6.14 Historical Perspective and Further Reading

The history of I/O systems is a fascinating one. This section gives a brief history of magnetic disks, RAID, databases, the Internet, the World Wide Web, and how Ethernet continues to triumph over its challengers.

Many of the most interesting artifacts of early computers are their I/O devices. Magnetic tape was the first low-cost magnetic storage and today persists as the lowest-cost storage medium. Early tape drives used reel-to-reel technologies and linear recording, which were eventually replaced by tape cartridges and helical recording. As disks became cheaper, tapes were relegated primarily to archival purposes, causing additional focus on density, as opposed to speed, and on large-scale archival technologies such as tape robots.

The earliest random access storage devices were drums and fixed-head disks. A drum had a cylindrical surface coated with a magnetic film. It used a large number of read/write heads positioned over each track on the drum (see Figure 6.14.1). Drums were relatively high-speed I/O devices often used for virtual memory paging or for creating a file cache to slower-speed devices. Drums, which had no seek time, survived into the 1970s in higher-speed applications, such as paging or use in high-end machines. Eventually, improvements in disk speed and the significant cost advantage of disks eliminated drum technology. Large (2 to 3 feet in diameter) single-platter, fixed-head disks were also in use in the 1950s.

Disk Storage

In 1956, IBM developed the first disk storage system with both moving heads and multiple disk surfaces in San Jose, helping to seed the development of the magnetic storage industry in the southern end of Silicon Valley. Reynold B. Johnson led the development of the IBM 305 RAMAC (Random Access Method of Accounting and Control). It could store 5 million characters (5 MB) of data on 50 disks, each 24 inches in diameter. The RAMAC is shown in Figure 6.14.2. Although the disk pioneers would be amazed at the size, cost, and capacity of modern disks, the basic mechanical design is the same as the RAMAC.

Moving-head disks quickly became the dominant high-speed magnetic storage, though their high cost meant that magnetic tape continued to be used extensively until the 1970s. The next key development for hard disks was the removable hard disk drive developed by IBM in 1962; this made it possible to share the expensive

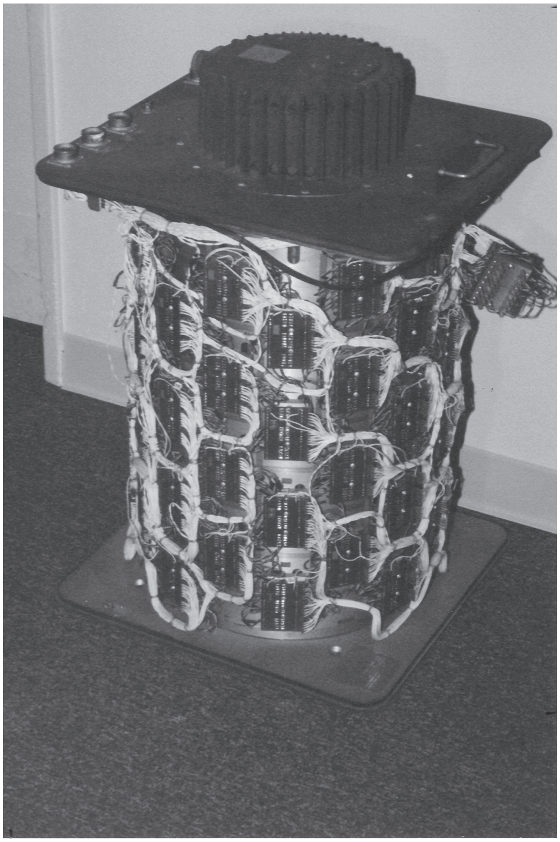

FIGURE 6.14.1 A magnetic drum made by Digital Development Corporation in the 1960s and used on a CDC machine. The electronics supporting the read/write heads can be seen on the outside of the drum.

drive electronics and helped disks overtake tapes as the preferred storage medium. Figure 6.14.3 shows a removable disk drive and the multiplatter disk used in the drive. IBM also invented the floppy disk drive in 1970, originally to hold microcode for the IBM 370 series. Floppy disks became popular with the PC about 10 years later.

