation before proceeding. Commercial voice recognition systems include software drivers that let them pretend to be keyboards.

Input Generated by Other (Nonhuman) Devices

As mentioned, an information delivery workstation may also be a more general-purpose computer system. In a growing category of CD ROM applications, the delivery system responds to calls from word processors (dictionary or thesaurus lookup), computer-aided design (CAD) systems (standards, symbol, or template lookup), and file servers on networks (historical or reference information).

In these cases, the other applications input data directly, in digital form. Human input goes to these other applications by the customary means.

Display Equipment

A delivery system can display information it finds and processes in a variety of ways. Common output devices include:

CRT displays. Most TV sets and computer monitors display data on a cathode ray tube (CRT). CRT technology has been improving for decades and, despite its many problems, remains the best method for displaying large amounts of text or images.

CRTs are bulky and fragile. They can consume lots of power. They often don't preserve the exact geometry of images, tending to stretch and bend them. They are hard to read in bright light. If used incorrectly, they can emit radiation. The electronics driving them can make high-pitched sounds. If you consider any of these characteristics an absolute disqualification, investigate one of the alternatives discussed below. Chapter 8 deals at length with display technology and applications, so in this discussion we address functional issues.

Most PCs use CRT displays specially designed to show a moderate amount of text (usually 24 rows of 80 characters) clearly, although with no graphic variety. Such monitors usually can show graphics, provided the appropriate circuit card is installed in the host. Graphics cards treat the screen as a mosaic of dots (called *picture elements* or *pixels*). The more pixels per screen, the more detail (*resolution*) is delivered.

Manufacturers' descriptions of screen resolutions sometimes are reminiscent of olive size nomenclature (the smallest is "giant"). The lowest resolution commonly used is a mosaic 320 pixels across by 200 high (the IBM Color Graphics Adapter low-resolution mode). This is coarse enough for a standard TV receiver to display, but too coarse to be satisfactory in most applications. Other common pixel maps include 640 by 200 (IBM CGA high-resolution), 640 by 350 (IBM EGA), and 720 by 348 (Hercules).

None of these is sufficient to show large areas of maps, business documents, or engineering drawings. For these demanding applications, other graphics standards exist with pixel maps such as 1024 by 800, 1280 by 800, 1280 by 1024, and even 2200 by 1728. Displays and adapters for these standards are available for PCs, but cost from \$1500 to \$5000. They also usually require the developer to write or buy special application software.

This range of resolutions can be found in monochrome (not necessarily black and white, because many screens produce only green or amber light) or in color. Color costs more because of the more complex CRT, the additional electronics needed to drive it, and the more complex driver card required to address it.

Most color computer displays are limited to a small number of colors at each pixel (for example, 16), poorly suiting them to depict natural scenes. Even if the same pixel counts are specified, monochrome screens usually deliver more detail and sustain their image quality with less maintenance.

The circuit card, sometimes called a *video* or *display adapter card*, is installed in the host computer. This card, which drives the display, consists primarily of semiconductor RAM that stores the brightness and color of each pixel on the screen. The memory organization can be thought of as a map of the screen, hence the widely used term *bit map* to describe pixel-oriented graphics displays.

Television sets and audio output. As CD ROM matures, interest in interactive applications that present TV-quality images overlaid with simple computer graphics and accompanied by sound (ranging from very high fidelity down to telephone squawk) is growing. This idea is at the heart of the CD-I (Compact Disc Interactive) standard, but probably will be implemented by some vendors before CD-I stabilizes.

TV sets are as omnipresent a display device as can be imagined, so it makes sense to tailor the delivery systems for mass applications to use them. Compared with computer monitors, TV sets usually are better at delivering natural images but much worse at showing crisp graphics or lots of letters and numbers.

LCDs. Many portable computers deliver information using liquid crystal displays (LCDs). These are addressed as dot matrices, similar to those in CRTs, and produce somewhat more blurred images and text. Brightness and the direction of light falling on the screen are critical. Recently developed backlit LCDs, however, can be read under less stringent lighting conditions.

LCDs are attractive purchases because they are almost flat (this really means thin compared with a CRT) and draw very little power. Color LCDs are beginning to appear commercially.

GPDs, ELDs, et al. Other "flat" image displays employ gas plasma (GPD) and electroluminescent (ELD) technologies. Their display and usage characteristics are somewhere between those of CRTs and LCDs. Military users like the GPD's ruggedness.

