

◆ The Impact of Solid-State Electronics on Computing and Communications

Arun N. Netravali

The transistor has become the fundamental building block of all computing and communications systems. A remarkable progression of regular and persistent reductions in size and cost has resulted in ever-greater functionality per unit of silicon. This has enabled designers to achieve enormous increases in the performance of computing and communications systems by applying innovative software, hardware, and architectural improvements. No area of modern life is untouched by the astonishing progress of microelectronics and its powerful, closely allied field, photonics. Examples of such progress in this paper include the evolution of computers from massive, unreliable, experimental models to today's personal computer; the evolution of computer subsystems, such as magnetic and optical mass storage media; and the deployment of new, highly functional operating systems and other powerful software applications. In telecommunications, computers and stored program systems have been deployed ubiquitously to enable today's global network infrastructure. While voice traffic still predominates in telecommunications, the growth of data traffic and the promise of networked video and multimedia are spurring the evolution toward broadband networks, where current and emerging microelectronic and opto-electronic technologies are vastly expanding data rates in local, wide area, and long distance networks. Solid-state electronics also is addressing society's need for mobility, helping to fuel the rapid growth of portability, increase functionality, and extend battery life. Additionally, solid-state electronics is enabling the explosive growth in personal communications, making it possible for people to reach out via networks to databases and to each other. This paper discusses these subjects, provides some brief historical perspectives, and speculates about future developments.

Introduction

The roots of solid-state technology date to the late nineteenth century, but the theoretical roots were greatly strengthened with the development of quantum mechanics through the 1920s and 1930s. Quantum mechanics enabled an understanding of various solid-state structures, which in turn created sufficient insight to help guide more specific experiments. The efforts of Bell Labs researchers William Shockley, John Bardeen, and Walter Brattain that culminated in December 1947 in the discovery of the transistor were carried out in this context. Stimulated

by the Communications Act of 1934, which mandated universal, affordable nationwide telephone service, the three researchers were seeking an alternative to bulky, unreliable vacuum tubes and relays that could detect, amplify, rectify, and switch current in telecommunications equipment.

The researchers discovered a principle that was based on specific material properties and operating voltages. The discovery would substantially affect the development of solid-state electronics in the years that followed, and their research would extend far beyond

Panel 1. Abbreviations, Acronyms, and Terms

A/D—analog to digital	GUI—graphical user interface
AM—administrative module	I/O—input/output
AMPS—Advanced Mobile Phone Service or Analog Mobile Phone Service	IC—integrated circuit
ARPANET—Advanced Research Projects Agency NETWORK	IF—intermediate frequency
ASIC—application-specific integrated circuit	InGaAsP—indium gallium arsenide phosphide
ATM—asynchronous transfer mode	ISDN—integrated services digital network
BiCMOS—bipolar complementary metal-oxide semiconductor	ISI—intersymbol interference
CCS—common channel signaling	KIPS—thousands of instructions per second
CD—compact disk	LAN—local area network
CD-E—compact disk erasable	LSI—large scale integration
CDMA—code division multiple access	MAC—media access controller
CD-R—compact disk recordable	MIMD—multiple instruction, multiple data
CD-ROM—compact disk read-only memory	MIPS—millions of instructions per second
CD-RW—compact disk rewritable	ML—maximum likelihood
CMOS—complementary metal-oxide semiconductor	MR—magnetoresistive
CO—central office	NC—network computer
CPU—central processing unit	NDRO—nondestructive readout
D/A—digital to analog	OS—operating system
DC—direct conversion	PBX—private branch exchange
DCS—distributed communications system	PC—personal computer
DRAM—dynamic random access memory	PCS—personal communications services
DS1—digital signal level 1, a transmission rate of 1.544 Mb/s	PE—processing element
DSL—digital subscriber line	PMOS—positive (P-type) metal-oxide semiconductor
DSP—digital signal processor	PRML—partial response maximum likelihood
DVD—digital versatile disk	PR—partial response
DVD-RAM—digital versatile disk random access memory	RF—radio frequency
DVD-R—digital versatile disk (write once version)	RISC—reduced instruction set computer
DWDM—dense wavelength division multiplexing	RLL—run-length limitation
ECC—error-correcting coding	ROM—read only memory
ENIAC—Electronic Numerical Integrator and Calculator, the world's first general-purpose computer	RS—Reed-Solomon
ESS—electronic switching system	SIMD—single instruction, multiple data
FPGA—field-programmable gate array	SMP—symmetric multiprocessor
GaAs—gallium arsenide	SPC—stored program control
GIPS—billions of instructions per second	TCP/IP—transmission control protocol/Internet protocol
GSM—Global System for Mobile Communications	TDMA—time division multiple access
	TMS—time multiplexed switch
	TSI—time-slot interchanger
	T-S-T—time-space-time
	VLIW—very long instruction word
	VLSI—very large scale integration
	WAN—wide area network
	WORM—write once read many

their laboratory at Murray Hill, New Jersey, bringing about vast and mostly unforeseen societal changes.

In the course of time, two solid-state devices—the transistor and the laser (invented about ten years later)—would have profound effects on the ways we use and think about computers, as well as on the speed, ease of use, and cost of communications—and the networks that connect people and computers. These devices also precipitated the trend toward miniaturization of the various building blocks of computer systems, such as magnetic and optical storage devices. This trend has made the laptop personal computer, personal digital assistant, and cellular telephone possible and no doubt will enable other as yet unexpected and unanticipated products. The laser is a companion device to the transistor. Together, they are creating a veritable revolution in ultra high-speed interconnection and communication over long distances, regardless of whether voice, data, video, or images are being transmitted.

Without the underlying support of solid-state devices, such rapidly growing technologies as wireless communications and the Internet would not have been possible. This paper reviews a number of key topics in the past, present, and future of solid-state electronics as applied to computing and communications.

Microprocessors, Computers, and Communications

Computers have become so common that we take them for granted, and it is easy to forget that the first electronic digital computer was built in 1946—well within the lifetime of many adults—by J. Presper Eckert and John Mauchly of the Moore School, University of Pennsylvania. This machine, called the Electronic Numerical Integrator and Calculator (ENIAC), was used for generating artillery firing tables for the U. S. Army.

John Von Neumann, working at Princeton University in the Institute for Advanced Studies, was attracted to the ENIAC project while the group was investigating ways to improve the method used to enter programs into the machine. Von Neumann summarized the thinking of the group in a paper that proposed entering and storing programs as numbers, and he is generally credited with conceiving the stored pro-

gram computer. In fact, this technology was made possible by a sizable group of engineers, including both Eckert and Mauchly.

Of course, all the early computers used vacuum tubes for their switching elements, making the machines physically large, electrically inefficient, and very unreliable. After the invention of the transistor, these limitations were greatly reduced, resulting in the rapid growth in both the use and complexity of computers.

Integrated circuits (ICs) incorporating transistors, resistors, capacitors, and associated wiring on a monolithic chip of silicon became available in the mid-1960s. ICs initially integrated formerly discrete computer logic functions and contained only elementary logic gates and flip-flops. These early ICs further reduced the size of computers, thereby allowing an even greater increase in their complexity.

At the beginning of the era of large scale integration (LSI) in the early 1970s, fairly complex microprocessors and solid-state memories were introduced. Initially, these devices integrated existing computer circuits that had been implemented with discrete logic functions. More important, monolithic microprocessors made it practical to put many processing functions on ICs to meet demands of many emerging markets, including consumer goods such as pocket calculators and portable stereos.

Simple single-chip dedicated processors containing their own control and data store soon emerged for use in a variety of industrial applications. Later, they found their way into household appliances and automobiles. These dedicated processors are generally known as *microcontrollers*. Today, the trend is to embed a microcontroller along with other complex logic to produce a complete monolithic system solution—the proverbial “system on a chip.”

Estimates are that in the course of a day, the average person has more interactions with various computers than with people. Most of the various computer interactions are with the silicon chips (microcontrollers) that act as invisible agents operating, for instance, in our automobiles (a luxury car contains up to 60 microcontrollers), cell phones, fax machines, and home appliances (see **Figure 1**). These microcon-

trollers (small yet powerful stand-alone computers) represent a fast-growing segment of the semiconductor industry.

From its earliest days, the computer found applications in communications systems. Teletype terminals (along with punched cards) were one of the first input/output (I/O) devices for the computer. It did not take long for early computer users to realize that the teletype unit did not have to be attached to the computer. In 1939, George Stibitz, inventor of a relay-based electrical digital computer, conducted an experiment using telephone lines to link the machine in New York to Dartmouth University in New Hampshire. By the late 1950s, the military's anti-missile systems had target- and missile-tracking radar units located some distance from each other. Using modems, each radar unit fed data to a computer located in yet another location via dedicated telephone lines. The idea of networked computers grew as users installed more and more machines in disparate locations that needed to share such information as order, inventory, and payroll data.

Users soon realized that a network could help enhance computing reliability, share resources, and connect them in communities of interest. Both public telephone networks and private leased lines were used for computer interconnection. Shortly thereafter, local area networks (LANs) and wide area networks (WANs) were implemented as the cost of computers declined and the desirability for sharing data and printers increased. Communication networking systems—for example, Ethernet, Datakit®, and the Advanced Research Projects Agency NETWORK (ARPANET), later to become the Internet—were developed to facilitate the reliable transfer of data between computers.

These developments demonstrate an important synergy: From the beginning, the marriage of computers and communications was a natural and necessary interdependency. The linchpin that both binds and enables communications and computers is the continuous and relentless cost reduction in electronic circuits provided by shrinking silicon technology. **Figure 2** shows the cost reduction trend for high-performance complementary metal-oxide semiconductor (CMOS)

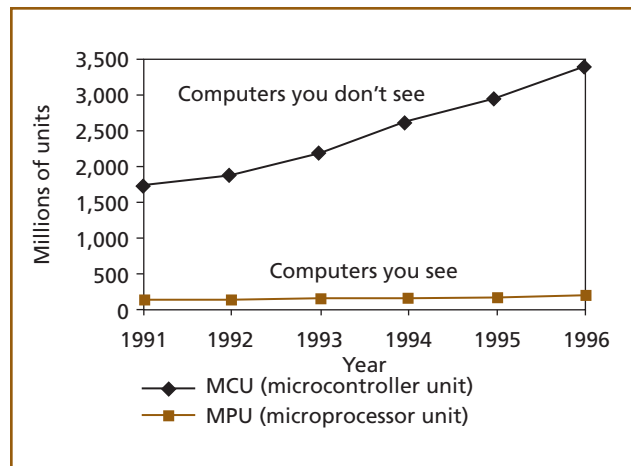


Figure 1.
Microcontroller growth versus microprocessor growth.

transistors of the type so commonly used in today's silicon chips. The data show that the cost of a transistor of the type used to build microprocessors has decreased by two orders of magnitude in just two decades. Furthermore, projections indicate that this trend will continue at the same rate for at least another decade. As **Figure 3** shows, this reduction in cost partly accounts for the dramatic decrease in the cost of computer processing power measured in dollars per millions of instructions per second (MIPS). As **Figure 4** depicts, the MIPS per chip also have increased at a similarly impressive rate.

Figure 5 shows the dramatic reduction in cost per bit for dynamic random access memory (DRAM), which is used extensively in computer systems. The DRAM costs per bit have fallen even faster than the cost of the transistors used in processors, plunging almost two orders of magnitude in a single decade—a direct result of the steady reduction in the feature size of silicon technology. **Figure 6** illustrates the astounding worldwide growth of DRAM use, which is currently increasing at the rate of 65% to 75% per year. This rate is much greater than that for computers alone, because the use of much more sophisticated software, as well as the rapid increase in applications requiring wider bandwidth, such as graphics and video, requires a dramatic increase in the memory needed for program and data storage.

How does this progress in silicon technology and computing enable the revolution that is occurring in

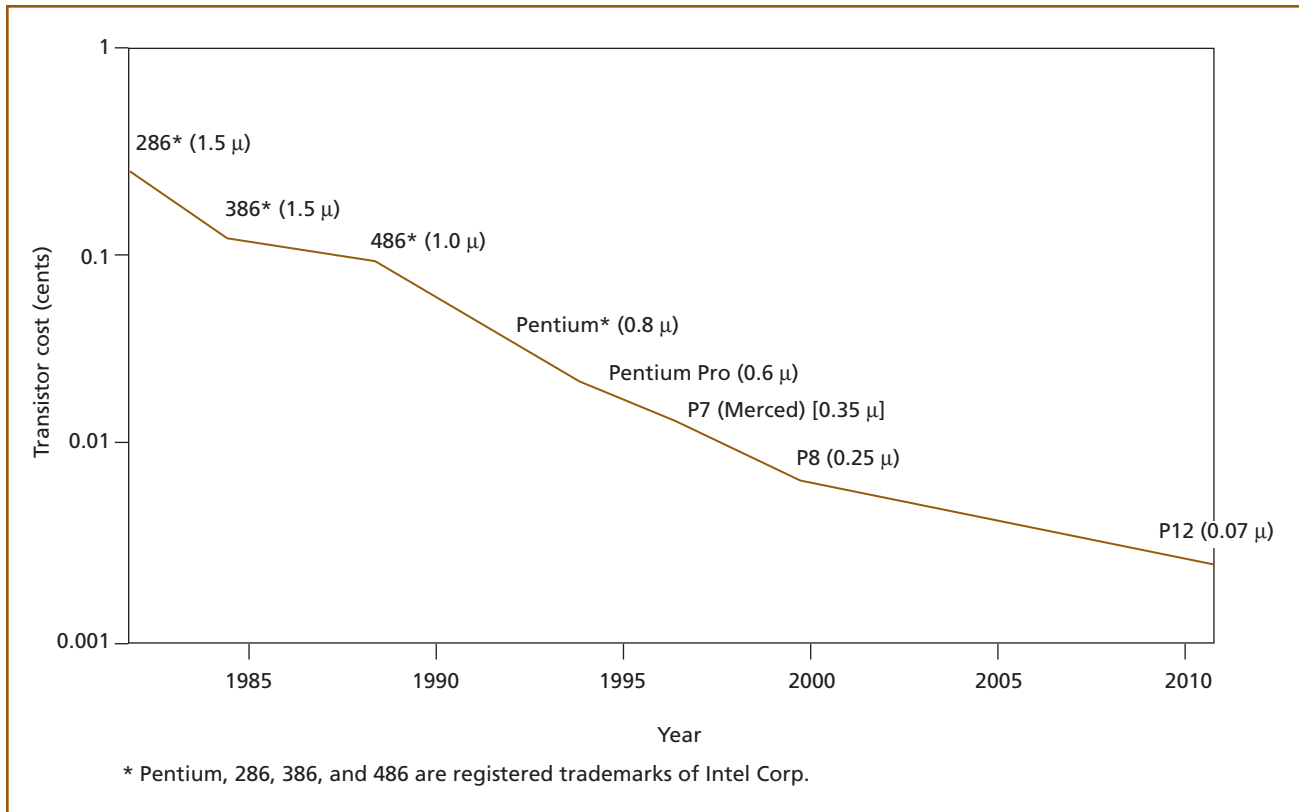


Figure 2. Cost reduction trend for high-performance CMOS transistors of the type used in silicon chips.

communications? **Table I** shows the general requirements for digital communications. Notice that going from audio to video communications requires improving the critical components of communication—data rate, processing power, and memory storage—by a factor of 1,000. This large upward step in requirements for video will absorb all the impressive gains in cost and performance that silicon technology can provide. Indeed, we could not even dream of many current and future products and services apart from the marvel of constantly improving silicon technology.