The sealed Winchester disk, which was developed by IBM in 1973, completely dominates disk technology today. (All the disks shown in Figure 6.4 are Winchester disks.) Winchester disks benefited from two related properties. First, reductions in the cost of the disk electronics made it unnecessary to share the electronics and

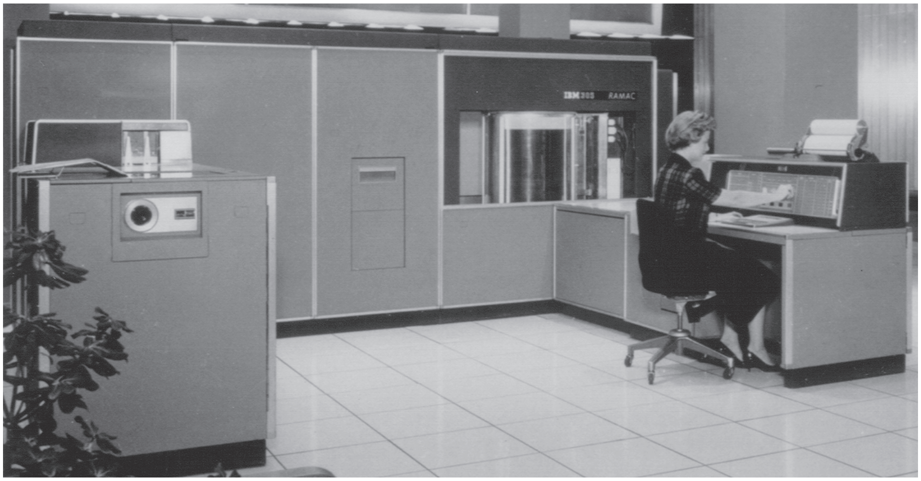

FIGURE 6.14.2 The RAMAC disk drive from IBM, made in 1956, was the first disk drive with a moving head and the first with multiple platters. The IBM storage technology Web site has a discussion of IBM's major contributions to storage technology.

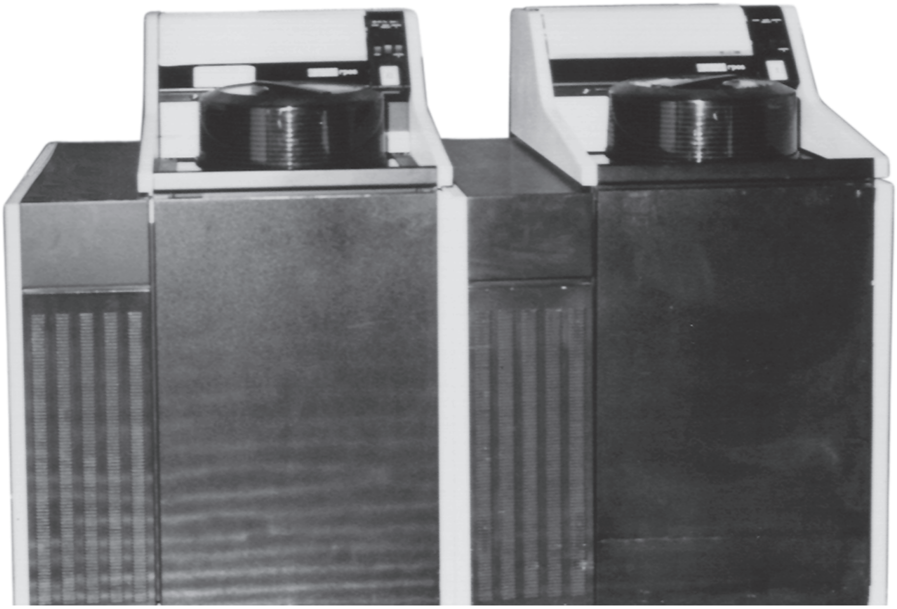

FIGURE 6.14.3 This is a DEC disk drive and the removable pack. These disks became popular starting in the mid-1960s and dominated disk technology until Winchester drives in the late 1970s. This drive was made in the mid-1970s; each disk pack in this drive could hold 80 MB.

thus made nonremovable disks economical. Since the disk was fixed and could be in a sealed enclosure, both the environmental and control problems were greatly reduced, allowing significant gains in density. The first disk that IBM shipped had two spindles, each with a 30 MB disk; the moniker "30-30" for the disk led to the name Winchester. Winchester disks grew rapidly in popularity in the 1980s, completely replacing removable disks by the middle of that decade.

