# THE UNIVERSITY OF NOTTINGHAM
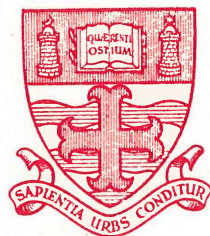
# Applications

# of

# Computers

LECTURE 4.

"BASIC PRINCIPLES - STORAGE"

by

A. S. DOUGLAS, B.Sc., M.A., Ph.D., A.Inst.P.

## "BASIC PRINCIPLES - STORAGE"

### For arithmetic and control operations.

The extreme speed of operation of the arithmetic unit would be of little practical use unless adequate provision is made for the storage of numbers and the control of their flow to and from the arithmetic unit. For this reason some part of the storage must be sufficiently rapidly accessible that no delay is occasioned to the arithmetic operations by drawing upon its contents for information. In principle it is desirable that all storage should be of this nature, but this is not generally practicable for economic and other reasons. It is usual, therefore, to provide at least two (and often more) different 'levels' of storage, each having its own characteristics, and each being of different speed of access. We shall distinguish here between three fundamental types of storage, viz., the storage directly contained within the arithmetic or control units, the storage connected directly to either or both of these units, which we shall call the 'working' store or stores, and the storage connected only to the working store or some part of it, which we shall call the 'backing' store or stores.

In this section we shall consider the properties and nature of any storage required within the control and arithmetic units. Inside these units there is usually some need for storage of information during the execution of fundamental computing instructions either to retain numbers, such as the multiplier multiplicand and partial product during multiplication, or the divisor, dividend and quotient during division, or to hold parts of instruction specifications or special information, such as indications of overflow conditions, sign values and so on. This storage

must as a rule be rapidly accessible in order to match the speed of operation of the fundamental logical circuitry, unless slow overall operation is acceptable.

The most rapidly accessible form of storage is provided by a collection of electronic components known as a 'flip-flop'. These components are normally arranged in an Eccles-Jordan circuit, which employs two thermionic valves connected in such a way that when one valve is conducting the other is shut off, and the configuration is stable in either of the two possible states. The device is thus suitable for the storage of one binary digit or 'bit' of information. The particular setting of the system can be readily detected and the information read out by 'gating' it to other units in times of the order of $10^{-7}$ seconds.

Whilst present day machines use valve circuitry almost exclusively, they do not necessarily employ the Eccles-Jordan circuit. A variant of this circuit using beam-switching in a specially constructed valve has been made to operate (including 'gating') at a speed of $2 \times 10^{-8}$ seconds. Furthermore analogous circuits can be constructed using transistors, and the latter time has been matched with circuitry using specially selected transistors (with over 250 Mc/s cut-off). It would appear theoretically possible, and, in the long run, cheaper to attain even higher speeds with transistors. At present it is not practical to use this type of circuitry on a large scale, since transistors of suitable performance are expensive and in short supply. Nevertheless it would be possible now to provide high performance storage for the arithmetic and control units of a fast machine.

If sufficiently slow overall operation is acceptable, then storage of the kind described can be dispensed with, and a more economical but slower substitute employed. However, it is alternatively possible to

use the faster access store and reduce the requirements for amount of storage in the units discussed. To a considerable extent both in engineering design and, as will be seen later, in programming, storage and speed are interchangeable commodities.

## The working store.

Apart from the storage required during the carrying out of fundamental operations such as multiplication, a large storage is required for numbers and instructions used during the course of computation. For simplicity in programming it is desirable that this store should be as large as possible and that access to all parts of it should be equally easy. It has been found extremely inconvenient to reduce its size below about 50 numbers; there is no known upper limit.

As has been indicated in an earlier lecture, it is possible, though at present unusual (for reasons which will appear later), for instructions and numbers to be stored separately within the computer. In this case there will be two sections of the working store, the numerical section being connected to the arithmetic unit and the instruction section being attached to the control unit. For any particular problem it is always possible to regard the working store as divided in this way. For the purpose of description of the hardware available for storage the essential difference between the two situations is that, whereas it is necessary to write numbers into the store almost as often as they are read out, this is not true of instructions (or of constants), which, once written, may change seldom or never during the course of the computation which they control. A writing system substantially slower than that connected with reading may therefore be tolerable.

It has already been pointed out that a desirable form of working store uses the fastest elements available, but that there are reasons

why this is not practical. Several possible media are available for working storage, and we shall now examine these. They are (1) magnetic cores, (2) regenerative systems and (3) magnetic drums and discs.