General-purpose processors of the type used in personal computers, workstations, and microcontrollers are not the only complex logic devices showing dramatic growth and performance improvements. **Figure 7** depicts the cost and performance trends for digital signal processors (DSPs) from Lucent Technologies' Microelectronics Group. DSPs are high-performance special-purpose logic devices widely used in communications equipment. As shown in the lower half of

Figure 7, an important DSP feature is the exceedingly low power consumption of today's products relative to general-purpose processors. Typically, such high-performance DSPs are incorporated into the design of wireless communications products, contributing to their low power consumption, long battery life (maximizing talk time between recharges), small size, and light weight (for easy portability).

Evolving Microprocessor Architecture

The first microprocessor, Intel's 4004, was shipped in 1971. This chip contained 2,300 transistors on a 12 square millimeter die fabricated using 10.0-μm PMOS technology, and it ran from a 108-kHz clock. Today, Digital's 21164 Alpha chip, one of the most aggressive desktop processors, contains 9.3 million transistors on a 209 square millimeter die fabricated using 0.35-μm CMOS technology, and it runs at a clock speed of 500 MHz.

A detailed analysis¹ of microprocessor integer performance shows that microprocessor performance is

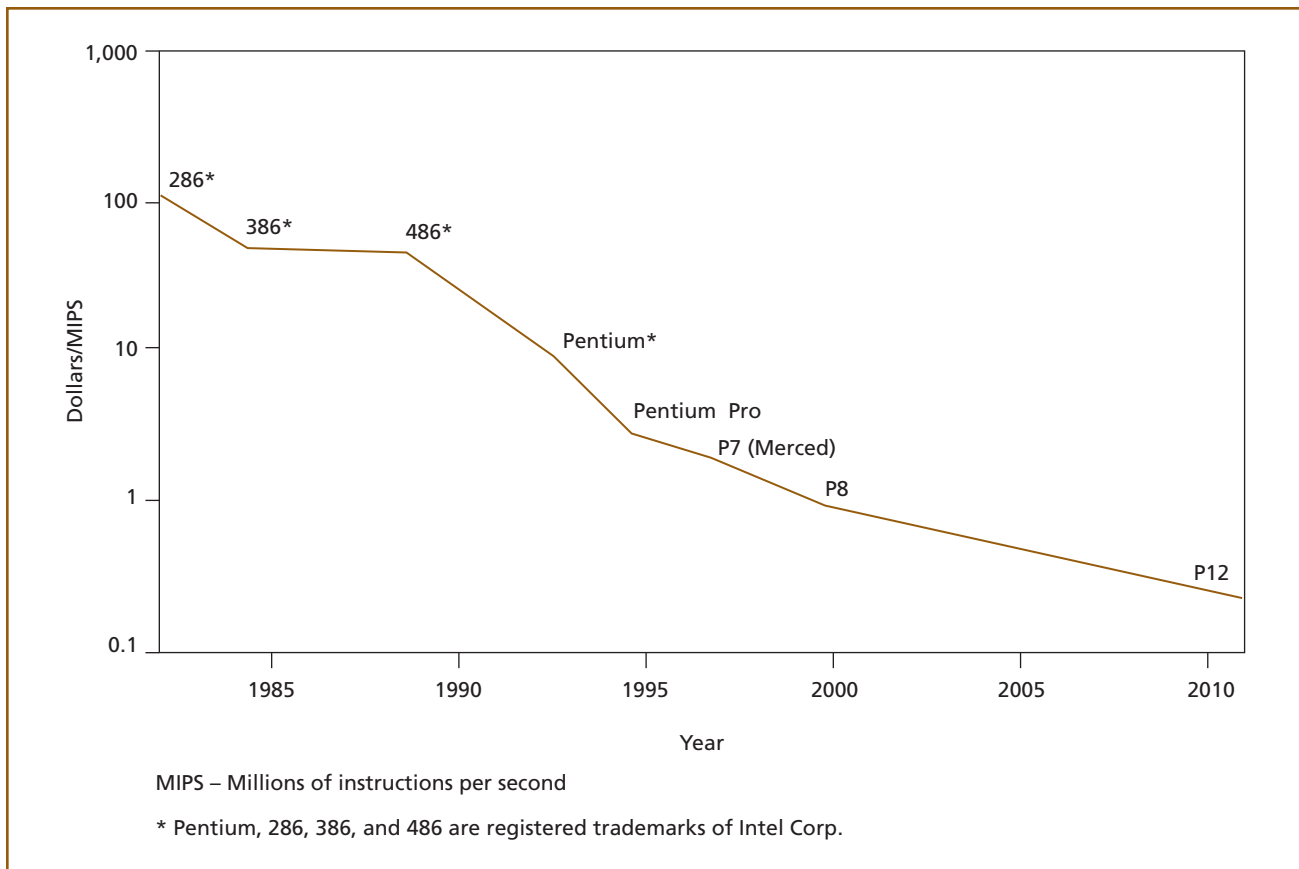


Figure 3.
The dramatic decrease in the cost of computing power.

more than doubling every two years. The increase in transistor speed due to faster silicon technologies (smaller feature size) is about 20% per year,² explaining part of the performance improvement. The remainder is due to architecture and circuit improvements. The exponentially increasing number of transistors available per chip (Moore's Law)² is the driving force behind these architectural improvements.

The following subsections briefly discuss the key architectural concepts that have improved microprocessor performance so dramatically over the past few years and that will continue to do so in the future.

Cache Memory

The first microprocessor to use on-chip cache memory was the Motorola 68020 in 1984. Today, most microprocessors are using fast on-chip cache memory to keep frequently used instruction sequences and data items close to the microprocessor core. Cache memory improves performance by reducing the number of

accesses into the slower external memory (often DRAM). In today's desktop processors, cache memory consumes as much as half the chip area (more than 100 KB of total cache on Digital's 21164 chip). To further improve performance, the width of the bus connecting the on-chip cache to external memory has been increased over time; it is currently 64 bits or more (Digital's 21164 chip features a 128-bit bus).

Datapath Width

The integer size that can be processed with a single instruction has been growing from 4 bits (Intel's 4004 chip) to 8, 16, and 32 bits, and now it has doubled again to 64 bits (Digital's 21164 chip). For example, on a 32-bit processor, a 32-bit addition requires only one instruction, whereas on an 8-bit processor it requires four. Memory addresses (pointers) are a particularly important integer data type, and the need to address memories larger than 4 GB is the driving force behind the current transition to 64-bit processors.

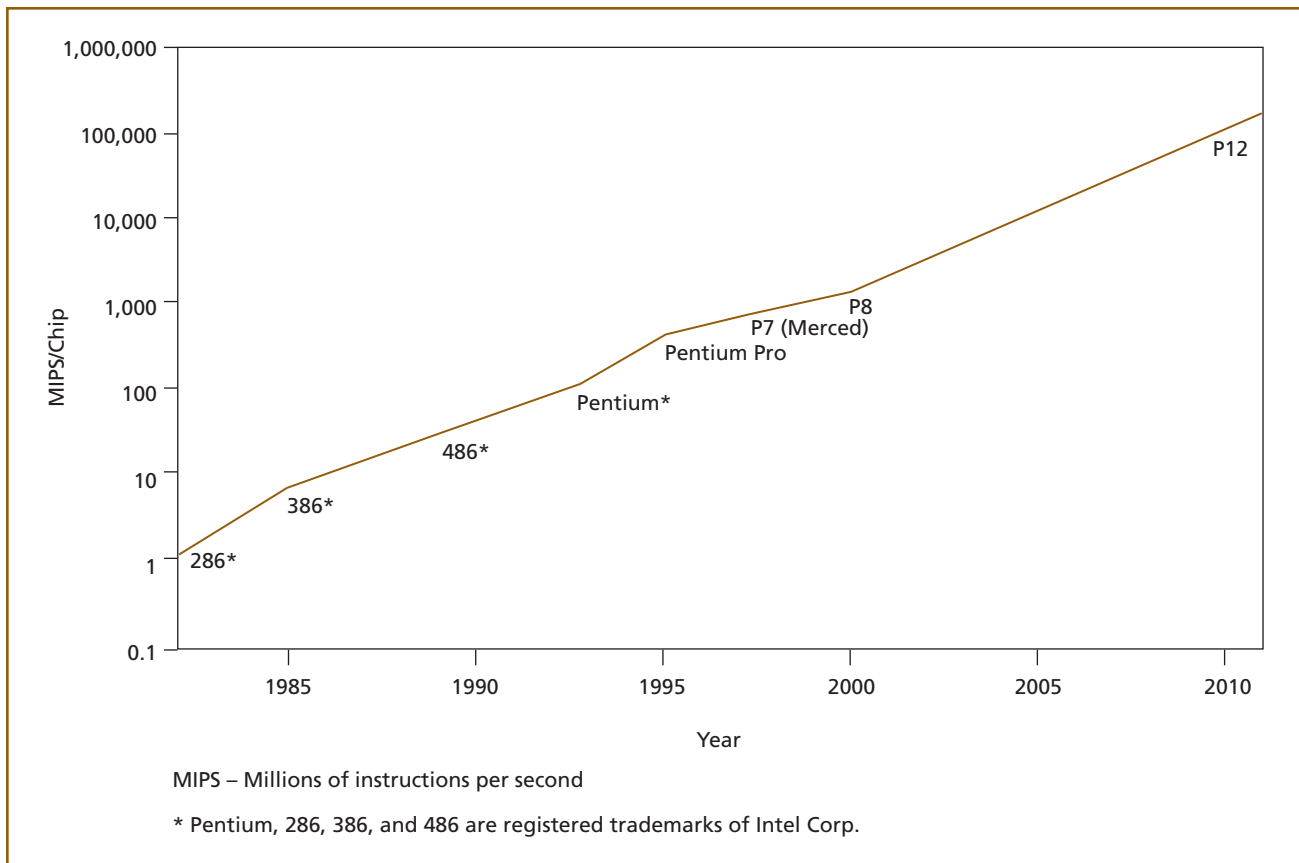


Figure 4.
Increase in MIPS per chip.

Pipelining

Pipelined microprocessors became popular in the early 1980s through the reduced instruction set computer (RISC) designs from the University of California at Berkeley and Stanford University. A pipelined processor works according to the principle of an assembly line, where instruction processing is broken down into stages. At the end of every clock cycle, each stage feeds its result into the next stage and starts working on the subsequent instruction.

In an ideal pipelined design, a new instruction is started every clock cycle, whereas in a non-pipelined design, the equivalent time of 5 clock cycles would be required for each instruction. Today's desktop processors have between 5 and 14 pipeline stages.

To promote pipelining in microprocessors, the instruction sets have been simplified (RISC). For example, all instructions have a fixed length (such as 32 bits), and instructions perform either a load/store

operation or an operation on the register file, but never both. To come as close as possible to executing one instruction per cycle, many architectural enhancements, such as data bypassing (forwarding), delayed branches and loads, speculative execution, and out-of-order execution, have been introduced.³

Non-RISC architectures, such as Intel's x86, are also taking advantage of RISC techniques. For instance, Intel's Pentium* Pro processor is translating x86 instructions into RISC-like instructions before executing them on a heavily pipelined processor.

Superscalar

A simple pipelined processor can execute up to one instruction per clock cycle whereas a superscalar processor can execute several. A superscalar processor analyzes the sequential instruction stream for independent instructions and dispatches them in parallel to the available execution units. The first truly superscalar processor, Intel's i960CA, appeared in late 1989

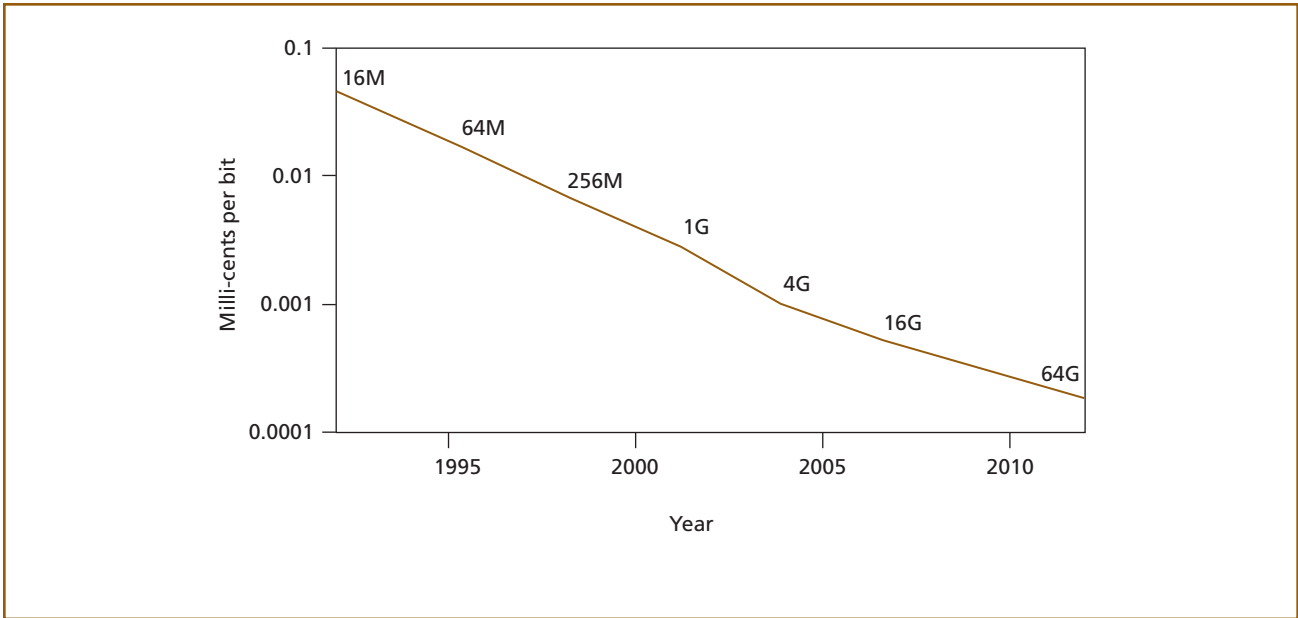


Figure 5.
Cost per bit for DRAM.

