systems fundamental laws; it is a generalisation of the *second law of thermodynamics*, which states that the energy in the universe degrades irreversibly as time passes.

The concept of entropy is a fairly difficult one to understand conceptually. Though it is possible to define it more rigorously (i.e. mathematically!) the following examples illustrate the essence of the concept. A hot plate of food or an ice-cold drink placed on the table will tend to take on the ambient room temperature. A spring-driven clock or watch will slowly unwind and stop. A house or garden that is not maintained will slowly become more and more disorderly. Music or video signals on a magnetic tape will slowly fade away until, after many decades, nothing but noise remains. As a log fire keeps radiating heat and light energy, it slowly cools down until nothing but ashes remain. In every one of these examples, the input of energy serves to restore the system temporarily to a lower level of entropy (although this usually implies an increase in the overall entropy of the supra-system).

3.4   South African Perspective

Careful analysis of an intended system (its inputs, outputs, interfaces, processes etc) can prove invaluable in identifying unforeseen problems at the planning stages. Apparently this did not happen in the case of the renovation and expansion of the facilities served by the upper cable station on Table Mountain – no process was included to dispose of the increased volume of sewage that resulted, until a major problem had become apparent!

Until the end of 2001, a mixture of partially treated sewage from lavatories and waste water from the restaurant's sinks and dishwashers was allowed to seep down the mountain, killing vegetation and causing an outcry among environmentalists. Since then, effluent has been stored in a tank on the summit and is removed by cable car, but with the expansion of facilities and increasing numbers of visitors, the cableway company is concerned that this is not a sustainable solution.

3.5   Beyond the Basics

A number of system rules, which have been developed based on general systems theory, are particularly relevant to business. Examples of these are:

- Law of requisite variety: Control can be obtained only if the variety of the controller is at least as great as the variety of the situation to be controlled. In other words, every possible situation should provide feedback, which can be acted upon if required; this can be applied to controllers ranging from smoke detectors to computer error messages.

- Law of requisite hierarchy: The weaker and more uncertain the regulatory capability, the more hierarchy is needed in the organisation of regulation and control to get the same result – where systems (or people) are not self-regulating, then strong management supervision becomes essential.

- Sub-optimisation: If each subsystem, regarded separately, is made to operate with maximum efficiency, the system as a whole will not operate with utmost efficency. In practice, the goal of the entire organisation must be the first priority, even if this is achieved at the expense of some projects or departments.

- 80/20 principle: In any large, complex system, eighty per cent of the output will be produced by only twenty per cent of the system. This applies to people, equipment, products, etc, and it is worth trying to identify the most productive 20% to make sure that it continues to function effectively.

- Redundancy of resources principle: Maintenance of stability under conditions of disturbance requires redundancy of critical resources. So make sure that you have spare parts for critical equipment, and staff who are able to fill in if a key employee is off sick!

- The environment-modification principle: To survive, systems have to choose between two main strategies. One is to adapt to the environment, the other is to change it. And after all, changing the (customer) environment is what marketing is all about.

3.6 Exercises

3.6.1   Systems concepts

Consider the following two systems:

1. *McDonalds hamburger franchise in Rondebosch*

2. *A large hospital, e.g. Groote Schuur*

- Give examples for *each* system of the system goal or purpose, 2 inputs, 2 components, 2 processes, 2 outputs, one sub-system, one supra-system and an external feedback mechanism.

- How would you measure whether the ***entropy*** of each system increases or decreases?

3.6.2   Systems Thinking

"Systems thinking" refers to trying to solve problems by looking at their wider context, and examining the root causes, interactions, feedback mechanisms, etc.

- Think of an example of a problem that cannot be solved effectively by addressing only its apparent symptoms, and explain why this is the case.

3.6.3   The Systems View

Make a list of at least 10 different systems that you are a part of in everyday life.

CASE STUDY: GREENFINGERS GARDEN SERVICES

Greenfingers Garden Services was established in Cape Town in 1990, starting as a small suburban garden maintenance service (mowing lawns, weeding flowerbeds, etc), employed by a limited number of clients on an annual contract basis. Over the last decade the business has expanded considerably, and it now has three branches (in the Southern Suburbs, Northern Suburbs, and Camps Bay area) which offer consulting and landscaping services, one-off "spring cleans" as well as regular garden maintenance. The company has both residential and corporate clients, and ascribes its success to the efficient and personal service that it has always provided.

The main Southern Suburbs branch is run by the owner, Alice Cooper, and is responsible for overall management of the business, including all the accounting and general administration functions, marketing, consulting, and stock. Each branch manages its own supplies of gardening equipment and schedules its labour force (permanent and casual staff) to ensure that jobs are completed on time.

In more detail, the business operates as follows:

- The **marketing** function, headed by Steve Carlton, includes newspaper, yellow pages and radio advertising, colour brochures which can be left with prospective clients, and regular phone calls to existing clients to check that they are happy with the service they have received. Garden service foremen are also supposed to suggest additional services to clients, such as pruning, replanting or fertilizing, but in many cases the householder is not home at the time of the garden service. Steve relies on information received from the various branches to let him know when they have spare capacity, in which case flyers are distributed in letter boxes within a restricted area, offering special deals on garden cleanups or new contracts for a limited time. Steve deals with all responses and queries, provides quotations, draws up contracts, and then passes on new client information to the accounts department.

- The **accounts** department receives new client information from Steve Carlton, and receives a daily listing from each branch of jobs completed for invoicing. In addition, they order equipment and supplies for all branches, schedule vehicle maintenance, attend to creditor payments and client receipts, and maintain the payroll for permanent and casual staff.

- Alice Cooper oversees the business **administration** herself, and does her best to keep costs down by having equipment repaired rather than replaced, and by ensuring that the permanent staff is kept as small as possible by utilising casual labour. However, this increases the complexity of management, since she needs to know a week ahead of time how many casuals will be required, so as to ensure that she will not be short-staffed. Also, while equipment is being repaired, it is often necessary to reschedule jobs or try to shift lawnmowers etc between branches, involving a number of telephone calls.