## Magnetic Cores

Let us first consider magnetic cores. These consist of toroids of ferrite material with diameters of from 2 to about 13 millimetres. When suitably wound with wire they exhibit characteristic magnetic properties when a current is passed through the wire. A sufficient current causes magnetisation of the core in a given direction and some of this magnetisation remains after the current is switched off. If the current is then reversed, magnetisation occurs in the opposite sense, and this too is retained if the current is cut off. The core thus possesses two stable states of magnetisation and has the properties of storage of a flip-flop. The particular state of a given core can be detected by applying a current in a standard direction. If the core is magnetised in the opposite sense then energy is required to 'turn it over', whereas if it is already magnetised in the standard direction, no such energy is absorbed. This energetic change can be detected by measuring the current in an independent winding round the core. The information thus read off can be used to restore the state of the core later, since this method of reading destroys the information originally held. It is also possible to read from a core system without destroying the information on the cores, but this technique has not so far been applied successfully to construction of a large scale storage unit. It is probable that this technique will be more suited to the storage of constants and instructions than to storage of working numbers, since the speed of writing is likely to be relatively slow.

-4-

The main advantage of magnetic core storage is that it
is easy to assemble in large units, single physical systems often holding
of the order of 16,000 40-bit words. It is relatively cheap, and access
is fairly rapid, normal access times to a word being of the order of
$5 \times 10^{-6}$ seconds, including restoration of the original information on the
cores.

## Regenerative systems

A number of regenerative systems have been used for
working storage in computers. The earliest types were the cathode ray
tube system, due to Williams and Kilburn, and the mercury delay line
system, taken over from radar applications. The basis of any system of
the type is to make use of transient electrical charge, or electromagnetic
sound, or mechanical pulses which exist for long enough that there is time
to regenerate them continuously. Thus the cathode ray tube storage device
makes use of the charge on the surface of the tube when the electron beam
is played upon it. This charge decays slowly after the electron beam is
switched off, but remains detectable for a sufficient period that over 1000
different positions can be scanned between each inspection of the position
originally charged up. On each such inspection the charge is renewed if
necessary, and the state of charge can be read off. Thus on the face of
one tube over 1000 bits can be recorded, each position being either re-
charged or left uncharged on each inspection. Inspections take place
about once every $2 \times 10^{-5}$ seconds in a typical system, and the average
access time is therefore about $10^{-5}$ seconds if the bits of a word are held
in parallel on several synchronised tubes. If all the bits of a word are
on a single tube, then access may be somewhat, though not greatly slower.

Mercury delay lines make use of the piezzoelectric effect
of quartz crystals. The system comprises a column of mercury closed at

-5-

each end by a crystal. Electrical stimulation of one crystal causes it to vibrate and so to transmit a sound wave through the mercury column. This can be detected by the other crystal, and subsequently regenerated. The attenuation of the original pulse is such that it is necessary to reshape the pulse on each entry, and the received pulse is normally only used to trigger a new pulse from an oscillator system.

This new pulse can, of course, be sent out simultaneously to other parts of the machine and this is controlled by an electronic 'gate', itself opened or shut by the control unit as required. Both of these types of storage are becoming obsolete, but are still in use on some machines. However, a similar type of storage is possible on nickel delay lines, using the magneto-strictive effect. It is found that if a suitably arranged wire is stimulated by a magnetic field a vibration can be imparted to the wire at ultrasonic frequency. The wire is then used in a manner similar to the mercury column, pulses being supplied by an electromagnet at one end, and being received by a similar device at the other end.

The actual time of transit of a pulse is determined by the length of the wire or mercury column used. Normal use is in the serial mode, and, if a single word only is stored on one delay line, access is available in times of the order of $10^{-4}$ seconds. In some machines several words are stored on a single delay line, and access is then substantially slower. It is possible, but not usual to employ delay lines in the parallel mode; the average access time would then be approximately 1/2 of the cycle time, i.e. about $5 \times 10^{-5}$ seconds. Special arrangements to make use of the storage more efficiently when used in serial can be made and will be discussed elsewhere.

## Magnetic drums and discs

Whilst it is more usual to employ magnetic drums and discs in the role of backing stores, and this is discussed below, some

machines do use them as working storage; it is, therefore, appropriate to include their description at this stage. A magnetic drum consists of a cylinder of metal, coated on the outside with iron oxide, rotated about an axis parallel to a generator of the cylinder through the centre of its circular cross section. A disc is similar to a drum but has a short linear dimension, and is coated only on its two circular ends. The coated surfaces of disc or drum are such that small areas can be magnetised permanently by the application of suitable magnetic fields. Supported near the surfaces are a set of fixed 'heads', devices for writing onto or reading from the surface in terms of the magnetised areas, which, by virtue of the rotation of the drum or disc, pass close to them. A head traces out a circular path over the rotating surface, this path being known as a 'track'. All drum tracks are of equal length and of sufficient width that there is no mutual interference between adjacent information; all disc tracks are of different radius depending on the distance of the head from the axis of rotation.