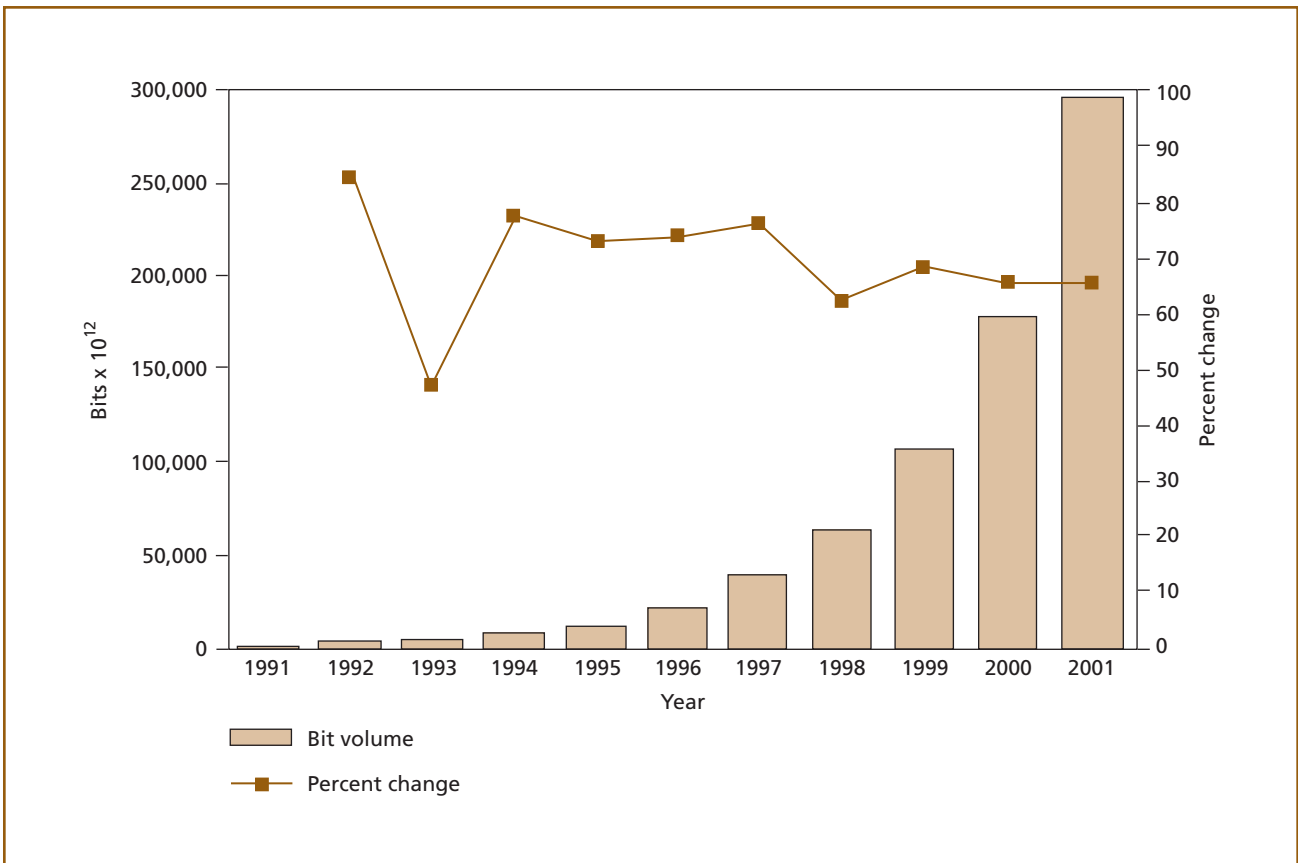


Figure 6.
Astounding worldwide growth of DRAM use and percent change per year.

Table I. General requirements for digital communications.

Data	Text	Audio	Video
Communication rate	10 b/s	10 kb/s	10 Mb/s
Processing	KIPS	MIPS	GIPS
Memory	KB	MB	GB

and could issue two instructions per cycle. Today's desktop processors can issue between two and six.

A study⁴ shows that as the scalarity of a particular processor is increased from 1 to 2, 3, and 4 instructions per cycle, performance increases from 1.00 to 1.46, 1.78, and 1.83. In general, superscalar processors are fairly large and complex, and they achieve only modest performance gains. However, they have the advantage of remaining object-code compatible with the previous generation, a very important consideration in the desktop processor market.

Promising Architectural Extensions

Processor architectures have been evolving since the ENIAC. The primary objective has been to reduce cost, increase performance, and/or extend functionality. One example is the single instruction, multiple data (SIMD) form, which appears as Sun's VIS,* a coprocessor for the UltraSPARC* or as Intel's MMX* coprocessor for the Pentium (1997). Like a single processor, a SIMD parallel processor is controlled by a single program, but it operates on multiple data sets in parallel. The same instruction is applied to all data sets. For example, one MMX instruction can operate on four 16-bit or eight 8-bit data sets in parallel. The cost/performance ratio of MMX-type extensions is excellent: The additional circuitry increases a chip's overall size by less than 10% while the performance gain on image/video processing increases by a factor of 2 to 4.

The next step is a more complicated architecture called multiple instruction, multiple data (MIMD). A MIMD parallel processor consists of several processing elements (PEs) plus some means for communication between the PEs (for example, shared memory or message passing). In a MIMD parallel processor, each PE executes a different thread (or task) of the overall task. Unlike the SIMD PE, the MIMD PE is a complete self-contained processor with its own instruction and data memories. Texas Instruments' (TI's) TMS320*C8x family of signal processors (1994) is

based on the MIMD concept. These multiprocessor chips contain one RISC processor and two to four DSPs communicating over a crossbar switch.

The third architecture that is getting major attention is the very long instruction word (VLIW) processor. It is similar to the superscalar processor in that it executes more than one instruction per clock cycle. Unlike the superscalar, the instruction-level parallelism is extracted and mapped to the available execution units at compile time. The compiler groups the instructions that can execute in parallel into a VLIW that, for example, can be 256 bits and can contain eight 32-bit instructions. Currently, the VLIW concept is used in such signal processors as Philips' TM-1* (1996) and TI's TMS320C6201 (samples in 1997). It is also speculated that Intel's and Hewlett-Packard's next-generation desktop processors will use VLIW technology. But effective tools for instruction scheduling and compilation must be demonstrated before commercial success is achieved. It is almost impossible to program a VLIW computer by hand. Furthermore, the program code for current VLIW machines is significantly larger than that for a DSP or RISC processor (low code density), which in turn causes higher memory cost. Both issues are the subject of current research.

The general industry view is that silicon will not be the limiting factor for the current computing and communications vision of the future. The various technology roadmaps produced by the Semiconductor Industry Association and SEMATECH show silicon technology easily carrying its share of the load for at least the next few decades.

Mass Storage

While solid-state memories have continued to show tremendous improvements in cost and capacity through the years, magnetic and optical mass storage media have also benefited from improvements in microelectronics and photonics technologies to retain a strong, competitive position for mass storage of data. In fact, the mass storage industry has become one of the world's largest, with a primary goal of increasing the amount of information that can be stored per disk unit area.

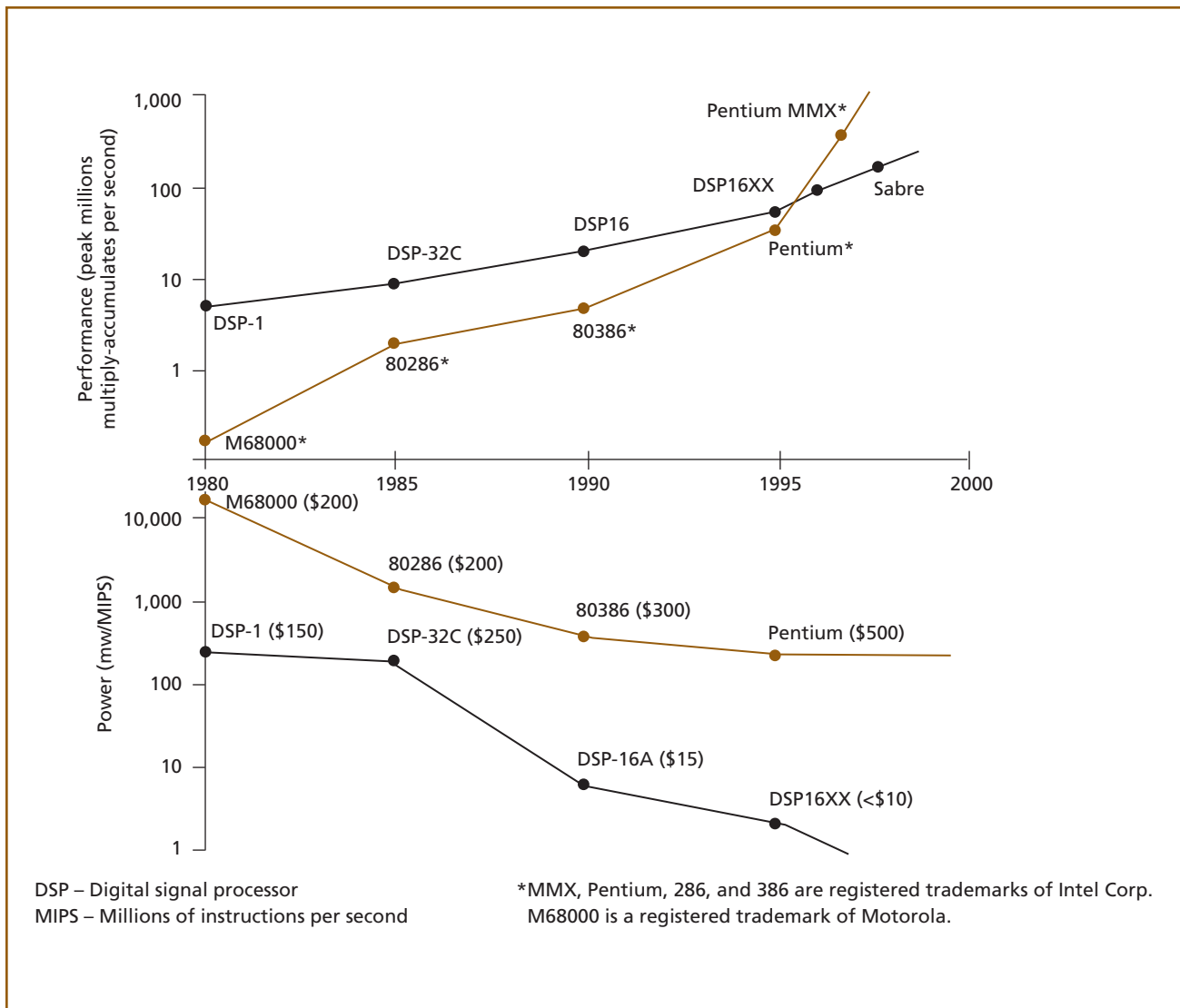


Figure 7. Cost and performance trends for Lucent DSPs used in communications terminals and base stations.

Magnetic Disk Storage

Magnetic disks remain the bulk storage medium of choice for PCs. The tremendous increase in magnetic disk storage density over the years has been attributed predominantly to technical improvements in the physical parts of recording systems, such as the storage medium, the read and write heads, and mechanisms governing flying heights and positioning of the heads. IBM,* which established a leadership position in the recording industry with its invention of the first digital recording system in the 1950s, is credited for most of these improvements.

The newer roles played by telecommunications and coding theory in adapting signal processing algorithms for recording systems, as well as the role of solid-state electronics in efficient implementation of these algorithms, are increasingly being acknowledged as essential for achieving enhanced storage density. Consequently, the prominence of Lucent, one of the leaders in signal processing, is growing in the magnetic recording industry.

New magnetic recording systems that are based on partial response signaling and maximum likelihood sequence estimation are referred to as PRML systems.

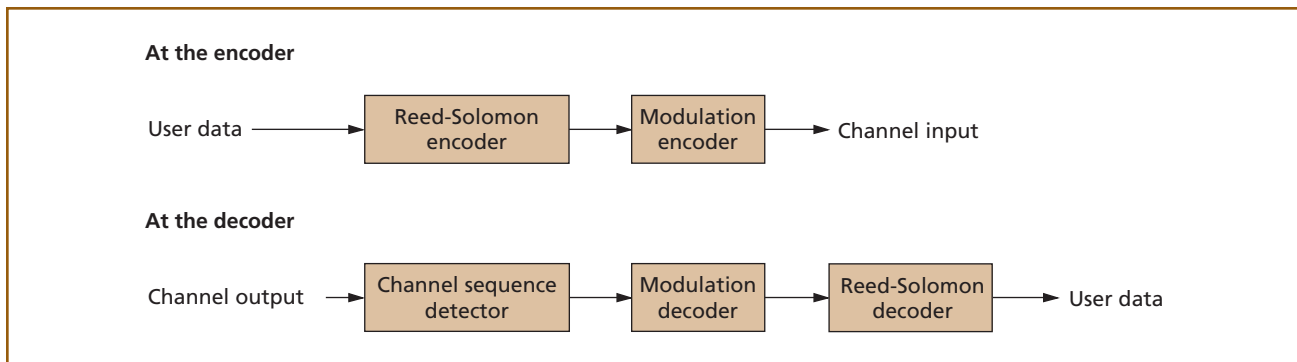


Figure 8.
Two-step channel coding in recording systems.

