Spring 3-1990

Core component choices in single-user computer systems : a home office user's perspective

James W. Violette

CORE COMPONENT CHOICES IN SINGLE-USER COMPUTER SYSTEMS:

A HOME OFFICE USER'S PERSPECTIVE

An Independent Research Project

by
James W. Violette

Executive MBA, Class of 1990
University of Richmond

March 19, 1990

Table of Contents

Section 1

Introduction

The home office is a rapidly growing segment of the business environment. The trend toward two-income families and concerns over quality of life have made the office at home an increasingly attractive alternative business style. The evolution of technology during the past ten years has opened up a broad array of choices.

The introduction of the IBM personal computer in the fall of 1981 provided the technological nucleus. Other office products aimed at the individual user such as personal copiers, facsimile machines, smart typewriters, and multi-function telecommunications products have grown around it.

The evolution of personal computer technology has been accelerating since its introduction; the home office user has a broad and confusing array of choices at varying levels of technological development and intercompatibility.

The objective of this study is to identify and explore the most significant choices faced by the home office computer user. In summary, these choices are: The operating system, the microprocessor, the memory configuration, the hard disk, the monitor, and the printer.

Each section will begin with an analysis of the current array of choices available and a brief historical perspective of the evolution of those choices. Sufficient technical explanation will be provided to illustrate the bases for the recommendation(s) that follow.

The personal computer market is evolving at a very rapid rate. Once clear distinctions between personal computers, engineering workstations, minicomputers, and mainframes in the hierarchy of computing platforms have become blurred. Power, speed, and flexibility once found only in large systems have migrated to the personal computer. The situation is further complicated by the fact that the matrix of choices changes weekly, changing the relationships between classes of systems and the available hardware and software tools that they employ.

The erosion of hardware and software prices caused by increased product proliferation and new product introduction further complicates decision making. It is nearly impossible to make a purchase decision that will not look as if it should have been postponed a month.

Given the rate of change, it is inappropriate (if not impossible) for the prospective user to make choices intended to be relevant over greater than a three year period. It is likely that the technology affecting these decisions in three years is substantially unknown today. The fact that such decisions cannot be expected to remain optimal does not alter the fact that they must be made. Therefore, the recommendations made in this paper acknowledge the dynamic nature of the situation. The intended useful life of the choices recommended is three years.

The choices addressed here begin with the operating system to be employed. While the operating system is not a hardware component, choices concerning the hardware configuration cannot be intelligently made with out taking the operating system, along with any accompanying environmental requirements, into account. The key determinants of system performance are generally acknowledged to be processor speed, hard disk access time, and video resolution and speed. The processor choice carries with it implications for system memory. Accordingly, the choice of microprocessor, memory, hard disk, and monitor will be considered in that order. Finally, since a printer is a necessary component of virtually every home-office computer system, it will be considered as well.

Following the individual sections will be a brief summary of the conclusions reached and recommendations made.

Section 2

Operating Systems

DOS

The operating system is the software that identifies and controls the hardware in the computer system on behalf of users. Users, in this context, are either application programs, people, or both. IBM introduced the PC in the fall of 1981 with an operating system called PC-DOS (for Personal Computer-Disk Operating System) developed by Microsoft Corporation. The generic version of the operating system for non-IBM machines was and is known as MS-DOS. DOS has undergone numerous revisions and improvements since it was first introduced to accommodate advancements in hardware and application software requirements.

Despite this evolution, DOS is constrained (in the interest of backward hardware and software compatibility) by the architectural limits of the 8088 microprocessor--the engine of the first PC. The essence of this constraint is that application software is required to run within a 640K space while system drivers occupy some portion of the space between 640K and the one megabyte address limit of the 8088 chip. Although subsequent microprocessors are free of this constraint, the attractiveness of maintaining compatibility with the large hardware and software population in place has limited exploitation of the additional flexibility.

DOS Shells

A DOS shell can be thought of as a user-friendly packaging of the operating system. Typically, the DOS shell loads during the initial boot process and presents the hard disk directory, file structure, and commonly used DOS command options in an easy to use menu structure. A DOS shell may be part of a composite utility program such as Norton Utilities or PC Tools Deluxe or as a stand-alone program such as One-Dir Plus.

The latest version of DOS, DOS 4.0, incorporates such a shell along with support for hard disks larger than 32MB (in previous versions, hard disks had to be partitioned into several hard disk volumes, none of which could be larger than 32MB).[4] The upper limit of hard disk size under DOS 4.0 is one gigabyte. DOS 4.0 also recognizes the EMS 4.0 specification, enabling users to allocate memory above 640K for RAM or disk caching (which is discussed in more detail in the memory section of this paper.) This capability has been available for some time in utility programs, but never before through the operating system itself. Early releases of DOS 4.0 had compatibility problems, particularly with TSRs (terminate and stay resident utility programs) and operating environments like Microsoft Windows; however, these problems have been solved.

Graphical User Interfaces

A graphical environment represents choices in terms of symbols know as icons. Instead of typing instructions from the keyboard, as would be the case in a character based environment, choices are highlighted using a mouse or alternative pointing device and selected by pushing a button. The advantages of the graphical environment are that navigation can be accomplished quickly and intuitively, in approximately the same way people think and work in a non-computer environment. The disadvantage of a graphical environment is inherent in its bit-mapped presentation (each bit of a screen display must be separately mapped in memory). Substantially more time and memory is required to paint a graphical screen. Consequently, the hardware requirements of a graphical operating environment are substantially greater than for a character based environment for any given screen change selection.

Windows

Microsoft Windows is a graphical user interface that deserves a special look because it is intended to move well beyond simply providing a quick, mouse-driven operating environment within which to work. Windows also attempts to standardize the input/output functions normally handled by individual application programs. This includes drivers for printers and video output which normally occupy part of the application program's memory requirement while in use.

Programs written for Windows not only defer these I/O functions to the environment software, but offer standardized menu management features that are common to all other Windows programs. The practical result of this feature is that the user need learn only one program; additional programs added to the environment use the same control panel and procedures and are thus much easier to master than their discrete program counterparts.

Also, relieving the application programs of individual driver sets reduces their memory requirement such that more application programs may be left open in a given memory space at any one time. This facilitates the DOS extender capabilities of Windows.

DOS Extenders

A DOS extender acts as an attachment to DOS which allows the application(s) to take advantage of extended memory available to the the 80286, 80386, and 80486 microprocessors in their protected mode. Such extenders may be part of discrete application programs such as Lotus 1-2-3 Version 3.0 or operating environments like Windows 386 or DESQview 386. Essentially, the DOS extenders create the appearance of an 8088 environment for the operating system. Among the advantages that accrue to these operating environments is the ability to work in several different application software packages during a single session, sharing relevant data between them.

DESQview

DESQview deserves special mention because it provides much of the utility of Windows in a much less hardware-hungry, character-based environment. Multiple applications can also be opened up under resizeable windows simultaneously, and ASCII data can be moved between the opened applictions. DESQview can also display both character- and graphic-based applications on the screen at the same time. A Windows application like Excel can be open in one DESQview window, while WordPerfect, a character-based application, is open in another.

OS/2

OS/2 (for Operating System/2) was developed by IBM and Microsoft Corporation to succeed DOS as the standard operating system on 80286- and 80386-based PCs. It was introduced in December, 1987 as OS/2 Standard Edition 1.0 and followed by OS/2 Standard Edition 1.1, released in October, 1988. Version 1.1 added a graphical user interface known as Presentation Manager. Both these versions are list priced at $325 and are sold by vendors of PC compatibles. IBM sells Extended Edition versions of both OS/2 releases for the same price; these versions have been optimized for IBM equipment.[17]

OS/2 uses the protected-mode (as opposed to the 8086/8088-like real-mode) of 80286-, 80386-, and 80486-based PCs. It was designed around the 80286 microprocessor and can directly address the 16MB physical address space of that chip. OS/2 can run several programs at once (multi-tasking) and use virtual memory (storing overflows from available memory to disk).[20]

Presentation manager programs have a look and feel that is virtually the same as those written for Microsoft Windows. This similarity is intended to allow a nearly seamless transition when users move from the DOS/Windows environment to OS/2.

Thus far, the migration from DOS to OS/2 has been less than enthusiastic. This is attributable partly to the high cost of conversion. The minimum RAM memory requirement of OS/2 with Presentation Manager is 4MB. Also, the reduction in screen sequencing speed caused by repainting a bit-mapped screen requires faster image processing to break even with character based screens. Another factor has been the relatively weak offering of OS/2 applications presently available. The device drivers required by OS/2 are similar but more complex than those required for windows. Postscript printer drivers, not currently available, are scheduled to be included in the next release of OS/2 (Version 1.2).

Another significant limitation of OS/2 is its inability to take advantage of the 32-bit registers and addressing available on the 80386 microprocessor (which will be discussed in more detail in the microprocessor section of this paper.) An OS/2 version for the 386 is scheduled for release sometime in 1990. It will support nonsegmented, 32-bit address space and will also support multiple DOS sessions.[17]

Unix

Unix is another operating system option competing with DOS and OS/2. It offers a number of powerful advantages, including faster, full 32-bit performance (OS/2 is a 16-bit operating system). Unix is capable of acting as a parent operating environment under which DOS may be a subsidiary operating system.

Past Unix users are likely to think it among the least likely candidates for a main-stream operating system because it is characteristically among the least user-friendly and most difficult to learn. It was first developed by Bell Labs about 20 years ago using a very terse (as few characters as possible) command language. One of its major virtues is its portability--it can be easily installed and used on computers of all sizes from mainframe to microcomputer.

In fact the Unix world has accepted the idea of a graphical user interface (GUI)--at least in principle. The problem is that no concensus has developed. AT&T, in concert with Sun Microsystems, developed a GUI called Open Look (Sun/Open Windows by Sun). A group of competitors, fearing an unfair advantage in the Sun/AT&T alliance, formed their own alliance. IBM, Digital Equipment (DEC), the Santa Cruz Operation (SCO), and Hewlett-Packard (HP) reacted by putting together their own power group, the Open Software Foundation (OSF). They have developed an interface with the style of Microsoft's Presentation Manager. Meanwhile, Microsoft, SCO, and HP are expected to deliver a version of the Presentation Manager for Unix. The NeXT computer from Apple cofounder Steven Jobs uses its own user-friendly Unix windowing system called NeXTStep, which has been licensed by none other than IBM. Clearly, some time will pass before the GUI of choice emerges from this scenario.

Recommendation

OS/2 is not yet fully developed in terms of its own capabilities. Its file management system is cumbersome and undergoing modification, the driver sets for mainstream devices such as laser printers are not yet available, and the available application program packages are limited in number and expensive. It is likely to be the operating system of choice for the future in the single-user environment, assuming the proposed enhancements that allow use of the 80386 32-bit instruction sets and multiple DOS applications are successfully implemented. Failure to implement the 386 applications will leave the door open for Unix.

Unix is likely to become a significant operating system in the multi-user, multi/system environment at some point. It has the ability to interface with different classes of systems on a client-server basis without distinguishing between machine architectures. Unix is capable of maintaining a virtual machine environment, under which DOS and OS/2 applications may both be run, but while it is the most flexible of the three operating systems, it has the least unified force supporting its implementation. It is therefore unlikely to become the operating system of choice for the single-user system in the next three years--if ever.

