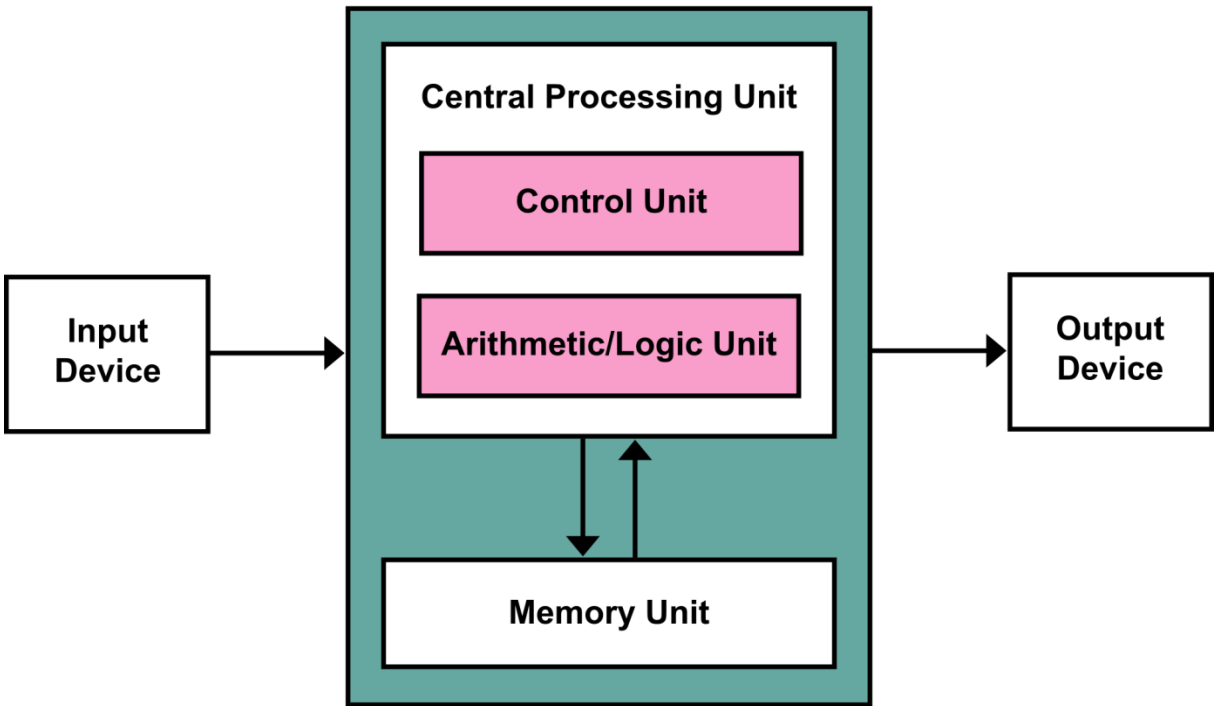
1 COMPUTER HARDWARE

The computer can be described as a device that is capable of performing the requested data transformation. From other simpler calculating machines, the computer differs by high speed performance, ability to utilize internal memory and work according to the program stored in the memory.
The revolutionary idea that computers could be controlled by program invented Charles ***Babbage.*** The specific method of implementation of this idea was developed gradually. The first idea was that the computer program is recorded on some external recording medium (punched tape, punch card, etc.) and the base unit sequentially reads, decodes and immediately executes the program. This approach leads to variety of problems, for example slow and difficult calling of subroutines, the need to rewind the tape when jump commands in the program called another place, etc. This was caused by the sequential memory type used to store the program.
Later, ***John von Neumann*** proposed to store a program in such a way that the entire program is permanently available. He suggested loading the program into a memory that is not sequential, but behaves like a random access memory. This type of computer is called ***computer with internal management*** that means computer memory stores not only data that are processed, also the program that manages the data processing. An important idea which prevails in within the von Neumann architecture is the principle that the computer should not be adapted to specific application needs by internal structure, but only programs. This means that the internal structure of the computer should not be changed and should be versatile in order to fulfill a wide variety of applications. All activities associated with the specific problem solution should be accomplished at the programing level. Von Neumann's concept is designed to be universal and it could be programmed for virtually everything.

1.1 Personal computers

The term "personal computer" was introduced in 1981 by IBM, marking the microcomputer ***PC (Personal Computer).*** These computers belong to a broader group of microcomputers. Their architecture is based on the ***Von Neumann concept,*** which is a computer with an internal management. A diagram of Von Neumann's concept computer represents Figure 3.1.
The main part that contains "control unit", "ALU" (ALU - *Arithmetic-Logic Unit)* and registers is called ***the processor,*** or **C**entral **P**rocessing **U**nit, in short **CPU.**
Microprocessor is a word consisting of two parts, *micro* and *processor,* which indicates that the processor as defined above is made utilizing miniaturization of the electronic circuits, microprocessor is an integrated circuit with a high degree of integration ***(VLSI*** - *Very Large Scale Integration,* ***ELSI*** - *Extra Large Scale Integration).* The microprocessor is constructed to communicate with the memory using memory bus and devices using input-output interface that is located usually outside the chip. The microprocessor in connection with memory ***(RWM*** - *Read Write Memory,* ***ROM*** - *Read Only Memory),* input-output interface and other auxiliary circuit (clock generator, power supply) is a ***microcomputer.***

Figure 3.1 Von Neumann concept



Microprocessors opened up new possibilities in the design and construction of digital circuits and systems. Initially, the microprocessor is used as one of the subsystems in the terminals, calculating and communication devices, but in 1974, two years after its production, there was a real boom in the use of microprocessors. Number of produced microprocessors exceeded the sum of all existing minicomputers, medium and large scale computers. Microprocessors have changed design and functionality of traditional digital systems because of its low cost, great flexibility and reliability. The main area of microcomputer utilization is computing.

3.2 Short overview of the development of microprocessors

The first use of the term "microprocessor" is attributed to Viatron Computer Systems describing the custom integrated circuit used in their System 21 small computer system announced in 1968.

Intel introduced its first 4-bit microprocessor 4004 in 1971 and its 8-bit microprocessor 8008 in 1972. During the 1960s, computer processors were constructed out of small and medium-scale ICs—each containing from tens of transistors to a few hundred. These were placed and soldered onto printed circuit boards, and often multiple boards were interconnected in a chassis. The large number of discrete logic gates used more electrical power—and therefore produced more heat—than a more integrated design with fewer ICs. The distance that signals had to travel between ICs on the boards limited a computer's operating speed.

In the NASA Apollo space missions to the moon in the 1960s and 1970s, all onboard computations for primary guidance, navigation and control were provided by a small

custom processor called "The Apollo Guidance Computer". It used wire wrap circuit boards whose only logic elements were three-input NOR gates.

The first microprocessors emerged in the early 1970s and were used for electronic calculators, using binary-coded decimal (BCD) arithmetic on 4-bit words. Other embedded uses of 4-bit and 8-bit microprocessors, such as terminals, printers, various kinds of automation etc., followed soon after. Affordable 8-bit microprocessors with 16-bit addressing also led to the first general-purpose microcomputers from the mid-1970s on.

Since the early 1970s, the increase in capacity of microprocessors has followed Moore's law; this originally suggested that the number of components that can be fitted onto a chip doubles every year. With present technology, it is actually every two years, and as such Moore later changed the period to two years.

3.2.1 Firsts

Three projects delivered a microprocessor at about the same time: Garrett AiResearch's Central Air Data Computer (CADC), Texas Instruments (TI) TMS 1000 (1971 September), and Intel's 4004 (1971 November).

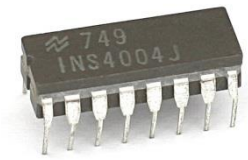
CADC

In 1968, Garrett AiResearch (which employed designers Ray Holt and Steve Geller) was invited to produce a digital computer to compete with electromechanical systems then under development for the main flight control computer in the US Navy's new F-14 Tomcat fighter. The design was complete by 1970, and used a MOS-based chipset as the core CPU. The design was significantly (approximately 20 times) smaller and much more reliable than the mechanical systems it competed against, and was used in all of the early Tomcat models. This system contained "a 20-bit, pipelined, parallel multi-microprocessor". The Navy refused to allow publication of the design until 1997. For this reason the CADC, and the MP944 chipset it used, are fairly unknown. Ray Holt graduated California Polytechnic University in 1968, and began his computer design career with the CADC. From its inception, it was shrouded in secrecy until 1998 when at Holt's request, the US Navy allowed the documents into the public domain. Since then people have debated whether this was the first microprocessor. Holt has stated that no one has compared this microprocessor with those that came later. According to Parab et al. (2007), "The scientific papers and literature published around 1971 reveal that the MP944 digital processor used for the F-14 Tomcat aircraft of the US Navy qualifies as the first microprocessor. Although interesting, it was not a single-chip processor, as was not the Intel 4004 – they both were more like a set of parallel building blocks you could use to make a general-purpose form. It contains a CPU, RAM, ROM, and two other support chips like the Intel 4004. It was made from the same P-channel technology, operated at military specifications and had larger chips -- an excellent computer engineering design by any standards. Its design indicates a major advance over Intel, and two year earlier. It actually worked and was flying in the F-14 when the Intel 4004

was announced. It indicates that today's industry theme of converging DSP-microcontroller architectures was started in 1971." This convergence of DSP and microcontroller architectures is known as a digital signal controller.

Intel 4004

The Intel 4004 ("four-thousand-four") is a 4-bit central processing unit (CPU) released by Intel Corporation in 1971. It was the first microprocessor as well as the first general purpose programmable microprocessor on the market.[1] The chip design started in April 1970, when Federico Faggin joined Intel, and it was completed under his leadership in January 1971. The first commercial sale of the fully operational 4004 occurred in March 1971 to Busicom Corp. of Japan for which it was originally designed and built as a custom chip.[2] In mid-November of the same year, with the prophetic ad "Announcing a new era in integrated electronics", the 4004 was made commercially available to the general market. The 4004 was history's first monolithic CPU fully integrated in one small chip. Such a feat of integration was made possible by the use of the then-new silicon gate technology which allowed twice the number of random-logic transistors and an increase in speed by a factor of five compared to the incumbent technology. The 4004 microprocessor was one of 4 chips constituting the MCS-4 chip-set, which included the 4001 ROM, 4002 RAM, and 4003 Shift Register. With these components, small computers with varying amounts of memory and I/O facilities could be built.