The first commercially available PRML systems, such as the Shiva* chip manufactured by the Lucent Microelectronics Group in 1994, used equalization to the so-called PR4 channel and maximum likelihood sequence detection matched to this channel.

Performance of magnetic recording systems is in various ways improved by channel codes. As **Figure 8** shows, channel coding in these systems is usually realized in two almost independent successive steps: error-correcting coding (ECC) and modulation coding. At the encoder side, user data is first encoded by ECC, usually by means of a Reed-Solomon (RS) code. Then, the output of the ECC encoder is encoded using a modulation encoder to limit, for example, the run-lengths of 1's. At the decoder side, the steps are reversed—first, the output of the channel sequence detector is decoded by a modulation decoder, which passes its estimate to the RS decoder. These RS codes, used in almost all contemporary storage products, are high-rate codes capable of correcting random errors affecting recorded bits, as well as burst errors affecting strings of successive recorded bits.

During the past few years, significant progress has been made in the design of high-rate distance enhancing codes for high-density magnetic recording channels. Magnetic recording channels are expected to become more complex as recording densities increase. More complex data recovery schemes will be required to handle these challenges. Their implementation may be inconceivable today, but they will become a cost-effective reality with anticipated advances in VLSI technology.

Optical Storage

An optical disk is a rotating storage device that is similar in many ways to a magnetic disk, although it stores data by physical effects involving laser light rather than magnetic fields. The essential idea is to store bits on an optical medium by forming dark spots called *pits* in a reflective surface. A pit is read by shining a pinpoint laser beam onto the disk and measuring the intensity of the reflected light.

Optical disk technologies can be classified by the number of times each portion of the surface can be written. Read-only disks, such as audio CDs and computer CD-ROMs, are manufactured with data already recorded on them. Write-once disks, also known as write once read many (WORM) disks, are left blank when manufactured. Each sector of a WORM disk can be written to by burning pits into the disk surface with a laser. Read/write optical disks, also known as rewritable optical disks, record data via reversible effects, such as a phase change between amorphous and crystalline states in an optical recording medium.

Magneto-optical storage is a hybrid read/write technology that uses laser light to enable a large magnetic head, far from the disk surface, to write small magnetic domains as bits. The bits are read by detecting the rotation of a polarized laser beam that occurs when the beam is reflected off the magnetized disk surface.

The compact disk (CD) is becoming a ubiquitous optical storage medium. A CD-ROM is the read-only version, a CD-R (recordable) is the write-once version, while CD-E (erasable) and CD-RW (rewritable) disks are read-write versions. CDs hold about 650 MB; they are robust, long lived, inexpensive, and well standard-

ized. The basic transfer rate of a CD is 150 KB/s, but in 1997, multispeed versions are common. Multispeed disks are driven at 8 to 24 times the basic rate.

Digital versatile disk (DVD) is a new optical disk family. The DVD-ROM, the read-only version, stores 4.7 GB on a one-sided 120-mm disk and has a basic transfer rate of 1.38 MB/s. DVD-R is the write-once version, holding 3.95 GB per side. The read/write DVD-RAM version currently holds 2.6 GB per side. A rapid evolution to higher densities for future versions is expected. In the optical storage market, a crossover is expected from CD-R to CD-RW in 1998 and from CD-RW to DVD-RAM in 1999.

Past predictions that optical disk technology will replace tape storage and magnetic disks have not come true. The huge PC market has stimulated rapid improvements to magnetic disks, while optical technology has advanced more slowly. A typical optical disk drive has a factor of 3 worse positioning time than a magnetic disk drive, a transfer rate 4 to 10 times worse, and a cost per gigabyte 5 times worse. However, the removable platter sells for approximately one-tenth the cost of a magnetic disk drive of equivalent capacity.

Future improvements in optical storage technology, driven in large part by solid-state technology, need not be incremental. In particular, researchers foresee the following areas of discontinuous improvements in the density and speed of optical storage:

- *Stacking of a few transparent or semitransparent optical disk layers*, resulting in a three-dimensional sandwich accessed by focusing a laser beam onto the desired storage layer.
- *Adoption of shorter wavelength lasers in blue or blue-green*, shrinking the size of the optical pits.
- *Near-field optical recording*, flying an optical head very close to the storage medium to enable the writing and reading of optical pits smaller than the wavelength of the laser light.
- *Holographic storage*, whose underlying idea is to accumulate binary data as the pixels of a monochrome image. As reported in the literature, it is possible to record many holograms in a small region of an appropriate storage medium.⁵ This technique promises to have

both a high data rate (because an entire image is transferred in a flash of laser light) and a high data density (because of the small spacing required between images stored on a holographic medium).

Just as Lucent is making substantial contributions to magnetic storage technology—particularly in microelectronic innovations and data coding—it is also advancing the techniques and materials used in optical storage. Two notable areas of this work are near-field optical recording and holographic storage.

In near-field optical recording, Bell Labs researchers have attached a laser diode to a magnetic disk slider to obtain a flying height comparable to that of a magnetic disk. Using a phase-change storage medium, they have demonstrated a recording density several times higher than DVD.

In holographic storage, Bell Labs has demonstrated 100 times the storage density of current magnetic disks, as well as media transfer rates far in excess of magnetic storage. Their advances and those of their collaborators in polymer storage media and related technologies promise to support data rates and storage capacities substantially greater than DVD.

Software Evolution

Advances in IC and computer system technology shifted the computer industry from primarily large mainframe computers to minicomputers and, finally, to high-performance PCs. As the computing industry evolved, the bulk of computer use shifted from program development to executing applications.

The availability of inexpensive, high-capacity, solid-state-driven microprocessor and storage technologies not only created the PC, it also provided a tremendous stimulus to development of desktop software applications that applied these new capabilities in highly innovative ways.

The software business landscape has changed dramatically in the past few years, creating companies that specialize in creating software directed specifically at desktop applications. It is well to remember that the success of Microsoft* operating systems and software is a phenomenon of the past decade or so, made possible by the rise and proliferation of PCs. Other major play-

ers have created huge businesses through software designed for desktop users. Users who were not even computer literate a decade ago now associate companies with their products, such as Oracle for database applications and Lotus for spreadsheets and groupware, in much the same ways as earlier generations associated Jello with desserts. The emergence of the World Wide Web as an alternate form of distributed computation also has stimulated development of highly sophisticated browsers, spurring the fortunes of newer companies such as Netscape and entirely new products such as search engines created by companies like Yahoo. All of this progress runs on the fuel of microelectronics.

The software business can be segmented in many ways—for example, software for programmers versus application software for a broad audience. Software for programmers and large servers continues to be a turbulent market, made up of variants of the UNIX* operating system, Microsoft Windows NT,* and other similar products. Elegance of expression, innovative design, and program development environments attract the best programmers. Ease of use and portability/compatibility attract application users.

Although Xerox and Apple* were first, IBM is often credited with launching commercially viable and trusted personal computing. Both Intel* and Microsoft benefited from IBM's entry into personal computers through their selection as microprocessor and operating system (OS) suppliers. Both companies have become leaders in their respective hardware and software markets.

Word processors and financial spreadsheets are two widely used software applications. Microsoft was not the first to market with either product, but the company was able to create and then refine a series of word processing and spreadsheet programs. In doing so, Microsoft redefined the market and gained market share. Shortly thereafter, the programs were combined with presentation, scheduling, and other popular applications and marketed as a so-called "office suite" with a common look and feel. Microsoft's office applications have been available on Apple systems. However, the suite of applications combined with Windows 95* and Windows NT operating systems provides a workable

office environment that many corporations mandate as a standard. Recently, Microsoft has made strong efforts to assert its products in Web computing, but it is far from alone. In addition, Java* from Sun Microsystems and Inferno™ from Lucent Technologies will play key roles in this developing World Wide Web environment.

While technical and marketing excellence have underpinned Microsoft's growth, making it a major force in today's software industry (as **Table II** shows), it is important to remember that the solid-state revolution created the PC milieu that Microsoft and other relatively new companies have helped to define; the ever-improving performance of microelectronic devices has provided powerful support for new, more-functional suites of PC software.

Another major area of impact for microelectronics-based computing is workstations, powerful (often networked) desktop computers made by newer companies like Sun Microsystems* using sophisticated software for higher-level applications, such as computer-aided design and automation. These workstations, running application software from companies such as Pixar,* are now standard equipment in motion-picture studios, where they provide animation and graphics capabilities undreamed of less than a decade ago. Much of today's entertainment industry—films, recorded music, live performances—runs on microchips in terminals, control consoles, and electronic instruments, and recording equipment unavailable a dozen years ago. The CD-ROM, the first widespread application of photonics in consumer technology, is now the standard for music playback, having replaced vinyl discs in less than five years.

Evolution of Switching Systems

While solid-state technology created the modern computing industry, it also revolutionized the telecommunications business—the industry that stimulated the development of the transistor. It is no accident that the first solid-state computers were being designed at the same time that the first electronic switching systems were being developed.

Switching can be considered the interconnection of access inputs via centralized nodes and the further interconnection of these nodes into a network. The

Table II. Market capitalization comparison among the leading companies.

The New Giants			
Company/Totals	Sales 1996 (\$ billions)	Employees 1996 (thousands)	Market Capitalization 1Q97 (\$ billions)
Intel	20.8	49	116
Microsoft	8.7	21	120
Cisco	4.1	10	34
Totals	33.6	80	270
The Old Giants			
General Motors	164.1	647	43
Ford	147.0	345	38
Chrysler	61.4	114	22
Totals	372.5	1,106	103

nodes and network are implemented by means of the appropriate control, signaling, and connectivity—functions that are, not surprisingly, like those found in large computers. Initially, switching was a mechanism for providing interconnection for a multiplicity of voice telephone lines and was implemented by means of electromechanical devices. Although its application to voice is still paramount, switching has been extended to include the interconnectivity of data, image, and video. Today, switching systems use the most advanced semiconductor, microprocessor, and fiber-optic technology, including:

- Processors and microprocessors for control;
- Semiconductor memories and logic circuits for switching;
- DSPs for tone generation, touch-tone dialing, voice announcement systems, echo cancellation, and subscriber line functions;
- Light emitting diodes for fiber-optic interfaces and displays; and
- Semiconductors for subscriber-line and interoffice trunk interfaces, as well as power supplies.

Electromechanical Systems

In the early days of switching, interconnectivity for telephone subscribers was provided by telephone operators who manually inserted patch cords into centrally located switchboards. Early automatic switching equipment made use of Strowger step-by-step technology,⁶ followed in later decades by step-by-step, panel, and crossbar technology.⁷

Automatic switching required not only the interconnecting networks but also a control structure with its need for signaling and audible tones. The interconnecting networks evolved through electro-mechanical devices, such as rotary switches, relay networks, and cross-connect networks. Control of a telephone call evolved from progressive control, where it was passed from one electromechanical stage to another (step-by-step) to centralized common control by relay networks (crossbar). In the early 1950s, a card translator to obtain nationwide routing information was developed for the No. 4 crossbar switch using phototransistors and transistor amplifiers.⁸

Semiconductor Penetration

Less than ten years after the discovery of the transistor, Bell Labs researchers began work on experimental switching systems using transistor and diode control.⁹ Subsequently, in the early 1960s, Bell Labs made a major breakthrough in the use of transistors for commercial switching systems with the introduction of the 101ESS PBX and the 1ESS central office (CO) switch.¹⁰ Both these systems incorporated essentially the same number of semiconductors and made practical the use of a stored program processor.

A special-purpose computer for switching applications, called the 1ESS central processor, was designed using discrete transistors and semiconductor diodes. Its memory system consisted of a nondestructive readout (NDRO) permanent magnet twistor memory for storing the program and a ferrite sheet memory for tem-

porary call-related features. Both types of memory used transistor drivers and amplifiers for accessing and reading out the information.

The space division (analog) switching network for interconnection consisted of an array of magnetically latched devices called *ferreeds*, which were driven by a pulsed semiconductor device. Semiconductors were also used in power supplies and tone generators. Line and trunk circuits did not use any transistors. The 1ESS was highly successful and widely deployed, and it still is in service in some areas. However, the 1ESS now is generally being replaced by time division (digital) switches.

Digital Switching

Although there were other switches, such as the 2ESS and 3ESS, as well as stored program control for such specialized applications as suburban, rural, and operator services, the next major breakthrough was the 4ESS switch,¹¹ the first in the world to implement digital switching.

Because the 4ESS is a toll switch, it provides connectivity *between* local COs, as well as *between* other toll switches. It does not terminate subscriber lines. The 4ESS switch proved ideal for the implementation of digital switching technology because of the high volume of digitally multiplexed trunks that connect local and toll switches. By digitizing the voice and assigning a fixed synchronized time slot to each conversation, multiple connections can be supported cost effectively on a given transmission medium, such as wire pairs, radio channels, or fiber optics.

A DS1 facility providing twenty-four 64-kb/s channels was the initial standard for digital transmission connectivity. The 4ESS switch used semiconductor circuitry to interface to these DS1 facilities and implemented both the time-slot interchangers (TSIs) and time multiplexed switches (TMSs) using semiconductor memory and logic circuits. TSIs were interconnected via the TMS. Analog trunks were also terminated, but semiconductor A/D and D/A converters were used to multiplex and demultiplex the signals into and out of the TSI.