ELDs use a powdery, electrically sensitive material that glows when electronically stimulated. Their colors range from a soft blue to yellow, green, and white. Electroluminescent display brilliance usually is about the same

as that of CRT screens. By their nature, ELDs involve none of the glass, vacuum, and high-voltage hazards associated with CRTs; nor do they produce radiation.

Electroluminescent displays at present are monochrome. Displays that are gray-scale capable are rare, and experimental color EL panels are years away from manufacture. Still, these displays are popular with several portable computer manufacturers and may eventually find their way into CD ROM display workstations.

Vacuum Fluorescent (VF) displays are miniature CRT systems. However, like electroluminescent displays, these thin, glass-enclosed displays use point-addressable (bit-image) matrices to create images instead of the raster-scanned electron beam system typical of large CRT displays. Vacuum fluorescent screens operate at lower voltages than do electroluminescent panels and CRTs and therefore are appropriate for portable and hazardous-environment CD reader systems.

Vacuum fluorescent displays usually are monochrome. The normal color for these "fluorescent" displays is a soft blue-green. However, some manufacturers import other colors to the system using colored overlays. Red, yellow, blue, and green can be obtained, but white cannot.

Color VF displays still are experimental. They lend themselves to portable and battery-powered applications, as do EL displays. Vacuum fluorescent display systems' greatest limitation is screen size. The largest VF glass packages are about the size of a 7-inch TV screen. That's a safety measure; it keeps atmospheric pressure from crushing the vacuum glass package. But the total display thickness, less than an inch, still makes color VF displays appropriate for portable CD reader applications.

Printers

A number of reference applications, such as those for air navigation and medical investigation, may require "hard copy" printouts of information. Several types of printers are available commercially.

Daisy-wheel printers produce truly high-quality text at PC prices. These speedy descendants of electric typewriters spin wheels or thimbles covered with molded representations of the characters in a typeface along a ribbon, striking the appropriate character at the appropriate time to make the requested letter. Daisy-wheel printers are rarely used for graphics, and changing their typefaces is a tedious manual operation.

Dot-matrix printers lay down small dots in a fine mosaic pattern to form characters or images. As with pixel-oriented displays, the resolution of

dot-matrix printers varies widely. Resolutions are usually expressed in dots per inch (dpi). They range from 72 dpi for the basic printer often attached to a PC, through 120 to 240 dpi for higher quality *near letter quality* or *graphics* printers, through the 300 dpi and higher typical of laser printers, and climbing to 2500 dpi and beyond for phototypesetters and for drum plotters.

Daisy-wheel and dot-matrix printers are mostly mechanical, moving a print head across the paper. The head usually contains an array of pins, which it drives against a ribbon to transfer ink to the paper. This *impact* method is valuable for printing multicopy forms, but the head wears quickly, and it is noisy. Prices for such printers can range from \$200 to more than \$1000.

Other print head technologies, such as thermal transfer, ink jet, and direct thermal writing, are in the low resolution range. These printers are relatively quiet, but can mark only one surface at a time. Some new models can print in a limited range of colors. The direct thermal writing printer requires special paper and is used widely in facsimile (*fax*) machines working at 200 dpi. These printers are not quite as well developed as the impact models mentioned above, and the least expensive is about \$300.

Laser printers make it easy to print a page in a variety of typefaces and to print graphics and are part of the hot new area of *desktop publishing*. And, with potential resolutions of 300 dpi and higher, they are the most versatile hard-copy units for CD ROM applications. Unfortunately, laser printers still cost more than \$2000, are evolving rapidly, and require considerable maintenance.

Video printers are a new entry into the market and faithfully reproduce video images. Most rely on some sort of thermal printing technology. Color video printers use either instant camera film packs or color thermal ribbons on smooth paper. The photographic cameras produce color screen images about the size of a snapshot, but the resolution is too poor to recreate text or detailed images. Color-thermal transfer ribbons allow page-sized printing. These systems can reproduce several colors and some shades of gray. Printing an image can cost as much as a dollar per copy.

Additional Components

Many applications require special delivery system components.

Data-encryption devices. Data on a CD ROM disc is often encrypted for several reasons. One is security. Many companies encrypt their private databases for this reason. Another is commercial protection.