The historic role of IBM in the disk industry came to an end in 2002, when IBM sold its disk storage division to Hitachi. IBM continues to make storage subsystems, but it purchases its disk drives from others.

A Very Brief History of Flash Memory

Flash memory was invented by researchers at Toshiba in the 1980s. They invented both the NOR-based flash memory in 1984 and the denser NAND-based flash memory in 1989. The first use was in digital cameras, starting with the CompactFlash form factor for NOR flash memory and the SmartMedia form factor for NAND flash memory. Today, all digital cameras, cell phones, and music players rely on flash memory.

Flash has successfully pushed the 1-inch diameter disk drives out of the storage market. In 2008, the first laptops are being shipped using flash memory instead of disk drives, and the first examples are both much more expensive and have less capacity. It will be interesting to see how well flash memory competes against the 1.8-inch disk drives in the next few years.

A Brief History of Databases

Although there had been data stores of punch cards and later magnetic tapes, the emergence of the magnetic disk led to modern databases.

In 1961, Charles Bachman at General Electric created a pioneering database management system called Integrated Data Store (IDS) to take advantage of the new magnetic disks. In 1971, Bachman and others published standards on how to manage databases using Cobol programs, named the Codasyl approach after the standards committee on which they served. Many companies offered Codasyl-compatible databases, but not IBM. In 1968 IBM had introduced IMS, which was derived from IBM's work on the NASA Apollo project. Both Codasyl databases and IMS are classified as navigational databases, because programs had to navigate through the data.

Ted Codd, a researcher at IBM, thought the navigational approach was wrongheaded. He recalled that people didn't write programs when dealing with the old punch card databases. Instead, they set up data flows through a series of punch card machines that would perform simple functions like copy or sort. Once the card machines were set up, you just pushed all the cards through to get your results. In his view, users should only declare the type of data they were looking for and leave it up to computers to process it. In 1970, he published a new way to organize and access data called the relational model. It was based on set theory; data was independent of the implementation and users described what they were looking for in a declarative, nonprocedural language.

This paper led to considerable controversy within IBM, because it already had a database product. Codd even arranged a public debate between him and Bachman, which led to internal criticism at IBM that Codd was undermining IMS. The good news was that the debate led researchers at IBM and U.C. Berkeley to try to demonstrate the viability of relational databases by building System R and Ingres.

System R in 1974–79 demonstrated its feasibility and, perhaps more importantly, created the Structured Query Language (SQL) that is still widely used today. However, these results were not sufficient to convince IBM, and some of the researchers left IBM to build relational databases for other companies.

Mike Stonebraker and Gene Wong were interested in geographic data systems, and in 1973 they decided to pursue relational databases. Rather than build on IBM mainframes, the Ingres project was built on DEC minicomputers and UNIX. Ingres was important because it led to a company that tried to commercialize the ideas, because 1000 copies of its source code were openly distributed, and because it trained a generation of database developers and researchers. The code and people led to many other companies, including Sybase. Larry Ellison started Oracle by first reading the papers from the System R and Ingres groups and then by hiring people who worked on those projects. Microsoft later purchased a copy of Sybase sources that became the foundation of its SQL Server product.

Relational databases matured in the 1980s, with IBM developing its own relational databases, including DB2. The 1990s saw both the development of object-oriented databases to address the impedance mismatch between databases and programming and the evolution of parallel databases for analytic processing and data mining.

ACM showered awards on this community. The ACM Turing Award went to Charles Bachman in 1973 for his contributions via IDS and the Codasyl group. Codd won it in 1980 for the relational model. In 1988, the developers of System R (Donald Chamberlin, Jim Gray, Raymond Lorie, Gianfranco Putzolu, Patricia Selinger, and Irving Traiger) shared the ACM Systems Software Award with the developers of Ingres (Gerald Held, Michael Stonebraker, and Eugene Wong). Jim Gray won the Turing Award in 1998 for his contributions to transaction processing and databases. Finally, the first two ACM SIGMOD Innovations Awards went to Stonebraker and Gray, and the 2002 and 2003 editions went to Selinger and Chamberlin.