The 4ESS switch was conceived at a time when solid-state technology was maturing rapidly, and it was designed to receive regular infusions of improved microelectronics and software technologies. Now in its

twenty-first year of use, the 4ESS switch is still the backbone of AT&T's long distance network, where it has evolved to handle a doubling in calling volume over the past seven or eight years, as well as a large number of new customer services.

Digital Switching For Subscriber Lines

The next big step in switching systems was the implementation of digital switching for local COs. Digital switches interfaced directly to subscriber lines and were marketed by such companies as Northern Telecom, AT&T, Siemens, and Alcatel beginning in the early 1980s. In contrast to the digital trunks that interface with toll switches, local COs primarily dealt with analog line inputs from telephone subscribers and thus had to provide extensive A/D and D/A conversions for interfacing with the digital switching fabric. Typical architectures were those of Northern Telecom and AT&T, which used a time-space-time (T-S-T) TSI to TMS to TSI architecture that interfaced to subscriber-line and trunk circuits for interoffice connections.

The Lucent 5ESS[®] switch¹² incorporated an innovative distributed architecture based on microprocessor technology that has withstood the test of time and has allowed it to combine new technologies gracefully and to provide multiple applications on the same switch. Instead of depending on a large central processor, the 5ESS switch used a cost-effective switching module that aggregated a microprocessor-based call processing computer, a TSI based on VLSI memory, DSPs for tone generation and service circuits, and peripheral units consisting of subscriber-line and interoffice trunk circuits.

The switch modules were interconnected into a switching system by means of fiber-optic connections to a TMS, which incorporated control channels to a message switch. Overall administrative and maintenance functionality was provided by a special-purpose computer called the *administrative module* (see **Figure 9**). Switch modules and groupings of subscriber lines could be remoted either over fiber-optic links or digital T1 facilities.

The 5ESS switch designers periodically have upgraded the call processing computer, TSI, DSPs, line and trunk circuits, TMS, message switch, and administrative module to keep pace with the ever-improving

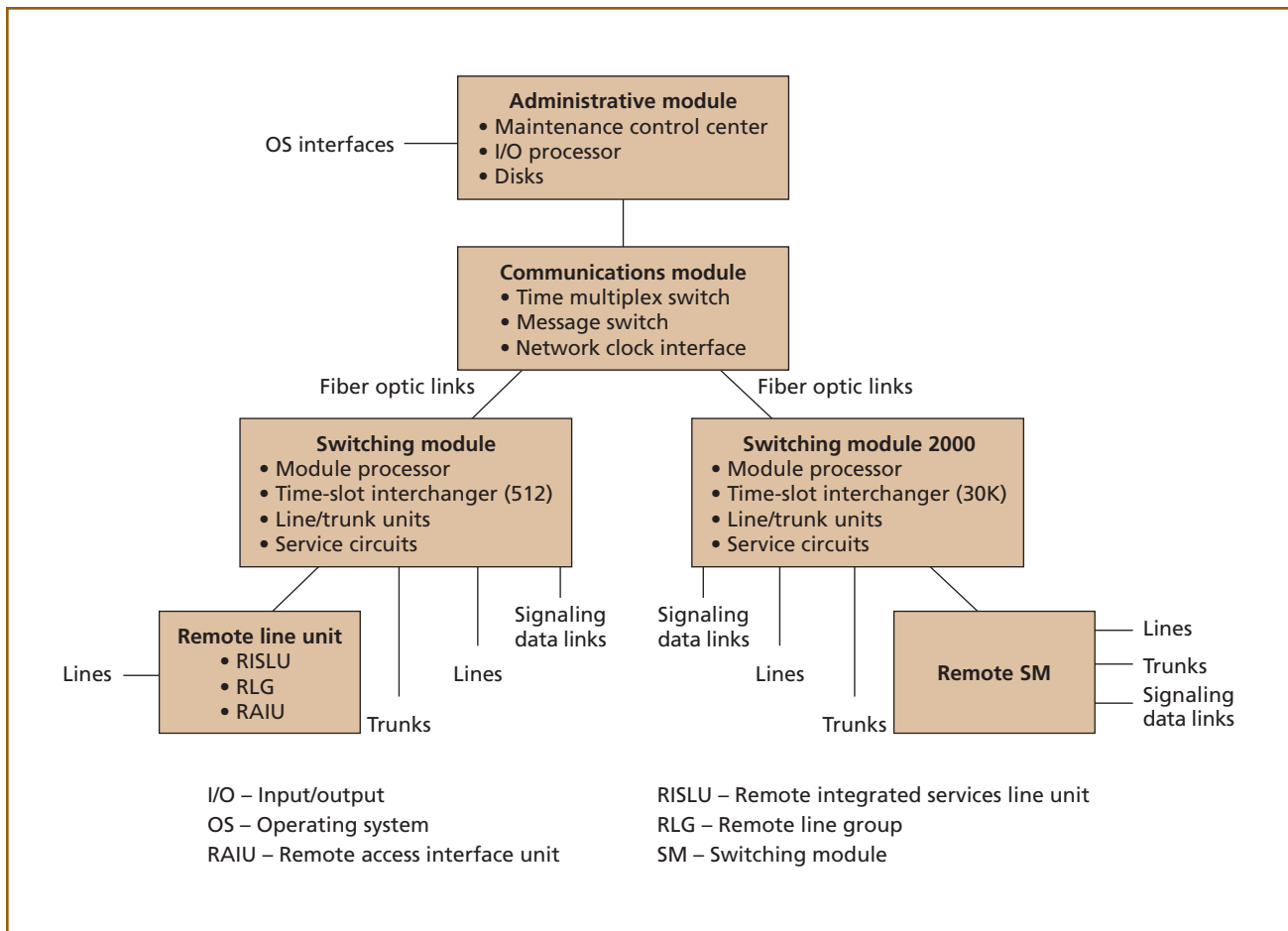


Figure 9. Overview of the hardware used in the Lucent 5ESS®-2000 digital switch.

semiconductor and microprocessor technologies. In 1982, the 5ESS switch used a handful of application-specific integrated circuits (ASICs), which had a 1K to 2K logic gate density and operated at rates of from 2 to 4 MHz/s. Today, hundreds of ASICs are deployed throughout the switch, and new designs are being implemented that have up to 300K logic gates and operate at speeds of up to 150 MHz. Field programmable gate arrays (FPGAs) are widely used in specialized areas, such as in the module processor. In addition, the switch has incorporated new functionality, such as a packet switching unit (PSU).

Regular infusions of new semiconductor technology in the 5ESS switch support multiple applications on the same switch simultaneously (see Figure 10). This attribute makes the 5ESS switch unique and is largely due to its distributed architecture (see Figure 9). The applications include local and toll call switch-

ing, operator and directory services, ISDN, intelligent network, international gateway, as well as wireless TDMA, CDMA, and GSM switching. The PSU has played a key role in supporting ISDN and common channel signaling (CCS), as well as DSPs for wireless applications using compressed voice. The increasing capacity of the microprocessor-based call processing computers have allowed for the support of the millions of lines of software required to provide these multiple applications.

Switching Evolution Summary

Solid-state technology has driven the evolution of switching systems by providing extensive amounts of processing power and memory. The 1ESS processor, for example, used 55,000 transistors and 160,000 diodes for a 10,000 subscriber line office.¹³ Today the Motorola MC68060 microprocessor has 2.5 million

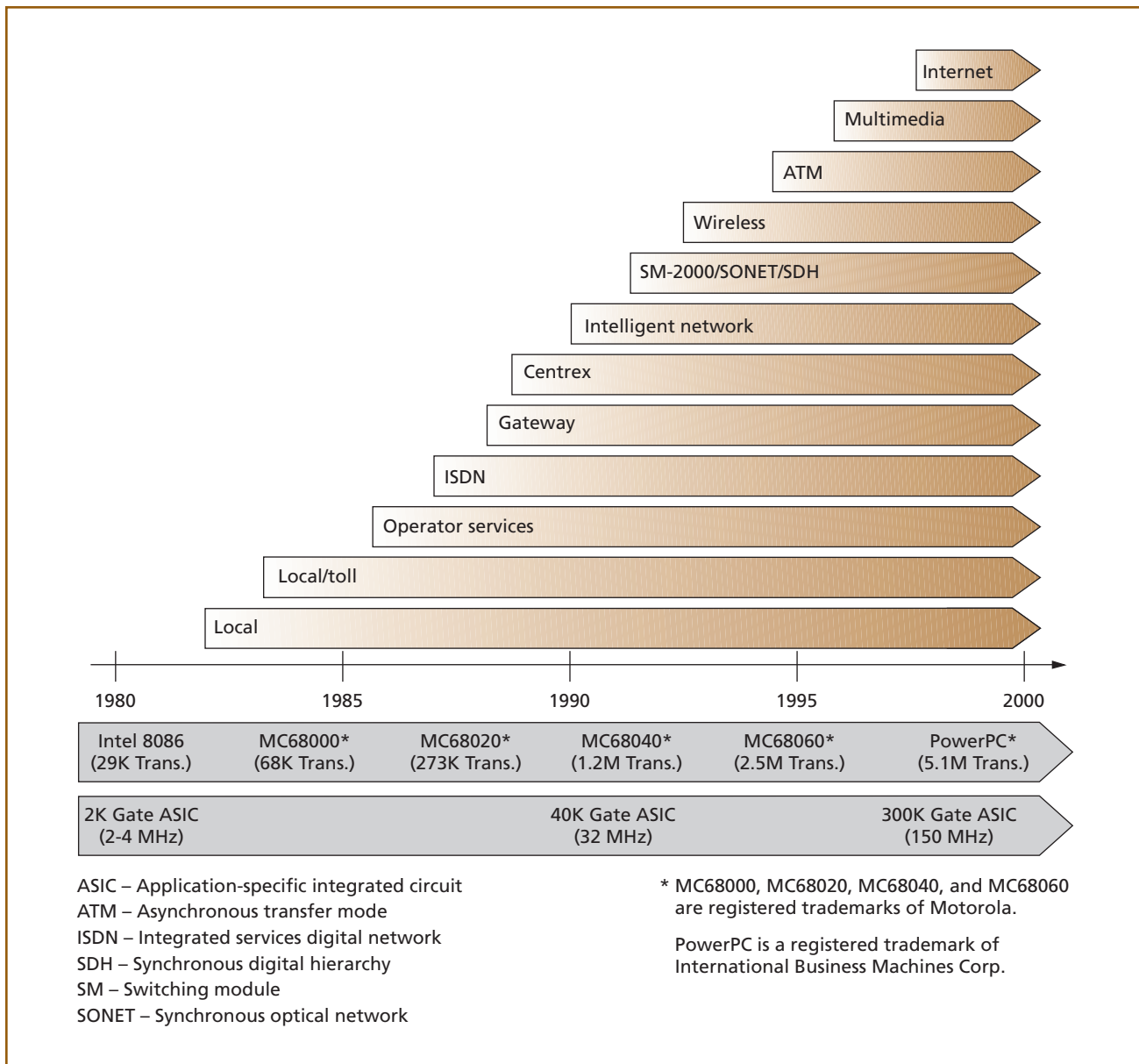


Figure 10. New simultaneous applications on Anymedia™ 5ESS®-2000 switch fueled by semiconductor technology.

transistors^{14,15} and a PowerPC* 604e RISC microprocessor has more than five million transistors.^{14,16} Many such microprocessors are used in implementing a modern switching system.

As a result of this extraordinary progress, electronic switches have created, through the networks that host them, a vast range of new services, such as toll-free calling, caller ID, custom calling services, and premium-billing services, that exist because of the computing power resident in switching systems and

associated databases. It is not an exaggeration to say that telecommunications networking has transformed and interlinked national economies into a global economy, with markets that are open and accessible 24 hours a day. The very concept of nationhood is being redefined by the subsuming of local and national time-tables and marketplaces by a vital, teeming global marketplace. Network-based intelligence has the ability to provide people with personal telephone numbers, enabling them and their communications appliances

to be reached directly anywhere in the world. As a result, many businesspeople have become true “citizens of the world.”

Solid-state technology also has greatly enhanced communications functionality at the customer premises level. For example, private branch exchanges (PBXs) have evolved from electromechanical systems to modern stored program electronic systems in lock step with the advance of microelectronic devices. The Dimension PBX, which debuted in 1975, used newly developed microprocessors and solid-state memories to implement customer switching features previously only available through large, CO-based Centrex systems. Less than a decade later, the System 75 and System 85 PBX systems used advanced microprocessor architectures to implement digital switching technology on customers’ premises, with rich feature sets serving several dozen to thousands of office stations. Today, such PBX’s as the DEFINITY® provide custom user-programmable features and networking capabilities that help companies create global information networks.

The impact of solid-state electronics reaches into equipment for smaller offices and homes as well. Sophisticated microelectronics-based key telephone systems now bring some of the features and functionalities of PBXs to small businesses. DSP-based intelligence has created ever more capable telephone sets, often combining two line appearances with speakerphone, caller ID, repertory dialing, and fully addressable memories reached via built-in keyboards. DSPs with nonvolatile flash memory also provide the intelligence in digital answering machines, which store messages on chips instead of bulky, unreliable tapes. Many of these products are providing vital support to the growing small office/home office market, which now includes nearly 20 million Americans and is growing at double-digit rates.