3.2.2 8-bit designs

The Intel 4004 was followed in 1972 by the Intel 8008, the world's first 8-bit microprocessor. The 8008 was not, however, an extension of the 4004 design, but instead the culmination of a separate design project at Intel, arising from a contract with Computer Terminals Corporation, of San Antonio TX, for a chip for a terminal they were designing,[32] the Datapoint 2200 — fundamental aspects of the design came not from Intel but from CTC. In 1968, CTC's Vic Poor and Harry Pyle developed the original design for the instruction set and operation of the processor. In 1969, CTC contracted two companies, Intel and Texas Instruments, to make a single-chip implementation, known as the CTC 1201.[33] In late 1970 or early 1971, TI dropped out being unable to make a reliable part. In 1970, with Intel yet to deliver the part, CTC opted to use their own implementation in the Datapoint 2200, using traditional TTL logic instead (thus the first machine to run "8008 code" was not in fact a microprocessor at all and was delivered a year earlier). Intel's version of the 1201 microprocessor arrived in late 1971, but was too late, slow, and required a number of additional support chips. CTC had no interest in using it. CTC had originally contracted Intel for the chip, and would have owed them $50,000 for their design work.[33] To avoid paying for a chip they did not want (and could not use), CTC released Intel from their contract and allowed them free use of the design.[33] Intel marketed it as the 8008 in April, 1972, as the

world's first 8-bit microprocessor. It was the basis for the famous "Mark-8" computer kit advertised in the magazine Radio-Electronics in 1974. This processor had an 8-bit data bus and a 14-bit address bus.[34]

The 8008 was the precursor to the very successful Intel 8080 (1974), which offered much improved performance over the 8008 and required fewer support chips, Zilog Z80 (1976), and derivative Intel 8-bit processors. The competing Motorola 6800 was released August 1974 and the similar MOS Technology 6502 in 1975 (both designed largely by the same people). The 6502 family rivaled the Z80 in popularity during the 1980s.

A low overall cost, small packaging, simple computer bus requirements, and sometimes the integration of extra circuitry (e.g. the Z80's built-in memory refresh circuitry) allowed the home computer "revolution" to accelerate sharply in the early 1980s. This delivered such inexpensive machines as the Sinclair ZX-81, which sold for US$99. A variation of the 6502, the MOS Technology 6510 was used in the Commodore 64 and yet another variant, the 8502, powered the Commodore 128.

3.2.3 16-bit designs

The first multi-chip 16-bit microprocessor was the National Semiconductor IMP-16, introduced in early 1973. An 8-bit version of the chipset was introduced in 1974 as the IMP-8.

Other early multi-chip 16-bit microprocessors include one that Digital Equipment Corporation (DEC) used in the LSI-11 OEM board set and the packaged PDP 11/03 minicomputer—and the Fairchild Semiconductor MicroFlame 9440, both introduced in 1975–1976. In 1975, National introduced the first 16-bit single-chip microprocessor, the National Semiconductor PACE, which was later followed by an NMOS version, the INS8900.

Another early single-chip 16-bit microprocessor was TI's TMS 9900, which was also compatible with their TI-990 line of minicomputers. The 9900 was used in the TI 990/4 minicomputer, the TI-99/4A home computer, and the TM990 line of OEM microcomputer boards. The chip was packaged in a large ceramic 64-pin DIP package, while most 8-bit microprocessors such as the Intel 8080 used the more common, smaller, and less expensive plastic 40-pin DIP. A follow-on chip, the TMS 9980, was designed to compete with the Intel 8080, had the full TI 990 16-bit instruction set, used a plastic 40-pin package, moved data 8 bits at a time, but could only address 16 KB. A third chip, the TMS 9995, was a new design. The family later expanded to include the 99105 and 99110.

The Western Design Center (WDC) introduced the CMOS 65816 16-bit upgrade of the WDC CMOS 65C02 in 1984. The 65816 16-bit microprocessor was the core of the Apple IIgs and later the Super Nintendo Entertainment System, making it one of the most popular 16-bit designs of all time.

Intel "upsized" their 8080 design into the 16-bit Intel 8086, the first member of the x86 family, which powers most modern PC type computers. Intel introduced the 8086 as a cost-effective way of porting software from the 8080 lines, and succeeded in winning much business on that premise. The 8088, a version of the 8086 that used an 8-bit external data bus, was the microprocessor in the first IBM PC. Intel then released the 80186 and 80188, the 80286 and, in 1985, the 32-bit 80386, cementing their PC market dominance with the processor family's backwards compatibility. The 80186 and 80188 were essentially versions of the 8086 and 8088, enhanced with some onboard peripherals and a few new instructions. Although Intel's 80186 and 80188 were not used in IBM PC type designs, second source versions from NEC, the V20 and V30 frequently were. The 8086 and successors had an innovative but limited method of memory segmentation, while the 80286 introduced a full-featured segmented memory management unit (MMU). The 80386 introduced a flat 32-bit memory model with paged memory management.

The 16-bit Intel x86 processors up to and including the 80386 do not include floating-point units (FPUs). Intel introduced the 8087, 80187, 80287 and 80387 math coprocessors to add hardware floating-point and transcendental function capabilities to the 8086 through 80386 CPUs. The 8087 works with the 8086/8088 and 80186/80188,[35] the 80187 works with the 80186 but not the 80188,[36] the 80287 works with the 80286 and the 80387 works with the 80386. The combination of an x86 CPU and an x87 coprocessor forms a single multi-chip microprocessor; the two chips are programmed as a unit using a single integrated instruction set.[37] The 8087 and 80187 coprocessors are connected in parallel with the data and address buses of their parent processor and directly execute instructions intended for them. The 80287 and 80387 coprocessors are interfaced to the CPU through I/O ports in the CPU's address space, this is transparent to the program, which does not need to know about or access these I/O ports directly; the program accesses the coprocessor and its registers through normal instruction opcodes.

3.2.4 16-bit designs

16-bit designs had only been on the market briefly when 32-bit implementations started to appear.

The most significant of the 32-bit designs is the Motorola MC68000, introduced in 1979.[dubious – discuss] The 68k, as it was widely known, had 32-bit registers in its programming model but used 16-bit internal data paths, three 16-bit Arithmetic Logic Units, and a 16-bit external data bus (to reduce pin count), and externally supported only 24-bit addresses (internally it worked with full 32 bit addresses). In PC-based IBM-compatible mainframes the MC68000 internal microcode was modified to emulate the 32-bit System/370 IBM mainframe.[38] Motorola generally described it as a 16-bit processor, though it clearly has 32-bit capable architecture. The combination of high performance, large (16 megabytes or 224 bytes) memory space and fairly low cost made it the most popular CPU design of its class. The Apple Lisa and Macintosh designs made

use of the 68000, as did a host of other designs in the mid-1980s, including the Atari ST and Commodore Amiga.

The world's first single-chip fully 32-bit microprocessor, with 32-bit data paths, 32-bit buses, and 32-bit addresses, was the AT&T Bell Labs BELLMAC-32A, with first samples in 1980, and general production in 1982.[39][40] After the divestiture of AT&T in 1984, it was renamed the WE 32000 (WE for Western Electric), and had two follow-on generations, the WE 32100 and WE 32200. These microprocessors were used in the AT&T 3B5 and 3B15 minicomputers; in the 3B2, the world's first desktop super microcomputer; in the "Companion", the world's first 32-bit laptop computer; and in "Alexander", the world's first book-sized super microcomputer, featuring ROM-pack memory cartridges similar to today's gaming consoles. All these systems ran the UNIX System V operating system.

The first commercial, single chip, fully 32-bit microprocessor available on the market was the HP FOCUS.

Intel's first 32-bit microprocessor was the iAPX 432, which was introduced in 1981 but was not a commercial success. It had an advanced capability-based object-oriented architecture, but poor performance compared to contemporary architectures such as Intel's own 80286 (introduced 1982), which was almost four times as fast on typical benchmark tests. However, the results for the iAPX432 was partly due to a rushed and therefore suboptimal Ada compiler.[citation needed]

Motorola's success with the 68000 led to the MC68010, which added virtual memory support. The MC68020, introduced in 1984 added full 32-bit data and address buses. The 68020 became hugely popular in the Unix supermicrocomputer market, and many small companies (e.g., Altos, Charles River Data Systems, Cromemco) produced desktop-size systems. The MC68030 was introduced next, improving upon the previous design by integrating the MMU into the chip. The continued success led to the MC68040, which included an FPU for better math performance. A 68050 failed to achieve its performance goals and was not released, and the follow-up MC68060 was released into a market saturated by much faster RISC designs. The 68k family faded from the desktop in the early 1990s.

Other large companies designed the 68020 and follow-ons into embedded equipment. At one point, there were more 68020s in embedded equipment than there were Intel Pentiums in PCs.[41] The ColdFire processor cores are derivatives of the venerable 68020.

During this time (early to mid-1980s), National Semiconductor introduced a very similar 16-bit pinout, 32-bit internal microprocessor called the NS 16032 (later renamed 32016), the full 32-bit version named the NS 32032. Later, National Semiconductor produced the NS 32132, which allowed two CPUs to reside on the same memory bus with built in arbitration. The NS32016/32 outperformed the MC68000/10, but the NS32332—which arrived at approximately the same time as the MC68020—did not have enough

performance. The third generation chip, the NS32532, was different. It had about double the performance of the MC68030, which was released around the same time. The appearance of RISC processors like the AM29000 and MC88000 (now both dead) influenced the architecture of the final core, the NS32764. Technically advanced—with a superscalar RISC core, 64-bit bus, and internally overclocked—it could still execute Series 32000 instructions through real-time translation.

When National Semiconductor decided to leave the Unix market, the chip was redesigned into the Swordfish Embedded processor with a set of on chip peripherals. The chip turned out to be too expensive for the laser printer market and was killed. The design team went to Intel and there designed the Pentium processor, which is very similar to the NS32764 core internally. The big success of the Series 32000 was in the laser printer market, where the NS32CG16 with microcoded BitBlt instructions had very good price/performance and was adopted by large companies like Canon. By the mid-1980s, Sequent introduced the first SMP server-class computer using the NS 32032. This was one of the design's few wins, and it disappeared in the late 1980s. The MIPS R2000 (1984) and R3000 (1989) were highly successful 32-bit RISC microprocessors. They were used in high-end workstations and servers by SGI, among others. Other designs included the Zilog Z80000, which arrived too late to market to stand a chance and disappeared quickly.