RAID

The small-form-factor hard disks for PCs in the mid-1980s led a group at Berkeley to propose redundant arrays of inexpensive disks (RAID). This group had worked on the reduced instruction set computer effort, and so expected much faster processors to become available. Their two questions were: What could be done with the small disks that accompanied their PCs? What could be done in the area of I/O to keep up with much faster processors? They argued that it was better to replace one large mainframe drive with 50 small drives, as you could get much greater performance with that many independent arms. The many small drives even offered savings in power consumption and floor space.

The downside of many disks was much lower MTTF. Hence, on their own the Berkeley group reasoned out the advantages of redundant disks and rotating parity to address how to get greater performance with many small drives yet have reliability as high as that of a single mainframe disk.

The problem they experienced when explaining their ideas was that some researchers had heard of disk arrays with some form of redundancy, and they didn't understand the Berkeley proposal. Hence, the first RAID paper [Patterson, Gibson, and Katz, 1987] is not only a case for arrays of small-form-factor disk drives, but also something of a tutorial and classification of existing work on disk arrays. Mirroring (RAID 1) had long been used in fault-tolerant computers such as those sold by Tandem. Thinking Machines had arrays with 32 data disks and 7 check disks using ECC for correction (RAID 2) in 1987, and Honeywell Bull had a RAID 2 product even earlier. Also, disk arrays with a single parity disk had been used in scientific computers in the same time frame (RAID 3). Their paper then described a single parity disk with support for sector accesses (RAID 4) and rotated parity (RAID 5). Chen, et al. [1994] survey the original RAID ideas, commercial products, and other developments.

Unknown to the Berkeley group, engineers at IBM working on the AS/400 computer also came up with rotated parity to give greater reliability for a collection of large disks. IBM filed a patent on RAID 5 shortly before the Berkeley group submitted their paper. Patents for RAID 1, RAID 2, and RAID 3 from several companies predate the IBM RAID 5 patent, which has led to plenty of courtroom action.

EMC had been a supplier of DRAM boards for IBM computers, but around 1988, new policies from IBM made it nearly impossible for EMC to continue to sell IBM memory boards. The Berkeley paper crossed the desks of EMC executives, and so they decided to go after the market dominated by IBM disk storage products. As the paper advocated, their model was to use many small drives to compete with mainframe drives, and EMC announced a RAID product in 1990. It relied on mirroring (RAID 1) for reliability; RAID 5 products came much later for EMC.

Over the next year, Micropolis offered a RAID 3 product; Compaq offered a RAID 4 product; and Data General, IBM, and NCR offered RAID 5 products.

The RAID ideas soon spread to the rest of the workstation and server industry. An article explaining RAID in *Byte* magazine led to RAID products being offered on desktop PCs, which was something of a surprise to the Berkeley group. They had focused on performance with good availability, but higher availability was attractive to the PC market.

Another surprise was the cost of the disk arrays. With redundant power supplies and fans, the ability to "hot-swap" a disk drive, the RAID hardware controller itself, the redundant disks, and so on, the first disk arrays cost many times the cost of the disks. Perhaps as a result, the "inexpensive" in RAID morphed into "independent." Many marketing departments and technical writers today know of RAID only as "redundant arrays of independent disks."

In 2004, more than 80% of the nondesktop drive sales were found in RAIDs. In recognition of their role, in 1999 Garth Gibson, Randy Katz, and David Patterson received the IEEE Reynold B. Johnson Information Storage Award "for the development of Redundant Arrays of Inexpensive Disks (RAID)."

Wide Area Networks

The earliest of the data interconnection networks are WANs. The forerunner of the Internet is the ARPANET, which in 1969 connected computer science departments across the United States that had research grants funded by the Advanced Research Project Agency (ARPA), a U.S. government agency. It was originally envisioned as using reliable communications at lower levels. It was the practical experience with failures of underlying technology that led to the failure-tolerant TCP/IP, which is the basis for the Internet today. Vint Cerf and Robert Kahn are credited with developing the TCP/IP protocols in the mid-1970s, winning the ACM Software Award in recognition of that achievement.