DOS remains the most versatile and cost-effective choice between the operating systems for the 36 month life of this decision. The software library is prolific and relatively inexpensive. The large population of users suggests that this will continue to be the case for at least the next five years and probably longer. The use of DOS extenders in concert with the 80386 microprocessor extend many of the intended OS/2 benefits to the DOS environment. Operating environments such as Windows and DESQview provide a flexible framework for moving between applications that transcends the application data-sharing issue. This capability very simply allows the user to work with a computer in a manner similar to any other desktop business tool such as a calculator, a telephone, or a reference book--no additional overhead is required to switch from one to another.

DESQview offers a more flexible (in that programs need not be written specifically for it) environment with smaller hardware requirements (in processing speed, video resolution, and memory) but it is character based--and each program remains a separate entity to be separately learned.

Windows applications, once the first application is learned and with the appropriate hardware needs met, become easier to learn and more intuitive to use. Additional applications can be added without the need to learn a new set of controls. It has the additional virtue of similarity to the most likely operating system of the future--OS/2. The transition to OS/2 from Windows is likely to be the most painless.

References:

1. What's Behind The New EISA Standard?, _Personal Computing_, Jan '89, p297(2).

2. EISA Sets The Stage For Open Standards, John Blackford, _Personal Computing_, Jan '89, p301(2).

3. OS/2: Just Say No, Patrick Honan, _Personal Computing_, Dec '88, p107(2).

4. DOS 4.0: One Step Beyond, Patrick Honan, _Personal Computing_, Dec '88, p111(5).

5. Charting A Windows Strategy, Christopher O'Malley, _Personal Computing_, Nov '88, p128(6).

6. The Venerable DOS Lives On, Christopher O'Malley, _Personal Computing_, Nov '88, p240(2).

7. UNIX: Tomorrow's Operating System?, Christopher O'Malley, _Personal Computing_, Jun '88, p100(9).

8. Bus Wars, Eric Bender, _PC World_, Feb '89, p158(4).

9. Which Bus to Take?, Jim Nech, _PC World_, Jan '89, p290.

10. Presenting The Future--What Benefits Will The OS/2 Presentations Manager Bring, And When?, Eric Bender, _PC World_, Jan '89, p134(3).

11. Windows: Shortcut or Detour?, Daniel Brogan, _PC World_, Jan '89, p140(2).

12. Faces of UNIX, Ross M. Greenberg, _PC Magazine_, Sep 12 '89, p143(14).

13. EISA Under The Glass, Patrick Honan, _Personal Computing_, Aug '89, p67(4).

14. _The Winn Rosch Hardware Bible_, Winn L. Rosch, 1989, Simon & Schuster, Inc.

15. Taking A Realistic Look At DOS 4.0, Ray Duncan, Jan 17 '89, _PC Magazine_, p329(5).

16. What's Wrong With Windows, Bill Howard, _PC Magazine_, Sep 12 '89, p116(2).

17. GUIs for DOS and OS/2, Charles Petzold, Luisa Simone, & Tami D. Peterson, _PC Magazine_, Sep 12 '89, P111(26).

18. The GUI--An Interface You Won't Outgrow, Jim Seymour, PC Magazine, Sep 12 '89, p97(12).

19. The Great Divide--EISA vs. Micro Channel, Will L. Rosch, PC Magazine, Dec 27 '88, p165(21).

20. OS/2 Extended Edition--System Integration The IBM Way, Charles Petzold, PC Magazine, Jan 31 '89, p141(10).

21. Extending DOS, Peter D. Varhol, Personal Workstation, v2-Mar '90, p85(4).

Section 3

Microprocessors

The engine of the personal computer is the microprocessor. It is constructed from a thin slice of pure silicon that has been treated with impurities that affect how the silicon conducts electricity. The impurities change the metal silicon into a new class of materials called semiconductors. The term refers to the altered silicon resisting electrical current flow more than conductors (like the copper in wires) but not as much as insulators (like the plastic wrapped around the wires). The microprocessor is an elaborate arrangement of circuits implemented on a single small slab of silicon. This integrated circuit is sometimes called a chip because of the construction from a single small piece of silicon, a chip off the crystal from which it comes.[13]

The signals going into the microprocessor consist of a collection of digital pulses arriving nearly simultaneously--in parallel--on several wires. Each pattern is a command to perform a certain function and each has a specific name. The entire repertory of these functions and their names is called the command set of the microprocessor.

The microprocessor stores and manipulates digital bits in special areas called registers; data moves into and out of the microprocessor on a set of connections called the data bus. In addition, microprocessors use another bus called the address bus to indicate to the rest of the computer which memory areas it needs to access. Microprocessors differ in the resources they devote to each of these facilities, and this in turn influences the speed at which the microprocessor can work.[13]

Not only do microprocessors have different numbers of registers, but the registers may be of different sizes. Registers are measured by the number of bits they can work with at one time. For example, a 16-bit microprocessor must have one or more registers that can hold 16 bits of data at a time.

The number of bits in the data bus of a microprocessor directly influences how quickly it can move information. The more bits a chip can use at a time, the faster it is. Microprocessors with 8-, 16-, 32-, and 64-bit data buses are used in various IBM personal computers. The number of bits available on the address bus influences how much memory a microprocessor can address. For example, a microprocessor with 16 address lines can directly work with 2^{16} addresses or 65,536 (or 64K) different memory locations.[13]

Evolution

The evolution of microprocessors has been largely a matter of increasing register size and bus width. By 1978, Intel had developed the 8086 chip which had 16-bit registers and a 16-bit data bus--both double the size of the predecessor 8080. The 20-bit address bus was also substantially larger, allowing the 8086 to directly control over one million bytes (a megabyte) of memory. At the time, very little 16-bit support hardware existed and using the 8086 forced engineers to design custom devices which were not cost effective at the time.[13]

The 8086/8088

Intel introduced the 8088 a year after the introduction 8086 in reponse to this problem. It was the same as the 8086 in every respect except one--its data bus was reduced to 8 bits, allowing the 8088 to take advantage of readily available hardware. It might have had a short life, but when IBM built its first PC around the 8088 in 1981, its future was secure.[13]

The 80286

In 1984, IBM introduced the PC AT, using Intel's 80286 chip. A true 16-bit chip, the 286 was much faster that the 8088, and it could address 16MB of RAM. The 286 was also downward compatible. Although the chip included some new instructions and memory management circuits that could give PCs new powers (such as running more than one program at a time), DOS pretty much ignored these abilities and treated the 286 as a very fast 8088.[11]

Although the AT ran only 25 percent faster than the PC, its 16-bit bus and registers allowed it to achieve throughput about five times greater. Most important in the long run, was to be the superior memory handling ability of the 80286. Instead of the 20 address lines of the 8088/8086, the 80286 had 24. The four extra lines increased the maximum amount of memory the chip was able to address by 15 megabytes, up to a total of 16 megabytes.[13]

The 80286 allowed the use of virtual memory. Virtual memory is not made up of physical (real) memory chips; instead it is information stored elsewhere, in a mass storage system, that can be transferred into physical memory when needed. The 286 has special provisions for distinguishing each memory byte that is in real and virtual memory, although it requires additional circuitry to handle the actual swapping of bytes. The chip can track up to one gigabyte (1024 megabytes) of total memory--16 megabytes physical, 1008 megabytes of virtual memory.

Theoretically, the one megabyte addressing barrier of the earlier chips should have been a thing of the past. The success of the earlier PC and the substantial base of software that had built up to support it prevented this from happening.

Compatibility with the earlier chips would speed acceptance of the 286, so Intel's engineers endowed it with two operating modes: Real mode was designed to duplicate the operation of the 8086, including the one megabyte limit on memory. The limit was needed for the 80286 to be able to recognize memory addresses just as an 8086 would.

Intel created the protected mode to take advantage of the improved memory handling of the 286 architecture. This allowed the full range of 286 real and virtual memory capabilities to be actively used by programs written specifically to take advantage of it. As a practical matter, protected mode was not accepted quickly. Almost three years after the introduction of the AT, IBM introduced its protected mode operating system, OS/2-it has not yet achieved widespread acceptance.[13]

There were two reasons for the slow acceptance. First, Intel designed the process of shifting from real to protected mode as a one-way process. The chip was designed to start up only in the real mode, and allowed an upshift to the protected mode; but, a downshift back to the real mode was not possible. Once in protected mode, the only way back was to reboot the computer. The second reason for slow acceptance was that, although it allowed more memory to be used, it still operated with memory segments of 64 kilobytes. Instead of a free range of memory within which the software could operate, programmers got a bunch of little boxes that they had to shift their numbers between.[13]

The 80386

The dust had not settled on the introduction of the AT before Intel began shipping the 80386. It was anounced in October of 1985 and began appearing in computers in the fall of 1986. First, the 386 doubled the size of registers and data buses to a full 32 bits. Information could be moved into the chip and processed twice as fast as with 16-bit chips like the 286. Secondly, the clock speeds were designed to start where the 286 left off--12.5 and 16 MHz initially. 386 chips are currently available with clock speeds of 33 MHz.

Complementing the expansion of the data bus of the 386 to 32 bits, the number of available address lines was also increased to 32. By itself, that expansion allows the 80386 to directly address up to four gigabytes of physical memory. In addition, the chip can handle up to 16 terabytes (trillion bytes) of virtual memory. Perhaps most importantly, the chip has full facilities for managing all this memory built into its circuitry.[13]

Dividing the memory is an option in the 386 chip--segments can be virtually any size that is convenient for a program to work with (as long as they are smaller than four gigabytes). The 386 also stores the next 16 bytes of instructions for the program being executed in a waiting area, requiring less waiting as code is retrieved from system memory.

The 386 also has a real mode in which the chip boots up and operates as if it were an 8086, complete with one megabyte addressing limit. From real mode, it can be switched to protected mode where it can operate like a 286 except for the additional flexibility of manipulating the memory segment size to suit its application.

A new mode called Virtual 8086 mode gives the 386 particular freedom in running DOS programs. In this mode, the chip simulates not just one 8086 but an almost unlimited number of them, all at the same time. This mode allows the 386 to divide its memory up into many virtual machines, each one acting like it is an entirely separate computer equipped with an 8086 microprocessor. This property of the 386 makes multitasking control of software almost trivial because all the difficult work is done in hardware. Off the shelf DOS programs work without modification in most 386-based multitasking environments.[13]

The 80386SX

The 80386SX is a scaled-down 80386. It has all the features of the 386 in terms of memory management and multitasking, but less power and speed. Two differences separate the 386SX from the 386: The 386SX uses a 16-bit memory bus instead of a 32-bit bus; its registers must be filled in two steps from a 16-bit I/O channel. The 386SX is also less expensive. The 386SX still works faster than a 286 with an equivalent speed rating, because it can process instructions twice as fast with 32-bit registers. Like the 386, it is also backwardly compatitible with the earlier chips.