The ARM first appeared in 1985.[42] This is a RISC processor design, which has since come to dominate the 32-bit embedded systems processor space due in large part to its power efficiency, its licensing model, and its wide selection of system development tools. Semiconductor manufacturers generally license cores and integrate them into their own system on a chip products; only a few such vendors are licensed to modify the ARM cores. Most cell phones include an ARM processor, as do a wide variety of other products. There are microcontroller-oriented ARM cores without virtual memory support, as well as symmetric multiprocessor (SMP) applications processors with virtual memory.

In the late 1980s, "microprocessor wars" started killing off some of the microprocessors.[citation needed] Apparently,[vague] with only one bigger design win, Sequent, the NS 32032 just faded out of existence, and Sequent switched to Intel microprocessors.[citation needed]

From 1993 to 2003, the 32-bit x86 architectures became increasingly dominant in desktop, laptop, and server markets and these microprocessors became faster and more capable. Intel had licensed early versions of the architecture to other companies, but declined to license the Pentium, so AMD and Cyrix built later versions of the architecture based on their own designs. During this span, these processors increased in complexity (transistor count) and capability (instructions/second) by at least three orders of magnitude. Intel's Pentium line is probably the most famous and recognizable 32-bit processor model, at least with the public at broad.

3.2.5 64-bit designs

While 64-bit microprocessor designs have been in use in several markets since the early 1990s (including the Nintendo 64 gaming console in 1996), the early 2000s saw the introduction of 64-bit microprocessors targeted at the PC market.

With AMD's introduction of a 64-bit architecture backwards-compatible with x86, x86-64 (also called AMD64), in September 2003, followed by Intel's near fully compatible 64-bit extensions (first called IA-32e or EM64T, later renamed Intel 64), the 64-bit desktop era began. Both versions can run 32-bit legacy applications without any performance penalty as well as new 64-bit software. With operating systems Windows XP x64, Windows Vista x64, Windows 7 x64, Linux, BSD, and Mac OS X that run 64-bit native, the software is also geared to fully utilize the capabilities of such processors. The move to 64 bits is more than just an increase in register size from the IA-32 as it also doubles the number of general-purpose registers.

3.2.6 Multi-core designs

A different approach to improving a computer's performance is to add extra processors, as in symmetric multiprocessing designs, which have been popular in servers and workstations since the early 1990s. Keeping up with Moore's Law is becoming increasingly challenging as chip-making technologies approach their physical limits. In response, microprocessor manufacturers look for other ways to improve performance so they can maintain the momentum of constant upgrades.

A multi-core processor is a single chip that contains more than one microprocessor core. Each core can simultaneously execute processor instructions in parallel. This effectively multiplies the processor's potential performance by the number of cores, if the software is designed to take advantage of more than one processor core. Some components, such as bus interface and cache, may be shared between cores. Because the cores are physically close to each other, they can communicate with each other much faster than separate (off-chip) processors in a multiprocessor system, which improves overall system performance.

In 2005, AMD released the first native dual-core processor, the Athlon X2. Intel's Pentium D had beaten the X2 to market by a few weeks, but it used two separate CPU dies and was less efficient than AMD's native design. As of 2012, dual-core and quad-core processors are widely used in home PCs and laptops, while quad, six, eight, ten, twelve, and sixteen-core processors are common in the professional and enterprise markets with workstations and servers.

1.3 Personal computer architecture

The motherboard is the main component of computer. It is a large rectangular board with integrated circuitry that connects the other parts of the computer including the CPU,

the RAM, the disk drives(CD, DVD, hard disk, or any others) as well as any peripherals connected via the ports or the expansion slots.

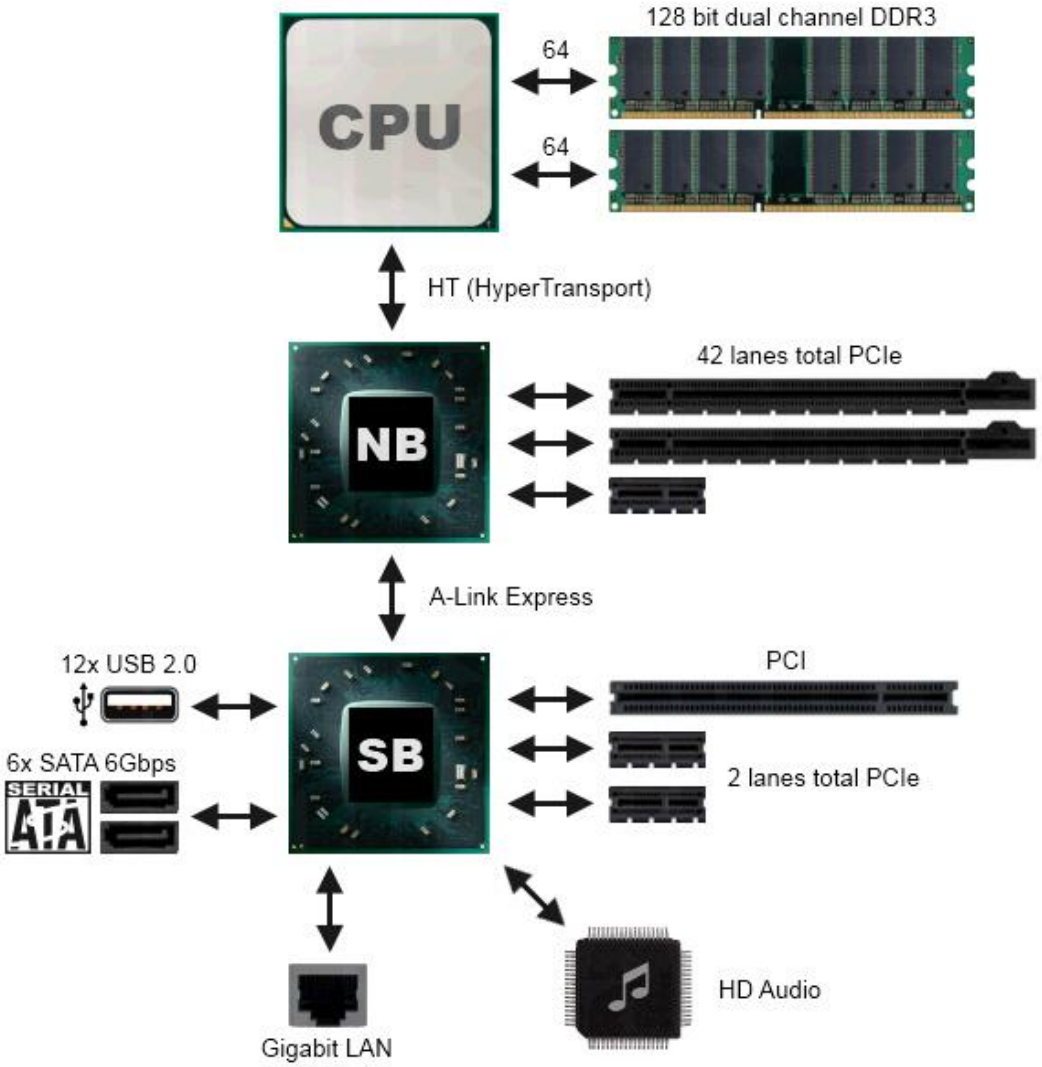
Components directly attached to or part of the motherboard includes:

- The CPU (Central Processing Unit) performs most of the calculations which enable a computer to function, and is sometimes referred to as the "brain" of the computer. It is usually cooled by a heat sink and fan. Most new CPUs include an on-die Graphics Processing Unit (GPU).
- The Chipset, which includes the north bridge, mediates communication between the CPU and the other components of the system, including main memory.
- The Random-Access Memory (RAM) stores the code and data that are being actively accessed by the CPU.
- The Read-Only Memory (ROM) stores the BIOS that run when the computer is powered on or otherwise begins execution, a process known as Bootstrapping, or "booting" or "booting up". The BIOS (Basic Input Output System) includes boot firmware and power management firmware. Newer motherboards use Unified Extensible Firmware Interface (UEFI) instead of BIOS.
- Buses connect the CPU to various internal components and to expansion cards for graphics and sound.
- The CMOS battery is also attached to the motherboard. This battery is the same as a watch battery or a battery for a remote to a car's central locking system. Most batteries are CR2032, which powers the memory for date and time in the BIOS chip.

The BIOS (an acronym for Basic Input/Output System and also known as the System BIOS, ROM BIOS or PC BIOS) is a type of firmware used during the booting process (power-on startup) on IBM PC compatible computers. The BIOS firmware is built into personal computers (PCs), and it is the first software they run when powered on. The name itself originates from the Basic Input/Output System used in the CP/M operating system in 1975. Originally proprietary to the IBM PC, the BIOS has been reverse engineered by companies looking to create compatible systems and the interface of that original system serves as a de facto standard.

In isolation, the microprocessor, the memory and the input/output ports are interesting components, but they cannot do anything useful. In combination, they can form a complete system if they can communicate with each other. This communication is accomplished over bundles of signal wires (known as buses) that connect the parts of the system together.

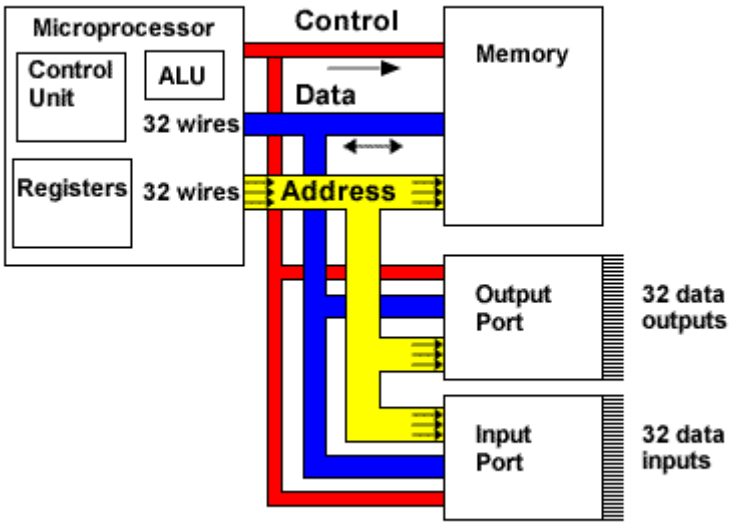
Figure 3.2 Motherboard scheme



There are normally three types of bus in any processor system:

- An address bus: this determines the location in memory that the processor will read data from or write data to.
- A data bus: this contains the contents that have been read from the memory location or are to be written into the memory location.
- A control bus: this manages the information flow between components indicating whether the operation is a read or a write and ensuring that the operation happens at the right time.