In 1975, there were roughly 100 networks in the ARPANET; in 1983, only 200. In 1995, the Internet encompassed 50,000 networks worldwide, about half of which were in the United States. That number is hard to calculate for 2000, but the number of IP hosts grew by a factor of 20 in five years. The key networks that made the Internet possible, such as ARPANET and NSFNET, have been replaced by fully commercial systems, and yet the Internet still thrives.

The key application of the Internet is the World Wide Web. Tim Berners-Lee, a programmer at the European Center for Particle Research (CERN), coined the term in 1989 and invented the URL for information access. In 1992, a young programmer at the University of Illinois, Marc Andreessen, developed a graphical interface for the Web called Mosaic. It became immensely popular. He later became a founder of Netscape, which popularized commercial browsers.

In May 1995, at the time of the second edition of this book, there were 30,000 Web pages, which represented less than one gigabyte, but the number was doubling every two months. In 2000 there were about 2.5 billion static Web pages, yielding a total of 20 to 50 terabytes, and that number was growing by 7 million pages a day. By August 2003, in the time frame of the third edition, the static Web had expanded to about 167 terabytes. The "deep Web," which consists of dynamic pages and intranet sites, is estimated to be 400 to 550 times larger than the "surface Web" of static pages [Lyman and Varian, 2003].

Local Area Networks

ARPA's success with wide area networks led directly to the most popular local area networks. Many researchers at Xerox Palo Alto Research Center had been funded by ARPA while working at universities, and so they all knew the value of networking. In 1974, this group invented the Alto (see Chapters 1 and 7) *and* the Ethernet [Metcalfe and Boggs, 1976], today's LAN.

This first Ethernet provided a 3 Mbit/sec interconnection, which seemed like an unlimited amount of communication bandwidth with computers of that era. It relied on the interconnect technology developed for the cable television industry. Special microcode support gave a round-trip time of 50 μs for the Alto over Ethernet, which is still a respectable latency. It was Boggs's experience as a ham radio operator that led to a design that did not need a central arbiter, but instead listened before use and then varied back-off times in case of conflicts.

The announcement by Digital Equipment Corporation, Intel, and Xerox of a standard for 10 Mbit/sec Ethernet in 1978 was critical to the commercial success of Ethernet. This announcement short-circuited a lengthy IEEE standards effort, which eventually did publish IEEE 802.3 as a standard for Ethernet.

There have been several unsuccessful candidates in trying to replace the Ethernet. The Fiber Distributed Data Interface (FDDI) committee, unfortunately, took a very long time to agree on the standard, and the resulting interfaces were expensive. It was also a shared medium when switches were becoming affordable. ATM also missed the opportunity, due in part to the long time it took them to standardize the LAN version of ATM. The editions of our books often introduce a challenger to Ethernet that must be removed by a subsequent edition.

Because of failures of the past, LAN modernization efforts have been centered on extending Ethernet to lower-cost media, to switched interconnect, to higher link speeds, and to new domains such as wireless communication.

Further Reading

Bashe, C. J., L. R. Johnson, J. H. Palmer, and E. W. Pugh [1986]. *IBM's Early Computers*, MIT Press, Cambridge, MA.

Describes the I/O system architecture and devices in IBM's early computers.

Brenner, P. [1997]. *A Technical Tutorial on the IEEE 802.11 Protocol* found on many Web sites.

A widely referenced short tutorial that outlives the start-up company for which the author worked.

Chen, P. M., E. K. Lee, G. A. Gibson, R. H. Katz, and D. A. Patterson [1994]. "RAID: High-performance, reliable secondary storage," *ACM Computing Surveys* 26:2 (June) 145–88.

A tutorial covering disk arrays and the advantages of such an organization.

Gray, J. [1990]. "A census of Tandem system availability between 1985 and 1990," *IEEE Transactions on Reliability* 39:4 (October) 409–18.

One of the first papers to categorize, quantify, and publish reasons for failures. It is still widely quoted.