- The garden **equipment** itself (mowers, trimmers, sprayers, clippers, spades & forks, etc) is kept in store at each branch, and issued to each team at the start of the day. At

the end of the day it is checked back in, and any problems or breakages are reported to central administration at the main branch. Central admin checks whether or not the item is still under guarantee, and arranges for any necessary repairs to be undertaken. Each branch also has a number of light delivery vehicles for transporting work teams and equipment to the clients' premises.

- Greenfingers does not directly employ any **landscaping** consultants, but maintains a short list of landscapers and other garden specialists who can be called on when required. When a landscaping job is undertaken, the consultant is paid a commission based on the value of the job, but Greenfingers make use of their own labour and source the plants themselves. They have arrangements with a wholesale nursery for the provision of shrubs etc at a discount price. Heavy machinery for earthmoving or tree removal is hired when needed.

- The actual **work teams** who carry out the garden maintenance jobs, are managed by their specific branches. Based on the orders received via marketing, and the regular scheduled contract jobs, the admin department informs each branch on Mondays of their jobs for the following week. The branch manager must then check that labour and equipment will be available, and inform central admin of any potential problems so that they can be resolved in good time. Each work team has a foreman and at least one gardener who are on the permanent staff, and additional casual workers are hired by the day. Once the work teams have left for the client properties each morning, the branch manager compiles a list of the previous day's completed jobs, the staff employed, and any damaged equipment, for submission to central admin.

With the growth of the business, its administration is turning into a nightmare, and Alice finds that her evenings and weekends are being spent collating information, working out schedules, and trying to overcome problems by sharing equipment and even staff between branches. More and more clients are complaining that special instructions given to Steve Carlton are not being carried out by their work team, or that discounts that were promised have been omitted from the invoice. Alice is also aware that opportunities are being missed for the provision of extra services such as spraying and pruning, which are highly profitable since the work team is already on site and no additional time or cost is needed for transport.

(a) Construct a labelled diagram that will show the flow of information between central administration, marketing, accounts, each of the branches, the work teams, and the consultants.

(b) Explain how the automation of this information flow could improve the running of Greenfingers Garden Services.

(c) At a Transaction Processing level, what data would need to be captured in order to provide the input for an automated information system?

(d) Provide a comprehensive list of useful reports that could be produced by a Management Information System, and specify who each would be used by.

(e) What feedback mechanisms could be put in place to monitor and control business activities?

Section II: IS Technologies

An **information system** is a system that gathers and transforms data in order to produce information for its end-users. Typical technologies that are incorporated in this process are hardware, software, data and networks. (The fifth important component of an information system, people, is not classified as a technology!)

These same components are found in small, informal information systems such as the one you use to keep track of the contact details of your friends and family. You may have phone numbers stored in an address book, on your cell phone, and on the cover of your exam pad (hardware); you have particular rules about whose numbers you are going to keep and how you store them (software); the collection of numbers that you already have forms your database; and when you update a number in your address book from the back of the exam pad, you are transferring data between resources. But for the purposes of this book, we will focus on the components of larger, computerised information systems, operating within the environment of a commercial organisation.

Hardware

The most easily visible component of an information system is the *hardware*. One can typically distinguish between the media used to store the information and the capital equipment used to read, write and store the media. In *manual* information systems, paper is perhaps the major information medium, although microfilm, microfiche and audio/video tape are also still popular. Some of the capital equipment associated with manual systems are pens, typewriters, photocopiers, filing cabinets and shelves, binders, staplers, shredders, microfilm readers, recorders. Many of these manual technologies are still a significant component of c*omputerised* information systems but the latter also use various optical and magnetic disc media which can be read using only sophisticated computer input and output devices such as scanners, printers, monitors and projectors. Other hardware includes the computer itself, storage devices such as hard drives and a whole "alphabet zoo" of network or communication devices and cabling.

Software and Procedures

Software consists of the methods used to record, transform and extract information. The term is used mainly in the context of computer-based systems where it refers to the computer programs containing detailed instructions that govern their data processing. Although computer hardware used to receive most attention in the past, software has since become far more important than hardware, both in terms of monetary investment and management attention.

Generally considered part of the information systems software are the documentation, instructions and training that are needed to use the software.

The various *organisational procedures* for humans also form part of an information system's software, though this is extending the concept of software to the realm of people, not just computers. This would include the paper forms, procedures and guidelines for handling in-

coming telephone calls, dealing with mail, processing orders and purchases, internal communication, meetings, decision criteria etc. For conceptual purposes, it should make no difference whether a certain procedure is executed by people or by computers.

Data

Although a computer stores program software and data in the same way, from a system's viewpoint data is a very important and separate subsystem. Only recently have organisations (and individuals) started to realise that their databases represent huge investments, which potentially contain significant value.

Corporate data should have its own guardians such as database administrators who control access, integrity, format and security. Careful analysis of organisational data requirements is needed to be able to produce useful management information to support both current and future decision-making. In addition, the data storage mechanisms that are implemented are likely to influence the ease with which current and historical data can be retrieved and processed.

Networks

The *networking and communications infrastructure* has recently been considered to be a separate component of information systems, although in some cases this distinction can be regarded as a matter of personal opinion. There are definitely a number of specific decision and management issues associated with the network infrastructure. On the other hand, the network infrastructure also consists of hardware and software components, most of which share attributes with the mainstream information systems hardware and software mentioned above.

Organisations have realised that *stand-alone* computers present many problems: fragmentation of data, lack of control, insufficient integration and limited opportunity for teamwork. One of the major trends over the last decade was the move not only to have a personal computer on the desk of virtually every knowledge worker, but to have that computer linked to the other computers in the organisation.

Although these different technologies will be described separately in the chapters that follow, it is important to remember that they all need to work together in order to achieve the goals of any particular organisation. Compatibility of the various system components is vital, since the outputs of one technology may well provide the inputs for another. And above all, the purpose of an information system is to provide useful information to **people** (the non-technical component of the system), so the selection of technologies to be used should be based on the need to meet human requirements.

4. Hardware

Hardware is the most visible part of any information system: the equipment such as computers, scanners and printers that is used to capture data, transform it and present it to the user as output. Although we will focus mainly on the personal computer (PC) and the peripheral devices that are commonly used with it, the same principles apply to the complete range of computers:

- Supercomputers, a term used to denote the fastest computing engines available at any given time, which are used for running exceptionally demanding scientific applications.

