FACTORS AFFECTING THE CHOICE OF MEMORY

Claude F. King
Logicon Inc.
Palos Verdes Estates, Calif.

Introduction

One of the fundamental choices in the design of a computer or data processing system is concerned with the medium for the storage of information. Once the system requirements impose the need for information storage that exceeds a certain level, typically several hundred bits, a "hierarchy" principle of memory sets in. As expressed by Von Neumann, the system may call for an overall storage of \( N \) words at an access time \( t \). For economic or other reasons it may be more practical to provide for some smaller quantity of storage, \( N_i \), with an access time \( t_i = t \) and obtain the total storage \( N \) at some longer access time depending on the system needs. Extending this reasoning then to the more general case there would be a sequence of capacities \( N_1 < N_2 < \ldots < N_k < N \), with access times \( t_1 < t_2 < \ldots < t_k \), so that \( N_i \) words are required for each access time \( t_i \).

Then each value of \( i \) would represent one level in the hierarchy of memories, and the hierarchy has \( k \) such levels. In many modern computers there are also quasi-levels that don't fall into Von Neumann's ordered sequence of levels. Certain bodies of the information storage may have special characteristics or requirements that call for different storage media that would yield greater capacities at shorter access times or vice versa that may not fall strictly into the ordered sequence. In making the choice of memories for a given system the characteristics for each level should be considered. The starting point should be the set of system requirements. These would be examined to determine the different levels that may be accommodated to satisfy the requirements. A set of requirements would then be generated for each level. An initial choice might then be made for each level. Since the choice of one level could strongly influence another they should be examined together and appropriate changes made. It should be borne in mind that several levels of the hierarchy may exist within one medium, for example the complete tracks as compared to the short recirculating loops on a magnetic drum would represent different levels.

References to specific types of memories in this discussion such as core or drum will usually be in the generic sense. For example, the family of core memories includes twistors, thin films and multi-aperture cores, as well as the conventional torroids. Also, the comments on drum memories will usually apply to the disc, as well as the cylindrical configuration.

The Determination Of The Hierarchical Levels From The System Requirements

For certain applications, such as the use of the computer in a control system for a well-defined purpose, the procedure can be straightforward. The computer's task can be described by a set of equations along with the required computation rates. The equation set should be separated into each of its modes and the needed capacity and access times evaluated separately. For each mode the information should be divided into different kinds that may impose different requirements on the storage. For example, the intermediate results of computations have different requirements from the storage of the instructions. Considerations should be made at each level as to the need for special protection for the information storage. For example, can volatility be tolerated for the section of the storage under consideration. The need and frequency of information change should be taken into account for each level. Some of the storage levels may be able to tolerate different degrees of unreliability than others. The tolerance to environmental conditions may be more critical for some levels than others. Where power consumption may pose problems it may be possible to operate at different power levels, depending on the mode. When each of the levels has been defined by these considerations, the possible choices for each level can then be made.

For many applications the computer's use is multi-purpose and considerations must be made on the basis of worst-case expected problems and highest-use types. Different types of problems can be looked on as different modes of the system. The flexibility to handle many different kinds of tasks is an overriding consideration itself, that often dictates maxima in both
speed and capacity limited only by economic considerations. This type of application leads to the ordered type of level identification that is based primarily upon access time and capacity.

Sample Set of Levels

As an example of a set of levels that might be derived from a hypothetical computer system, consider the following:

1. Level #1 consists of the active storage elements of a system that would be changed at computer clock rate. These are the single bit elements that take part in the computer control, arithmetic, input and output buffering and buffering into other levels of storage. In most modern computers they are transistor flip-flops, and for most controls computers the total capacity of this kind of storage runs from 50 to 500 bits depending upon how the machine is organized and the task it performs. Access times for this level of storage is usually from 1 to 10 microseconds. Since this level of storage is usually the most costly in terms of numbers of components per bit, an attempt is usually made to restrict the number of bits of this type to as small a number as possible, perhaps at the expense of some of the other levels. Loss of memory with removal of power is usually not a consideration at this level, although occasionally it is protected by providing for a complete dump into another level on the sensing of a power drop-out. Protection against loss of memory from a power drop-out can also be obtained by the use of magnetic cores either alone or in conjunction with flip-flops arranged to hold a state on sensing a power drop-out. Since at this level the storage elements communicate directly with the logic elements and through the logic elements to other storage elements they must have a power amplification greater than unity which is not generally true for other storage levels.