The semiconductor technology evolution also has fueled the large growth of data switching capability, supported by LANs and computer centers in the business market. The increasing popularity of Internet access and functionality is responsible for the widespread deployment by Internet service providers of routers made by companies such as Cisco Systems, Bay Networks, Cabletron, and 3COM. Very strong technol-

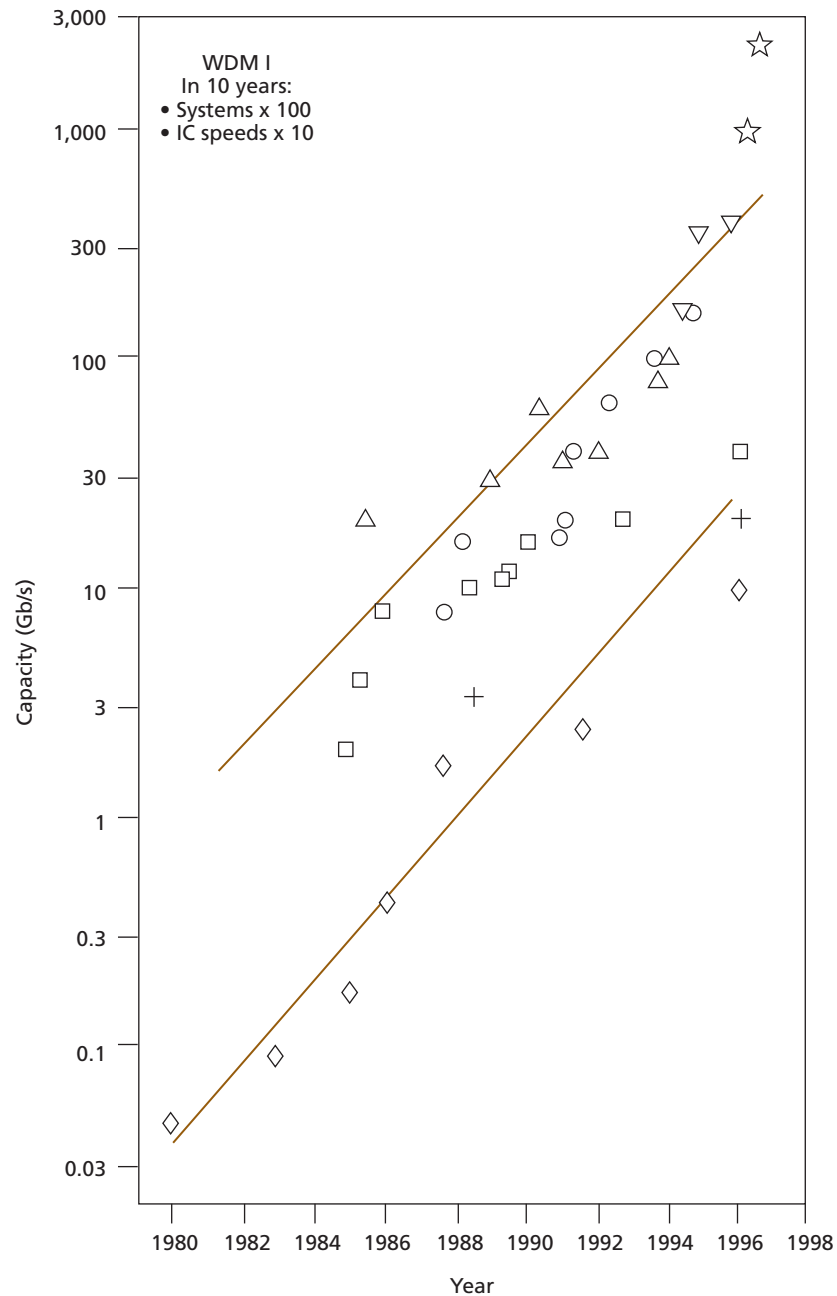
ogy and competitive pressures are pushing to cost-effectively combine voice, data, and video into universal switch routers. Today, asynchronous transfer mode (ATM) is finding widespread acceptance as a method of transporting voice and data on backbone networks. Researchers are evaluating different approaches for providing quality-of-service enhancements to TCP/IP to enable voice communications on data networks. Telephone switch vendors have the ability to integrate ATM and TCP/IP into their switches, and they are exploring voice interfacing on the Internet. Concurrently, router vendors are introducing methods to provide voice communications on data networks.

The rise of worldwide data networking enabled by solid-state electronics has supported the increasing globalization of commerce. Toll-free 800 calling, which depends on network-based computers to store and act on database information and routing tables, has created a \$110 billion industry that enables companies to operate nationwide and even global businesses from a single electronic “storefront.” The World Wide Web looms as the next frontier of electronic commerce. Today, Web-based sales amount to less than \$1 billion. But with the Internet growing by 5,000 users per hour and with infusions of new, highly reliable routers, servers, and secure-transactions software, some predictions reach as high as \$250 billion within five years. Eventually, Internet shopping will be as commonplace as toll-free telephone shopping is now.

It is in the humble beginnings of the transistor fifty years ago that one finds the wellspring of these new applications, which have such a profound effect on how we live today. Equally important, microelectronic and photonic technologies are setting the stage for broadband multimedia networking, which is just over the horizon.

Enabling Broadband

Barely a decade after the invention of the transistor came another major development that revolutionized the information industry—the invention of the laser. The potential impact of this invention on communications was immediately recognized, and additional key developments of silica optical fiber and the double heterostructure semiconductor laser soon fol-



- | Experimental | Commercial |
|--------------------------|-------------------------|
| □ Single channel (ETDM) | ◇ Single channel (ETDM) |
| △ Multi-channel (WDM) | + Multi-channel (WDM) |
| ○ Single channel (OTDM) | |
| ▽ WDM + OTDM | |
| ☆ WDM + polarization MUX | |
| ☆ WDM + OTDM | |
| ☆ WDM | |
- ETDM – Electrical time division multiplexing
IC – Integrated circuit
MUX – Multiplexer
OTDM – Optical time division multiplexing
WDM – Wavelength division multiplexing

Figure 11.
Progress in lightwave transmission capacity.

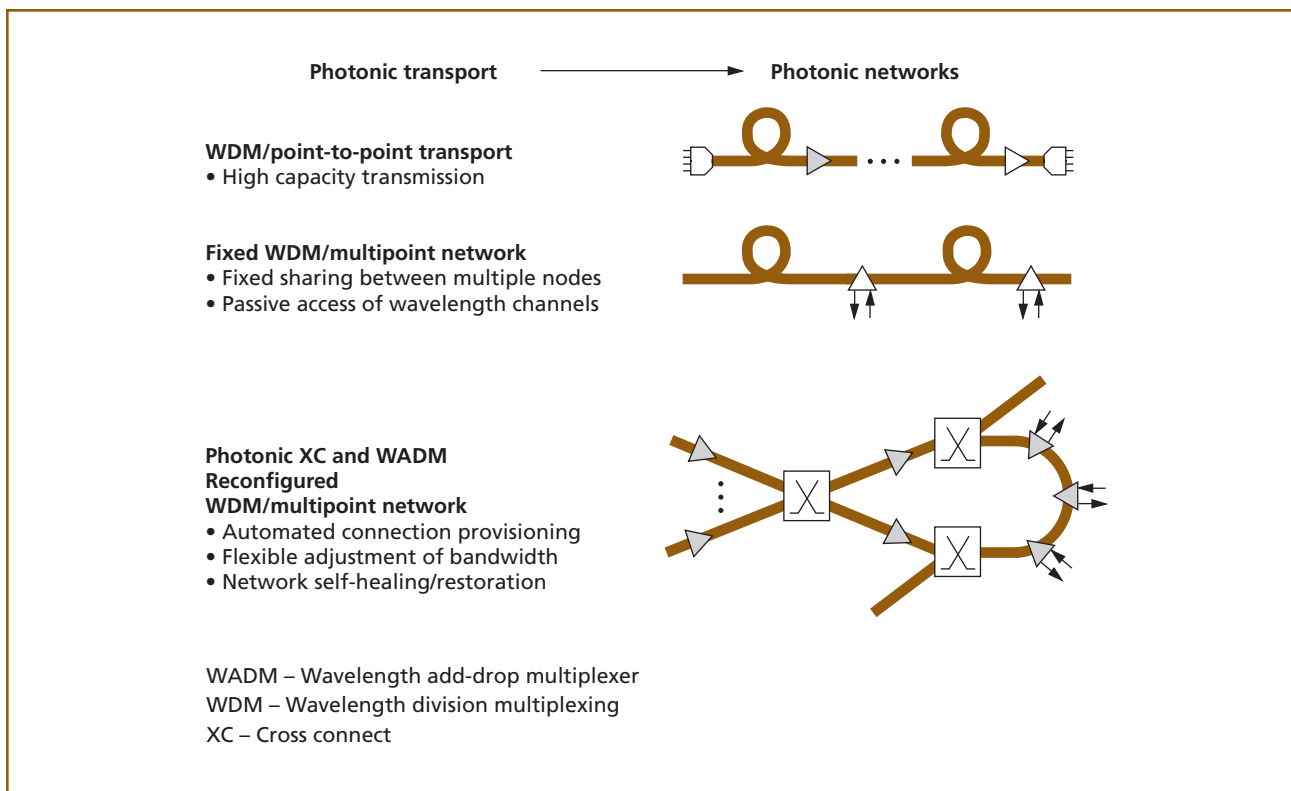


Figure 12.
Crucial changes in the evolution of optical networks.

lowed. These critical components benefited directly from many of the advances in physics, chemistry, and materials sciences and from processing technologies that had been developed for solid-state electronic devices. They form the basis of the optical communications systems as we know them today, and they have spawned a multibillion-dollar industry. Since the development of the first optical fiber communications system in the late 1970s, the capacity of lightwave systems has doubled each year, as **Figure 11** shows. This rate of increased capacity considerably exceeds the rate at which IC speed has increased. In 1996, experimental demonstrations of communications at terahertz (THz) rates were achieved on a single optical fiber, thereby heralding a new era of THz optical networking.

Such high data rates exceed the speed of commercial electronic circuits and are made possible by dense wavelength division multiplexing (DWDM), in which multiple wavelength channels—for example, 8, 16, or more wavelengths—separated by approximately 100 GHz are sent down the same optical fiber, indepen-

dently modulating and detecting the individual wavelengths. Users at one node can tune to the wavelengths of interest while the remaining wavelength channels continue to the next node.

Figure 12 shows the evolution of optical networks. The initial focus of DWDM systems was on increasing the capacity of installed fiber between a transmitter and receiver. By the mid 1990s, add/drop multiplexers were developed, allowing certain wavelengths to be dropped and added at intermediate nodes along the link. In 1997, the first optical cross-connect was demonstrated, raising the prospects for full terabit networking at the optical layer. Researchers predict that this evolution will greatly enhance optical networking.

Enabling this system’s evolution was the development of many novel optical and optoelectronic devices, which included:

- *High-capacity optical fiber (TrueWave® fiber)*. This medium is capable of carrying multiple wave-

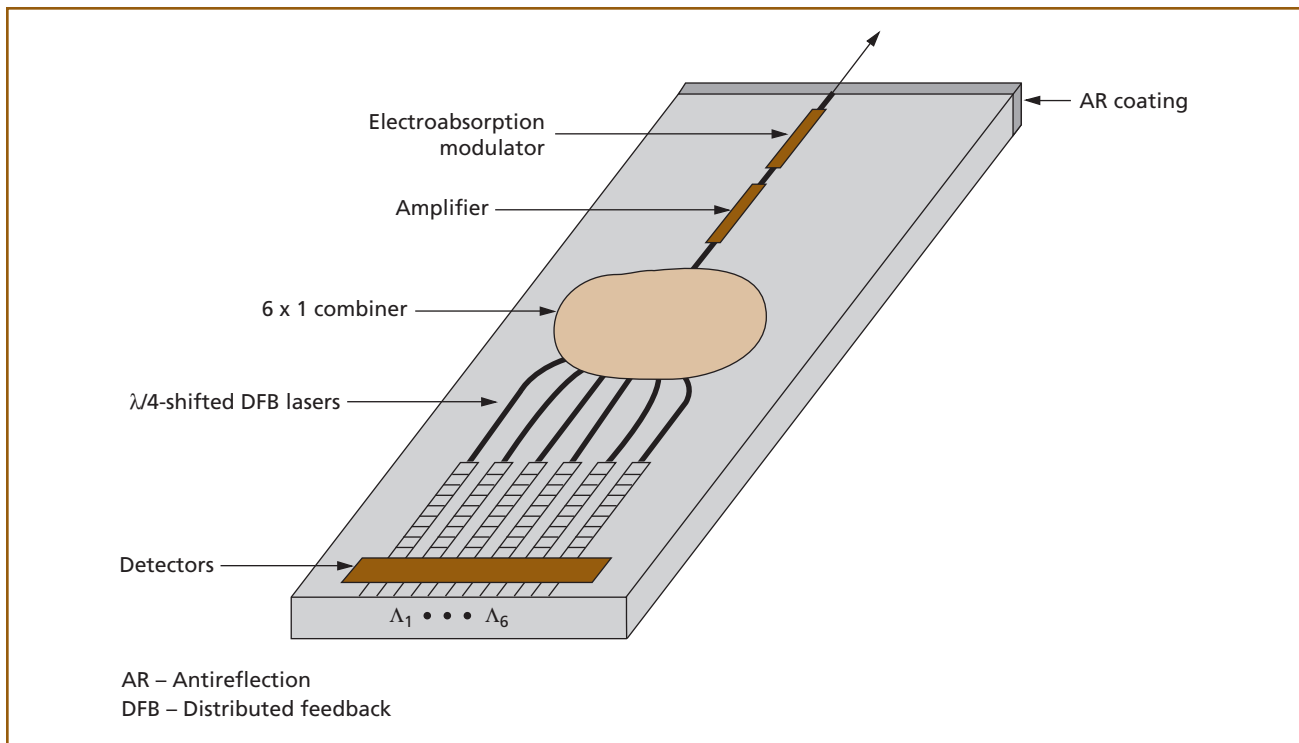


Figure 13. Next-generation laser featuring multiple wavelengths individually selectable by network management systems.

lengths over long distances without cross-channel interference.

- *Broadband amplifiers having the same gain for all optical wavelengths.* An 80 nm-wide amplifier capable of 100 wavelength channels was demonstrated recently.
- *Wavelength-selectable lasers.* Currently, commercially available lasers operate at a single (very precise) wavelength capable of modulation up to 10 GHz. Next-generation lasers now being developed will be capable of multiple wavelengths individually selectable by network management systems. **Figure 13** illustrates an example of such a laser.
- *Optical monitoring systems.* Currently being developed, these systems monitor all aspects of network operation at the optical layer and permit rapid network reconfiguration in the event of a fault.
- *Optical add/drop filters.* Fixed frequency add/drop filters based on fiber grating technology are now commercially available. The

next generation will be wavelength reconfigurable, thereby allowing maximum flexibility.