- Mainframe computers, which provide high-capacity processing and data storage facilities to hundreds or even thousands of users operating from (dumb) terminals.

- Servers, which have large data storage capacities enabling users to share files and application software, although processing will typically occur on the user's own machine.

- Workstations, which provide high-level performance for individual users in computationally intensive fields such as engineering.

- Personal computers (including laptop/notebook computers) have a connected monitor, keyboard and CPU, and have developed into a convenient and flexible business tool capable of operating independently or as part of an organisational network.

- Mobile devices such as personal digital assistants or the latest generation of cellular telephones, offer maximum portability plus wireless connection to the internet, although they do not offer the full functionality of a PC.

And we are already moving into the age of wearable computers for medical or security applications, embedded computers in appliances ranging from motor cars to washing machines, and the smart card which will provide identification, banking facilities, medical records and more!

4.1 Input devices

Data may enter an information system in a variety of different ways, and the input device that is most appropriate will usually depend on the type of data being entered into the system, how frequently this is done, and who is responsible for the activity. For example, it would be more efficient to scan a page of typed text into an information system rather than retyping it, but if this happens very seldom, and if typing staff are readily available, then the cost of the scanner might not be justified. However, all of the input devices described in this chapter have at least one thing in common: the ability to translate non-digital data types such as text, sound or graphics into digital (i.e. binary) format for processing by a computer.

4.1.1 The keyboard

A lot of input still happens by means of a keyboard. Usually, the information that is entered by means of a keyboard is displayed on the monitor. The layout of most keyboards is similar

to that of the original typewriter on which it was modelled. Ironically, this "QWERTY" keyboard layout was originally designed to slow the operator down, so that the keys of the typewriter would not get stuck against each other. This layout now works counter-productively since a computer can process keyboard input many times faster than even the fastest typist can manage. A number of attempts have been made to design alternative layouts by rearranging the keys (the Dvorak keyboard) or by reducing the number of keys. None of these alternative designs has really caught on. Special keyboards have also been designed for countries that use a non-Roman alphabet, and also for disabled people.

4.1.2   Pointing devices

The now ubiquitous electronic *mouse* is an essential input device for use with any *graphical user interface*. It consists of a plastic moulded housing, designed to fit snugly in the palm of the hand, with a small ball at its bottom. Moving the mouse across a flat surface will translate the movements into a rolling action of the ball. This is translated into electronic signals that direct the corresponding movement of a cursor on the computer monitor. Buttons on the mouse can then be used to select icons or menu items, or the cursor can be used to trace drawings on the screen.

The less popular *trackball* operates exactly like an "upside-down" mouse except that the ball is much larger and, instead of the mouse being moved over a surface, the user manipulates the ball directly. Since the trackball can be built into the side of the keyboard, it obviates the need for a free surface area and is therefore handy in situations where desktop surface area is at a premium or not available. Originally popular in educational laboratory settings and for laptop computers, trackballs are now mainly confined to exhibition displays and other public terminals.
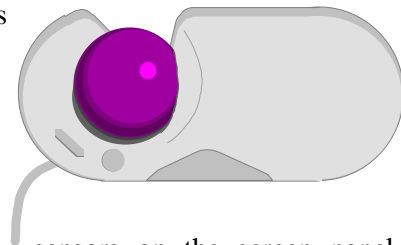
Touch-screens are computer monitors that incorporate sensors on the screen panel itself or its sides. The user can indicate or select an area or location on the screen by pressing a finger onto the monitor. *Light and touch pens* work on a similar principle, except that a stylus is used, allowing for much finer control. Touch pens are more commonly used with handheld computers such as personal organisers or digital assistants. They have a pen-based interface whereby a stylus (a pen without ink) is used on the small touch-sensitive screen of the handheld computer, mainly by means of ticking off pre-defined options, although the fancier models support data entry either by means of a stylised alphabet, which resembles a type of shorthand, or some other more sophisticated handwriting recognition interface.

Digitiser tablets also use a pressure sensitive area with a stylus. This can be used to trace drawings. A similar conceptual approach is used for the *touch pad* that can be found on the majority of new notebook computers, replacing the more awkward joystick or trackball. The user controls the
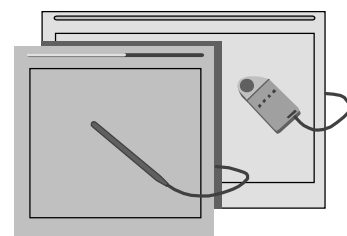
cursor by moving a finger across a fairly small rectangular touch-sensitive area below the keyboard, usually about 5 cm by 7 cm.

A large number of *game interfaces* have been developed to provide a more realistic and natural interface in various gaming situations and simulations: the joy stick, steering wheel, foot pedal and other gaming devices. They all perform functions similar to the mouse in that they allow the user to control a cursor or simulate generally real-time motion control. Contact your nearest game arcade for details.

Although the *data glove* also fits under the previous category, it is technically a lot more complex. It looks like a hand glove but contains a large number of sensors and has a data cable attached; though the latter is being replaced by means of infrared cordless data transmission. Not only does the data glove allow for full three-dimensional movement but it also senses the position of individual fingers, translating this into a grip. The glove is currently used in virtual reality simulators where the user moves around in an artificially rendered environment projected onto tiny LCD screens fitted into vision goggles. The computer generates various imaginary objects, which the user can "pick up" and manipulate by means of the glove. Advanced models even allow for tactile feedback by means of small pressure pockets built into the glove.

4.1.3   Optical scanners and readers

There are a number of different optical scanner technologies on the market.