The 80486

The 486 uses the same basic instruction set as the 386, but a number of extra circuits are added to process floating-point math and move data quickly and efficiently. Benefiting most from what it has to offer are processor-intensive applications such as engineering, computer-aided design or manufacturing (CAD/CAM), and complex financial analysis.

An important advance introduced with the first non-IBM 486 PCs is the Extended Industry Standard Architecture (EISA). Previously, the only bus capable of transmitting data in 32-bit chunks--a necessity if the rest of the computer is to keep up with the 486's 32-bit internal transfer rate--was the Micro Channel (MCA) circuitry introduced by IBM in 1987.[23]

MCA was not well received for two reasons. First, IBM tightly controlled the licensing terms for the MCA to restrict cloning of its newer machines. Second, IBM made the Micro Channel incompatible with expansion boards built for the earlier AT and PC buses. Users were reluctant to pay the cost of replacing their old expansion cards.

The result was the formation of the Gang of Nine, a consortium of computer manufacturers led by Compaq, which agreed on a standard for a 32-bit wide bus which would accept old 8- and 16-bit cards as well as new expansion boards designed specifically for EISA's 32-bit channel. And, the specifications for EISA are available to any manufacturer who wants to use them.

The 486 has powerful multitasking capabilities. For example, it can run more than one processor simultaneously; one 486 chip could be dedicated to running a network, and another in the same machine would be free to multitask the applications used on the network.[23]

The 486 can use drive arrays, a concept common to a mainframe environment and never before implemented in PCs. The controller splits data in half and writes the halves simultaneously on two different drives, effectively cutting the transfer rate in half.

The primary importance of the 486 to the home-office user is the effect that its introduction will have on the price of computing at every level. 486 capabilities, given the currently available applications, are well beyond the needs of the home-office user. The benefits of the 486 will be felt in price reductions of the computing platforms beneath it. Those effects are already being observed and they will be magnified as the supply of 486 chips is widened to include a broader population of competitive manufacturers.

Recommendation

The price erosion currently being experienced has pushed the 286 into the realm of 8088/8086, effectively making the 286 the entry level processor. For a single-application user environment, where that condition can be expected to remain constant, the 286 may well be the most cost-effective choice.

For the multi-application home-office environment though, the choice should be a version of the 386. The memory management capabilities of the 386 and 386SX allow multiple applications flexibility and control that outweigh the relatively modest remaining differential in cost.

In a graphical operating environment, the faster 386 should be chosen over the 386SX. It will be a worthwhile investment in the near term and be much more likely to survive the transition to an OS/2 environment at the end of the three year system life if such a move is indicated. If the processing requirements of the CPU are not substantial and/or if the near term financial contraints are severe, the 386SX may be the best choice.

References:

1. The 80286: Unsafe At Higher Speeds?, Edward Mendelson, PC Magazine, Dec 27 '88, p195(27).

2. 25MHz 386s--Eclipsed but Still Shining, Catherine D. Miller, PC Magazine, Nov 28 '89, p139(49).

3. The Many Faces of a 386, Russ Lockwood, Personal Computing, Jul '88, p90(7).

4. The New Breed of 386 Machines...And Where They Stand in the Microcomputing Spectrum, John C. Blair, Jr., Online, May '88, p103(12).

5. Built For Speed--The Role of System Architecture, Lisa Kleinman, Personal Computing, Sep '88, p112.

6. Micro Channel: Forging a New Path, Russ Lockwood, Personal Computing, Sep '88, p117.

7. 80386SX Performance Trials, Russ Lockwood, Personal Computing, Aug '89, p87(11).

8. IBM Kicks Off the 486 Race, Eric Bender, PC World, Sep '89, p66(2).

9. Why You'll Want a 486, Richard Landry, PC World, Sep '89, p17(7).

10. The 32-Bit Question: 286 or 386SX?, John Diebold, PC World, Sep '89, p106(2).

11. The Next Chip off Intel's Block, Phillip Robinson, PC Computing, Apr '89, p95(6).

12. What's Old Is New, Winn L. Rosch, PC Magazine, Jan 17 '89, p259(13).

13. The Winn Rosch Hardware Bible, Winn L. Rosch, 1989, Simon & Schuster, Inc.

14. RISC Will Eventually Dominate the PC Industry, Observers Say, Carole Patton, InfoWorld, Apr 11 '88, p26.

15. 286s--No Frills, Few Thrills, Mitt Jones, PC Magazine, Sep 12 '89, p159(68).

16. First 486 Machines Offer Minicomputer Power, John Dickinson, PC Magazine, Dec 12 '89, p33(4).

17. The Size is Right--Packing 386 Power Into Sleek PCs, Catherine D. Miller, PC Magazine, Nov 15 '88, p92(35).

18. The Controversy Continues--286 vs. 386, John Dickinson, PC Magazine, Jun 27 '89, p175(55).

19. The Midrange Has Improved--386s Become Affordable, Bruce Brown & Stephanie K. Losee, PC Magazine, Feb 28 '89, p145(27).

20. When One PC Equals Four--386 Multitasking, Steven R. Davis & Winn L. Rosch, PC Magazine, Feb 28 '89, p94(47).

21. Multitasking Now!, Paul Bonner, PC Computing, Oct '89, p98(5).

22. High-Rise Computing: 22 Powerhouse PCs Top Off at 33 MHz, John Dickinson, PC Magazine, Oct 31 '89, p97(79).

23. Speed Demons: The New Fast 486s, Ron White, PC Computing, v3-Mar '90, p88(11).

Section 4

Memory

The IBM PC was introduced during the fall of 1981 using an Intel 8088 microprocessor. The addressable limit of the 8088 is one megabyte. Since the 8088 has only 20 address lines, it can directly manipulate only one megabyte of memory.[1] At the time of the IBM PC's introduction, the architectural limits of the 8088 chip were well beyond the requirements of the anticipated applications. In those days, Visicalc (the spreadsheet of [severely limited] choice) ran on the Apple II using 48K of RAM.[9]

The first 640K of the one megabyte limit was (and is) available to DOS and the application programs. The remaining 384K was used for "memory-mapped" peripherals and the system's BIOS.[4] "Memory-mapping" is the PC's method of gaining fast and flexible access to its peripheral devices. For example, the PC's display devices are memory mapped, occupying chunks of memory space above 640K. The PC writes to the display screen by depositing information at the proper location in this memory.[9]

IBM used a more advanced microprocessor, the Intel 80286, when it introduced the PC/AT. The 80286 chip is capable of addressing 16 megabytes of physical memory directly. But, since the 80286 has to emulate the 8088 in order to run a DOS application, it is effectively confined to the same one megabyte limitation. The same constraint applies to the Intel 80386 microprocessor; the 80386 chip can directly address up to 4 gigabytes (4,000 megabytes) of physical memory.[10]

The Need for More

Application programs grew in sophistication and potential power at a much faster rate than the developers of the IBM PC had anticipated. Microcomputers could handle data base and spreadsheet applications by 1985 that required a minicomputer in 1980.[12] Along with this power came the need for more memory for processing code and data storage.

Attachment programs (add-ons and add-ins) that enhanced the capabilities of major application programs were introduced. Convenient stand alone programs were designed to load and then "terminate and stay resident." TSRs allowed users to access a variety of useful features still resident in memory from any application they happened to be using. The price of these attachments and TSRs was more and more of the increasingly precious 640K of available RAM.

Developers of software, attempting to reduce anxiety and make personal computers more accessible to computer illiterate users, began offering icons as substitutes for keyboard character entry commands (the Graphic Environmental Operating System or GEOS and similar features in application programs). These features also demanded a memory price be paid.

Expanded Memory

The first problem was finding a way to add memory to a system that could not access any more--the 8088 based PC. The solution was to add a new memory-mapped peripheral device in the upper 384K of memory called an EMS (Expanded Memory Specification) RAM card. Where a display controller provides its memory-mapped RAM only for the sake of storing the image to be displayed, the Expanded (not extended) Memory Specification card provided RAM for storing whatever information the software decided to store there.[9]

The memory available to this new memory-mapped peripheral device in the upper 384K was very limited (64K at any one time) but a special access scheme made it effectively much larger. (This initial solution was called LIM EMS 3.2 and it was initially announced in July, 1985; the LIM stands for the developing/endorsing companies Lotus, Intel, and Microsoft.) The EMS was allocated 64K in the upper memory called a "page frame." The memory in this page frame was further divided into smaller chunks called "pages" of 16K each. The memory board was likewise divided up into corresponding pages of the same size. Portions of expanded memory were switched in and out of the page frame as needed. While only 64K of expanded memory was accessible at any given moment, the circuitry that performed the memory page mapping operated very quickly, making other 16K pages available almost instantaneously.[9]

Added to this "smart board" was an EMS software driver, written for the hardware, which defined and implemented a standard software protocol to let multiple resident and transient application programs control portions of EMS memory while preventing usage conflicts.[9] For this to work, the application program had to accept the 16K page size and overall 64K limit that was available at any one time, but it was a system that generally did work to allow the 8088 to access up to 8 megabytes of additional RAM.

Note: The use of the past tense in this paper reflects only the fact that the process being described is no longer "state of the art" and is not intended as a comment on the extent of its continued use.

Enhanced Expanded Memory

AST Research developed an Enhanced Expanded Memory Specification (EEMS) and introduced it in the fall of 1985 (Only a few months after the introduction of EMS 3.2.) It was promptly endorsed by board maker Quadram and by software maker Ashton-Tate. EEMS allowed information to be swapped in larger segments by effectively removing the limits on the size of the page frame. It theoretically allowed the movement of up to 64 16K pages (the entire one megabyte) in and out of the DOS area as well as the IBM reserved area above 640K.[1] (Respect for the territory occupied by the ROM BIOS, video memory, COMMAND.COM, and other necessities made the theoretical limit impractical. A range of 16 pages was more realistic.)

But, the effect of this additional flexibility was to make it easier to swap whole programs (with considerably greater requirements than the 64K allowed under EMS 3.2) in and out of expanded memory, not just their data.[8] This allowed the programs themselves to run their code in expanded memory and opened up multitasking as an option under some operating environments such as DESQview.[11] Control software that used EEMS (such as DESQview) had the responsibility of making sure programs had all their code available when they needed it.[1] AST referred to the EEMS as a "superset" of the LIM EMS board, meaning that it could do everything that the EMS board could do and more.[5] In fact, it worked exactly like EMS 3.2 if no DOS memory was available to it.[1]

Software support for the two standards was uncertain. The marketing clout of Lotus Development, Intel, and Microsoft, plus the head start in software development, favored the EMS 3.2 standard but the AST standard offered the advantage of combining with DESQview and offering multitasking capability, a powerful functional edge.

LIM/EMS 4.0

In August, 1987, a solution to the uncertainty was announced called EMS 4.0 or LIM 4.0. It is completely compatible with both EMS 3.2 and EEMS so it will work with software that takes advantage of either specification. And it provides access to 32 megabytes of memory, twice the direct addressing range of the 80286 microprocessor.[1]

While both EMS 3.2 and LIM 4.0 call for addressing memory in page frames of 64K, LIM 4.0 offers the ability to map any application that resides in memory from 0K to 884K. (Areas above 884K are reserved for ROM.) In other words, LIM 4.0 allows for multiple simultaneous page frames.[1] While EMS 3.2 allowed the addressing of no more than two megabytes per expansion slot--eight megabytes total, LIM 4.0 seems virtually unlimited with a capacity of 32 megabytes. (Given the historical rate of change in this industry, it will be interesting to reevaluate that impression in five years.)