Widely distributed publications often are encrypted to ensure that only customers who buy the decoding software can use the publication.

Either way, the delivery system must decode the data. In most systems, after the file manager makes the data available to the operating system, decryption software decodes the data. However, software decryption creates problems with speed and security that can be overcome only by adding a hardware decryptor.

A decryptor may be a plug-in circuit board, but more typically it is a metal box with a key-locked switch, or a card reader. The U.S. government has set very tough standards for these devices' security and dependability, and the manufacturers of government-approved devices are happy to let you know this.

Image compression and decompression devices. Images (such as business documents, engineering drawings, maps, or printed pages) can quickly consume space on a CD ROM. Many system developers use compression techniques to increase the number of images that can be placed on a disc. Compression involves analyzing a stream of data and rewriting it to be more compact.

Images are compressed before they are mastered to a disc. The combination of hardware and software compression techniques is discussed elsewhere in this book. A compressed image must be decompressed to its original form before it is displayed or printed. This requires a software or hardware decompression device.

CD-I (Compact Disc-Interactive)

CD ROM's inherent flexibility enables you to do all sorts of impressive and surprising things with data, and it lets you attach the drives to many kinds of computers and other data systems. But it also requires application developers to learn more about the technology than they might wish to know. In addition, the open format of CD ROM tends to cause incompatible system architectures to proliferate and keeps the minimum cost of a delivery system rather high—typically, the cost is at least as much as that of a PC and CD ROM drive.

Philips and Sony early in 1986 announced a new CD ROM-based standard that outlines a completely self-contained delivery system that should be a low-cost alternative. The standard is called CD-I for *Compact Disc-Interactive*. Commercial availability of such systems, aimed at the consumer market and selling for less than \$1000, is expected during 1988.

The CD ROM standard tells you how to record 2048-byte *blocks*, or *sectors*, of data on a disc in such a way that each has its own explicit address. This enables you to retrieve data randomly or in big, serial chunks. The standard also provides for error correction and detection sufficient to protect the stored data at least as well as does any other available storage medium. (See the following chapter for more information about the CD ROM standard format.)

That, however, is the extent of the specification. All decisions about system design and packaging are left to the product's author or vendor. These design decisions involve determining such things as:

- A logical file format for the disc (such as the proposed High Sierra format we describe elsewhere in this book).
- The disc's directory and file structures.
- Graphics and audio formats.
- The operating system on which the application will run.
- The type of processor or computer system the application will require.
- The display system to be used.

Philips/Sony's CD-I standard is based on the CD ROM format, but provides a tightly specified software and hardware framework for developing and presenting data, images, and sound. Designed for minimum hardware cost and maximum compatibility of discs and delivery systems, the CD-I standard will specify everything about:

- Digital representation of pictorial and audio information.
- Data organization on the disc.
- Hardware and software system components.

CD-I delivery systems will work as adjuncts to television sets and home hi-fi systems, which will be their output devices. They are intended to have computing power sufficient to support rapid, complex interaction; presentation of good-looking graphics; and information searching. This comes from a dedicated computer based on a 68000 microprocessor and 512 KB of RAM, running a ROM-based variant of the OS9 operating system. Input could be from a keyboard, but more likely will come from some sort of pointing device, because the design calls for simplicity.

People now selling CD ROM systems to professional, technical, and industrial customers look forward to using low-priced CD-I players to

broaden their markets—although this may be at the expense of features or performance value that their current products deliver.

Creative people are now developing CD-I multimedia programs for use in the home market. They hope to share in some of the initial success that CD-I's grandfather, CD audio, is experiencing. Time will tell whether CD-I achieves the omnipresence of CD audio and PCs, or whether it will be just another special-purpose information tool like the videodisc.

Specialized CD ROM Systems

Some CD ROM publishers design applications to run only on specially designed hardware. Often they believe that offering a dedicated system makes good business sense. But just as often, developers discover they cannot find the right combination of equipment to satisfy their design specification.

Two commercial systems demonstrated recently are examples of dedicated, portable CD ROM delivery systems. One is a navigation system for cars that uses a form of dead reckoning to determine current position. This information in turn is used to decide which part of a map of the area stored in the system will be displayed on a CRT that the driver views.