Figure 3.3 CPU buses types



Arithmetic logic unit (ALU)

The Arithmetic and Logic Unit (ALU) is responsible for performing most of the computation in a modern processor.

This includes basic arithmetic such as binary addition and subtraction, operations to shift and rotate the bits in a binary number, comparison operations (such as testing for zero, negative numbers, etc.) and logical operations such as AND, OR, XOR (exclusive OR) and NOT (negation).

What constitutes "basic" arithmetic varies according to the processor architecture. Many ARM processors include a coprocessor hardware unit which is used to perform much more complex mathematical operations such as arcsine, cosine, floating-point division, etc.

Inside the ALU is a vast array of logic gates arranged in various subcircuits (such as ripple carry adders and comparators) to perform the necessary operations. In computer design, significant effort is expended in ensuring the ALU is efficiently implemented.

Means that part of the microprocessor, in which an individual elementary arithmetic (e.g. addition) and logical (e.g. logical product) operation. Their implementation is controlled by the microprocessor controller. An important part of the ALU adder, which serves to sum the two input operands of a length typically long as the word the microprocessor. On adder can perform the logical operations.

The Program Counter

The Program Counter (or PC) is a register inside the microprocessor that stores the memory address of the next instruction to be executed. In ARM processors, the Program Counter is a 32-bit register which is also known as R15.

The processor first fetches the instruction from the address stored in the PC. The fetched instruction is then decoded so that it can be interpreted by the microprocessor. Once decoded, the instruction can then be executed and the PC incremented so that it contains the address of the next instruction. This is known as the fetch-decode-execute cycle.

In ARM processors, all instructions take up one word (4 bytes). Hence incrementing the PC actually adds 4 to its value as memory addresses are given in bytes but aligned on word boundaries.

Microprocessor register
Registers are used as temporary storage for instructions and data within the microprocessor. In ARM processors:

- Registers R0 to R14 are 32-bit general-purpose registers. These can be used by programmers for almost any purpose.
- R15 is the Program Counter and is also 32-bits wide.
- The Current Program Status Register contains conditional flags and other status bits that reflect computational results (such as arithmetic overflows, carry out results from the ALU, etc.) It is also 32 bits wide but only the first four and last eight bits are currently used. The other bits are reserved for future developments
- The address register is an internal 32-bit register which can store either a future Program Counter address (so that the next instruction can be fetched in advance) or the address of a value (an operand needed for a computation
- The data registers ("data in register" and "data out register") are used to hold data read from memory and data written to memory respectively. Naturally these are 32-bit registers. Other processor architectures have specialized registers known as accumulators which are used to store intermediate arithmetic results and their assembly languages have commands to enable programmers to utilize them.

3.4 Basic concepts of processors - architecture RISC and CISC

Today's x86 processor designs are an amalgamation of features and functionality from the last 30 years, right up to today's Intel-VT and AMD-V instructions to support hardware-assisted virtualization.

But there's a problem with this complex instruction set computing (CISC) approach; every new instruction or feature adds tens of thousands of transistors to the processor die, adding power demands and latency even if the instructions are rarely used. The chip is extremely versatile, but it runs hot and sucks power with ever-increasing clock speeds.

Processors run much more efficiently when tailored to a specific task. Reduced instruction set computing (RISC) strips out unneeded features and functionality, and builds on task-specific capabilities. Simpler, more reliable RISC processors provide the same effective computing throughput at a fraction of the power and cooling.

The question in CISC vs. RISC arguments is versatility vs. efficiency. Traditional x86 CISC processors can tackle almost any computing task using an extraordinarily comprehensive instruction set. This made CISC the preferred chip design for general-purpose computing platforms: enterprise servers, desktop PCs and laptop/notebook systems.

Purpose-built RISC processors sacrifice versatility for efficiency. Removing unneeded instructions dramatically reduces the processor's transistor count. Tackling fewer tasks in hardware means those tasks are performed faster, even at lower clock speeds (less power) than a full x86 CISC counterpart.

Printers, home routers, and even multifunction telephones and remote controls use RISC processors, and the concept is growing dramatically for fully featured computing platforms. A tablet or smartphone's RISC processor can deliver smooth video playback, fast webpage display and a responsive user interface for many hours on a battery charge, with no cooling devices. This same chip design paradigm is systematically finding traction in data center systems.

3.4.1 Computer buses

In computer architecture, a bus (related to the Latin "omnibus", meaning "for all") is a communication system that transfers data between components inside a computer, or between computers. This expression covers all related hardware components (wire, optical fiber, etc.) and software, including communication protocols.

Internal bus

The internal bus, also known as internal data bus, memory bus, system bus or Front-Side-Bus, connects all the internal components of a computer, such as CPU and memory, to the motherboard. Internal data buses are also referred to as a local bus, because they are intended to connect to local devices. This bus is typically rather quick and is independent of the rest of the computer operations.

External bus

The external bus, or expansion bus, is made up of the electronic pathways that connect the different external devices, such as printer etc., to the computer.

Buses can be parallel buses, which carry data words in parallel on multiple wires, or serial buses, which carry data in bit-serial form. The addition of extra power and control connections, differential drivers, and data connections in each direction usually means that most serial buses have more conductors than the minimum of one used in 1-Wire and UNI/O. As data rates increase, the problems of timing skew, power consumption, electromagnetic interference and crosstalk across parallel buses become more and more difficult to circumvent. One partial solution to this problem has been to double pump the bus. Often, a serial bus can be operated at higher overall data rates than a parallel bus, despite having fewer electrical connections, because a serial bus inherently has no timing skew or crosstalk. USB, FireWire, and Serial ATA are examples of this. Multidrop connections do not work well for fast serial buses, so most modern serial buses use daisy-chain or hub designs.

Network connections such as Ethernet are not generally regarded as buses, although the difference is largely conceptual rather than practical. An attribute generally used to characterize a bus is that power is provided by the bus for the connected hardware. This emphasizes the busbar origins of bus architecture as supplying switched or distributed

power. This excludes, as buses, schemes such as serial RS-232, parallel Centronics, IEEE 1284 interfaces and Ethernet, since these devices also needed separate power supplies. Universal Serial Bus devices may use the bus supplied power, but often use a separate power source.

3.5 PC interface ports

In computer hardware, a port serves as an interface between the computer and other computers or peripheral devices. In computer terms, a port generally refers to the female part of connection. Computer ports have many uses, to connect a monitor, webcam, speakers, or other peripheral devices. On the physical layer, a computer port is a specialized outlet on a piece of equipment to which a plug or cable connects. Electronically, the several conductors where the port and cable contacts connect provide a method to transfer signals between devices.

Serial Port

- Used for external modems and older computer mouse
- Two versions : 9 pin, 25 pin model
- Data travels at 115 kilobits per second

Parallel Port

- Used for scanners and printers
- Also called printer port
- 25 pin model
- Also known as IEEE 1284-compliant Centronics port

PS/2 Port

- Used for old computer keyboard and mouse
- Also called mouse port
- Most of the old computers provide two PS/2 port, each for mouse and keyboard
- Also known as IEEE 1284-compliant Centronics port

Universal Serial Bus (or USB) Port

- It can connect all kinds of external USB devices such as external hard disk, printer, scanner, mouse, keyboard etc.
- It was introduced in 1997.
- Most of the computers provide two USB ports as minimum.
- Data travels at 12 megabits per seconds
- USB compliant devices can get power from a USB port

VGA Port

- Connects monitor to a computer's video card.
- Has 15 holes.
- Similar to serial port connector but serial port connector has pins, it has holes.

Power Connector

- Three-pronged plug
- Connects to the computer's power cable that plugs into a power bar or wall socket

Firewire Port

- Transfers large amount of data at very fast speed.
- Connects camcorders and video equipments to the computer
- Data travels at 400 to 800 megabits per seconds
- Invented by Apple
- Three variants : 4-Pin FireWire 400 connector, 6-Pin FireWire 400 connector and 9-Pin FireWire 800 connector

Modem Port

- Connects a PC's modem to the telephone network

Ethernet Port

- Connects to a network and high speed Internet.
- Connect network cable to a computer.
- This port resides on an Ethernet Card.
- Data travels at 10 megabits to 1000 megabits per seconds depending upon the network bandwidth.

Game Port

- Connect a joystick to a PC
- Now replaced by USB.

Digital Video Interface, DVI port

- Connects Flat panel LCD monitor to the computer's high end video graphic cards.
- Very popular among video card manufacturers.

Sockets

- Connect microphone, speakers to sound card of the computer

Figure 3.4 PC interface ports examples

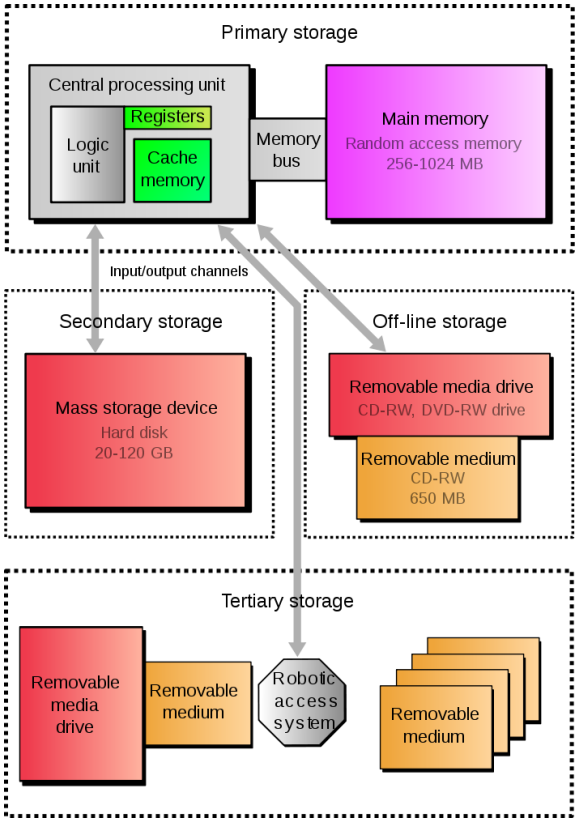
3.6 Computer memory

Computer data storage, often called storage or memory, is a technology consisting of computer components and recording media used to retain digital data. It is a core function and fundamental component of computers. The central processing unit (CPU) of a computer is what manipulates data by performing computations.
In practice, almost all computers use a storage hierarchy, which puts fast but expensive and small storage options close to the CPU and slower but larger and cheaper options farther away. Often the fast, volatile technologies (which lose data when powered off) are referred to as "memory", while slower permanent technologies are referred to as "storage", but these terms are often used interchangeably.