In addition to the support for multitasking of DOS programs under operating environments such as Windows and DESQview, LIM 4.0 offers a provision that allows multiple applications to share the same data.[4]

Programs can execute within expanded memory and TSRs can operate with only a kernel in conventional memory; when activated by the user, the kernel can transfer control to the bulk of the program located in expanded memory. LIM 4.0 includes a function that allows developers of software to write programs that utilize all types of expanded memory hardware including EMS 3.2, EEMS, and 80386 memory using a single programming procedure.[4]

LIM 4.0 behaves as a set of guidelines that takes advantage of the additional memory provided by an expanded memory card rather than requiring specific hardware. The specification details the proper use of a variety of special function calls that can be used with a program to access and manage the extra memory. Linking the software with the expanded memory is a special device driver called the Expanded Memory Manager (EMM). The EMM is loaded when the system boots, and performs the actual hardware interactions needed to carry out EMS function calls initiated by the application software or the operating environment. (The EMM driver is usually provided by the board maker since the driver has to match the hardware design of the board.)[4] When a program wants to claim expanded memory, it notifies the EMM driver which takes pages from the memory board and reserves them for the program to use.

To access the expanded memory that has been reserved, the program tells the EMM to "map" the expanded memory pages it needs into the page frame. Once the mapping is complete, when the microprocessor accesses a memory address with a mapped page, it will really be accessing the corresponding expanded memory page.

Conversion from EEMS to LIM 4.0 is easier than from EMS 3.2, at least to the extent that achieving all of the LIM 4.0 capabilities is the objective. Conventional EMS 3.2 boards are equipped in hardware with a single set of registers for storing the map data that determines which page of expanded memory is mapped to which page in conventional memory. Whenever a new expanded-memory page is mapped into the conventional-memory space, the contents of the corresponding register must be changed via input-output (I/O) instructions from the microprocessor.

EEMS boards have an additional register set, allowing the board to store two complete alternate memory maps, enough for two applications. When a multitasking operating environment such as DESQview shifts processing control from one program to the other, the EEMS driver can switch almost instantly to the corresponding register set, rather than having to recreate the entire map, register by individual register.[4] This means faster multitasking performance for a program that has exclusive control of one of the hardware register sets.

Under LIM 4.0, board manufacturers can add multiple alternative register sets at their own discretion, extending the performance benefits to many applications during multitasking.[4]

In addition, LIM 4.0 allows expanded-memory pages to conduct background direct-memory access (DMA) data transfers, such as those that occur when reading from, or writing to, a disk, while other portions of expanded-memory interact with the CPU. (This feature requires special hardware registers, but these are not otherwise required by the new standard.)[4]

Extended Memory

Extended memory is the memory above one megabyte that can be addressed by the 80286 and 80386 microprocessors but cannot be addressed by DOS. The 80286 and 80386 must be running in their "protected mode" to access extended memory above one megabyte and this conflicts with the "real mode" compatibility of DOS and its application programs.[10] Although extended memory cannot be directly addressed by DOS, it can still be used.

It can be used as a virtual disk to speed up disk-access-intensive applications. DOS treats the RAM disk as just another logical device name, but because all of the manipulations are electronic, there is no lag time for a read/write head to seek the appropriate data sector. The downside is that the memory is volatile and is not retained after the power is cut off.[2]

Extended memory can also be used for printer spooling. By setting up a print spooler (a print buffer in extended memory), a long document can be put into extended memory for use by the printer, freeing the microprocessor for more productive activities.

Disk caching is another application of extended memory. It is similar to a RAM disk but different in that it is controlled entirely by the system. A software utility monitors the actions of a program and stores frequently used files in memory so that they will be accessed more quickly and system performance will be improved by the reduction in access time.

80386 Memory

The 80386 microprocessor supports multiple "virtual 8086" tasks from within its protected mode. In effect, this provides existing real mode DOS application with some of the best of both worlds. The 80386 can remap the system's actual physical memory from above the one megabyte limit to the address space below one megabyte where current DOS applications can gain access to it. It emulates the EMS card's expanded memory hardware to allow DOS applications that use EMS to have as much EMS memory as they could want.[10] This capacity has been looked at more fully in the microprocessor section.

Summary of Expanded-Memory Variations

In the EMS 3.2 specification, the page frame was limited to a 64K space located in the region of conventional-memory above the DOS limit of 640K. A maximum of four 16K expanded-memory pages could be mapped into the page frame at any one time. The

EEMS specification, by contrast, allowed the page frame to be larger, occupying 256K or more. In addition, expanded-memory pages could be mapped into any address within conventional memory, including the lower 640K bytes used by DOS.

As far as the user was concerned, the primary difference between the EMS 3.2 and EEMS specifications is that with EMS you could only use expanded memory to store data. On an EEMS board, you could also run multiple programs under an operating environment like DESQview.

Because of the growing demand for larger programs and for multitasking, LIM 4.0 has incorporated the key features of EEMS (enlarged page frame size and ability to map into the lower 640K of memory). AST Research has fully supported the new standard and indicated that there is no longer a need for EEMS as a separate standard, which, at least for the time being, unifies the standard and eliminates the uncertainty.[4]

Memory Speed

If the memory chip is not fast enough to keep up with the processing speed of the microprocessor, a wait state is imposed on the system. This means that the microprocessor suspends operation for one or more clock cycles until the memory circuits catch up. A number of techniques can be used to insure that memory is fast enough to keep up with processor speed, but they necessarily result in a zero wait states description by the vendor.

Recommendation

Although DOS functions in a 640K environment, this section has illustrated a number of means by which the limit may be circumvented and a number of reasons for doing so. The quickest and most effective means of utilizing memory above 640K is 80386 memory management of onboard system memory. Onboard system memory is readily available to the processor without input/output overhead imposed by accessing memory on a card.

Economical packaged computer systems based on the 80386 microprocessor are typically offered in 1MB, 2MB, and 4MB onboard memory configurations. The smallest of these is adequate only for a single-application user environment, because part of the one megabyte space is allocated to BIOS and video memory. Choosing between 2MB and 4MB configurations depends partly on the nature of the applications (large and/or multiple applications that can remain simultaneously in memory--reducing I/O time). The operating environment selected is also a factor. A character-based environment, like DESQview, requires less memory to accomplish its management tasks and provides substantial flexibility with two megabytes of system memory; a graphical environment, like Windows 386 (which requires a minimum of two megabytes), will require more memory to accomplish the same tasks because of the need to maintain bit-mapped screens. Significant use of terminate-and-stay-resident (TSR) programs will also influence the memory configuration choice. Either management environment is capable of using separately identified extended and expanded memory.

Individual user applications determine the need, but memory requirements are more likely to grow than to shrink; if the decision is a close one, it should be noted that extra memory in a packaged system will probably be less expensive than it will be if purchased separately.

References:

1. Leveraging RAM with LIM 4.0, Winn L. Rosch, PC Magazine v7-Oct 11 '88, p223(16).

2. Taking Your AT's Memory to the Maximum, Frank Bican, PC Magazine v6-Aug '87, p255(18).

3. AT Multifunction Boards: Putting More Memory to Work, Winn L. Rosch, PC Magazine v6-Jan 27 '87, p237(10).

4. That Stubborn 640K-Byte Limit May Be History, Steve Cummings, PC Week v4-Nov 3 '87, p109(5).

5. Expanded Memory Boards, John Greitzer, PC Week v3-July 8 '86, p539(14).

6. Users Must Choose Sides in EMS Board Struggle, Paul Gillin and Bruce Stephen, PC Week v3-May 6 '86, p140(3).

7. Expanded Memory Solution at Hand, Pat Bellamah, PC Week v4-June 2 '87, p1(2).

8. Beyond the 640K Barrier: The Search for an EMS Simulator That Does the Job, Michael J. Miller, InfoWorld v8-Oct 13 '86, p44.

9. Entering Twilight Zone of EMS Expanded Memory, Steve Gibson, InfoWorld v10-Feb 15 '88, p34.

10. The Extended Memory Game and How to Play It, Steve Gibson, InfoWorld v10-Feb 22 '88, p34.

11. 16-Bit EMS Memory: RAM Beyond the DOS Barrier, Stephen Satchell, InfoWorld v9-Sep 14 '87, p35(5).

12. Augmenting the Power of Personal Computers, James Yaremko, Office v106-Aug '87, p38(4).

13. EMS Boards Bring New Life to DOS, Steve Cummings, PC Week v4-Nov 3 '87, p109(5).

14. Expanded PC Memory for Multi-Tasking, Maciek Brzeski, Systems International v15-Mar '87, p75(2).

Section 5

Hard Disks

There are other ways of storing data, but for the time being, a hard disk is by far the most practical. The size and complexity of current major application programs (word processor, spreadsheet, database manager, and graphics package) make running a computer from floppy disks prohibitively cumbersome if not impossible.

Many popular spreadsheet and word processing programs take up at least half a dozen diskettes; running such programs from floppies, if it can be done at all, usually means spending a great deal of time changing diskettes. A hard disk can provide fast access to information because it spins constantly (unlike a floppy disk which spins only upon I/O instruction) and at a speed roughly ten times that of a floppy disk. A wide assortment of application programs and data can be controlled at relatively low cost.

Alternate Technology

It should be noted that there are new, high performance alternatives to the hard disk arriving on the scene. The optical disk is one receiving a good deal of attention. Optical disk machines are now available with 'read and write' capability--an improvement from the 'write once, read many' (worm) configurations to which the first offerings were limited. But, the technology is still very young. It is not yet widely distributed, which means it is expensive and not thoroughly tested by a substantial population of users. Because of the high cost and the real possibility of encountering expensive surprises, the optical disk remains the technology of the future.

Background

The hard disk is an electro-mechanical device (with moving parts.) It is a metal casing that contains rotating platters that store data, read/write heads that transfer data to and from the platters, and the motors to move both. One read/write head is assigned to each side of every platter.

The casing protects the delicate interior from dust particles which could block the narrow gaps between the heads and platters. Such contact would result in plowing a furrow in the platter's magnetic coating--an event known as a head crash.

The operating system software manages the disk by storing a catalog of the files contained on the disk (a File Allocation Table or FAT). The heads used to read and write each side of a platter are fixed at the end of a head arm and all move in unison regardless of which head is to perform the next read or write.

The data is stored on the disk in concentric circles called tracks. The heads align magnetic particles on the surface of the disk to write to it and detect particles that have already been aligned to read from it. An up to date table of contents is maintained in the FAT.

The operating system stores a file in the first sector it finds free, even if that means splitting a file up among noncontiguous sectors. A single file may be strewn across a hard disk in hundreds of sectors. The FAT keeps a record of each link in such a chain. Once the directory data have passed through the drive electronics and hard disk controller back to the operating system, the operating system sends the read/write heads across the surface of the disk to the appropriate spot while the platters spin at 60 revolutions per second.[3] After the operating system writes a new file to the disk, it returns the read/write heads to the FAT long enough to record a list of the file's sectors.