Magnetic tape cassettes at present store the digitized map information for this system, but their limited capacity restricts coverage to part of a single metropolitan area. With CD ROM storage, one disc could store maps representing a large part of the United States.

Another system, now in prototype, customizes navigation maps for pilots. The pilot works with the system, either on the ground or in the air, to put in a flight plan and other relevant information, and receives an appropriate map from the system's printer. Since this system must be carried onto small aircraft, it must be small and light.

Multiuser Systems

Most of this discussion of CD ROM delivery systems deals with single-user systems. In several special cases, one CD ROM disc can be shared among many users. For instance, as local area networks (LANs) connecting PCs become more common, we can expect CD ROM servers to appear on the networks along with print servers and file servers.

Sharing a CD ROM drive among many users is a dangerous economy, because the CD ROM standard limits performance to a level that will frustrate several users simultaneously requesting information. If many users will be active at once, it probably is wiser to give each a CD ROM drive

and a copy of the CD ROM disc containing the information of interest. Regardless of the drawbacks, a CD ROM file server might consist of the following elements.

A specialized *search server* (or *library server*, or *teacher server*) could contain:

- A dedicated hardware search engine.
- A CD ROM drive and device interface.
- A network node.
- A hard disk to store the index temporarily or to buffer requests and responses.

This machine responds to queries by sending back blocks of information (text or, with proper linking of protocols, images) that the user's computer or terminal displays. As with the single-user system, the user can invoke information directly, or another application can do so.

The CD ROM delivery system also can be integrated into a general-purpose file server, letting users call for information that is either on the file server's hard disk or on the CD ROM disc. In this case, the search engine may reside in the user's computer.

Mainframes and Minicomputers

CD ROM can be useful when attached to bigger computers, but the same caution applies here. If a network of users communicates through terminals by means of a central computer that stores all storage peripherals, CD ROM applications will require either multiple CD ROM drives, to avoid head contention, or the transfer of CD ROM data to one or more high-performance hard-disk drives. Under these circumstances, CD ROM becomes purely a distribution medium, rather than one accessed directly.

CD ROM is well suited to deliver system and application software to large computers. Many large system software packages are tens of megabytes in size and are sent out on magnetic tape. Tape is expensive, bulky, and fragile compared with CD ROM and does not allow random access to data.

General Concerns

To this point, we have focused on the components of a delivery system. We will now deal with a few issues you should address when you begin analyzing your delivery system requirements and choosing components.

Frequency of access. CD ROM, a disc medium, can be accessed directly. However, because access and transfer are slow, very active users (or networks of users) will demand more throughput (measured, for example, in records per hour) than a CD ROM drive delivers.

Three remedies (ordered below according to the degree of throughput improvement and cost) are available:

- Provide multiple CD ROM drives, each loaded with a copy of the same disc, controlled to spread access requests among the drives.
- Transfer the database index on the CD ROM disc to a hard disk to support the many accesses typical of a query session. The CD ROM drive then supplies only the relatively large blocks of information called for once the retrieval list has been refined.
- Transfer the CD ROM disc's entire contents to one or more hard disk drives, and let them support the accesses. This can be a costly proposition since one CD ROM disc's contents would require about five 100 MB hard disks, costing about \$1500 apiece.

Number of users. Giving many users access to a CD ROM drive can cause the kind of access overload mentioned above. The same remedies are available, but so is the straightforward remedy of distributing CD ROM drives among the users. As the cost of a drive decreases, this will make even more sense.

Updating. At the current state of the technology, CD ROM discs must be mastered and replicated in special facilities. Thus, when a database changes or evolves rapidly, this medium may be inappropriate unless an updating strategy is in place.

Many developers are exploring new ways to update. Some strategies involve periodic reissue of an updated disc. Others require an additional hard-disk drive to store the new information. Still others may require modems, which receive new information by telephone or drives for removable magnetic media. Every method, no matter how simple it seems, also will require software designed to unify the directories and data structures of the updated material and the CD ROM data. Since the subject of updating is rather complex, the options are discussed in more detail in Chapter 14.

Security. Distributing information via CD ROM eliminates the "hacker" vulnerability of networks and telecommunication systems. However, the organization distributing the CD ROMs must deal with unauthorized use of a CD ROM disc. Most commonly, data encryption,

building unconventional directory structures and file formats, and keeping the discs under lock and key prevent this.