Hierarchy of storage

Generally, the lower a storage is in the hierarchy, the lesser its bandwidth and the greater its access latency is from the CPU. This traditional division of storage to primary, secondary, tertiary and off-line storage is also guided by cost per bit. In contemporary usage, "memory" is usually semiconductor storage read-write random-access memory, typically DRAM (Dynamic-RAM) or other forms of fast but temporary storage. "Storage" consists of storage devices and their media not directly accessible by the CPU (secondary or tertiary storage), typically hard disk drives, optical disc drives, and other devices slower than RAM but non-volatile (retaining contents when powered down).

Figure 3.5 Scheme of storage hierarchy



Primary storage

Primary storage (also known as main memory or internal memory), often referred to simply as memory, is the only one directly accessible to the CPU. The CPU continuously reads instructions stored there and executes them as required. Any data actively operated on is also stored there in uniform manner.

Historically, early computers used delay lines, Williams's tubes, or rotating magnetic drums as primary storage. By 1954, those unreliable methods were mostly replaced by magnetic core memory. Core memory remained dominant until the 1970s, when advances in integrated circuit technology allowed semiconductor memory to become economically competitive.

This led to modern random-access memory (RAM). It is small-sized, light, but quite expensive at the same time. (The particular types of RAM used for primary storage are also volatile, i.e. they lose the information when not powered).

As shown in the diagram, traditionally there are two more sub-layers of the primary storage, besides main large-capacity RAM:

Processor registers are located inside the processor. Each register typically holds a word of data (often 32 or 64 bits). CPU instructions instruct the arithmetic and logic unit to perform various calculations or other operations on this data (or with the help of it). Registers are the fastest of all forms of computer data storage.
Processor cache is an intermediate stage between ultra-fast registers and much slower main memory. It's introduced solely to increase performance of the computer. Most actively used information in the main memory is just duplicated in the cache memory, which is faster, but of much lesser capacity. On the other hand, main memory is much slower, but has a much greater storage capacity than processor registers. Multi-level hierarchical cache setup is also commonly used primary cache being smallest, fastest and located inside the processor; secondary cache being somewhat larger and slower.
Main memory is directly or indirectly connected to the central processing unit via a memory bus. It is actually two buses (not on the diagram): an address bus and a data bus. The CPU firstly sends a number through an address bus, a number called memory address, that indicates the desired location of data. Then it reads or writes the data in the memory cells using the data bus. Additionally, a memory management unit (MMU) is a small device between CPU and RAM recalculating the actual memory address, for example to provide an abstraction of virtual memory or other tasks.

Secondary storage
Secondary storage (also known as external memory or auxiliary storage), differs from primary storage in that it is not directly accessible by the CPU. The computer usually uses its input/output channels to access secondary storage and transfers the desired data using intermediate area in primary storage. Secondary storage does not lose the data when the device is powered down it is non-volatile. Per unit, it is typically also two orders of magnitude less expensive than primary storage. Modern computer systems typically have two orders of magnitude more secondary storage than primary storage and data are kept for a longer time there.

In modern computers, hard disk drives are usually used as secondary storage. The time taken to access a given byte of information stored on a hard disk is typically a few thousandths of a second, or milliseconds. By contrast, the time taken to access a given byte of information stored in random-access memory is measured in billionths of a second, or nanoseconds. This illustrates the significant access-time difference which distinguishes solid-state memory from rotating magnetic storage devices: hard disks are typically about a million times slower than memory. Rotating optical storage devices, such as CD and DVD drives, have even longer access times. With disk drives, once the disk read/write head reaches the proper placement and the data of interest rotates under

it, subsequent data on the track are very fast to access. To reduce the seek time and rotational latency; data are transferred to and from disks in large contiguous blocks.

When data reside on disk, block access to hide latency offers a ray of hope in designing efficient external memory algorithms. Sequential or block access on disks is orders of magnitude faster than random access, and many sophisticated paradigms have been developed to design efficient algorithms based upon sequential and block access. Another way to reduce the I/O bottleneck is to use multiple disks in parallel in order to increase the bandwidth between primary and secondary memory.

Some other examples of secondary storage technologies are: flash memory (e.g. USB flash drives or keys), floppy disks, magnetic tape, paper tape, punched cards, standalone RAM disks, and Iomega Zip drives.

The secondary storage is often formatted according to a file system format, which provides the abstraction necessary to organize data into files and directories, providing also additional information (called metadata) describing the owner of a certain file, the access time, the access permissions, and other information.

Most computer operating systems use the concept of virtual memory, allowing utilization of more primary storage capacity than is physically available in the system. As the primary memory fills up, the system moves the least-used chunks (pages) to secondary storage devices (to a swap file or page file), retrieving them later when they are needed. As more of these retrievals from slower secondary storage are necessary, the more the overall system performance is degraded.

Tertiary storage

Tertiary storage or tertiary memory provides a third level of storage. Typically it involves a robotic mechanism which will mount (insert) and dismount removable mass storage media into a storage device according to the system's demands; this data is often copied to secondary storage before use. It is primarily used for archiving rarely accessed information since it is much slower than secondary storage (e.g. 5–60 seconds vs. 1–10 milliseconds). This is primarily useful for extraordinarily large data stores, accessed without human operators. Typical examples include tape libraries and optical jukeboxes.

When a computer needs to read information from the tertiary storage, it will first consult a catalog database to determine which tape or disc contains the information. Next, the computer will instruct a robotic arm to fetch the medium and place it in a drive. When the computer has finished reading the information, the robotic arm will return the medium to its place in the library.

Off-line storage

Off-line storage is a computer data storage on a medium or a device that is not under the control of a processing unit. The medium is recorded, usually in a secondary or tertiary storage device, and then physically removed or disconnected. It must be inserted or

connected by a human operator before a computer can access it again. Unlike tertiary storage, it cannot be accessed without human interaction.

Off-line storage is used to transfer information, since the detached medium can be easily physically transported. Additionally, in case a disaster, for example a fire, destroys the original data, a medium in a remote location will probably be unaffected, enabling disaster recovery. Off-line storage increases general information security, since it is physically inaccessible from a computer, and data confidentiality or integrity cannot be affected by computer-based attack techniques. Also, if the information stored for archival purposes is rarely accessed, off-line storage is less expensive than tertiary storage.

In modern personal computers, most secondary and tertiary storage media are also used for off-line storage. Optical discs and flash memory devices are most popular, and to much lesser extent removable hard disk drives. In enterprise uses, magnetic tape is predominant. Older examples are floppy disks, Zip disks, or punched cards.

3.7 Input-output devices

Before a computer can process your data, you need some method to input the data into the machine. The device you use will depend on what form this data takes (be it text, sound, artwork, etc.).

Similarly, after the computer has processed your data, you often need to produce output of the results. This output could be a display on the computer screen, hardcopy on printed pages, or even the audio playback of music you composed on the computer.

The terms “input” and “output” are used both as verbs to describe the process of entering or displaying the data, and as nouns referring to the data itself entered into or displayed by the computer.

Below we discuss the variety of peripheral devices used for computer input and output.

3.7.1 Input devices

Keyboard
The computer keyboard is used to enter text information into the computer, as when you type the contents of a report. The keyboard can also be used to type commands directing the computer to perform certain actions. Commands are typically chosen from an on-screen menu using a mouse, but there are often keyboard shortcuts for giving these same commands.

In addition to the keys of the main keyboard (used for typing text), keyboards usually also have a numeric keypad (for entering numerical data efficiently), a bank of editing keys (used in text editing operations), and a row of function keys along the top (to easily invoke certain program functions). Laptop computers, which don’t have room for large

keyboards, often include a “fn” key so that other keys can perform double duty (such as having a numeric keypad function embedded within the main keyboard keys).

Improper use or positioning of a keyboard can lead to repetitive-stress injuries. Some ergonomic keyboards are designed with angled arrangements of keys and with built-in wrist rests that can minimize your risk of RSIs.

Most keyboards attach to the PC via a PS/2 connector or USB port (newer). Older Macintosh computers used an ABD connector, but for several years now all Mac keyboards have connected using USB.

Mouse

The mouse pointing device sits on your work surface and is moved with your hand. In older mice, a ball in the bottom of the mouse rolls on the surface as you move the mouse and internal rollers sense the ball movement and transmit the information to the computer via the cord of the mouse.

The newer optical mouse does not use a rolling ball, but instead uses a light and a small optical sensor to detect the motion of the mouse by tracking a tiny image of the desk surface. Optical mice avoid the problem of a dirty mouse ball, which causes regular mice to roll unsmoothly if the mouse ball and internal rollers are not cleaned frequently.

A cordless or wireless mouse communicates with the computer via radio waves (often using BlueTooth hardware and protocol) so that a cord is not needed (but such mice need internal batteries).

A mouse also includes one or more buttons (and possibly a scroll wheel) to allow users to interact with the GUI. The traditional PC mouse has two buttons, while the traditional Macintosh mouse has one button. On either type of computer you can also use mice with three or more buttons and a small scroll wheel (which can also usually be clicked like a button).

Figure 3.6 On the left is an example of a wireless optical mouse, trackball is in the middle and tablet on the right

Graphics Tablet

A graphics tablet consists of an electronic writing area and a special "pen" that works with it. Graphics tablets allow artists to create graphical images with motions and actions similar to using more traditional drawing tools. The pen of the graphics tablet is pressure sensitive, so pressing harder or softer can result in brush strokes of different width (in an appropriate graphics program).

Trackball
The trackball is sort of like an upside-down mouse, with the ball located on top. You use your fingers to roll the trackball, and internal rollers (similar to what's inside a mouse) sense the motion which is transmitted to the computer. Trackballs have the advantage over mice in that the body of the trackball remains stationary on your desk, so you don't need as much room to use the trackball. Early laptop computers often used trackballs (before superior touch pads came along).

Trackballs have traditionally had the same problem as mice: dirty rollers can make their cursor control jumpy and unsmooth. But there are modern optical trackballs that don't have this problem because their designs eliminate the rollers.

Trackpoint
Some sub-notebook computers (such as the IBM ThinkPad), which lack room for even a touch pad, incorporate a trackpoint, a small rubber projection embedded between the keys of the keyboard. The trackpoint acts like a little joystick that can be used to control the position of the on-screen cursor.

Touchpad
Most laptop computers today have a touch pad pointing device. You move the on-screen cursor by sliding your finger along the surface of the touch pad. The buttons are located below the pad, but most touch pads allow you to perform "mouse clicks" by tapping on the pad itself.