The number of platters and the composition of the magnetic material coating them determine the capacity of the hard disk. The coatings used on hard disks have become less fragile recently with the use of a cobalt alloy three millionths of an inch thick replacing ferric oxide. The cobalt alloy offers much denser storage and far tougher surfaces.

Hard disks are most vulnerable to head crashes when they are turned off. When the power is cut off, the air flow that keeps the read/write heads 'flying' gradually decreases. The heads come to a controlled landing on the surface of the disk. Most hard disks reserve a landing area specifically for that purpose at one edge or the other of the actual data storage area.

A software command is usually required to bring the head to the landing area and hold it there while the disk spins down. This process is called <u>head parking</u>. Some disks are designed so that whenever power is switched off, the head automatically retracts to the landing area before the disk spins down. This feature is called automatic head parking.

Even when the head touches down in the proper landing area, there is the potential for trouble. A shock to the system can jar the head out of the landing zone, and bounce it across vulnerable media in the process. Some hard disks lock their heads in the landing area during power-down to guard against such a possibility. This feature is generally called park-and-lock.[13]

Understanding the differences between hard disks seems confusing at first, but there are really only three important issues to consider: Speed, capacity, and price. This section will help to clarify the terms used to describe hard disks.

Speed

Hard disks are 3 1/2 or 5 1/4 inches wide, and either half- or full-height. Full-height drives are usually first generation offerings; subsequent generations of a model generally benefit from more compact designs.

Hard disks are connected to the rest of the computer system through a controller with connections for power, data, control, and electrical ground. A floppy disk has an average access time of 300 milliseconds (3/10 second). The original XT specification required hard disk access time of 85 ms; the AT specification required 40 ms. Currently, because of the disk-intensive nature of applications, satisfactory hard disk access time is typically considered 28 milliseconds or faster.

There are three common interfaces used in hard disks; they are methods of getting the hard drive to communicate with the rest of the system.

The majority of PC hard disks in existence today operate under the Seagate ST-506/412 interface standard. Two encoding methods (methods of storing data on the disk) come under this standard: Modified Frequency Modulation (MFM) and Run Length Limited (RLL). MFM is by far the more common, although RLL controllers, which cost about the same, speed data transfer and increase disk storage by up to 50 percent. RLL drives pack in 26 sectors per hard disk track compared with MFM's 16, and data is transferred at a faster rate: 7.5 megabits per second instead of 5.

Although technically any hard disk can run with an RLL controller, the drive should be certified for RLL use; RLL disk drives must withstand extra strain.[4] This interface provides satisfactory performance with disk drives of 70MB or less.

ESDI (Enhanced System Device Interface) controllers are used for large capacity hard disks. ESDI is basically an upgrade to the ST506 standard which permits much higher data transfer speeds. They are more expensive than other types of controllers, but are better suited for high-capacity and high-data-transfer rate hard disk applications. Most ESDI drives lay 34 sectors on each track and transfer rates of 10 megabits per second are considered common.

ESDI has the additional virtue of storing configuration and bad track information on the drive itself, which frees the system from storing the information in a BIOS extension or in configuration memory. Each ESDI drive tells the controller how many tracks and cylinders it has and stores bad track information in a standardized form that can be accessed by the controller directly.

SCSI (Small Computer Systems Interface) technology has been borrowed from the Macintosh world for use on high-capacity disks. SCSI is a standard interface, recognized and defined by the American National Standards Institute (ANSI). (ESDI is also a standard of sorts, but one manufacturer's drive may not work with another's controller.)

SCSI Drives are just as fast as ESDI drives, and they offer several practical advantages. A SCSI interface means more accurate time between the controller and hard disk because the SCSI controller is built right onto the drive. More accurate timing results in more accurate reading and writing of data.

Another benefit is the SCSI interface's ability to accept other devices, including floppy drives, optical disks, and printers. This is because SCSI is a system-level interface, while ST506 and ESDI are device-level interfaces. In effect, SCSI provides its own expansion bus to plug into peripherals. An SCSI interface can control up to seven devices with a single controller. SCSI drives are also less expensive than ESDI drives: A SCSI drive with its imbedded controller is approximately the same price as an ESDI drive without a controller.

The abilities of the controller are just as important as the interface in determining hard disk performance. Interleaving is a term that describes how data is placed in sectors and is designated during low-level formatting of the disk.

An interleave of 1:1 means that data is placed in sequential sectors. This is difficult to do because many controllers are not capable of getting the data to the read/write head fast enough to keep pace with the revolution speed of the disk. If the disk is set for a 1:1 interleave and the controller cannot accomplish it, the disk has to spin around again to come back to that next track. This can slow hard disk throughput by as much as 80 percent. A 1:3 interleave means that data is recorded every third sector, and an interleave scheme of up to 1:16 is possible. It is important to know what the optimum interleave is for the hard disk controller to get maximum performance out of the hard disk.

Capacity

While the appropriate capacity varies according to the application of the individual user, it must be noted that application software packages are generally growing in sophistication and demand for hard disk space. Recent versions of popular DOS-oriented spreadsheet and database applications have recommended minimums of two and four megabytes respectively. Mathematical and statistical analysis packages have greater requirements. Commonly offered capacities range from 40MB to 165MB.

Despite the less vulnerable construction of current disk drives, use of a single high-capacity disk drive amounts to putting all the eggs in one basket. Some protection can be achieved by using multiple drives of lesser capacity to meet the aggregate storage needs. Good (and frequent) data backup procedures should be utilized in either case, but the latter approach assures that some hard drive capacity can be maintained if one drive fails.

Recommendation

ST506 provides a satisfactory and economical level of performance in drives with a capacity of 70MB or less. Multiple ST506 drives can also provide a measure of security when capacity requirements are high. This configuration is not typically offered as a package, however; thus is not the among the most economical.

SCSI provides high capacity performance, a flexible bus configuration, and is generally less expensive than ESDI when the hard disk and controller are purchased outside a fully configured system. The advantages of the extended bus configuration are not sufficient to warrant payment of a substantially higher price by the typical home-office user.

High capacity needs can generally be met more economically by acquiring a fully configured system with a high-capacity ESDI drive. ESDI has emerged as the interface of choice in personal computer systems, partly because it has been chosen by both IBM and Compaq for highest capacity/highest performance applications.

The low failure rate of current hard disk technology justifies a single ESDI high-capacity disk drive when storage requirements are high. For smaller requirements, the 28ms ST506 remains an adequate performer. When appropriate backup safeguards are employed, widely available rental equipment provides an adequate ability to return to interim full operation quickly.

One final caution is that capacity requirements are easily underestimated. Adding capacity to an installed system is considerably more expensive than it would have been if acquired in an initially configured computer system package. Acquiring hard disk capacity that is double the currently anticipated requirements is a reasonable approach to meeting the three year life span requirements of the system.

References:

1. A Look at Hard Drives and How to Unscramble the Meaning of MFM, RLL, ARLL, SCSI, and ESDI, Anita Amirrezvani, PC World, Feb '89, p197(3).

2. Tune Up Your Hard Disk, Winn L. Rosch, PC Computing, Dec '89, p116(4).

3. How it Works: The Hard Drive, Stephen Sagman, PC Computing, Sep '89, p226(2).

4. Buying Hard Disks By Mail, Henry Fersko-Weiss, PC Magazine, v8-Jun 27 '89, p95(15).

5. Removable Mass Storage: You CAN Take It With You, Mitt Jones, Stephanie K. Losee, Bill O'Brien, & Jennifer Zaino, PC Magazine, v7-Sep 13 '88, p213(22).

6. What's New In Optical Storage, Patrick Honan, Personal Computing, Feb '89, p111(6).

7. What's New In Storage Media, Patrick Honan, Personal Computing, Nov '88, p163(5).

8. Backup Devices & Software, Patrick Honan, Personal Computing, Feb '88, p169(7).

9. Tape Backup on a Budget, T.J. Byers, PC World, Feb '89, p110(7).

10. How to Prevent Hard Disk Disaster, Lewis Perdue, PC World, Feb '89, p186(2).

11. Removable Mass Storage, Randy Ross, PC Computing, Aug '89, p100(5).

12. Microscience HH-3120, T.J. Byers, PC World, Jan '89, p110(3).

13. The Winn Rosch Hardware Bible, Winn L. Rosch, 1989, Simon & Schuster, Inc.

Section 6

Video Displays

At first glance, the choice of a monitor for the computer system seems a simple one. A closer look reveals that it is anything but simple. The technological standards for displays are among the fastest moving and most confusing in a fast moving personal computer industry.

This section begins with a brief description of the factors that determine display quality followed by the chronology of standards since IBM introduced its PC in 1981. This discussion will be limited to the standards or near standards. Although many more combinations of resolution, color presentation, and/or shading exist among video boards and monitors, their effective use requires driver sets not commonly available for the major software packages. Next, important characteristics and available choices in video boards and monitors are highlighted. Many of these characteristics are subjective. There is no substitute for personal experience in the selection of this product. The prospective user should try the board/monitor combination in conditions approximating those in which they will be used before making a commitment. Nevertheless, these highlights will provide some guidance about what to seek. Finally, given the evolution of video standards, reasonable prospects for continued development in the immediate future, and relative costs, conclusions will be drawn and recommendations made about the most cost effective choices for the home-office environment.

The Basics

The sharpness of a color monitor's display is determined by three factors. The first factor, **bandwidth,** is the range of video frequencies that can be handled by the monitor. The video frequency determines how closely the display dots can be spaced on the screen. The bandwidth may be fixed or, in the case of multiscanning monitors, capable of displaying a range of frequencies.

Every pixel (light/color source) on the screen is made from three projected dots of color, sorted out by the shadow mask of the picture tube. The three primary colors (red, green, and blue) combine to produce the combination of displayed colors of which the monitor is capable. The second factor, **dot pitch,** is the measure of pixel spacing produced by the holes in the shadow mask. The smaller the dot pitch rating, the sharper the on-screen image the monitor can produce.

The third factor, **convergence,** is a measure of how precisely the colored dots making up each pixel converge on the same point. When the dots do not converge properly, the pixels are enlarged and the image is blurred.[2]

A Little History

When IBM introduced its PC in the fall of 1981, two display alternatives were offered. The color alternative was CGA (Color Graphics Adapter) and offered a palette of 16 colors and two modes of operation: 320 by 200 pixels of resolution with four colors displayed or 640 by 200 pixels with two colors displayed. The monochrome alternative was MDA (Monochrome Display Adapter) and it offered higher resolution (720 by 350 pixels), but it could only display text in a single color with no more than 25 lines of 80 characters each. By 1982 power users were using both adapters, MDA for text and CGA for the then-new Lotus 1-2-3 graphics.[2] Both CGA and MDA were TTL (digital) display adapters.

The Hercules Graphics Card was introduced in 1982. It was a TTL adapter capable of generating monochrome graphics with a resolution 720 by 348 pixels. Hercules became a de facto standard when Lotus supported it in release 1A of 1-2-3. It has the distinction of being the only non-IBM video standard in popular use.