2. Level #2 might consist of from 100 to 1000 bits with a word access rate of from 10 to 100 microseconds. This storage may be provided for special input or output functions where incremental techniques could be applied at rates beyond that necessary for other portions of the computations. Such things as delay lines, magnetic core stepping registers and drum recirculating registers are commonly applied for this storage level.

3. Level #3 could contain the registers that take part in the arithmetic operations. In most configurations this is not a separate identifiable level, the functions being provided by either level #1 or level #2 storage or part by both. However, it has its own characteristics that could lead to a choice of a separate media. Its access time requirements may be longer than that required for levels #1 and #2.

4. Level #4 in this set is taken to be that section in which the intermediate results of computations are stored. The access time requirements for this level are probably the same as for level #3 and in some cases levels #4 and #3 are combined. The capacity of this level usually varies between 100 and 1000 words. In some computers where levels #2 and #3 described here are combined with level #1 the intermediate results store represents the first level in which a bulk type store is utilized.

5. Level #5 in the scheme described here is the first level in which the instructions are stored. For this level one might conceive of a group of high-speed, wired-in subroutines in which a certain speed advantage may be derived at a minimum cost.

6. Level #6 might be a memory that would be combined with level #4 in which a large number of instructions are stored in a manner that is readily changed. This is the work horse of the large data processing computers and contains up to 100,000 words in some. Access times vary from a millisecond to a microsecond in present computers.

7. Level #7 could be another wired-in type of storage that could be varied by a plug-board arrangement or some other means to provide for a body of instructions that are protected against memory loss due to electrical phenomena yet may be changed in a reasonable (several days) span of time. For a drum memory there might be some instructions with write amplifiers active that would place them in level #6. Level #7 could be associated with tracks in which the write amplifiers are deactivated or no longer present after recording. The distinction between the two is sometimes referred to as "hot" storage for level #6 and "cold" storage for level #7.
8. Level 3 would be a bulk storage of instructions that would block-transfer routines into the working store on program command. This level would probably be provided by magnetic tape or drum and need not communicate directly with the outside world.

9. Level 9 would be similar to and perhaps combined with level 8 and would contain the bulk store of data. It is listed separately from 8 since it could have different access time requirements than 8.

10. Level 10 for this example is that body of storage associated with the "outside world" and may actually be a multiplicity of levels consisting of punched tape or cards or magnetic tape, or even printed pages.

The levels listed here are those that may be suggested by a particular set of requirements. It is expedient in the actual execution of the design to combine many of these, but in almost all computers at least three levels are easily identified.

**Example**

In order to understand more fully the implications of this discussion consider an application that would present an environment so harsh and unpredictable and require a reliability so great that the designer must be forced to carefully consider his choice of memory at each level in order to have any chance of success. For this example consider a computer used to control a spacecraft on a journey to a near planet. The difficulty of the design for this application arises from the need to obtain useful operation during and after the long span of time in transit. This could be from six months to a year to the nearest planets. In order to keep it as simple as practical we will not burden it with the tasks involved in the launch guidance, but it shall be the primary center of control and guidance functions from the point of final stage vehicle separation onwards. The functions it would be called upon to perform might include system sequencing, system diagnostic checking, command decoding and interpretation, telemetry sequencing and formatting, antenna pointing, star tracking, engine and control jet commanding, terminal guidance, payload data buffering, instrument calibration, environmental control, and system decision-making (to select alternate system modes in the event of a recognized subsystem failure).

The problem of operation without maintenance for a period as long as one year is a severe one when looked at in terms of obtaining a probability of a successful mission as high as 90%. This implies a mean time to failure of 10 years for the system. One approach to this problem involves a computer designed to periodically diagnose itself, as well as the rest of the system, and take alternate actions by decisions itself or by supplying the diagnosis data by telemetry to earth where a decision can be made and a command issued back to the vehicle via the computer to alter the system, or computer itself, on the discovery of a malfunction. Consider now what this problem suggests in terms of different levels of memory.

1. There should be one section of instruction storage that contains the overall executive routine that must be wired in and designed in the most reliable manner possible. Brute force redundancy may be applied here and the section should be kept reasonably small.

2. In order to provide for the required operational programs plus diagnostic programs plus alternate routines and subroutines, the amount of non-working instruction storage required will be quite large.

3. The number of instructions working at a given time may be kept small so that the variable instruction store need not be large.

4. There may be a need for the storage of a large amount of payload data for a slow transmission to earth at a convenient time.