- *Optical cross-connects.* The first 32-channel optical cross-connect (8 wavelengths on 4 fibers) based on electro-optic lithium niobate technology was demonstrated in 1997.
- *Optical routers.* Devices that combine 16, 32, and more wavelength channels onto a single fiber (or, conversely, demultiplex wavelengths on a fiber) have been developed based on passive waveguide designs fabricated in silica-on-silicon by conventional chemical vapor deposition and photolithography.

The design and fabrication of all these enabling components were made possible by the long-term development (approximately thirty years) of solid-state optoelectronic materials meeting the stringent requirements of ultra-high purity, atomic layer dimension control, and precise optoelectronic properties. For example, the critical impurity levels in silica optical fibers are less than one part per billion. The complex structure of the electro-absorption

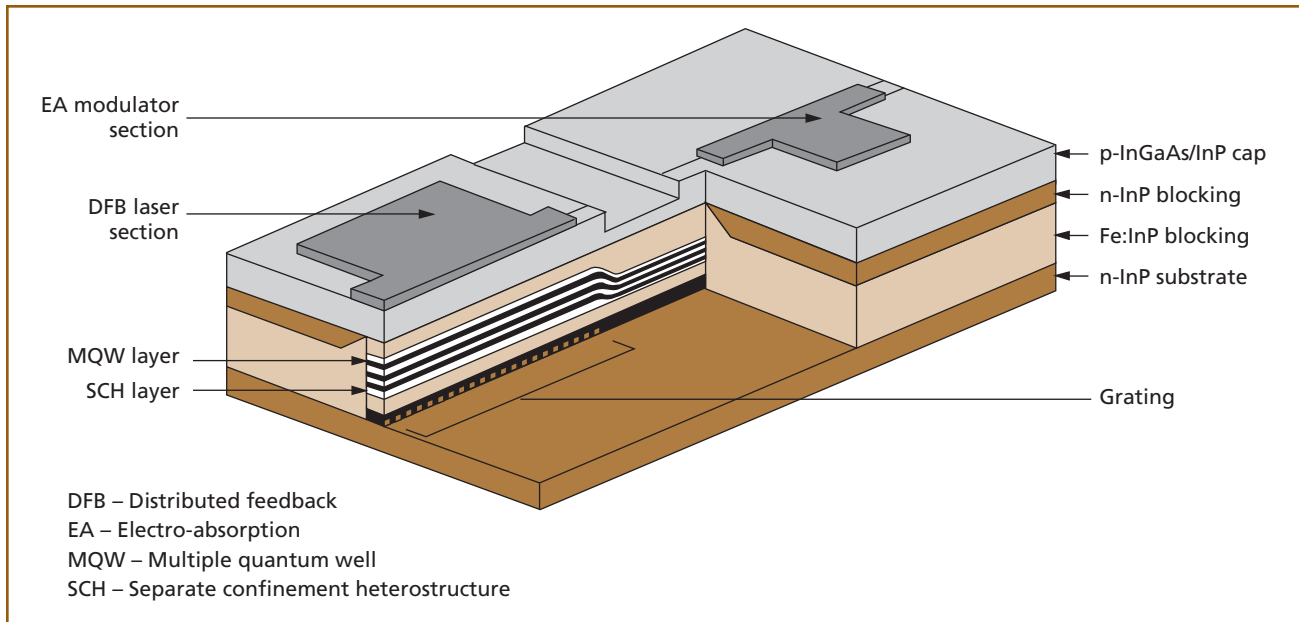


Figure 14.
Complex structure of the electro-absorption modulated InGaAsP laser.

modulated InGaAsP laser illustrated in **Figure 14** requires layer thickness control to within one atomic layer to achieve the necessary wavelength precision.

Solid-state electronics is primarily based on silicon technology. Because of its indirect band gap, silicon cannot be used for making light-emitting devices. Active devices are based on compound semiconductors, such as InGaAsP, that have direct bandgaps and efficiently convert electronic signals to photonic signals. Progress in photonics continues to be critically dependent on advances in materials processing, device design, testing, and packaging—the same factors that affect progress in microelectronic devices. These technologies are now becoming sufficiently mature to begin the integration of multiple photonic devices on a single chip.

The electro-absorption modulated laser (see Figure 14) is the first commercial photonic semiconductor device having two components: a laser and an external cavity modulator. In the near future, higher levels of integration will be realized in the wavelength-selectable laser (see Figure 13). At the same time, hybrid integration of electronic and photonic devices on silicon substrates will result in considerable cost reduction and increased functionality. This is an important factor in making

broadband multimedia communications a practical reality for millions of people.

Photonic technology already dominates long-haul communications, and commercial terabit-per-second photonic networks will begin to emerge around the turn of the century. The bandwidth requirements of today's homes and offices continue to increase as the vision of interactive multimedia systems becomes a reality and the same advantages offered by photonics in WANs become necessary in LANs and access networks. These emerging markets will put increasingly demanding requirements on low-cost devices. Optical fiber costs have already fallen below those of copper for new installations. Photonics integration and new manufacturing technology for associated devices soon will meet these demands for low cost. Thus, predictions are that photonics will become ubiquitous.

Photonic technology already has spread into areas that traditionally were the domain of electronics—for example, communications, storage, interconnection, and sensing. This proliferation and evolution will continue, and the photonic wall jack will perhaps become reality.

Figure 15 shows a prototype bidirectional transceiver for low-cost fiber-to-the-home applications in



Figure 15.
Prototype bidirectional transceiver for low-cost fiber-to-the-home applications integrating several functions on a single silicon wafer.

which several electro-optic and passive-optical functions are integrated on a single silicon wafer. As in electronics technology, the level of integration will continue to increase. However, the scale of integration will remain considerably below that of silicon technology because the wavelength of light limits minimum device dimensions to a few microns.

Enabling Mobility—A Historical View of the Mobility Trend

Solid-state electronics strongly supports society's need for ever-greater mobility in communications. This truly is a global phenomenon, with double-digit growth occurring in many regions simultaneously among developed and developing nations alike. In some countries, wireless provides a quick, economical alternative to building land-based telecommunications infrastructures. In others, wireless telephones, faxes, and pagers are augmenting and replacing wired appliances. The industries supporting mobile communications also are truly global, with Scandinavian manufacturers Nokia and Ericsson assuming leading positions in the manufacture of handsets while a variety of multinational companies including Lucent, Ericsson, Motorola, Siemens, and Northern Telecom are providing network and infrastructure equipment.

The astonishing growth of wireless communications (cellular telephony, paging, wireless data, cord-

less telephony) has been enabled by miniaturization and low power consumption and more effective uses of radio spectrum. **Figure 16** shows the size of the cellular phone as a function of time since the AMPS system trial in Chicago in the early 1980s. The same trend toward miniaturization also applies to the infrastructure equipment, such as base stations. For example, a compact base station today is about 2,500 cubic inches, about the same size as the user equipment in the first trial. (Size was considered such a limiting factor that the earliest estimates of cellular telephony use predicted fewer than a million customers in the first decade. In fact, almost 20 million customers signed on for cellular service in the first ten years—many of them because they could carry telephones in their pockets and purses. Growth continues in double digits today.)

The major driving force for size reduction is the increased level of integration. The impact of electronic functions on a portable system falls into several parts (see **Table III**):

- User interface,
- Baseband digital signal processing,
- Baseband analog signal processing,
- Intermediate frequency (IF) signal processing, and
- Radio frequency (RF) signal processing.

As in other communications and computing equipment, use of discrete transistors is giving way to integration on a single chip (IC). In addition, the technology of choice is moving away from the compound semiconductor and toward lower-cost silicon-based solutions. Thanks to the ever increasing level of integration, the overall size of today's wireless systems is decreasing while functionality is increasing.

Portability requires long battery life. This principle is related to another important enabler of mobility: *low-power electronics*. **Figure 17** shows the battery operation times of DSPs as a function of the years during which the various CMOS technologies were developed. The energy consumption of a transistor scales linearly with its capacitance and wiring while changing quadratically with its operating voltage. As transistor feature size is reduced, the total capacitance needed for a particular functional operation decreases.

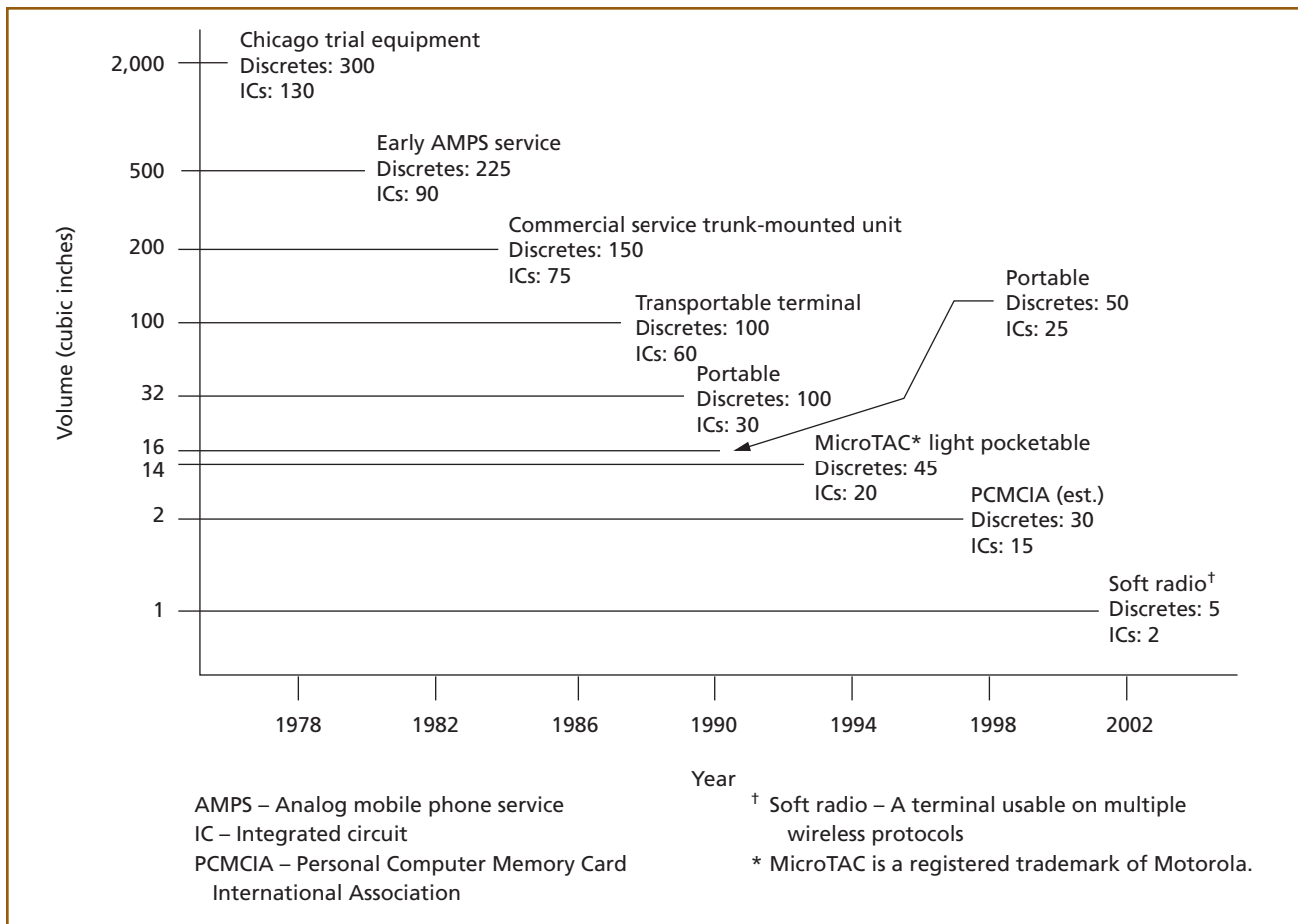


Figure 16. Semiconductor device use in cellular telephone terminal products.

Table III. Trends in increasing levels of integration in wireless systems.

Functions	Discrete	Modular	Subsystem	System
User interface	Simple logic gates	Microcontroller	Digital processor	Mixed-signal processor
Baseband digital	Simple logic gates	Custom circuits	Digital processor	Mixed-signal processor
Baseband analog	Discrete transistors	Analog IC	Mixed-signal ASIC	Mixed-signal processor
Intermediate frequency (IF)	Discrete transistors	GaAs IC	Silicon analog IC	Silicon RF/analog IC
Radio frequency (RF)	Discrete transistors	Discrete transistors	GaAs RF IC	Silicon RF/analog IC

As a device's dimensions are reduced, the operating voltage required for achieving a given speed also decreases. Both these factors contribute to power and energy reduction. In addition, innovative DSP designs and algorithms are being implemented to achieve additional power savings. Major power savings could also be obtained by choosing the right form of algorithms for the same function.

Future Challenges

The following major challenges must be addressed to ensure continuing explosive growth of wireless systems:

- *More MIPS per mW.* This challenge will remain as one of the key factors for the continuous increase in portability. Techniques ranging from processing technology to circuit and algo-

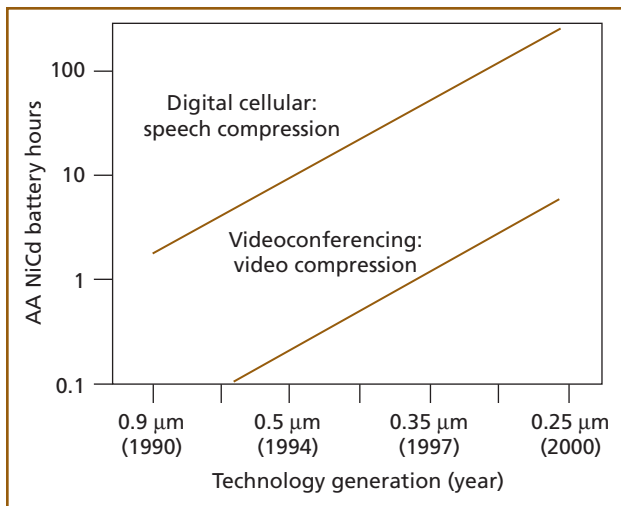


Figure 17.
Battery operation time for DSPs in wireless applications.

rithm design have been pursued. The challenge is to integrate and optimize them simultaneously. For example, different circuit design techniques require different processing technologies for ultimate optimization. Memory circuits might need a higher threshold voltage to turn off leakage currents to reduce standby power. Static circuits in the critical path might benefit from a zero-threshold voltage for faster operation if the power supply could be switched off when the circuits are idle to reduce standby currents.