IBM introduced the Enhanced Graphics Adapter (EGA) in 1984. When it first appeared, it was capable of 320 by 200 pixel resolution in 16 colors using a card with 64K of dynamic random access memory (DRAM). An optional 192K memory addition provided a 640 by 350 pixel resolution in 16 out of 64 color choices. The optional memory addition quickly became the standard configuration, even though it nearly doubled the price of the card. IBM also released the Professional Graphics Controller (PGC) at the same time as EGA. It was capable of 640 by 480 pixel resolution, but it was slow and expensive, and it failed to become a standard. While EGA was digital, PGC required analog signals.

In 1985, the first multiple frequency scanning monitor

(MultiSync) was introduced by NEC. It was capable of use with both digital and analog adapters and, because it was capable of displaying a range of scanning frequencies, it could keep pace with the evolving standards of video display signal generation without being obsoleted.

In 1987, IBM introduced the Video Graphics Array (**VGA**) with a color palette of 262,144 colors.[2] VGA provides 256 colors at 320 by 200 resolution or 16 colors at 640 by 480 resolution using an analog signal. The VGA text resolution is 720 by 400 pixels. It should be noted that although the VGA text resolution is higher than monochrome, the monochrome text image is sharper. This is because the VGA image is produced by three guns, one for each of the primary colors of the spectrum. No matter how good the monitor is, these three guns never line up perfectly, causing some fuzziness not found in a single gun monochrome monitor.[8] Originally, VGA was available only on IBM's PS/2 Micro Channel machines, but board makers were able to create cards to port the higher resolution to standard PC buses in less than a year.

A consortium of monitor and graphics board makers called VESA (Video Electronics Standards Association) got together to offer a higher resolution standard in 1988. It was called **Super VGA**, **Extended VGA**, and **VGA Plus** by the various vendors and is capable of 800 by 600 graphics resolution, with 56 percent more on-screen pixels than VGA. It uses 16 colors out of a palette of 256 and an analog signal.[15] Super VGA resolution is also available in monochrome with 64 shades of gray. The difference between VGA and Super VGA is more apparent than between EGA and VGA, and the VESA offering has begun to enjoy major software support, most notably from Microsoft.

Two major hardware players, IBM and Compaq, have declined to accept the Super VGA standard. Instead, they are offering products with a resolution of 1,024 by 768 pixels, but with a fundamental difference between their products.

The IBM product is the 8514A adapter, allowing simultaneous display of 16 colors using 500K of memory. It uses an 'interlaced' display, which means that the entire screen must be scanned twice to produce a single image, resulting in a discernable flicker. The Compaq product produces its high resolution with a noninterlaced or single scan display.

This situation presents board makers and software vendors with a dilemma. They can choose between support of a technologically inferior product or crossing swords with the mighty IBM. Neither choice is attractive. The boards are expensive and, given the doubt about their future, the result is that not much software exists to support either adapter. At present, Super VGA is the highest resolution enjoying widespread support and it appears to be on the way to becoming a standard.

A Question of Speed

As video technology developed, video signal generation shifted from digital to analog. Analog signals allowed a wide variety of colors or gray tones to be produced--an array of choices not available with the on-off digital signals. The ability to address more on-screen pixels in more colors requires more processing, more memory, and takes more time. Black and white 640 by 200 CGA requires 16K per screen; 16-color 640 by 350 EGA requires 109K; but, Super VGA's 16-color 800 by 600 needs 234K per screen.[1] This additional overhead is frustratingly apparent to users of graphic operating environments like Windows but, for graphics intensive applications like desktop publishing, graphic arts, and computer aided design (CAD), activities like redrawing or rotating images can be intolerably slow.

The original VGA cards used an 8-bit data path with DRAM memory, but the speed issue has led to evolution here too. Changes to meet this need include adding RAM to the board, using a different type of memory chip, and expanding the data path for video information.

One technique is the use of VRAM, a dedicated on-board video random access memory. It is about three times as fast as standard RAM (DRAM) because it uses a dual data port, where data moves to or from a source simultaneously. A VRAM board can send data to a monitor at the same time it receives data from a processor--something DRAM cannot do.[9] VRAM can also change a single pixel on a screen without redrawing other pixels that will remain the same; DRAM must redraw all the pixels in a one-byte segment whenever a single pixel is to be changed; then it must write the entire byte of information back to memory.

Some manufacturers have tried to get the benefits of the dual ported VRAM chip while using less expensive DRAM. The approach is to use a first in, first out (FIFO) buffer to process data like VRAM--moving data in one side and out the other, but using custom circuitry to manage the process so that it can be handled by standard DRAM chips. While sound in concept, tests of this process have not shown a discernable difference in speed between cards using the buffer approach and otherwise similar cards which did not.[1]

16-bit VGA cards can be used in 286 and 386 machines with 16-bit slots. The 16-bit data path can move information about twice as fast as the original 8-bit VGA cards. Tests comparing the performance of 16-bit DRAM and 16-bit VRAM cards show negligible speed differences between them.[1]

Another technique used by video display cards to increase the display speed is to relocate the video BIOS into RAM where it can perform faster than in its permanent storage location in ROM. Many 386 computers offer this BIOS 'shadowing' in RAM as part of their internal setup operations, which is redundant with BIOS relocation performed by these cards. The 386 relocation is actually superior in that it relocates the BIOS into extended memory rather than the system memory below 640K. (These drivers typically take up only 12K of memory and are not likely to cause RAM-cram by themselves. However, it is just one more demand on the valuable 640K real estate). It should also be noted that BIOS relocation has its most noticeable effect on 8-bit cards; tests have shown little measurable gain when relocation was performed with 16-bit cards.[1]

BIOS relocation does not help at all with programs that write directly to the display adapter hardware because they bypass the BIOS altogether. Programs like Lotus 1-2-3, Microsoft Windows, and many graphics programs do this to boost performance.

It should be noted that adding additional memory to the display card may increase the number of colors that can be simultaneously displayed on the screen but has no affect on the speed performance of the card.

Important Characteristics and Choices

The most significant choice to be made from an expense point of view is between a single-scanning and a multiscanning monitor. The digital monitors that provide CGA or EGA are single scanning as are analog VGA monitors. They are designed to work only within the fixed frequency ranges of their respective standards: 15.575KHz for CGA, 22.1KHz for EGA, or 31.5KHz for VGA.[2] The new Super VGA requires a horizontal scanning frequency slightly above 35KHz. Multiscanning monitors can have the advantage of forward and backward compatibility (and compatibility with both digital and analog input.) Note: Frequency ranges of multiscanning monitors vary--some multiscanning monitors do not cover the entire spectrum. A multiscanning monitor is more likely to work on a future video standard than a single-scanning one, although there is no guarantee. Changing to a new standard may require only the replacement of the video card--a considerably smaller expense. The old card can also serve as a backup in the event of a failure in the new one as well as protection against an unanticipated incompatibility between the new card and an old application program.

Single-scanning monitors do have two advantages: They tend to produce a better picture within a given standard, and they are generally less expensive than multiscanning monitors. The difference in picture quality is not great, and like price is attributable to higher complexity. It is easier and cheaper to do one thing well than to do many.[8]

The dot pitch of the monitor is an important indication of picture quality--it measures the distance between adjacent pixels. The greater the density of the pixels, the more vivid the display. Most 12-inch VGA monitors have a dot pitch of .28mm; most 14-inch VGA monitors have a dot pitch of .31mm. Anything lower than these numbers is exceptional.[8] (14-inch CGA monitors have a dot pitch of .41mm.)

The color convergence is determined by the alignment of the guns that project the three colors making up the display of a color monitor. That alignment can be affected by a rough ride from the factory to the point of sale. Ideally, the prospective monitor purchase should be 'test driven' before any money changes hands. Unfortunately, the sellers that can offer such security are not normally the most competitively priced. The price of this risk is another subjective choice.

The quality of the antiglare screen is an important factor in image quality. There are three techniques used to treat screen surfaces. The most effective (and most expensive) method of treatment is chemical etching. In the mid-range of cost and quality is coating the screen with a layer of silicon to reduce glare. The cheapest and least effective method is abrasive etching.

Flat screen design is also used along with a surface treatment to reduce glare. Deep bezels (the frame surrounding the screen) and tilt-and-swivel bases help keep surrounding light off the screen and reduce glare. These are subjective factors. To be properly evaluated, the prospective monitor should be tested under lighting conditions similar to those in which it will be used and with a software package that is familiar.

While the most common screen sizes supporting personal computers are 12- and 14-inch (measured diagonally as are TVs,) larger screen sizes are available. A large display will be much more expensive but, within a given resolution standard, will not look as sharp. The reason is that the same number of display dots is spread out over a wider display area. Achieving high resolution on a large display requires a higher pixel density, which must be generated by a non-standard adapter card requiring custom video drivers. Further, this additional display volume is processing overhead which slows the display process. Compensating for that is a graphics coprocessor, another expensive addition. The cost gap between 14-inch monitors and their larger alternatives makes the latter impractical for most home-office users.

Recommendations

The first consideration is a software compatibility constraint. More and more packages are being developed to run under a graphical environment resulting in a requirement of EGA resolution or better. Microsoft's Windows and Excel are significant examples, and this trend can reasonably be expected to accelerate. While support for the EGA standard by major software suppliers is expected to continue during this evolution, the primary market for EGA monitors themselves has all but dried up. Where EGA product is available, it is not offered at a significant price advantage compared to VGA monitors. Furthermore, economical bundled systems typically include a VGA monitor as part of the package, adding to the economic incentive. In terms of initial outlay, a VGA monitor is the least expensive safe alternative. It will be supported by the broadest assortment of important application programs over the 36 month decision window.

If graphics play a major role in your home-office environment, an alternate scenario should be considered. A multiscanning monitor combined with a versatile display card (such as ATI's VGA Wonder) can work together allowing you to utilize the highest resolution driver set in any given application program. One program may include a Super VGA driver, another may offer a 1024 by 768 resolution driver, while a third may only be capable of standard VGA; this combination can take advantage of all three. This choice requires a premium of about

50% additional cost for the card and monitor combined. Some of this may be recoverable as future standards change, if only a card change is required to keep pace.

There is a caution to be aware of in going the multiscanning route. The monitor should have an autosizing capability when moving from one display format to another. The display proportions that are appropriate to 640 by 480 VGA are different from those required for 800 by 600 Super VGA. Some multiscanning monitors require manual sizing adjustments. Without automatic size adjusting (autosizing), the monitor's size control switches have to be adjusted manually with each change of format--otherwise the images will appear distorted. And, monitors that offer autosizing vary in their ability to do the job well.

Finally, the importance of the test drive cannot be overemphasized, particularly if the display alternative you are considering is not part of a bundled package. Unanticipated incompatibilities can occur between the video package and the computer system it is connected to. These problems can usually be solved but the annoyance can be avoided altogether by careful pre-purchase evaluation. Use similar software under lighting conditions that approximate your own environment; look for sharpness of resolution, especially at the edges and corners; also look for true color representation. Taking the time and trouble up front can avoid some very unpleasant experiences.

References:

1. 16-bit VGA Cards Stretch The Standard, Alfred Poor, PC Magazine v8-Jul '89, p145(11).