Touch pads have the advantage over mice that they take up much less room to use. They have the advantage over trackballs (which were used on early laptops) that there are no moving parts to get dirty and result in jumpy cursor control.

Touch screen
Some computers, especially small hand-held PDAs, have touch sensitive display screens. The user can make choices and press button images on the screen. You often use a stylus, which you hold like a pen, to "write" on the surface of a small touch screen.

Scanner
A scanner is a device that images a printed page or graphic by digitizing it, producing an image made of tiny pixels of different brightness and color values which are represented numerically and sent to the computer. Scanners scan graphics, but they can also scan pages of text which are then run through OCR (Optical Character Recognition) software that identifies the individual letter shapes and creates a text file of the page's contents.

Microphone
A microphone can be attached to a computer to record sound (usually through a sound card input or circuitry built into the motherboard). The sound is digitized—turned into numbers that represent the original analog sound waves—and stored in the computer to later processing and playback.

3.7.2 Output devices

CRT Monitor
The traditional output device of a personal computer has been the CRT (Cathode Ray Tube) monitor. Just like a television set (an older one, anyway) the CRT monitor contains a large cathode ray tube that uses an electron beam of varying strength to "paint" a picture onto the color phosphorescent dots on the inside of the screen. CRT monitors are heavy and use more electrical power than flat panel displays, but they are preferred by some graphic artists for their accurate color rendition, and preferred by some gamers for faster response to rapidly changing graphics.

Monitor screen size is measured diagonally across the screen, in inches. Not all of the screen area may be usable for image display, so the viewable area is also specified. The resolution of the monitor is the maximum number of pixels it can display horizontally and vertically (such as 800 x 600, or 1024 x 768, or 1600 x 1200). Most monitors can display several resolutions below its maximum setting. Pixels (short for picture elements) are the small dots that make of the image displayed on the screen. The spacing of the screen's tiny phosphor dots is called the dot pitch (dp), typically .28 or .26 (measured in millimeters). A screen with a smaller dot pitch produces sharper images.

Flat Panel Monitor
A flat panel display usually uses an LCD (Liquid Crystal Display) screen to display output from the computer. The LCD consists of several thin layers that polarize the light passing through them. The polarization of one layer, containing long thin molecules called liquid crystals, can be controlled electronically at each pixel, blocking varying amounts of the light to make a pixel lighter or darker. Other types of flat panel technology exist (such as plasma displays) but LCDs are most commonly used in computers, especially laptops.
Older LCDs had slow response times and low contrast, but active matrix LCD screens have a transparent thin film transistor (TFT) controlling each pixel, so response, contrast, and viewing angle are much improved.
Flat panel displays are much lighter and less bulky than CRT monitors, and they consume much less power. They have been more expensive than CRTs in the past, but the price gap is narrowing. You will see many more flat panels in the future.
As with CRTs, the display size of a flat panel is expressed in inches, and the resolution is the number of pixels horizontally and vertically on the display.
InkJet Printer
For hardcopy (printed) output, you need some kind of printer attached to your computer (or available over a network). The most common type of printer for home systems is the color ink jet printer. These printers form the image on the page by spraying tiny droplets

of ink from the print head. The printer needs several colors of ink (cyan, yellow, magenta, and black) to make color images. Some photo-quality ink jet printers have more colors of ink.

Ink jet printers are inexpensive, but the cost of consumables (ink cartridges and special paper) makes them costly to operate in the long run for many purposes.

Laser Printer

A laser printer produces good quality images by the same technology that photocopiers use. A drum coated with photosensitive material is charged, and then an image is written onto it by a laser (or LEDs) which makes those areas lose the charge. The drum then rolls through toners (tiny plastic particles of pigment) that are attracted to the charged areas of the drum. The toner is then deposited onto the paper, and then fused into the paper with heat.

Most laser printers are monochrome (one color only, usually black), but more expensive laser printers with multiple color toner cartridges can produce color output.
Laser printers are faster than ink jet printers. Their speed is rated in pages per minute (ppm). Laser printers are more expensive than ink jets, but they are cheaper to run in the long term if you just need good quality black & white pages.

Other Printers

Multi-function printers are available that not only operate as a computer printer, but also include the hardware needed to be a scanner, photocopier, and FAX machine as well.
Dot matrix printers use small electromagnetically activated pins in the print head, and an inked ribbon, to produce images by impact. These printers are slow and noisy, and are not commonly used for personal computers anymore (but they can print multi-layer forms, which neither ink jet or laser printers can).

Sound output

Computers also produce sound output, ranging from simple beeps alerting the user, to impressive game sound effects, to concert quality music. The circuitry to produce sound may be included on the motherboard, but high quality audio output from a PC usually requires a sound card in one of the expansion slots, connected to a set of good quality external speakers or headphones.

Control questions

1) Which parts has Von Neumann concept?
2) What is BIOS?
3) What is computer memory?
4) What belongs to input and output devices?

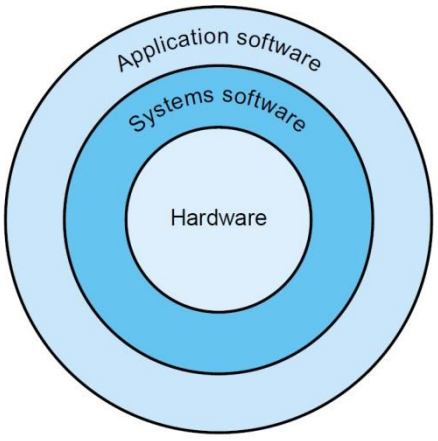
4 COMPUTER SOFTWARE

Software consists of computer programs, which are sequences of instructions for the computer. The process of writing (or coding) programs is called programming, and individuals who perform this task are called programmers.
Unlike the hardwired computers of the 1950s, modern software uses the stored program concept, in which stored software programs are accessed and their instructions are executed (followed) in the computer's CPU. Once the program has finished executing, a new program is loaded into main memory and the computer hardware addresses another task.

Computer programs include documentation, which is a written description of the functions of the program. Documentation helps the user operate the computer system and helps other programmers understand what the program does and how it accomplishes its purpose. Documentation is vital to the business organization. Without it, if a key programmer or user leaves, the knowledge of how to use the program or how it is designed may be lost.

The computer is able to do nothing until it is instructed by software. Although computer hardware is, by design, general purpose, software enables the user to instruct a computer system to perform specific functions that provide business value. There are two major types of software: systems software and application software. The relationship among hardware, systems software, and application software is illustrated in Figure 4.1.

Figure 4.1 Relationship among hardware, software and application software



Systems software is a set of instructions that serves primarily as an intermediary between computer hardware and application programs, and may also be directly manipulated by knowledgeable users. Systems software provides important self-regulatory functions for computer systems, such as loading itself when the computer is first turned on, managing hardware resources such as secondary storage for all applications, and providing commonly used sets of instructions for all applications to use. Systems programming is either the creation or maintenance of systems software.

Application software is a set of computer instructions that provide more specific functionality to a user. That functionality may be broad, such as general word processing, or narrow, such as an organization's payroll program. An application program applies a computer to a certain need. Application programming is either the creation or the modification and improvement of application software. There are many different software applications in organizations today, as this chapter will discuss. For a marketing application, for example, see the Market Intelligence box at the Web site.

In summary, application programs primarily manipulate data or text to produce or provide information. Systems programs primarily manipulate computer hardware resources. The systems software available on a computer system provides the capabilities and limitations within which the application software can operate. The next two sections of this chapter look in more detail at these two types of software.

4.1 System software

Systems software is the class of programs that control and support the computer system and its information-processing activities. Systems software also facilitates the programming, testing, and debugging of computer programs. It is more general than application software and is usually independent of any specific type of application. Systems software programs support application software by directing the basic functions of the computer. For example, when the computer is turned on, the initialization program (a systems program) prepares and readies all devices for processing. Other common operating systems tasks are following:

- Monitoring performance
- Correcting errors
- Providing and maintaining the user interface
- Starting ("booting") the computer
- Reading programs into memory
- Managing memory allocation to those programs
- Placing files and programs in secondary storage
- Creating and maintaining directories
- Formatting diskettes
- Controlling the computer monitor
- Sending jobs to the printer
- Maintaining security and limiting access
- Locating files
- Detecting viruses
- Compressing data

Systems software can be grouped into two major functional categories: system control programs and system support programs.

4.1.1 System Control Programs

System control programs control the use of the hardware, software, and data resources of a computer system. The main system control program is the operating system. The operating system supervises the overall operation of the computer, including monitoring the computer's status and scheduling operations, which include the input and output processes. In addition, the operating system allocates CPU time and main memory to programs running on the computer, and it also provides an interface between the user and the hardware. This interface hides the complexity of the hardware from the user. That is, you do not have to know how the hardware actually operates, just what the

hardware will do and what you need to do to obtain desired results. Specifically, the operating system provides services that include process management, virtual memory, file management, security, fault tolerance, and the user interface.

Process management means managing the program or programs (also called jobs) running on the processor at a given time. In the simplest case (a desktop operating system), the operating system loads a program into main memory and executes it. The program utilizes the computer's resources until it relinquishes control. Some operating systems offer more sophisticated forms of process management, such as multitasking, multithreading, and multiprocessing.

The management of two or more tasks, or programs, running on the computer system at the same time is called multitasking, or multiprogramming. The first program is executed until an interruption occurs, such as a request for input. While the input request is handled, the execution of a second program begins. Because switching among these programs occurs so rapidly, they appear to be executing at the same time. However, because there is only one processor, only one program is actually in execution mode at any one time. Multithreading is a form of multitasking that focuses on running multiple tasks within a single application simultaneously. For example, a word processor application may edit one document while another document is being spell-checked.

Time-sharing is an extension of multiprogramming. In this mode, a number of users operate online with the same CPU, but each uses a different input/output terminal. The programs of these users are placed into partitions in primary storage. Execution of these programs rotates among all users, occurring so rapidly that it appears to each user as though he or she were the only one using the computer.

Multiprocessing occurs when a computer system with two or more processors can run more than one program, or thread, at a given time by assigning them to different processors. Multiprocessing uses simultaneous processing with multiple CPUs, whereas multiprogramming involves concurrent processing with one CPU.

Virtual memory simulates more main memory than actually exists in the computer system. It allows a program to behave as if it had access to the full storage capacity of a computer, rather than just access to the amount of primary storage installed on the computer. Virtual memory divides an application program or module into fixed-length portions called pages. The system executes some pages of instructions while pulling others from secondary storage. In effect, primary storage is extended into a secondary storage device, allowing users to write programs as if primary storage were larger than it actually is. This enlarged capability boosts the speed of the computer and allows it to efficiently run programs with very large numbers of instructions.