- *Silicon superchip.* The next step in functional integration is to implement the RF function in monolithic silicon instead of in today's commercially available stand-alone components. For further power consumption reduction, engineers do not want to convert the signals to interface with a board-specific environment and then convert them back to interface with functions located on other components. Instead, a superchip is needed that integrates all circuit functions, thereby reducing power and size. The main challenge for integration is signal integrity management—understanding how different signals interact with each other, how to use different design techniques to reduce or control unwanted signal interac-

tions, and how to choose different architectures to optimize end solutions.

- *Packaging and passives.* The availability of high-quality passive components, either on-chip or in a form factor that is compatible with the final size constraint, is part of the superchip challenge. Equally important is the testing and yield control associated with the final packaging process. These issues are addressed using the methodology of circuit-packaging co-design.
- *Higher RF frequency.* Radio spectrum is a precious and limited commodity. As more people demand mobility-based services, higher frequencies will be opened for wireless services uses as the lower ones become saturated. Developers will need to create efficient and cost-effective integrated solutions that meet both size and power consumption requirements as they have at the lower frequencies. The answer might come from faster processing technology, better packaging techniques, and more clever ways of integrating these solutions.

Computer to Network Communications

The marriage of computing and communications brokered by solid-state electronics resulted in a fast-growing offspring: computer-to-network communications. This was formerly the domain of large mainframe-based computers and private-line networks. However, the number of PCs that are networked has grown rapidly over the past ten years.

Networking interfaces connect either inside the same premises (LANs) or through communications channels outside the premises (WANs). The evolution of desktop and networking interfaces has been quite different since the advent of the PC in the late 1970s. Networks and their associated peripheral interfaces have been on a performance treadmill since the days of PC networking in the early 1980s. Not only have higher-speed electronics, such as faster LAN transceivers and media access controllers (MACs), been an ever challenging objective, but higher performance-enabling and cost-effective media have also been important. Ethernet and its higher-bandwidth follow-on technologies (10 Base-T and

100 Base-T) continue to dominate the LAN and premises networks as standards. At the desktop, the networked PC interface is usually a plug-in board employing 10/100 Base-T Ethernet technology.

Progress in solid-state electronics has reduced the price of an individual 10-Mb/s card to as low as \$20. These LAN cards are attached to unshielded twisted-pair wire, enabling either 10- or 100-Mb/s speeds (over 100 m of copper). We will soon see 1.0-Gb/s speeds over twisted pairs (with the availability of new Gb/s transceivers and MACs). There is really no end in sight for ever-increasing speeds in networking interfaces. As consumers and businesses realize the benefits of retrieving and storing information outside the PC, the value of higher-speed connections will increase dramatically. Furthermore, information content that is increasingly video intensive will require higher-speed network connections.

Another fast-growing area of networking interface is the connection of a PC to a public network or WAN. Data rate price/performance for WAN interfaces have improved quite rapidly due to the ever-advancing state of the art in solid-state analog modems and the adoption of standards to encourage competition. Performance has increased for PC modems from the 1.2-Kb/s range in the early 1980s to the new 56-Kb/s modems available today. However, it remains to be seen how WAN network interfaces will progress further because modems can no longer advance in data rate without changing the analog local loop, or “last mile,” of most local telephone networks. Integrated services digital network (ISDN) technology, which has been available for ten years, is only now picking up momentum as a solution to the last-mile problem.

Other digital loop (or digital subscriber line [DSL]) technologies are emerging, but all of them require changes in the public switched network. If the past is any indicator of the future, the evolution of modem replacement technologies (ISDN and DSL, for example) will depend on both the level of ubiquity these interfaces achieve (standards) and the level of competition that develops. The Internet and all its derivatives, such as video streaming coupled with on-line video databases, are already providing the market pull.

The technology is ready, the customer need is apparent, but will the market environment be supportive?

People Interface and the Role of Computers in the Home

The major challenges to broader deployment and use of network computers and the future of PCs are to manage complexity and to provide greater functionality and ease of use. The computer has been moving steadily from a product that was aimed at helping us work smarter to a product category viewed as part of the family room entertainment system. Progress in solid-state electronics has enabled the addition of capabilities like digital television, Dolby* audio and, of course, Internet access. The inevitability is that the computer will move from its current main use as a productivity enhancement tool to firm entrenchment in our living rooms.

The computers of tomorrow will be friendlier and easier to use than today’s PCs (like the telephone and TV before them), and its interfaces will become ubiquitous. Computers in the family room will bridge the functions of a PC and a TV, with the ability to provide services that are *pushed* to us when we want them, according to our preferences, instead of the current *pull* technology that requires a user to search for the programs of choice.

To make this transition, the computer will have to migrate from a device whose user interface has been targeted at a known segment of the population (computer-literate adults and interested children) to a class of products intended for a much broader consumer market, with interfaces that fit the needs of all consumers. This market will require the computer to become more of an appliance—an intelligent tool—so important to our everyday lives (essentially, a natural extension of our family activities) that we use it without even thinking about it.

Intelligent appliances will allow users to break away from the traditional desktop computer in much the same way the PC freed us from the mainframe. These appliances will need to be as simple to use as a toaster but as powerful as a TV or telephone. While each appliance will serve a unique purpose, each will also link to the Internet, capturing additional Web data

and incorporating the data into its own capabilities. Today's network computer (NC) is a first step toward such an appliance. Like a PC, an NC has a processor and associated program memory, display, keyboard, and pointing device. Unlike PCs, NCs do not have disks with complex resident operating system and application software. To achieve such functionality, network architectures will need to become more robust and flexible to accommodate the extensive variety of devices.

Three types of interfaces are expected to become more widespread in the future:

- *Speech recognition.* This interface uses specialized computation and software, such as word-spotting algorithms, to automate simple tasks—for example, requests for operator-type services. Fueled by continuing progress in solid-state electronics, speech recognition will soon find much wider application. The challenge (as confirmed repeatedly in Lucent's human factors studies) is the necessity to attain better than 95% accuracy. Anything less and the user will either by-pass the feature and complete the interface manually or the product will come back in large numbers due to user dissatisfaction. Products capable of overcoming this hurdle are probably five years away for end-user applications. Once the algorithms can accommodate connected digits, mixtures of numbers and words, as well as varying background noise levels, voice recognition and control will be a viable—perhaps ubiquitous—human-to-PC user interface methodology.
- *Gesturing.* The use of gestures involves machine vision that not only can discern what a user is doing but that also can distinguish such action in various backgrounds accompanied by motion clutter caused by inadvertent user movement. Gesture-driven interfaces allow a user to employ the human-machine interface unconsciously in nonobtrusive ways for controlling household appliances, changing TV channels with thumb and forefinger movements, and for communicating through sign language to motion and text translators for

writing. Another obvious application is dialing the phone without the need for a keypad.

- *Graphical displays.* While the two foregoing interfaces still require several years of development and testing, improvements in graphical user interfaces (GUIs) are occurring at a logarithmic pace. It is possible that the icons used in GUIs will progress over the next three years from static pictures to animated cartoons that show what they are doing for the user and where they are in the execution process. These GUIs will follow the format that has been so successful in popular entertainment formats.

All these changes will greatly enhance the user experience and drive the PC's penetration into the home from its current 35% to 40% range up to the 95%-plus range typical of telephones and TVs.

Conclusion

It is not an exaggeration to say that the transistor is the single most important technical innovation of the twentieth century. Transistors and their progeny—large-scale integrated circuits and optoelectronic devices—have truly revolutionized every aspect of people's lives. They created new industries, new capabilities, and new economic paradigms that made "have-not" nations into "haves" and permitted global competition to be played out in marketplaces instead of on battlefields.

As discussed in this paper, computers in all forms owe their existence to silicon technology. Similarly, today's communications systems—though less visible to most people—have evolved based on the same fundamental technology. Amazingly, the limits of the use of silicon are today only vaguely in sight. We now see that the silicon-enabled hardware industry that provides the underlying technologies for LANs and WANs, as well as for mass storage, switching, transmission systems, and wireless communications, will grow dramatically during the next twenty-five years. Equally dramatic growth will be seen in the critically important software industry.

Driven by an impatient demand for bandwidth, the public communications networks with their strong circuit-switched voice base will evolve to multimedia

networks that handle voice, data, and video with equal facility. To accommodate this transformation, optical communications is critical.

Solid-state electronics created the information age, which has done more to enhance understanding and free individuals from drudgery and ignorance than any other development in world history. And yet, it is only in its infancy.

Acknowledgments

The author thanks the many individuals who provided the background information used in developing this paper. Bell Labs contributors include Bill Coughran (software and the Internet); Alastair Glass (enabling broadband); Bruce K. Hillyer (optical storage); Jim Mazo (mass storage); Edi Sackinger (architecture); and Ran Yan (wireless and mobility). Those contributing from the Lucent Microelectronics Group include Lynn Ditty and Tony Grewe (interface), as well as Douglas Haggan and Bill Troutman (silicon). Meyer Zola of Network Systems assisted with the section on switching. The author also appreciates the counsel of Bill Weiss, who provided editorial assistance, and Herb Burton, who read the manuscript and made a number of helpful suggestions.

*Trademarks

Apple is a registered trademark of Apple Computer.

Dolby is a trademark of Dolby Laboratories.

IBM and PowerPC are registered trademarks of International Business Machines Corp.

Intel, MMX, Pentium, 286, 386, and 486 are registered trademarks of Intel Corp.

Java, UltraSPARC, and VIS are trademarks of Sun Microsystems.

M68000, MC68000, MC68020, MC68040, MC68060, and MicroTAC are registered trademarks of Motorola.

Microsoft is a registered trademark and Windows NT and Windows 95 are trademarks of Microsoft Corp.

Pixar is a registered trademark of Pixar Animation Studios.

Shiva is a trademark of the Shiva Corporation.

TM-1 is a registered trademark of Phillips.

TMS320 is a trademark of Texas Instruments.

UNIX is a registered trademark of The Open Group.

References

1. L. Gwennap, "Processor Performance

- Climbs Steadily; Both RISC and CISC Improve at 50-55% Per Year—No Slowdown Yet," *Microprocessor Report*, Vol. 9, No. 1, Jan. 23, 1995, p. 17(4).
2. S. M. Sze, *VLSI Technology*, McGraw-Hill, New York, 1988.
3. D. A. Patterson and J. L. Hennessy, *Computer Architecture: A Quantitative Approach*, Morgan Kaufmann, San Mateo, California, 1990.
4. M. Tremblay and P. Tirumalai, "Partners in Platform Design", *IEEE Spectrum*, Vol. 32, No. 4, Apr. 1995, pp. 20-26.
5. D. Psaltis and F. Mok, "Holographic Memories," *Scientific American*, Vol. 273, No. 5, Nov. 1995, pp. 70-76.
6. B. Miller and S. G. McMeen, "Automatic Systems", *Cyclopedia of Applied Electricity*, Vol. 7, American School of Correspondence and American Technical Society, Chicago, Illinois, pp. 299-312.
7. *Engineering and Operations in the Bell System*, Call No. 621.385/B435e2, Bell Telephone Laboratories, 1980, pp. 225-245.
8. L. N. Hampton and J. B. Newsom, "The Card Translator for Nationwide Dialing", *Bell System Technical Journal*, Vol. 32, No.5, Sept. 1953, p. 1037.
9. A. E. Joel, "An Experimental Switching System Using New Electronic Techniques," *Bell System Technical Journal*, Vol. 37, No. 5, 1958, p. 1091.
10. R. W. Ketchledge and W. Keister, "No. 1 Electronic Switching System," *Bell System Technical Journal*, Vol. 43, No. 5, Parts 1 and 2, Sept. 1964.
11. A. E. Spencer, "No. 4 ESS," *Bell System Technical Journal*, Vol. 56, No. 7, Sept. 1977, p. 1015.
12. "The 5ESS® Switching System," Special Issue, *AT&T Technical Journal*, Vol. 64, No. 6, Part 2, July-Aug. 1985, pp. 1303-1564.
13. J. G. Ferguson, W. E. Grutzner, D. C. Koehler, R. S. Skinner, M. T. Skubiak, and D. H. Wetherell, "No. 1 ESS Apparatus and Equipment," *Bell System Technical Journal*, Vol. 43, No. 5, Part 2, Sept. 1964, p. 2355.
14. D. Spicer, Z. Allison, and L. Shustek, "Microprocessor Evolution (1971-1996)," poster from MicroDesign Resources, Sebastopol, California, <http://www.MDRonline.com/>
15. *Microprocessor Report*, Vol. 8, No. 5, MicroDesign Resources, Sebastopol, California, Apr. 18, 1994, p. 10.

16. *Microprocessor Report*, Vol. 9, No. 14, Micro-Design Resources, Sebastopol, California, Oct. 23, 1995.

(Manuscript approved October 1997)

ARUN N. NETRAVALI is vice president of Research at Bell Labs in Murray Hill, New Jersey. A fellow of the IEEE and the American Academy of Arts and Sciences, as well as a member of the U. S. Academy of Engineering, he holds a B. Tech degree from the Indian Institute of Technology in Bombay and, in addition, M.S. and Ph.D. degrees from Rice University in Houston, Texas, all in electrical engineering. He also is the recipient of an honorary doctorate from the École Polytechnique Fédérale in Lausanne, Switzerland. Dr. Netravali, who currently edits several journals, has authored more than 135 technical papers and coauthored three books, and he holds more than 60 patents in the areas of computer networks, human interfaces to machines, picture processing, and digital TV. He holds, in addition, numerous awards, including the Alexander Graham Bell Medal, the OCA National Corporate Achievement Award, the Engineer of the Year Award from the Association of Engineers in India, the Thomas A. Edison Patent Award, and the C&C Prize. ♦