5. In order to keep the speed requirements for the rest of the computer and memories to a sufficiently low value to conserve power and assure good reliability a special input-output buffer memory may be needed to cope with the high rate data from the terminal guidance sensors.
6. The intermediate results memory may be called upon to function at a greater speed than the working instruction store and thus might be accomplished by a memory of slightly different design.

Considering the logic portion of the computer as another level and the storage buffer on the earth for special instructions that may be sent over the command link we have described a system in which eight levels of memory may be a reasonable solution. Besides the extreme reliability requirement, the premium for size, weight and power consumption will be such that each level should be examined in detail and the design tailored to the requirements of that level and the communication with the other levels. Redundancy of different kinds might be applied at different levels. For example, for the bulk stores, present missile and satellite magnetic tape recorders capable of millions of bits each are available at weights of only 5 pounds each making complete unit redundancy attractive there. For the intermediate results and working stores a capacity capable of the most complex task may be sacrificed in the event of detected malfunction by programming around the trouble spots with simple programs constructed and entered from the earth via the command link.

General Comments On The Choice Between Cores and Drums

The trend in memories for the large data processing computers, where a maximum in speed, power, and flexibility is desired, has been to magnetic cores for the working storage supplemented by magnetic tape units for bulk storage. For applications in which economy has been paramount, magnetic drums and discs are the most prominent types of memories in use today. For computers used in the control of military systems the choice is not as obvious and the question at times is controversial. Let us look at this in some detail. Some of the general characteristics of this application are as follows:

1. There is a desire to protect a certain critical body of the storage against loss of information due to acts of man or nature over an extended period and yet have it ready at instant's notice. This set of instructions and constants are usually well-defined.

2. Because of the inherent power and flexibility of the computer it is desired to use it for other less critical and less well-defined tasks to obtain overall system simplification.

3. The desired reliability and mean time between maintenance is usually beyond that easily obtained.

These kinds of requirements have generally led designers to choose a drum, and this is partly because we tend to be purists when it comes time to choose the store. If we must choose just one kind, the drum is a reasonable choice. It satisfies the desire for permanence of storage of requirement # 1, since we can store the information and deactivate the write amplifiers and still feel quite secure about its staying there. It satisfies the flexibility desired in # 2 since we have the ability to readily change instructions and constants by electrical means. The drum has generally fallen short of the desire expressed in # 3 primarily because the memory is dependent on mechanical motion and is susceptible to wear and a reduction in overall computer margin due to timing shifts that arise from the effects of a rugged environment. While this has been true it still stacked-up favorably on # 3 against a pure wired-in approach which sacrifices # 2, or a complete variable approach which gives up # 1 because of a 3 to 1 ratio in numbers of components that usually occurred in the implementation. This is a result of the serial nature of the storage, the use of time selection afforded by the rotation and the use of recirculators for arithmetic registers, fast access storage and input-output special buffers and processors. However, the tide is beginning to turn. Applying the multiple level principles expressed earlier we can take care of # 1 by a wired core arrangement embodying magnetic switch selection to hold the electronic components down. This can be arranged such that programmability is not completely sacrificed, requiring a two-day operation to perform, which is not at all unreasonable when looked at in terms of the time it takes to verify a good program when a change has been made. We can retain the features of # 2 by supplying another section of random access core memory, again employing magnetic switch selection to hold the electronic components down. This memory would serve as intermediate results memory when in
the operational mode and also would be used for system service routines if given the ability to be addressed as instructions. These would be entered and operated piecemeal to keep the size of this section down. A third section of special input-output buffer of either core or magnetostrictive delay line would be added. To further make use of the speeds available with modern techniques a little application of microprogramming could drive the component count down still further. These things taken together will not still give us as small a component count as the drum, but brings the count perhaps within 2 to 1. With the best selection and utilization of modern electronic components this 2 to 1 ratio in component numbers in favor of the drum will not counterbalance the unfavorable mechanical aspects, and unless there is a significant and fundamental change in the way drums are constructed we can expect an ever-increasing percentage of military control computers with non-rotating memories in the future.

Conclusion

There are many techniques and media for storage of information described at this conference. Each should be considered in terms of its special qualities and characteristics. Be not quick to rule out one in favor of another for there may be one level in a given application that may be specifically tailored to the media discarded in light of requirements for another. As computer designers become more sophisticated there will be a greater tendency to multi-level memories and a wide application for the many techniques.