2. Multiscanning Monitors for VGA and Beyond, Winn L. Rosch & Kellyn S. Betts, PC Magazine, v8-May 16 '89, p94(29).

3. Monster Monitors, Winn L. Rosch, PC Computing v1-Oct '88, p168(8).

4. VGA Boards--The Second Generation, T.J. Byers, PC World, Dec '88, p146(10).

5. VGA Monitors--Cheap and Smart, T.J. Byers, PC World, Dec '88, p154(7).

6. NEC Aims to Establish Extended VGA Standard, Eric Bender, PC World, Feb '89, p62(3).

7. VGA Update--Fast Operators, Jeff Holtzman, PC Computing, v2-May '89, p133(5)

8. Buyer's Guide: High Resolution Monitors, Charles Bermant & John Blackford, Personal Computing, May '88, p157(32).

9. What's New in Video Boards, Patrick Honan, Personal Computing, Sep '88, p121(5).

10. Nec Breaks New Ground with First Color Portable, Mitt Jones, PC Magazine, v8-Oct 31 '89, p33(3).

11. Software VGA, Lori Grunin, PC Magazine, v8-May 16 '89, p27.

12. VDTs: Are They Safe?, Randy Ross, PC Computing, Mar '89, p146(2).

13. Get the Picture? Nine Popular 14-Inch VGA (Analog) Monitors, Gary Lee Hendershot, Gregory Smith, & Lauren Black, InfoWorld, v11-Apr 17 '89, p55(16).

14. Beyond VGA, Bob Ponting, InfoWorld, v11-Jun 26 '89, p45(5).

15. VESA: Looking Forward to a New Standard, Bill Machrone, PC Magazine, v8-May 16 '89, p108.

16. Now Showing on the Big Screen..., Alfred Poor, PC Magazine, v9-Feb 13 '90, p161(31).

17. Seeing the White: Paper-White Phosphors, Alfred Poor, PC Magazine, v9-Feb 13 '90, p163.

18. Brains on Board: The Role of Graphics Coprocessors, Alfred Poor, PC Magazine, v9-Feb 13 '90, p180.

Section 7

Printers

Variety

There are five basic types of printers: daisy-wheel, impact dot-matrix, ink-jet, thermal-transfer, and laser printers. Each has different characteristics with respect to print quality, speed, graphics capability, variability of type sizes and fonts (type style), and cost.

Of the hardware choices to be made in the makeup of the core home-office system, this one is probably the one that must be the most individually determined. The uses to which the printer will be put will determine the appropriate choice for you. It may be that one printer type will not meet all your needs--and that you should install more than one printer in your proposed system.

A Little Background

The daisy-wheel printer is the old-timer. It uses a print-wheel shaped like a daisy with print characters at the end of its 'petals.' To print a character, a hammer strikes the appropriate petal to press the type-face against the ribbon and make an impression on the paper. The action of these printers is noisy and slow; the speed range is typically between 12 and 90 characters per second. The letter quality of the print is excellent, but daisy wheel printers are the only type incapable of graphics (although individual printers of other types may be incapable of generating graphics). The daisy-wheel printer has essentially been obsoleted by other technologies.

Dot-matrix printers use a print head with a series of small pins that strike an inked ribbon and create an image made up of tiny dots.[1] The pins of the print head are mounted in a tube, and typically number either 9 or 24. The 9-pin heads usually arrange all nine pins in a single vertical column, while 24-pin heads usually have three offset columns of eight pins each.[21] The quality of print is determined partly by the number of pins used to define each character, and partly by the number of impressions. Both 9-pin and 24-pin printers can be operated in a draft mode (lower print quality, faster speed) or a near letter quality mode (NLQ), in which the print head makes a double impression of each letter. Speeds range from 25 to 240 characters per second in NLQ mode and 80 to 500 cps in draft mode. Either way, this is a noisy technology. Although dot-matrix printers cannot achieve perfect letter quality print definition, 24-pin printers can come close. Ink-jet printers create an image by firing dots of heated ink directly onto the paper. Ink-jet printers typically operate in the lower half of the impact dot-matrix speed range and can produce good to excellent print resolution. And they are quiet. Some ink-jet printers require the use of specially coated paper to avoid a 'wicking' effect (ink spreading through the fibers in the paper) although some can use high-grade photocopier or cotton-bond paper.[2] During the late 1980's ink-jet printers began to develop momentum as a relatively inexpensive alternative to laser printers (250 dots per inch compared to 300 dpi typically from a laser printer). There is some doubt that momentum will continue, given the fact that desk-top lasers are now available for under $1,000.

Thermal-transfer printers first appeared as a low-cost alternative to impact dot-matrix printers. They produce images on paper by firing small dots of heat at a heat-sensitive ribbon, or directly onto heat-sensitive paper. There are quiet, high quality thermal printers that can produce laser-quality resolution of as much as 350 dots per inch. Their future is not bright, because laser printers produce a very high quality resolution at a much faster speed. A very specialized niche remains available to thermal printers: the small, battery-powered printer used to support the rapidly growing laptop segment of the computer business.

The most powerful player on the printer scene today is the laser printer. It is by far the fastest of the printer types; it has very good print resolution; and, perhaps most importantly to the home-office user, prices have become affordable (under $1,000). The technology is similar in principle to a photocopier; a laser beam draws the image on a drum, creating an electrostatic charge that causes toner particles to adhere to it.[1] These are then offset onto paper. Most laser printers have a resolution of 300 dots per inch, but some are as high as 600 dpi.[15] They typically operate at speeds from 4- to 15-ppm (pages per minute.) Laser printers are virtually silent and have great flexibility in type-sizes, fonts, and graphics capability.

A Shakeout in Progress

The technology and pricing dynamics of the printer industry, at least that part of the industry that supports personal computers has reduced the viable choices to two types of printers: impact dot-matrix and laser.

Versatility and performance at a low price make dot-matrix printers appealing. Interestingly, the price erosion in laser printers is largely responsible for the strength of this position. Dot-matrix printers are the largest population of printers supporting PCs. As laser prices have declined, vendors of dot-matrix printers have also decreased prices to maintain the viability of the product. The resulting squeeze is forcing 9-pin dot-matrix printers into a very small entry-level position ($200 and under), with 24-pin dot-matrix dominating the mid-range up to $700, and laser occupying the upper tier. Ink-jet and high-end thermal printers compete with laser printers functionally--quiet, high-resolution, superior graphics--but, they lose on speed and cannot compete enough on price to make the difference.

The Right Tool for the Job

Dot-Matrix Printers

The dot-matrix printers offer versatility and flexibility. Wide-carriage models are available for maximum paper-handling range. They can be used to print any size documents from business cards, checks, and envelopes to standard 8 1/2 x 11 letters and 11 x 14 spreadsheets. Modern paper-handling features can make it relatively easy to 'park' two or three different forms in position ready to feed into the printer depending on the current application and/or allowing single sheets to be printed without reinstalling fan-fold paper.

Printer emulation is an important factor. Application software programs include drivers that send instructions to the printer, determining factors such as type size, line feed, and font. There are no formal standards for these codes; most software packages include those drivers required by the printers that are most significant in the marketplace. The Epson family and the IBM Proprinter are common to most software packages for that reason.[5] Printers that are not as common assure users of their universal compatibility with the ability to emulate a number of common models. Having a printer that cannot emulate a common model can lead to trouble.

Dot-matrix printers are typically capable of friction- or tractor-feed. Friction-feed is the same as on a typewriter; tractor-feed uses a sprocket to pull or push pin-feed computer paper past the print head. A pull-tractor grips the paper after it leaves the platen; a push-tractor pushes the paper through. Some printers come with push/pull tractors that can do either. It is worth mentioning, because pull-tractors tend to waste more paper, but they are more efficient for graphics and high-speed printing.[5]

A related consideration is the paper input feed path. Fan-fold paper may be fed into the printer in three basic ways: from the front, the back, or the bottom. A printer may be capable of only one path, two, or all three. The best path depends on how the home-office has been set up; the importance of multiple path options depends on how permanent that arrangement is expected to be.

The printer will have a buffer which is a reservoir of memory used to store the document being sent from the computer. When the document to be printed has been unloaded entirely into the buffer, the computer is free to perform other tasks. Most printers have an expandable buffer. For a relative perspective of the buffer size, think of a full single-spaced page of text as about 2,000 characters or 2K per page. (A 10K buffer would be expected to hold a full five page text document during the printing process.) Graphics jobs are typically much larger, since they require enough memory to address the individual dots on a page field.

Three factors affect the importance of the buffer size: print job size, job composition, and job scheduling. If print jobs are long, buffer size becomes an important factor because as long as the print job does not fit into the buffer, the computer is tied up. The graphic composition of print jobs adds exponentially to the memory requirement, because each dot in the print field takes space in memory. Job scheduling requirements determine the relative importance of the buffer size to your individual requirements. If long print jobs can be postponed until the end of the work day, they can be programmed to print sequentially when computer time is not required for other activities.

When long print jobs cannot be postponed, the buffer size becomes important. The average and peak job size should be anticipated using your typical text/graphic volume to determine the appropriate buffer requirement. There are also alternatives to the printer's own buffer capacity that accomplish the same objective as the buffer--freeing the computer for other tasks. Print spooling software that addresses expanded memory can be used to create an intermediate storage destination that will feed the print job to the printer. A print spooler can also be an outboard accessory device installed between the computer and the printer.

Another consideration that varies in importance according to individual need, is the availability of additional fonts. A font is a type style; examples of different styles include Roman, Sans Serif, and Gothic. Some printers offer a number of resident fonts in their standard configuration. Additional fonts can sometimes be added to printers, either in the form of cartridges or software downloaded into the printer's memory. The presence of alternative fonts allows the printer to produce, from text input, output in a number of type styles. Graphics software may also address the printer's graphics capability to produce varied type styles. But, doing it graphically requires addressing individual dots with a consequent overhead in memory and time.

The location of controls on printers varies widely and their accessibility in the home-office environment can be important. These include controls for on/off line, form/line feed, print density, form size, impact level (single or multi-part forms), access to font slots, etc.) Most of the current crop of computer printers have their control panels mounted in front, responding to competitive pressures. Nevertheless, personal tastes vary as to how this end is most effectively achieved and it is a consideration to keep in mind.

Finally, a small population of dot-matrix impact printers offers color printing capability. Early attempts at color printing were crude and cumbersome, but the current crop is capable of a high quality result. Dot-matrix color printing is more economical than alternatives in ink-jet and wax thermal printers. It costs from two to five cents to produce a page on a color dot-matrix printer, seven to 14 cents on an ink-jet printer, and 35 cents on a thermal printer.[4] Color printers, comparable in quality to their monochrome counterparts, cost two to four times as much.[11] Color ink-jet and dot-matrix printers are similar in cost. High quality wax thermal printers produce the best quality color but at about two and a half times the price. Transparency film is now available which is suited to the properties of dot-matrix printer ribbon ink--a problem in the past. Major word processing packages such as WordPerfect and Microsoft Word now have color printing options.

individual dots with a consequent overhead in memory and time.