§   *Optical Scanners* use light-emitting devices to illuminate the printing on paper. Depending on how much light is reflected, a light-sensor determines the position and darkness (or colour) of the markings on the paper. Special-purpose optical scanners are in use by postal services to read and interpret hand-written postal codes. General-purpose scanners are used with personal computers to scan in images or text. These vary from handheld devices (see picture) to flatbed scanners which feed input documents one sheet at a time. A common use of optical scanners is the scanning of black-and-white or colour images and pictures. When scanning text, it is necessary to load additional *optical character recognition (OCR)* software that converts the scanned raster-image of the text into the equivalent character symbols, so that they can be edited using word processing software.
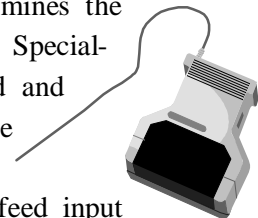
§   *Barcode scanners* detect sequences of vertical lines of different widths, the ubiquitous barcode as found also on the back of this book. These scanners have become very popular with retailers due to the fact that all pre-packaged products are now required to have a product bar code on their packaging, following the standard laid down by the South African Article Numbering Association (SAANA). Libraries and video shops now also commonly use bar code scanners. They are more generally used for tracking and routing large numbers of physical items such as for asset inventory purposes in many larger organisations, postal items by the postal services and courier services, or for luggage handling by airlines.

§   *Optical mark readers* are capable of reading dark marks on specially designed forms. The red multiple choice answer sheets in use at many educational and testing institutions are a good example.

4.1.4   Other input devices

A *magnetic card reader* reads the magnetised stripe on the back of plastic credit-card size cards. These cards need to be pre-recorded following certain standards. Although the cards can hold only a tiny amount of information, they are very popular for access (door) control and financial transactions (ATMs and point-of-sale terminals).

Magnetic ink character recognition (MICR) uses a special ink (containing magnetisable elements) and a distinct font type. It is used mainly in the banking sector for the processing of cheques.

Touch-tone devices can use a voice telephone to contact computer-based switchboards or enter information directly into remote computers. Many corporate telephone help-lines rely on the customer pressing the touch-tone telephone buttons to route his/her call to the correct operator by selecting through a menu of possible options. South African banks also enable their clients to perform a number of banking transactions via telephone.

Digital cameras allow you to make pictures of physical objects directly in a digital, i.e. computer-readable, format. Relatively low-cost digital still picture cameras are now available that capture images directly on electronic disk or RAM media instead of the traditional film. Apart from being very compact, most of these digital cameras can also interface directly with personal computers and are thus becoming a popular tool to capture pictures for e-mailing or loading on the world-wide Web.

Biometric devices are used to verify personal identity based on fingerprints, iris or retinal scanning, hand geometry, facial characteristics etc. A scanning device is used to capture key measurements and compare them against a database of previously stored information. This type of authentication is becoming increasingly important in the control of physical access.

Finally, *voice input devices* are coming of age. Voice-recognition has recently made a strong entry into the market with the availability of low-cost systems that work surprisingly well with today's personal computers. These systems allow for voice control of most standard applications (including the operating system). With voice control, the computer recognises a very limited number (50 or less) of frequently used, programmable system commands ("save", "exit", "print"…) from a variety of users. In fact, these systems are not only used for the interface of computer programs; they are also slowly making an appearance in consumer appliances, novelty items and even motor cars!

Much more difficult to achieve than voice control, is true *voice dictation* used to dictate e.g. a letter to your word processor. The difficulty is that the computer must not only distinguish between many tens of thousands of possible words, but it must also recognise the almost unnoticeable breaks in between words, different accents and intonations. Therefore, voice dictation typically requires a user to train the voice recognition software by reading standard

texts aloud. Nevertheless, for personal purposes and slow typists, voice recognition is rapidly becoming a viable alternative to the keyboard.

4.2 Central Processing Unit (CPU)

Once data has been entered into a computer, it is acted on by the CPU, which is the real brain of the computer. The CPU takes specific program instructions (usually one at a time), applies them to the input data and transforms the input into output.

4.2.1 Components of the CPU.

The CPU has two major components.

- The *Arithmetic and Logic Unit (ALU)* executes the actual instructions. It knows how to add or multiply numbers, compare data, or convert data into different internal formats.

- The *Control Unit* does the "housekeeping" i.e. ensures that the instructions are processed on time, in the proper sequence, and operate on the correct data.

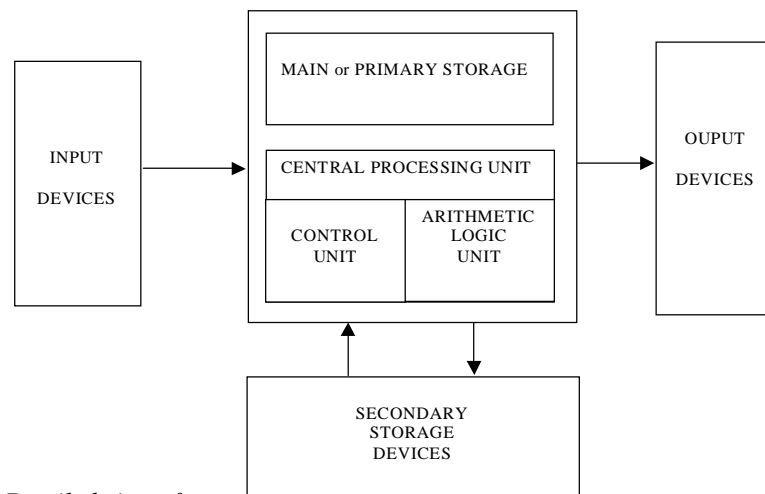Figure 4-1 gives a detailed representation of a computer system.

Figure 4-1: Detailed view of a   computer system.

4.2.2 Types of CPUs

The CPU is an electronic device based on microchip technology, hence it is also often called the *microprocessor*. It is truly the showcase and culmination of the state-of-the-art in the electronics industry: a tiny silicon-based chip occupying less than 1 square cm contains several millions of transistor elements, measuring less than a thousandth of a millimetre across. They operate at speeds way beyond our comprehension: a typical CPU can multiply more 7-digit numbers in one second than a human could do in ten lifetimes, but uses less energy than a light bulb!

Think of the motor car industry: there are different manufacturers or makes of cars (Volkswagen, Toyota, etc.), each with different models (Golf, Jetta, …), which come out in different versions (City Golf, Sports model, coupé, etc.). In addition, there exist *custom-made*

special-purpose cars. It is the same in the computer chip business. There are many different types of CPUs on the market. The best-known manufacturer is Intel, which produces the microprocessors for the IBM-compatible personal computer (*PC*). Some of its competitors produce clones or imitations (e.g. AMD), others manufacturers produce different types of microprocessors or concentrate on small volumes of highly specialised or very fast microprocessors. Intel has produced a large number of CPU types: the earliest model used in the Personal Computer was the 8088, followed by the 8086, the 80286, the 386, 486 and the line of Pentium processors.