The operating system is responsible for file management and security, managing the arrangement of, and access to, files held in secondary storage. The operating system creates and manages a directory structure that allows files to be created and retrieved by name, and it also may control access to those files based on permissions and access controls. The operating system provides other forms of security as well. For example, it must typically provide protected memory and maintain access control on files in the file system. The operating system also must keep track of users and their authority level, as well as audit changes to security permissions.

Fault tolerance is the ability of a system to produce correct results and to continue to operate even in the presence of faults or errors. Fault tolerance can involve error

correcting memory, redundant computer components, and related software that protect the system from hardware, operating system, or user errors.
Although operating systems perform some of their functions automatically, for certain tasks the user interacts directly with the computer through the systems software. The ease or difficulty of such interaction is to a large extent determined by the interface design. Older text-based interfaces like DOS (disk operating system) required typing in cryptic commands. In an effort to make computers more user-friendly, the graphical user interface was developed.
The graphical user interface (GUI) allows users to have direct control of visible objects (such as icons) and actions that replace complex command syntax. The GUI was developed by researchers at Xerox PARC (Palo Alto Research Center), and then popularized by the Apple MacIntosh computer. Microsoft soon introduced its GUIbased Windows operating system for IBM-style PCs. The next generation of GUI technology will incorporate features such as virtual reality, head-mounted displays, sound and speech, pen and gesture recognition, animation, multimedia, artificial intelligence, and cellular/wireless communication capabilities.

Types of operating systems
As previously discussed, operating systems are necessary in order for computer hardware to function. Operating environments, which add features that enable system developers to create applications without directly accessing the operating system, function only with an operating system. That is, operating environments are not operating systems, but work only with an operating system. For example, the early versions of Windows were operating environments that provided a graphical user interface and worked only with MS-DOS.

Operating systems (OSs) can be categorized by the number of users they support as well as by their level of sophistication (see the Operating Systems list on the Web site). Operating systems for mobile devices are designed to support a single person using a mobile, handheld device, or information appliance. Desktop operating systems are designed to support a single user or a small workgroup of users. Departmental server operating systems typically support from a few dozen to a few hundred users. Enterprise server operating systems generally support thousands of simultaneous users and millions or billions of simultaneous transactions.

Supercomputer operating systems support the particular processing needs of supercomputers. Supercomputer and enterprise server operating systems offer the greatest functionality, followed by departmental server operating systems, desktop operating systems, and finally EEM operating systems. An important exception is the user interface, which is most sophisticated on desktop operating systems and least sophisticated on supercomputer and enterprise server operating systems.

Desktop and notebook computer operating systems

The Windows family is the leading series of desktop operating systems. The MS-DOS (Microsoft Disk Operating System) was one of the original operating systems for the IBM PC and its clones. This 16-bit operating system, with its text-based interface, has now been almost totally replaced by GUI operating systems such as Windows 2000 and Windows XP. Windows 1.0 through Windows 3.1 (successive versions) were not operating systems, but were operating environments that provided the GUI that operated with, and extended the capabilities of, MS-DOS.
Windows 95, released in 1995, was the first of a series of products in the Windows operating system that provided a streamlined GUI by using icons to provide instant access to common tasks. Windows 95 is a 32-bit operating system that features multitasking, multithreading, networking, and Internet integration capabilities, including the ability to integrate fax, e-mail, and scheduling programs. Windows 95 also offers plug-and-play capabilities. Plug-and-play is a feature that can automate the installation of new hardware by enabling the operating system to recognize new hardware and install the necessary software (called device drivers) automatically.

Subsequent products in the Microsoft Windows operating system are:

- Windows 89
- Windows Millenium
- Windows NT
- Windows 2000
- Windows XP
- Windows Vista
- Windows 7
- Windows 8
- Windows 8.1
- Windows 10

UNIX provides many sophisticated desktop features, including multiprocessing and multitasking. UNIX is valuable to business organizations because it can be used on many different sizes of computers (or different platforms), can support many different hardware devices (e.g., printers, plotters, etc.), and has numerous applications written to run on it. UNIX has many different versions. Most UNIX vendors are focusing their development efforts on servers rather than on desktops, and are promoting Linux for use on the desktop.

Linux is a powerful version of the UNIX operating system that is completely free of charge. It offers multitasking, virtual memory management, and TCP/IP networking. Linux was originally written by Linus Torvalds at the University of Helsinki in Finland in 1991. He then released the source code to the world (called open source software, as discussed in the chapter opening case). Since that time, many programmers around the world have worked on Linux and written software for it. The result is that, like UNIX, Linux now runs on multiple hardware platforms, can support many different hardware devices, and has numerous applications written to run on it. Linux is becoming widely used by Internet service providers (ISPs), the companies that provide Internet connections.

The Macintosh operating system X (ten) (Mac OS X), for Apple Macintosh microcomputers, is a 32-bit operating system that supports Internet integration, virtual memory management, and AppleTalk networking. Mac OS X features a new Aqua user interface, advanced graphics, virtual memory management, and multitasking.

IBM's OS/2 is a 32-bit operating system that supports multitasking, accommodates larger applications, allows applications to be run simultaneously, and supports networked multimedia and pen-computing applications.

Sun's Java operating system (JavaOS) executes programs written in the Java language (described later in this chapter) without the need for a traditional operating system. It is designed for Internet and intranet applications and embedded devices. JavaOS is designed for handheld products and thin-client computing.

4.2 Application software

As defined earlier, application software consists of instructions that direct a computer system to perform specific information processing activities and that provide functionality for users. Because there are so many different uses for computers, there are a correspondingly large number of different application software programs available.

Types of Application Software

Application software includes proprietary application software and off-the-shelf application software. Proprietary application software addresses a specific or unique business need for a company. This type of software may be developed in-house by the organization's information systems personnel or it may be commissioned from a software vendor. Such specific software programs developed for a particular company by a vendor are called contract software.

Alternatively, off-the-shelf application software can be purchased, leased, or rented from a vendor that develops programs and sells them to many organizations. Off-the-shelf software may be a standard package or it may be customizable. Special purpose programs or "packages" can be tailored for a specific purpose, such as inventory control or payroll. The word package is a commonly used term for a computer program (or group of programs) that has been developed by a vendor and is available for purchase in a prepackaged form.

Types of Personal Application Software

General-purpose, off-the-shelf application programs that support general types of processing, rather than being linked to any specific business function, are referred to as personal application software. This type of software consists of nine widely used packages: spreadsheet, data management, word processing, desktop publishing, graphics, multimedia, communications, speech-recognition software, and groupware. Software suites combine some of these packages and integrate their functions.

Personal application software is designed to help individual users increase their productivity. Below is a description of the nine main types.

Spreadsheets

Computer spreadsheet software transforms a computer screen into a ledger sheet, or grid, of coded rows and columns. Users can enter numeric or textual data into each grid location, called a cell. In addition, a formula can be entered into a cell to obtain a calculated answer displayed in that cell's location. With spreadsheets, users can also develop and use macros, which are sequences of commands that can be executed with just one simple instruction.

Computer spreadsheet packages can be used for financial information, such as income statements or cash flow analysis. They are also used for forecasting sales, analyzing insurance programs, summarizing income tax data, and analyzing investments. They are relevant for many other types of data that can be organized into rows and columns. Although spreadsheet packages such as Microsoft's Excel and Lotus 1–2-3 are thought of primarily as spreadsheets, they also offer data management and graphical capabilities. Therefore, they may be called integrated packages.

Spreadsheets are valuable for applications that require modeling and what-if analysis. After a set of mathematical relationships has been specified by the user, the spreadsheet can be recalculated instantly using a different set of assumptions (i.e., a different set of mathematical relationships).

Data management

Data management software supports the storage, retrieval, and manipulation of related data. There are two basic types of data management software: simple filing programs patterned after traditional, manual data-filing techniques, and database management programs that take advantage of a computer's extremely fast and accurate ability to store and retrieve data in primary and secondary storage. File based management software is typically very simple to use and is often very fast, but it offers limited flexibility in how the data can be searched. Database management software has the opposite strengths and weaknesses. Microsoft's Access is an example of popular database management software. In Chapter 5, we discuss data management in much more detail.

Word processing

Word processing software allows the user to manipulate text rather than just numbers. Modern word processors contain many productive writing and editing features. A typical word processing software package consists of an integrated set of programs including an editor program, a formatting program, a print program, a dictionary, a thesaurus, a grammar checker, a mailing list program, and integrated graphics, charting, and drawing programs. WYSIWYG (an acronym for What You See Is What You Get, pronounced "wiz-e-wig") word processors have the added advantage of displaying the text material on the screen exactly—or almost exactly— as it will look on the final printed page (based on the type of printer connected to the computer). Word processing software enables users to be much more productive because the software makes it possible to create and modify the document electronically in memory.

Desktop publishing

Desktop publishing software represents a level of sophistication beyond regular word processing. In the past, newsletters, announcements, advertising copy, and other specialized documents had to be laid out by hand and then typeset. desktop software allows microcomputers to perform these tasks directly. Photographs, diagrams, and other images can be combined with text, including several different fonts, to produce a finished, camera-ready document.

Graphics

Graphics software allows the user to create, store, and display or print charts, graphs, maps, and drawings. Graphics software enables users to absorb more information more quickly and to spot relationships and trends in data more easily. There are three basic categories of graphics software packages: presentation graphics, analysis graphics, and computer-aided design software.

Presentation graphics software allows users to create graphically rich presentations. Many packages have extensive libraries of clip art—pictures that can be electronically "clipped out" and "pasted" into the finished image. One of the most widely used presentation graphics programs is Microsoft's PowerPoint.

Analysis graphics applications additionally provide the ability to convert previously analyzed data—such as statistical data—into graphic formats like bar charts, line charts, pie charts, and scatter diagrams. Both presentation graphics and analysis graphics are useful in preparing graphic displays for business presentations, from sales results to marketing research data.

Computer-aided design (CAD) software, used for designing items for manufacturing, allows designers to design and "build" production prototypes in software, test them as a computer object under given parameters (sometimes called computer-aided engineering, or CAE), compile parts and quantity lists, outline production and assembly procedures, and then transmit the final design directly to machines.