It is frequently said that pirating CD ROM is impossible because of the difficult technology and expensive equipment needed to replicate the discs. This, unfortunately, is not true. The commercial pressure to find simpler and less expensive ways to duplicate CDs carries over directly to CD ROM, making piracy fairly simple. You can find more information on data protection in Chapter 13.

Building Your Own System

This section is a one-minute guide to putting together a delivery system for resale. It provides a quick look over the shoulder of a database publisher that needs to equip its customers with adequate delivery hardware and that never really intended to be in the hardware business.

What to Buy

Everything you can. Shop hard and flexibly, looking for any way to use what others have designed, put into production, and committed themselves to support.

Hardware design and integration, even at the level of attaching drives to device interfaces and host computers, contain all kinds of technical traps. Getting the hardware fixed in the field usually is something you'd rather someone else did. The hypothetical publisher probably will use a PC with an internal-mount CD ROM drive containing an internal cable to a SCSI host adapter. The choice of SCSI frees the publisher to add Winchester, writable optical disc, or more CD ROM drives to the system without software gymnastics.

Writing and debugging system software (such as drivers for storage devices, displays, or input devices) is more dangerous than designing the hardware. Field support of system software is never effectively provided except by its authors, so buy your system software unless you enjoy traveling to visit angry clients and spending many painful hours fixing bugs. The publisher should choose a CD ROM drive manufacturer that also has a proven driver for the PC's operating system.

What Hardware to Build

For marketing reasons, you may wish to have a custom package designed for your system so it looks like a system and not like something that escaped from a computer store. Designed properly, this package also can make the system easier to use and harder to abuse.

What Software to Write

There are good reasons to write the application *front end* (that is, the software that controls communication between the human user and the heart of the application, be it a database management system (DBMS), full-text search, menu process, or whatever). You want to make your product look special when the user views it on the screen, and, more substantially, you want to suit the product to its targeted user or market.

This technical area, more than any other, is the one in which the publisher can best express its understanding of the market, its needs, and its culture. The perceived difference between a computer (that is, something fragile and scary) and a tool that does *my* job (that is, an ally, something appropriate and useful) lies mostly in this user interface.

How Much Marketing Value to Add

If you've designed (or had designed for you) a good, well-tailored user interface and physical package, they'll be unobtrusive in use, letting your customers feel they're interacting directly with the database you sold them. Drive this point home in your advertising and sales. Semantics are all-important here.

If customers understand that you are selling them a genuinely useful, ready-to-use solution to their very own problems, they'll like you more and pay you a higher price. Otherwise, they may feel you are trying to sell them the same kind of computer they could buy more cheaply through the mail, together with some general-purpose hardware they can't use until they've spent a lot of money on local "consultants" who are trying to finish the job.

Conclusion

Electronic publishing and database distribution are now in their early days. If the subject were cars, development would be at that point during the first decade or so of this century, when mass production of simple, straightforward machines had not yet begun. Inventors still are the main vendors of CD ROM hardware, and major revelations on how to do the job come several times a year.

Many publishers already have put together adequate delivery systems and are making money selling database products to markets that need and benefit from them to a degree that far more than compensates for their cost and awkwardness. Other publishers considered acting now, but decided to wait a bit for the frenzy of delivery hardware development, not to mention

program production, disc replication, and all other phases of this technology, to subside as the products mature.

This maturation will involve several steps. Along the way, some publishers will find that they need to act *now* or that their markets oblige them to put together delivery systems with special capabilities (such as very high resolution displays). The key to success remains understanding the market and applying technology judiciously in its service.

About the Author

Leonard Laub studied physics at the University of Chicago, Illinois Institute of Technology, and Northwestern University. In 1965 he began work in the research department of Zenith Radio Corporation where he conducted research in the fields of acousto-optics, signal and image processing, laser system design, and optical videodisc recording and playback. In 1976 he went to Xerox Electro-Optical Systems where he set up a program to develop read/write optical disc systems for computer and office system applications. Laub founded Vision Three, Inc., in 1981 to implement at full-scale the consulting and custom systems work that he has been involved in since 1972. Listed in *Who's Who in Frontier Science and Technology*, Laub is the author of numerous articles and reports, the holder of 12 patents, and a member of OSA, IEEE, ACM, SPIE, SPSE, and the Western Area Committee of the IEEE Computer Society.