4.2.3 Speed of processing

How does one measure the speed of, say a Porsche 911? One could measure the time that it takes to drive a given distance e.g. the 900 km from Cape Town to Bloemfontein takes 4½ hours (ignoring speed limits and traffic jams). Alternatively, one can indicate how far it can be driven in one standard time unit e.g. the car moves at a cruising speed of 200 km/hour.

In the same way, one can measure the speed of the CPU by checking the time it takes to process one single instruction. As indicated above, the typical CPU is very fast and an instruction can be done in about two billionths of a second. To deal with these small fractions of time, scientists have devised smaller units: a *milli*second (a thousandth of a second), a *micro*second (a millionth), a *nano*second (a billionth) and a *pico*second (a trillionth).

However, instead of indicating the time it takes to execute a single instruction, the processing speed is usually indicated by how many instructions (or computations) a CPU can execute in a second. This is exactly the inverse of the previous measure; e.g. if the average instruction takes two billionths of a second (2 nanoseconds) then the CPU can execute 500 million instructions per second (or one divided by 2 billionths). The CPU is then said to operate at 500 *MIPS* or 500 *million of instructions per second.* In the world of personal computers, one commonly refers to the rate at which the CPU can process the simplest instruction (i.e. the clock rate). The CPU is then rated at 500 *MHz* (megahertz) where mega indicates million and Hertz means "times or cycles per second". For powerful computers, such as workstations, mainframes and supercomputers, a more complex instruction is used as the basis for speed measurements, namely the so-called floating-point operation. Their speed is therefore measured in *megaflops* (*million of floating-point operations per second*) or, in the case of very fast computers, *teraflops* (billions of flops).

In practice, the speed of a processor is dictated by four different elements: the "clock speed", which indicates how many simple instructions can be executed per second; the word length, which is the number of bits that can be processed by the CPU at any one time (64 for a Pentium IV chip); the bus width, which determines the number of bits that can be moved simultaneously in or out of the CPU; and then the physical design of the chip, in terms of the layout of its individual transistors. The latest Pentium processor has a clock speed of about 4 GHz and contains well over 100 million transistors. Compare this with the clock speed of 5 MHz achieved by the 8088 processor with 29 000 transistors!

Moore's Law (see Figure 4-2) states that processing power doubles for the same cost approximately every 18 months.
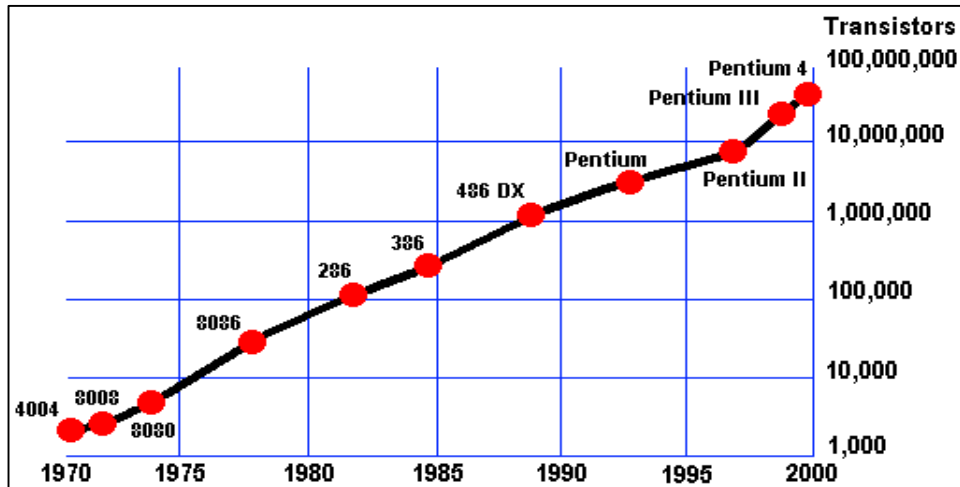
Figure 4-2. Illustration of Moore's Law

4.2.4 Von Neumann versus Parallel CPU Architecture

The traditional model of the computer has one single CPU to process all the data. This is called the Von Neumann architecture because he engineered this approach to computers in the days when computers were still a dream.

Except for entry-level personal computers, most computers now have two, four, or up to sixteen CPUs sharing the main processing load, plus various support processors to handle maths processing, communications, disk I/O, graphics or signal processing. In fact many CPU chips now contain multiple "cores" each representing an individual CPU.

Some super-computers that have been designed for massive parallel processing, have up to 64,000 CPUs. These computers are typically used only for specialised applications such as weather forecasting or fluid modelling. Today's supercomputers are mostly clusters (tight networks) of many thousands of individual computers.

4.2.5 Possible Future CPU Technologies

Perhaps the major future competitor of the microchip-based microprocessor is *optical computing*. Although the technology for developing electronic microchips suggests that CPUs will continue to increase in power and speed for at least the next decade or so, the physical limits of the technology are already in sight. Switching from electronic to light pulses offers a number of potential advantages: light (which consists of photons) can travel faster, on narrower paths and does not disperse heat. In theory, one can even process different signals (each with a different light frequency) simultaneously using the same channel. Although the benefits of optical processing technology have already been proven in the areas of data storage (CD-Rom, CD-R) and communication (fibre optics), the more complex all-optical switches required for computing are still under development in the research laboratories.
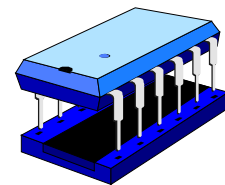
A very experimental alternative to optical and electronic technologies is the *organic computer*. Research indicates that, for certain applications, it is possible to let a complex organic molecule act as a primitive information processor. Since even a tiny container filled

with the appropriate solutions contains many trillions of these molecules, one obtains in effect a hugely parallel computer. Although this type of computer can attack combinatorial problems way beyond the scope of traditional architectures, the main problem is that the programming of the bio-computer relies entirely on the bio-chemical properties of the molecules.