Gray, J. and A. Reuter [1993]. *Transaction Processing: Concepts and Techniques*, Morgan Kaufmann, San Francisco.

A description of transaction processing, including discussions of benchmarking and performance evaluation.

Hennessy, J. and D. Patterson [2003]. *Computer Architecture: A Quantitative Approach*, third edition, Morgan Kaufmann Publishers, Chapters 7 and 8, San Francisco.

Chapter 7 focuses on storage, including an extensive discussion of RAID technologies and dependability. Chapter 8 focuses on networks.

Kahn, R. E. [1972]. "Resource-sharing computer communication networks," *Proc. IEEE* 60:11 (November) 1397–1407.

A classic paper that describes the ARPANET.

Laprie, J.-C. [1985]. "Dependable computing and fault tolerance: Concepts and terminology," *15th Annual Int'l Symposium on Fault-Tolerant Computing FTCS 15,* Digest of Papers, Ann Arbor, MI (June 19–21) 2–11.

The paper that introduced standard definitions of dependability, reliability, and availability.

Levy, J. V. [1978]. "Buses: The skeleton of computer structures," in *Computer Engineering: A DEC View of Hardware Systems Design*, C. G. Bell, J. C. Mudge, and J. E. McNamara, eds., Digital Press, Bedford, MA.

This is a good overview of key concepts in bus design with some examples from DEC machines.

Lyman, P. and H. R. Varian [2003], *"How much information? 2003," www.sims.berkeley.edu/research/ projects/how-much-info-2003/.*

This project estimates the amount of information in the world from all possible sources.

Metcalfe, R. M. and D. R. Boggs [1976]. "Ethernet: Distributed packet switching for local computer networks," *Comm. ACM* 19:7 (July) 395–404.

A classic paper that describes the Ethernet network.

Myer, T. H. and I. E. Sutherland [1968]. "On the design of display processors," *Communications of the ACM* 11:6 (June) 410–14.

Another classic that notes how building powerful coprocessors can be a never-ending cycle.

Okada, S., Y. Matsuda, T. Yamada, and A. Kobayashi [1999]. "System on a chip for digital still camera," *IEEE Trans. on Consumer Electronics* 45:3 (August) 584–90.

One of the few public descriptions of a camera chip.

Oppenheimer, D., A. Ganapathi, and D. Patterson [2003]. "Why do Internet services fail, and what can be done about it?," *4th Usenix Symposium on Internet Technologies and Systems*, Seattle, WA (March) 26–28.

A recent update on Gray's classic paper, this time focused on Internet sites.

Patterson, D., G. Gibson, and R. Katz [1988]. "A case for redundant arrays of inexpensive disks (RAID)," *SIGMOD Conference,* 109–16.

A classic paper that advocates arrays of smaller disks and introduces RAID levels.

Pinheiro E., W. D. Weber, and L. A. Barroso [2007] "Failure trends in a large disk drive population," *5th Usenix Conference on File and Storage Technologies* (FAST 2007).

Analysis of real failure data from tens of thousands of disks at Google, including many surprises to customers and manufacturers of disk drives. Largely corroborated by a different failure database in the paper by Schroeder and Gibson at the same conference.

Saltzer, J. H., D. P. Reed, and D. D. Clark [1984]. "End-to-end arguments in system design," *ACM Trans. on Computer Systems* 2:4 (November) 277–88.

A classic paper that defines the end-to-end argument.

Schroeder, B. and G. Gibson [2007]. "Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?" *5th Usenix Conference on File and Storage Technologies* (FAST 2007).

They mined data on component failures from supercomputing sites to come up with similar insights to the results by Pinheiro et al. from the same conference.

Smotherman, M. [1989]. "A sequencing-based taxonomy of I/O systems and review of historical machines," *Computer Architecture News* 17:5 (September) 5–15.

Describes the development of important ideas in I/O.

Talagala, N., R. Arpaci-Dusseau, and D. Patterson [2000]. "Micro-benchmark based extraction of local and global disk characteristics," *U.C. Berkeley Technical Report* CSD-99-1063, June 13.

Describes a simple program to automatically deduce key parameters of disks.