Manufacturers of all sorts are finding uses for CAD software. Computer-aided manufacturing (CAM) software uses digital design output, such as that from a CAD system, to directly control production machinery. Computer-integrated manufacturing (CIM) software is embedded within each automated production machine to produce a product. Overall, a design from CAD software is used by CAM software to control individual CIM programs in individual machines. Used effectively, CAD/CAM/CIM software can dramatically shorten development time and give firms the advantage of economies of scope.

Multimedia

Multimedia software combines at least two media for input or output of data. These media include audio (sound), voice, animation, video, text, graphics, and images. Multimedia can also be thought of as the combination of spatial-based media (text and images) with time-based media (sound and video).

Communications

Computers are often interconnected in order to share or relate information. To exchange information, computers utilize communications software. This software allows computers, whether they are located close together or far apart, to exchange data over

dedicated or public cables, telephone lines, satellite relay systems, or microwave circuits.
When communications software exists in both the sending and receiving computers, they are able to establish and relinquish electronic links, code and decode data transmissions, verify transmission errors (and correct them automatically), and check for and handle transmission interruptions or conflicting transmission priorities. E-mail and desktop videoconferencing rely on communications software.

Speech-recognition software
Two categories of speech-recognition software are available today: discrete speech and continuous speech. Discrete speech recognition can interpret only one word at a time, so users must place distinct pauses between words. This type of voice recognition can be used to control PC software (by using words such as “execute” or “print”). But it is inadequate for dictating a memo, because users find it difficult to speak with measurable pauses between every word and still maintain trains of thought. Software for continuous speech recognition can interpret a continuing stream of words. The software must understand the context of a word to determine its correct spelling, and be able to overcome accents and interpret words very quickly. These requirements mean that continuous speech-recognition software must have a computer with significantly more speed and memory than discrete speech software.

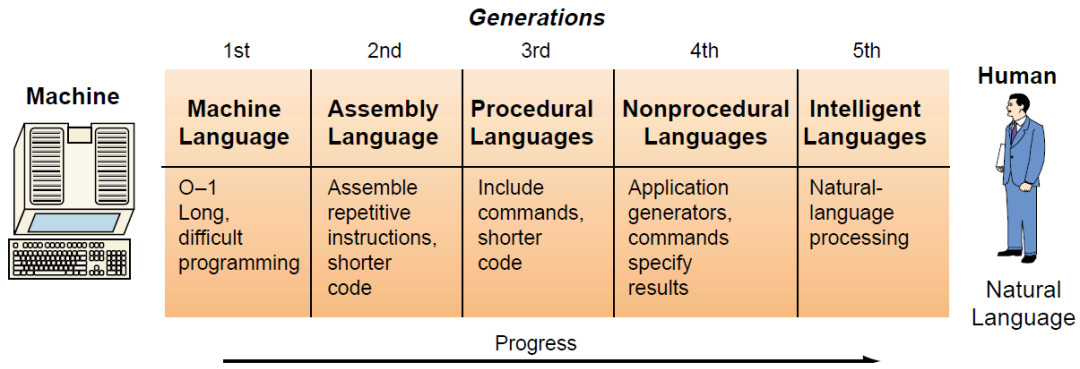
Groupware
Groupware is a class of software products that facilitates communication, coordination, and collaboration among people. Groupware is important because it allows workgroups—people who need to interact with one another within an organization — to communicate and share information, even when they are working together at a distance. Groupware can provide many benefits to businesses, including more efficient and effective project management, location independence, increased communications capability, increased information availability, and improved workflow.
Groupware comes in many varieties. The most elaborate system, IBM’s Lotus Notes/Domino, is a document-management system, a distributed client/server database, and a basis for intranet and electronic commerce systems, as well as a communication support tool. This class of groupware supplements real-time communications with asynchronous electronic connections (e.g., electronic mail and other forms of messaging). Thanks to electronic networks, e-mail, and shared discussion databases, group members can communicate, access data, and exchange or update data at any time and from any place. Group members might store all their official memos, formal reports, and informal conversations related to particular projects in a shared, online data store, such as a database. Then, as individual members need to check on the contents, they can access the shared database to find the information they need.

4.2.1 Programming languages

Programming languages provide the basic building blocks for all systems and application software. Programming languages allow people to tell computers what to do

and are the means by which software systems are developed. This section will describe the five generations of programming languages.

Figure 4.2 Generations of programming languages



Machine Language

Machine language is the lowest-level computer language, consisting of the internal representation of instructions and data. This machine code—the actual instructions understood and directly executable by the central processing unit—is composed of binary digits. Machine language is the only programming language that the machine actually understands. Therefore, machine language is considered the first-generation language. All other languages must be translated into machine language before the computer can run the instructions. Because a computer's central processing unit is capable of executing only machine language programs, such programs are machine dependent (nonportable). That is, the machine language for one type of central processor may not run on other types.

Machine language is extremely difficult to understand and use by programmers. As a result, increasingly more user-friendly languages have been developed. Figure 4.2 gives an overview of the evolution of programming languages, from the first generation machine language to more humanlike natural language. These user-oriented languages make it much easier for people to program, but they are impossible for the computer to execute without first translating the program into machine language. The set of instructions written in a user-oriented language is called a source program. The set of instructions produced after translation into machine language is called the object program.

Programming in a higher-level language (i.e., a user-oriented language) is easier and less time consuming, but additional processor time is required to translate the program before it can be executed. Therefore, one trade-off in the use of higher-level languages is a decrease in programmer time and effort for an increase in processor time needed for translation.

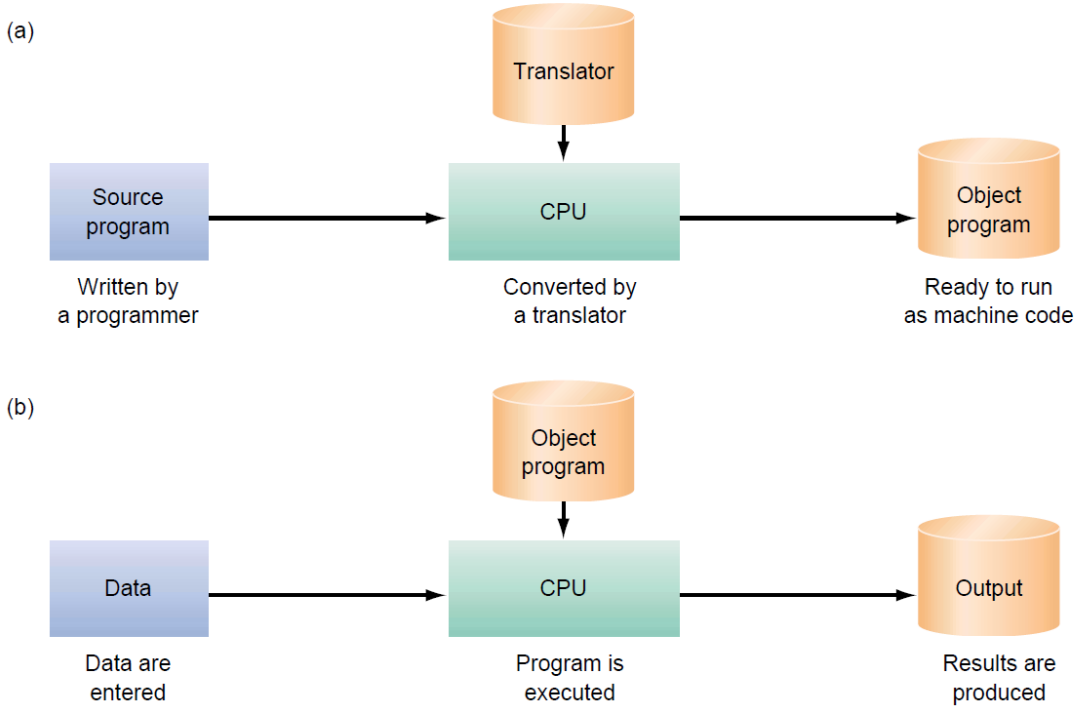
Assembly Language

An assembly language is the next level up from machine language. It is still considered a lower-level language but is more user-friendly because it represents machine language instructions and data locations in primary storage by using mnemonics, or memory aids,

which people can more easily use. Assembly languages are considered second-generation languages.
Compared to machine language, assembly language eases the job of the programmer considerably. However, each statement in an assembly language must still be translated into a single statement in machine language, and assembly languages are still hardware dependent. Translating an assembly language program into machine language is accomplished by a systems software program called an assembler.

Procedural Languages
Procedural languages are the next step in the evolution of user-oriented programming languages. They are also called third-generation languages, or 3GLs. Procedural languages are much closer to so-called natural language (the way we talk) and therefore are easier to write, read, and alter. Moreover, one statement in a procedural language is translated into a number of machine language instructions, thereby making programming more productive. In general, procedural languages are more like natural language than assembly languages are, and they use common words rather than abbreviated mnemonics. Because of this, procedural languages are considered the first level of higher-level languages.
Procedural languages require the programmer to specify, step by step, exactly how the computer must accomplish a task. A procedural language is oriented toward how a result is to be produced. Because computers understand only machine language (i.e., 0s and 1s), higher-level languages must be translated into machine language prior to execution. This translation is accomplished by systems software called language translators. A language translator converts the high-level program, called source code, into machine language code, called object code. There are two types of language translators—interpreters and compilers. Figure 4.3 shows the translation process for source code.
The translation of a high-level language program to object code is accomplished by a software program called a compiler, which translates the entire program at once. In contrast, an interpreter is a compiler that translates and executes one source program statement at a time. Because this translation is done one statement at a time, interpreters tend to be simpler than compilers. This simplicity allows for more extensive debugging and diagnostic aids to be available on interpreters. For examples of FORTRAN, COBOL, and C, see Examples of Procedural Languages on the Web site.

Figure 4.3 Translation process for source code



Nonprocedural Languages

Another type of high-level language, called nonprocedural languages, allows the user to specify the desired result without having to specify the detailed procedures needed for achieving the result. These languages are fourth-generation languages (4GLs). An advantage of nonprocedural languages is that they can be used by nontechnical users to carry out specific functional tasks. These languages greatly simplify and accelerate the programming process, as well as reduce the number of coding errors. The 4GLs are common in database applications as query languages, report generators, and data manipulation languages. They allow users and programmers to interrogate and access computer databases using statements that resemble natural language.

Natural Programming Languages

Natural programming languages are the next evolutionary step. They are sometimes known as fifth-generation languages, or intelligent languages. Translator programs to translate natural languages into a structured, machine-readable form are extremely complex and require a large amount of computer resources. Therefore, most of these languages are still experimental and have yet to be widely adopted by industry.

Control questions

1) What is system software?
2) What is application software?
3) Name the generations of programming languages.