Another exciting but currently still very theoretical development is the possible use of quantum properties as the basis for a new type of computer architecture. Since quantum states can exist in juxtaposition, a register of qubits (a bit value in quantum state) takes on all the possible values simultaneously until it is measured. This could be exploited to speed up extremely parallel algorithms and would affect such areas as encryption, searching and error-correction. To date, experimental computers with a few qubits have been built but the empirical validation of the actual usefulness of quantum computing still remains an open question.

4.3 Main Memory

The function of main memory (also referred to as primary memory, main storage or internal storage) is to provide temporary storage for instructions and data during the execution of a program. Main memory is usually known as RAM, which stands for Random Access Memory. Although microchip-based memory is virtually the only technology used by today's computers, there exist many different types of memory chips.

4.3.1   Random Access Memory (RAM)

RAM consists of standard circuit-inscribed silicon microchips that contain many millions of tiny transistors. Very much like the CPU chips, their technology follows to the so-called *law of Moore*, which states that they double in capacity or power (for the same price) every 18 months. A RAM chip easily holds hundreds of Megabytes (million characters). They are frequently pre-soldered in sets on tiny memory circuit boards called *SIMMS* (*Single In-line Memory Modules*) or *DIMMS (Dual ...)* which slot directly onto the *motherboard*: the main circuit board that holds the CPU and other essential electronic elements. The biggest disadvantage of RAM is that its contents are lost whenever the power is switched off.

There are many special types of RAM and new acronyms such as EDO RAM, VRAM etc. are being created almost on a monthly basis. Two important types of RAM are:

§   *Cache memory* is ultra-fast memory that operates at the speed of the CPU. Access to normal RAM is usually slower than the actual operating speed of the CPU. To avoid slowing the CPU down, computers usually incorporate some more expensive, faster *cache RAM* that sits in between the CPU and RAM. This cache holds the data and programs that are needed immediately by the CPU. Although today's CPUs already incorporate an amount of cache on the circuit itself, this on-chip cache is usually supplemented by an additional, larger, cache on the motherboard.

§   *Flash RAM or flash memory* consists of special RAM chips on a separate circuit board within a tiny casing. It fits into custom ports on many notebooks, hand-held computers

and digital cameras. Unlike normal RAM, flash memory is non-volatile i.e. it holds it contents even without external power, so it is also useful as a secondary storage device.

4.3.2   Read-Only Memory (ROM)

A small but essential element of any computer, ROM also consists of electronic memory microchips but, unlike RAM, it does not lose its contents when the power is switched off. Its function is also very different from that of RAM. Since it is difficult or impossible to change the contents of ROM, it is typically used to hold program instructions that are unlikely to change during the lifetime of the computer. The main application of ROM is to store the so-called *boot program*: the instructions that the computer must follow just after it has been switched on to perform a self-diagnosis and then tell it how load the operating system from secondary storage. ROM chips are also found in many devices which contain programs that are unlikely to change over a significant period of time, such as telephone switch boards, video recorders or pocket calculators. Just like RAM, ROM comes in a number of different forms:

§   **PROM** *(Programmable Read-Only Memory)* is initially empty and can be custom-programmed *once only* using special equipment. Loading or programming the contents of ROM is called burning the chip since it is the electronic equivalent of blowing tiny transistor fuses within the chip. Once programmed, ordinary PROMs cannot be modified afterwards.

§   **EPROM** (*Erasable Programmable Read-Only Memory*) is like PROM but, by using special equipment such as an ultra-violet light gun, the memory contents can be erased so that the EPROM can be re-programmed.

§   **EEPROM** (*Electrically Erasable Programmable Read-Only Memory*) is similar to EPROM but it can be re-programmed using special electronic pulses rather than ultra-violet light so no special equipment is required.

4.4   Secondary Storage Devices

Since the main memory of a computer has a limited capacity, it is necessary to retain data in secondary storage between different processing cycles. This is the medium used to store the program instructions as well as the data required for future processing. Most secondary storage devices in use today are based on magnetic or optical technologies.

4.4.1   Disk drives

The disk drive is the most popular secondary storage device, and is found in both mainframe and microcomputer environments. The central mechanism of the disk drive is a flat disk, coated with a magnetisable substance. As this disk rotates, information can be read from or written to it by means of a head. The head is fixed on an arm and can move across the radius of the disk. Each position of the arm corresponds to a "track" on the disk, which can be visualised as one concentric circle of magnetic data. The data on a track is read sequentially as the disk spins underneath the head. There are quite a few different types of disk drives.

In *Winchester hard drives*, the disk, access arm and read/write heads are combined in one single sealed module. This unit is not normally removable, though there are some models available where the unit as a whole can be swapped in and out of a specially designed drive bay. Since the drives are not handled physically, they are less likely to be contaminated by dust and therefore much more reliable. Mass production and technology advances have brought dramatic improvements in the storage capacity with Terabyte hard drives being state of the art at the end of 2006. Current disk storage costs as little R1 per gigabyte.

Large organisations such as banks, telcos and life insurance companies, require huge amounts of storage space, often in the order of many terabytes (one terabyte is one million megabytes or a trillion characters). This was typically provided by a roomful of large, high-capacity hard drive units. Currently, they are being replaced increasingly by *redundant arrays of independent disks (RAIDs)*. A RAID consists of an independently powered cabinet that contains a number (10 to 100) of microcomputer Winchester-type drives but functions as one single secondary storage unit. The advantage of the RAID is its high-speed access and relatively low cost. In addition, a RAID provides extra data security by means of its fault-tolerant design whereby critical data is mirrored (stored twice on different drives) thus providing physical data redundancy. Should a mirrored drive fail, the other drive steps in automatically as a backup.
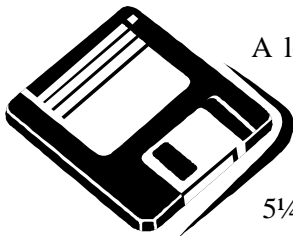
A low-cost, low-capacity version of the hard disk was popularised by the microcomputer. The *diskette* consists of a flexible, magnetic surface coated mylar disk inside a thin, non-removable, plastic sleeve. The early versions of the diskette were fairly large (8" or 5¼") and had a flexible sleeve, hence the name floppy diskette. These have rapidly been replaced by a diskette version in a sturdier sleeve, the stiffy disk, that despite its smaller size (3 ½") can hold more data. Although the popular IBM format only holds 1,44 megabytes, a number of manufacturers have developed diskette drives that can store from 100 to 250 megabytes per stiffy. An alternative development is the *removable disk cartridge*, which is similar in structure to an internal hard drive but provides portability, making it useful for backup purposes.