For purposes of access, information on a drum or disc is similarly situated to that in a delay line . It is first necessary to locate the information required, an average of half a complete rotation being required to achieve this, and then the information must be read off. Owing to mechanical considerations drums and discs cannot be spun too rapidly nor made too small. Typically over 100 40-bit words are stored on a track, and rotation speeds are of the order of 4,000 r.p.m. Thus location of a word may take about $8 \times 10^{-3}$ seconds, and word by word access is correspondingly slow.

Mention has been made above only of a drum or disc used in its normal operating mode with 100 words or so recorded permanently on each track. It is also possible to use any track as a form of regenerative storage. In this system the reading head is set at a

-7-

distance from the writing head such that the drum surface between can hold the number of digits it is proposed to store (normally 40 or more, i.e. one or more words). Information is written on the drum continuously, the digit written being that just read by the reading head. The word pattern occupying the sector of the track between the two heads is thus repeated around the track continuously, and the word or words stored can be read out as required. The access time will be on average $1^{1/2}$ times the time taken for a point on the drum to rotate through the sector in question. For a single word store this would be of the order of $2 \times 10^{-4}$ seconds.

Backing storage

It is at present expensive to provide even a thousand words of storage of access as rapid as is possible on magnetic cores or single word regenerative systems. Many problems of both scientific, engineering and, more particularly, accounting background involve processing many thousands of numbers. It is not essential to rapid solution of the problems concerned that all these numbers should be available equally easily and rapidly, and in most machines a compromise is sought between the conflicting claims of speed and convenience on the one hand and economy and reliability on the other. Thus it is frequently arranged that a working store of about 50 to 1000 words is provided and coupled to this are one or more backing stores. The latter are normally made to supply information in bulk, so as to replace at a single operation a large part of the working store contents. However, facilities are also sometimes provided for the transference of individual words between the stores.

The two most common forms of backing store are the magnetic drum or disc and magnetic tape. The former stores have been described in the preceding section. In the role of a backing store it is usual to arrange transference of several words at a time, a track, half a

track, or a block of 8 words being typical aggregates of information move at a single operation. These are as a rule read into a selected part of the working store and replace the information contained there, which is lost on being overwritten. In the same way information from the working store can be written over information on the drum or disc.

Magnetic tape consists of a plastic tape coated with magnetisable material. It is normally wound on spools or 'reels' in the same way as cinematograph film, each 'reel' holding from 600 to 3000 or more feet of tape. Information is stored on the surface of the tape by making magnetic marks in a manner similar to that used for storage on a drum surface. The width of the tape varies from 1/4" upwards, 1/2-3/4" being most common. Across this width are set a number of heads for reading and writing, usually about 8. As the tape is moved forward or backward under these heads, they trace out a straight line or 'channel' on the surface of the tape. Digits may be recorded on all of the channels simultaneously, or one or more of them may be reserved for special purposes, such as for recording the address of positions on the tape or for checking. Information is normally recorded 5 or more bits at a time, such an aggregate of bits being called a 'character'. The bits of a word are thus spread over 6 or more characters. Since only one character can be read at a time, it is necessary to provide storage attached to the magnetic tape unit large enough to allow complete assembly of a word or words before transference to the working store, and also to hold words which are split up during writing. Such storage is known as a 'buffer' store. This can, of course, form part of the working store, but is more usually a separate storage unit. Magnetic tape buffers are of various sizes, typical ones containing 16 or 32 words. An aggregate of information recorded on tape is called a 'section' or 'block' (if the latter term is not in use for other purposes); it is usually addressed as a unit

and often corresponds to the capacity of the buffer.

Tape movement speeds vary, but are usually of the order of 50-100"/sec., during reading or writing. The tape can move backwards and forwards, but can normally only be read in one direction. 'Free' running during rewinding can sometimes be done more rapidly than reading or writing and the speed often is doubled under these circumstances.

Characters are packed 100 or more to the inch, and thus the time to read off a word into the buffer is about $8 \times 10^{-4}$ seconds. The transference to and from buffer and working store depends on the characteristics of these stores, which are often similar in general construction, except in so far as the requirements for the buffer as regards access are different. Transfer from buffer to working store is thus sometimes done by interchange of information rather than by overwriting the information in one by that in the other. In this way the buffer can be used as a rapidly accessible extension of the working store as well as in its primary role as a buffer for the magnetic tape system.