The location of controls on printers varies widely and their accessibility in the home-office environment can be important. These include controls for on/off line, form/line feed, print density, form size, impact level (single or multi-part forms), access to font slots, etc.) Most of the current crop of computer printers have their control panels mounted in front, responding to competitive pressures. Nevertheless, personal tastes vary as to how this end is most effectively achieved and it is a consideration to keep in mind.

Finally, a small population of dot-matrix impact printers offers color printing capability. Early attempts at color printing were crude and cumbersome, but the current crop is capable of a high quality result. Dot-matrix color printing is more economical than alternatives in ink-jet and wax thermal printers. It costs from two to five cents to produce a page on a color dot-matrix printer, seven to 14 cents on an ink-jet printer, and 35 cents on a thermal printer.[4] Color printers, comparable in quality to their monochrome counterparts, cost two to four times as much.[11] Color ink-jet and dot-matrix printers are similar in cost. High quality wax thermal printers produce the best quality color but at about two and a half times the price. Transparency film is now available which is suited to the properties of dot-matrix printer ribbon ink--a problem in the past. Major word processing packages such as WordPerfect and Microsoft Word now have color printing options.

Ink-jet models have had problems with clogged jets and wax thermals have had problems with wax bleeding slightly during the color transfer process. Dot-matrix printers have been criticized for producing colors that lacked the depth and richness of ink-jet and thermal, requiring larger font sizes to compensate. New technologies are being developed to solved these problems. Meanwhile, color laser products have arrived on the scene--at very high prices.

If color constitutes a major part of the intended home-office applications, those particular applications should be carefully considered before choosing the technology better suited to them. But, if color is an ancillary feature, dot-matrix dominates the choices on the bases of economy and versatility of applications.

Laser Printers

This is the elite end of the printer spectrum. Until very recently, laser printers were prohibitively priced for most home-office users. That is no longer the case, and the biggest names in printers are leading the charge to lower prices: laser leaders, Hewlett-Packard and Canon, and dot-matrix giant, Epson. These companies now offer entry level units that list for about $1,500 and are available through various outlets for about $1,000. Lesser known offerings can be had for even less, and prices have not stopped moving. Why the excitement?

24-pin dot-matrix printers took the business away from daisy-wheel printers because of higher print speed and a graphical printing capability. They did not have the letter quality of daisy-wheels, but they came close. A 24-pin printer can produce one to two pages of text per minute in its high-quality print mode. The inexpensive lasers mentioned above produce four to six pages per minute in true letter quality.[24] They are also capable of graphics and, unlike dot-matrix printers, they are quiet.

It should be noted that their standard 512K memory is not enough to maximize their resolution quality of 300 dpi--an 8 by 10 inch printable area holds 879K dots.[24] This high resolution is necessary to desktop publishing activities, but spreadsheet chart graphics and the like can be very satisfactorily reproduced at 150 dpi resolution.

One of the current crop of entry level laser printers in its standard configuration can hold its own in a cost benefit comparison with a high quality 24-pin dot matrix printer. The fact that the laser is quiet, produces text output at two to three times the dot-matrix speed, and has true letter quality print resolution are three factors that are significant. Their worth and weight in the individual home-office environment will vary.

Printer emulation is a crucial factor to be considered. The standard in laser printers, since its introduction in 1987, has been Hewlett-Packard's LaserJet Series II. Most software publishers include a LaserJet driver, and even IBM features

LaserJet compatibility when it releases a new laser printer.[24] An inexpensive laser, out of the box, can do a great job with word processing or spreadsheets and with accompanying charts including sideways or landscape orientation. This may be plenty, but to take the laser printer to its desktop publishing limits, expensive accessories must be added.

PostScript

Hewlett-Packard's Printer Command Language (PCL) is the LaserJet command set and while it sets the standard, it falls short of the state of the art of which these entry level lasers are capable. Adobe Systems Inc.'s PostScript page description language has been dominant in the Apple Macintosh world--the historical world of choice for desktop publishing. PostScript supports scalable fonts (type styles that can be reduced or enlarged to fit any size); PCL printers have bitmapped (fixed size) fonts, with font changes available only through software downloads and/or plug-in cartridges. Addition of a PostScript emulation cartridge plus a memory upgrade of an additional 2-4MB to allow 300 dpi resolution can bring the printer up to its potential: high resolution scalable fonts, forms, graphics, scanned images, and desktop publishing with near-typeset quality output. Such a package seems expensive; a Hewlett-Packard LaserJet IIP with PostScript emulation cartridge and 2MB memory upgrade can be put together for $2,000.[25] But, an equivalent Apple LaserWriter IINT lists for $4,999.

Recommendations

The choice boils down to one of economy and versatility, favoring the 24-pin dot-matrix or speed and presentation quality favoring the laser.

The dot-matrix printer is adequate for most business correspondence, and is flexible enough to address labels and envelopes, print checks, and generate over-sized spreadsheets. It may even be capable of parking two or three different types of forms in a ready position while another is in use. The dot-matrix can also print on multi-part forms. It is noisy and it does not produce the best quality output, but it can do everything and do everything passably well.

Where presentation-quality output is needed, the laser printer is the tool for the job. It will produce very high quality output at high speed and one can still use the phone while it is printing. But, because it is limited to letter and/or legal sized paper, it will not do it all--a second device will be required.

If mailing activity is limited and the computer is not used to generate non-standard forms (multi-part checks, labels, etc.), an electronic typewriter is an option; it may be a particularly appropriate choice if many different jobs are done for short periods of time, requiring frequent changes. A very few typewriters exist with a built in parallel/centronics interfaces so they can be connected to a personal computer. The typical emulation in such typewriters is Diablo 630, a standard daisy-wheel printer emulation.

A better solution is a 24-pin dot-matrix printer complemented by software that allows it to be addressed as if it were a typewriter. This capability allows single items of various sizes to be friction-fed and printed on with the print head moved one space or line at a time. An inexpensive two position centronics switch can be used to gate the output to the appropriate device--either the laser printer or the dot-matrix printer. Or, a second parallel interface board can be installed in an open slot in the computer and identified to software as LPT2. (External logic-sensing switches are also available, but they are prohibitively expensive for this application.) Using a 24-pin dot-matrix printer as the second device has the additional virtue of being able to serve as a backup when the laser requires repair or maintenance.

References:

1. Printers, Corinne De Palma, 1988 INC. Office Guide, Jul '88, p64.

2. Ink-Jet Challenges Laser, 24-Pin Technologies, Mike Hildum & Sally J. Douglas, InfoWorld, v10-Jun 27 '88, p58(2).

3. What's New in Color Printers and Plotters, Patrick Honan, Personal Computing, Apr '89, p107(8).

4. Buyer's Guide: Laser Printers, John Bell, Emily Berk, & Joseph Devlin, Personal Computing, v12-Dec '88, p153(36).

5. What's New in Dot Matrix Printers, Barbara Krasnoff, Personal Computing, v12-Dec '88, p143(6).

6. Who Needs a Laser Printer?, Ron Bel Bruno, Personal Computing, Oct '88, p71.

7. What's New in Laser Printers, Michael Antonoff, Personal Computing, Feb '88, p157(8).

8. A Look at the New Laser Printers, Ted Needleman, Computers in Accounting, v4-Dec '88, p85(13).

9. What's New in Impact and Nonimpact Office Printers, Rick Friedman, The Office, May '89, p75(3).

10. Unlock Your Printer's Power, Miriam Liskin, Personal Computing, v12-Feb '88, p61(6).

11. The Rainbow Connection: Four Color Printers, Phillip Robinson, PC Computing, Jun '89, p84(10).

12. The Elements of Type Style, Edward Mendelson, PC Magazine, June 13 '89, p206(18).

13. The 1989 Printer Roundup, John Dickinson, PC Magazine, v8-Nov 14 '89, p95(15).

14. The Best Output in the Land, Mary Kathleen Flynn, PC Magazine, v8-Nov 14 '89, p98(2).

15. Laser Printers: A Broader Spectrun, Alfred Poor, PC Magazine, v8-Nov 14 '89, p115(39).

16. How to Get the Most Out of Your Laser Printer, Edward Mendelson, PC Magazine, v8-Nov 14 '89, p138(2).

17. Printer Languages: The Next Generation, M. David Stone, PC Magazine, v8-Nov 14 '89, p152(2).

18. Lights! Camera! Print It! Laser Printer Technology Explained, Alfred Poor, PC Magazine, v8-Nov 14 '89, p168(2).

19. Color Page Printers: Affordable At Last, Gerard Kunkel, PC Magazine, v8-Nov 14 '89, p215(18).

20. Monochrome Dot Mtrix Printers: Not Dead Yet, Stephanie K. Losee, PC Magazine, v8-Nov 14 '89, p245(47).

21. Dot's Good Enough For Me! Dot Matrix Technology Explained, Alfred Poor, PC Magazine, v8-Nov 14 '89, p264(2).

22. Portable Printers: On-The-Road Output, Steve Lewiski, PC Magazine, v8-Nov 14 '89, p325(6).

23. Color Dot Matrix Printers: The Inexpensive Path to Color, Bruce Brown, PC Magazine, v8-Nov 14 '89, p343(16).

24. Lasers At Bread-and-Butter Prices, Eric Grevstad, PC Resource, Feb '90, p52(9).

25. Half-Price PostScript Quality, Peggy Wallace, PC Computing, Feb '90, p38.

Section 8

Conclusion

The recommended system resulting from this study is based on an 80386 microprocessor and runs under the DOS operating system. This combination is actually more powerful than the currently available OS/2 environment in its current state of development. For users who desire a graphical operating environment, Microsoft Windows 386 provides most of the advantages of OS/2 with half the minimum memory requirement. The available library of DOS applications is large and inexpensive relative to OS/2 applications.

OS/2 is currently limited by its design in support of the 80286 microprocessor. Improvements in OS/2 will allow full utilization of 80386 memory management capabilities but will also include the capability of managing multiple DOS sessions. DOS applications will therefore have a future in either an OS/2 or a Unix operating environment.

A two to four megabyte onboard memory configuration provides the system with fast access to multiple applications in either a character or graphical environment. An ESDI high-capacity hard drive provides fast access to a large and growing volume of data in secondary storage. A VGA monitor provides the most cost-effective display of increasingly graphically oriented programs. The choice of 24-pin dot matrix, laser printer, or both is user application dependent; but, the home-office user will most likely find the most power per dollar expended among these choices.

Purchasing a completely configured system (excluding the printer) at the time of initial acquisition is more economical than selecting individual components separately and numerous options are readily available from credible sources--both mail-order and store-front.

It must be recognized and accepted that the costs of the recommended system will continue to decline; but, except for the speed with which price changes are happening, this has always been true. Postponement of the decision to acquire until the market stabilizes cannot be considered a viable strategy. Because of the rapid evolution of technology, the rate of change is more likely to increase than to decrease in the foreseeable future.

Some comfort can be taken from the fact that a system like the one recommended here was far beyond the reach of a single user five years ago; ten years ago, it would have required a large, environmentally controlled room for its installation. The evolution of personal computer technology and the power it has put under the control of the individual user has been largely responsible for making the home-office a widespread reality.