4.4.2 Magnetic tape

While disk and optical storage have overtaken magnetic tape as the most popular method of storing data in a computer, tape is still used occasionally – in particular for keeping archive copies of important files.
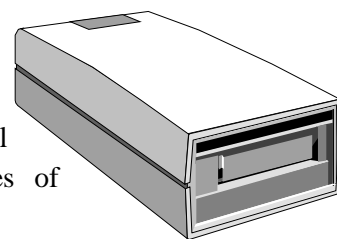
The main drawback of magnetic tape is that it is not very efficient for accessing data in any way other than strictly sequential order. As an illustration, compare a CD player (which can skip to any track almost instantly) with a music tape recorder (which has to wind the tape all the way through if one wants to listen to a song near the end). In computer terms, the ability to access any record, track, or even part within a song directly is called the *direct access* method. In the case of the tape recorder one may have to wind laboriously through the tape until one reaches the song required – this is referred to as the *sequential access* method.

The high-density diskette and recordable optical disk have all but eroded the marginal cost advantage that tape storage enjoyed. This technology is therefore disappearing fast.

4.4.3 Optical disk storage
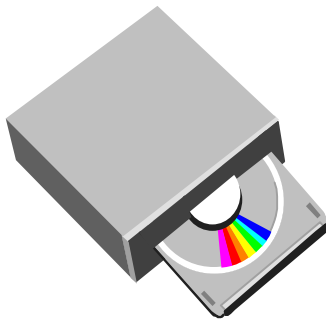
Optical disks, on the other hand, are rapidly becoming the storage medium of choice for the mass distribution of data/programs and the backup of data. Similar to disk storage, information is stored and read from a circular disk. However, instead of a magnetic read head, a tiny laser beam is used to detect microscopic pits burnt onto a plastic disk coated with reflective material. The pits determine whether most of the laser light is reflected back or scattered, thus making for a binary "on" or "off". In contrast to hard disks, data is not stored in concentric cylinders but in one long continuous spiral track.

Trivial fact: The spiral track used to store data on a CD is over six kilometres long.

A popular optical disk format is the 12-cm *CD-ROM*. The widespread use of music compact discs has made the technology very pervasive and cheap. Production costs for a CD-ROM are less than R1, even for relatively small production volumes. The drive reader units themselves have also dropped in price and are now hardly more than the cost of a diskette drive. A standard CD-ROM can store 650 megabytes of data and the data can be transferred at many megabytes per second, though accessing non-sequential data takes much longer.

The CD-ROM is a read-only medium. Data cannot be recorded onto the disk. The low cost and relatively large capacity makes the CD-ROM ideally suited to the distribution of software. They are also ideal for the low-cost distribution of large quantities of information such as product catalogues, reference materials, conference proceedings, databases, etc. It is indispensable for the storage of multimedia where traditional textual information is supplemented with sound, music, voice, pictures, animation, and even video clips.

The limitation of the read-only format lead to the development of low-cost recordable optical disks. The *compact disk recordable (CD-R)* is a *write-once, read-many (WORM)* technology. The CD-R drive unit takes a blank optical disk and burns data onto it using a higher-powered laser. This disk can then be read and distributed as an ordinary CD-ROM, with the advantage that the data is non-volatile i.e. permanent. The rapid drop in the cost of drive units and blank recording media (less than R2 per CD-R) is making this a very competitive technology for data backup and small-scale data distribution.

Although the 650 megabytes initially seemed almost limitless, many multimedia and video applications now require more storage. A new format, the *Digital Video Data (DVD)* standard increased the capacity of the CD-ROM by providing high-density, double-sided and double-layered CDs. By combining the increased storage capacity with sophisticated data compression algorithms, a DVD disc can easily store 10 times as much as a CD, sufficient for a full-length high-quality digital motion picture with many simultaneous sound tracks.

Even the DVD is not sufficient storage capacity and currently two optical technologies have

been developed to increase storage capacity even further. The basic specification of both HD-DVD and Blu-Ray provide for more than 25 GB of storage on a disc although multi-layer Blu-Ray discs with capacities of more than 200 GB have already been developed.

A promising research area involves the use of *holographic disk storage* whereby data is stored in a three-dimensional manner. Though in its infancy, early prototypes promise a many-fold increase in storage capacity and it could become the answer to the ever increase storage requirements of the next decade

Device	Access Speed	Capacity	Cost
RAM	< 2 nanosec	256 MB (chip)	<R1/MB
Tape	serial only	500 MB-4 GB	<10c/MB
Diskette (3 ½")	300 ms	1,44 MB	R1/MB
PC hard disk	10 ms	40-750 GB	<2c/MB
M/F hard disk	25 ms	100+ GB	R2/MB
CD-ROM	< 100 ms	660 MB	<0.1c/MB
CD-R	< 100 ms	660 MB	<0.2c/MB
DVD	< 100 ms	8 GB	<0.1c/MB
HD-DVD	< 100 ms	30 GB	?
Blu-Ray	< 100 ms	25 GB-200GB	?

*Figure 4-3: Comparison of secondary storage devices*

4.5 Output Devices

The final stage of information processing involves the use of output devices to transform computer-readable data back into an information format that can be processed by humans. As with input devices, when deciding on an output device you need to consider what sort of information is to be displayed, and who is intended to receive it.

One distinction that can be drawn between output devices is that of hardcopy versus softcopy devices. Hardcopy devices (printers) produce a tangible and permanent output whereas softcopy devices (display screens) present a temporary, fleeting image.