Magnetic tape apparatus is also provided with a method of searching for information on the tape. This is essential, since the tape is normally at rest, and the character immediately under the reading head is read unless the tape is first moved. The tape is normally searched under control of the programme in the machine, but the searching facility can sometimes be utilised whilst the rest of the machine is in operation, and is then independently controlled, although initial selection of the place required is made by the computer control. Normally, suitable arrangements are made to record the address of each section or block on the tape, in such a way as to enable the searching system to locate rapidly a required address. Individual tape spools may also be addressed and the searching system arranged to call for a change of reel if necessary, though this feature is not usually provided.

## Automatic checking of storage

A serious failure of any storage is to drop out information or include spurious digits. All types of storage are liable to this type of failure from time to time, and it is usually considered worthwhile to provide some means of detecting the most common situations; the loss or gain of a single digit in a word or character, or total failure of a one-word storage unit. It is, therefore, very usual for words to include an extra digit, called the 'parity' digit, which is inserted in such a way that every word has an odd (or even) number of 'ones' (or zeros). Thus the number 001 would be recorded as 0,001 whereas 011 would be recorded as 1,011. On reading out of the store, a test is applied to ensure that parity has been preserved correctly. This is done by counting ones and ensuring that an odd number is read out in each word. It may be readily seen that satisfaction of this check indicates that either the word is correct or that at least two compensating errors have occurred - a very unlikely happening. Furthermore the check will indicate complete failure of a storage location, since this will usually yield zero in every position, and will also detect a less usual result of total failure called 'clodding' (the filling up of all digit positions with ones) provided that the total number of digits in a word, including the parity digit, is even.

On magnetic tape it is usual to provide a parity digit in each character, and this performs the same function as the parity digit in a word. As a further safeguard an automatic arithmetic check is also sometimes provided. During recording a summation is made, in some predetermined but arbitrary way, of the words or characters being recorded. This 'check sum' is then recorded at the end of the block, section or other aggregate of information. On reading off the information the summation procedure is repeated and a comparison is made at the end of the sum so formed with the check sum read from the storage medium.

Agreement indicates that the likelihood of error having been introduced by the recording and reading is small. Disagreement, of course, indicates a failure in some part of the process of recording, reading, summation and comparison.

Whilst the methods of checking described are of great value, it must be clearly understood that no method is foolproof. The existence of built in checks of any kind do not, therefore, absolve the programmer from applying suitable checks to his results.


Programming implications of storage systems.

For the purposes of programming, the important characteristics of storage systems are concerned with the relative access times to words, and the methods of interconnection between the various stores, i.e. in what role they are being used. If a set of one-word stores can all be referred to equally easily in the programme and the access time to each word is the same, then we shall call this 'isomeric' (or 'equal delay') storage. It is usual to make the working store isomeric, especially if storage is on magnetic cores. This is, of course, the easiest type of storage to use, especially if the access time is of little consequence relative to arithmetic operating speeds, since no special arrangement of data within the store alters the fundamental speed considerations.

Anisomeric storage can be divided into two classes, cyclic and acyclic, the first embracing drums and regenerative devices used to store more than one word per cycle, and the second concerning magnetic tape alone. In using a cyclic store, the programmer is mainly concerned with problems of timing. Any required information on a drum used as a working store for instance, passes the reading heads every 16 milliseconds or so and thus we must wait an average of 8 milliseconds for the word we want. However, by a careful choice of the moment at which we consult

the drum (or of the placing of words in the store) it is possible to reduce the waiting time very considerably. If a drum is used as a backing store and the 'blocks' read from it do not completely fill a track, then similar considerations apply to waiting time between blocks, and this can be ameliorated automatically by suitable choice of block numbering around the drum. On the reasonable assumption, for instance, that programmers will normally call for blocks in sequence and that some simple counting procedure will take place between each such call, a suitable spacing can be worked out so that as little time is wasted as may be. Alternatively some special application may be emphasized.

If a cyclic store is used to hold instructions, then the timing problems of programming are made more acute, especially if the cyclic store is also used as a working store for numbers. Control must necessarily consult an instruction before it obeys it, and must subsequently draw the numbers associated with it from the store and put back the result in the store on completion. In these circumstances, much time can be lost in waiting unless careful placing of numbers and instructions is carried out, and this operation is described as 'optimum' or 'minimal latency' programming.

The only acyclic anisomeric storage used with computers at present is magnetic tape. The information most readily available on tape is always the same at any time, namely that which will be read by moving the tape in a reading direction. Once that information has been read it will not normally be accessible again without some delay due to searching and rewinding. It is important, therefore, so to arrange the information required by the programme in such a sequence that is is readily available on the tape as required. More than one tape reel may, of course, be attached to the buffer at any one time, although, as a rule, only one can be read from at once, so that the restriction is not

necessarily severe.    However, as much or more time can be wasted

in unnecessary searching on magnetic tape, as by waiting for information

contained in a cyclic store.