4.5.1 Display screens

The desk-based *computer screen* is the most popular output device. The standard *monitor*

works on the same principle as the normal TV tube: a "ray" gun fires electrically charged particles onto a specially coated tube (hence the name *Cathode-Ray Tube* or *CRT*). Where the particles hit the coating, the "coating" is being "excited" and emits light. A strong magnetic field guides the particle stream to form the text or graphics on your familiar monitor.
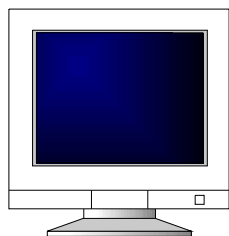
CRTs vary substantially in size and resolution. Screen size is usually measured in inches diagonally across from corner to corner and varies from as little as 12 or 14 inches for standard PCs, to as much as 40+ inches for large demonstration and video-conferencing screens. The screen resolution depends on a number of technical factors.

A technology that has received much impetus from the fast-growing laptop and notebook market is the *liquid crystal display (LCD)*. LCDs have matured quickly, increasing in resolution, contrast, and colour quality. Their main advantages are lower energy requirements and their thin, flat size. Although alternative technologies are already being explored in research laboratories, they currently dominate the "flat display" market.

Organic light-emitting diodes (OLED) can generate brighter and faster images than LED technology, and require thinner screens, but they have less stable colour characteristics, making them more suitable for cellular telephone displays than for computers.

Another screen-related technology is the *video projection unit*. Originally developed for the projection of video films, the current trend towards more portable LCD-based lightweight projectors is fuelled by the needs of computer-driven public presentations. Today's units fit easily into a small suitcase and project a computer presentation in very much the same way a slide projector shows a slide presentation. They are rapidly replacing the flat transparent LCD panels that needed to be placed on top of an overhead projection unit. Though the LCD panels are more compact, weigh less and are much cheaper, their image is generally of much poorer quality and less bright.

### 4.5.2   Printers and plotters

Printers are the most popular output device for producing permanent, paper-based computer output. Although they are all *hardcopy* devices, a distinction can be made between impact and non-impact printers. With impact printers, a hammer or needle physically hits an inked ribbon to leave an ink impression of the desired shape on the paper. The advantage of the impact printer is that it can produce more than one simultaneous copy by using carbon or chemically-coated paper. Non-impact printers, on the other hand, have far fewer mechanically moving parts and are therefore much quieter and tend to be more reliable.

The following are the main types of printers currently in use.

§   *Dot-matrix printers* used to be the familiar low-cost printers connected to many personal computers. The print head consists of a vertical row of needles each of which is individually controlled by a magnet. As the print head moves horizontally across the paper, the individual needles strike the paper (and ribbon in between) as directed by the control mechanism to produce text characters or graphics. A close inspection of a dot-matrix printout will reveal the constituent dots that make up the text. Although it is one of the cheapest printer options, its print quality is generally much lower that that of laser and ink-jet printers. However, today's models are quick and give a much better quality by increasing the number of needles.
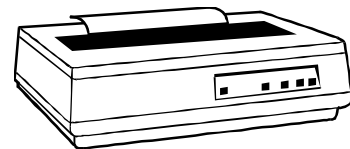
§   *Laser printers* are quickly growing in market share. They work on the same principle as the photocopier. A laser beam, toggled on and off very quickly, illuminates selected areas on a photo-sensitive drum, where the light is converted into electrical charge. As the drum rotates into a "bed" of carbon particles ("*toner*") with the opposite charge, these particles will adhere to the drum. The blank paper is then pressed against the drum so that the
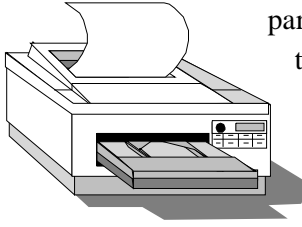
particles "rub off" onto the paper sheet. The sheet then passes through a high-temperature area so that the carbon particles are permanently fused onto the paper. Current high-end laser printers can cope with extremely large printing volumes, as is required e.g. by banks to print their millions of monthly account statements. The laser technology continues to develop in tandem with photocopier technology. Laser printers can now handle colour printing, double-sided printing or combine with mail equipment to perforate, fold, address and seal automatically into envelopes. At the lower end of the scale are the low-cost "personal" laser printers, which give a very good printing quality at a relatively modest cost.

§   *Thermal printers* use heat to print. The older thermal printers used *heat-sensitive* paper, similar to the special fax paper. A slight heat or pressure will leave a darker area. This produced very cheap but low-quality output. Currently, thermal-printing technology is used mainly for high-quality colour printing. These new thermal printers use coloured *wax* sticks and melt the wax onto the paper. Although they are slower than competing colour laser and inkjet technologies, they give a much more vibrant, colour-saturated image.

§   *Inkjet printers* are probably the most popular low-cost printing technology. Liquid ink is squirted onto the paper in the form of tiny droplets. These printers are about the same price as dot-matrix printers, albeit more expensive in terms of consumables. Their quality is close to that of the laser printers. Their great advantage is that the printers can easily be adapted to use coloured ink, thus making popular colour printers.

§   *Plotters* are mainly used for engineering and architectural drawings. A plotter consists of one (or several – in the case of colour plotters) pen(s) affixed to an arm. As the arm moves across the sheet of paper, the pen draws lines onto the paper. It is ideal for line drawings such as plans, especially in cases where the paper size exceeds that which can be accommodated by the other types of printers.
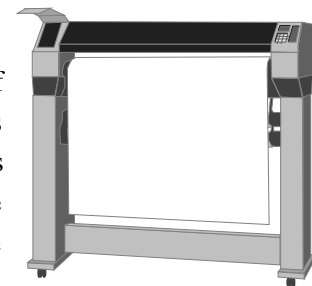
§   *Chain and line printers* are still popular in mainframe environments for the quick production of large volumes of internal printing. The line printer consists of a horizontal, rotating "drum" with 132 cylinders, each containing a full character set. As the 132-column wide paper moves up past the drum, a line at a time, each one of the 132 hammers on the other side of the paper strikes at the exact moment that the corresponding cylinder "shows" the correct character. The hammer hits the drum (and ink ribbon) and leaves an imprint of the character on the paper. The chain printer works on the same principle, but uses a horizontally rotating chain with engraved characters, instead of a drum. As anyone with some working experience in a large organisation knows, the print quality of these "computer printouts" is not very high.