# Some considerations regarding arrangement of records on magnetic tape

*By* P. Giles*

**This paper discusses the structure required for a set of life assurance records on magnetic tape, with the aim of picking out some of the basic features of organization of business records. These features must reflect both the elements of record present and the dynamic problems of processing them. There should be no arbitrary separation between the records and the procedures which operate upon them as has often been the case in past commercial languages (Cormack, 1965).**



One of the first problems in the employment of a computer for large-scale file maintenance is that of finding the specification which can perform the work required at the lowest cost (Giles, 1964). A major item in this cost, and one which affects the size of fast store required, is the magnetic tape on which the records are held. Efficient use of tape will at least reduce the time required for file maintenance, and may enable a cheaper tape system to be employed.

This problem has been studied in the context of records of policies of life assurance, where updating actions and enquiries must be able to access the file once in each working day in order to provide a satisfactory service to the policyholder. The length of the file under study is about 100 million digits, which demonstrates that a reading speed of 10,000 digits per second might be sufficient—especially if reading and writing could be performed simultaneously. Internal fast-store transfer speeds are often nearly a hundred times this figure, and this ratio is a strong incentive towards careful packing of records before writing to tape. If this packing requires one internal transfer of each component part of the records, and a time allowance of 100% in addition for the organization of packing, then it is justified if it produces a saving in record length of over 2%. This figure must then be divided by the number of blocks which are normally inactive for each active block. This is a very general conclusion which must now be examined in relation to the particular forms which records take and the methods of recording provided by the computer.

At the lowest level, records consist of alphabetical characters—principally in the name and address which form a large part of many commercial records—and figures. Alphabetical characters are recorded in six bits in many computers, and in eight bits in others. In the latter case a considerable saving can be made by conversion to six-bit characters before writing to tape, particularly if the activity rate is fairly low—e.g. if one in twenty records require unpacking each day. In the ¼-inch tape under consideration, an unusual method of recording is used by I.C.T. which results in eight bits being used but being so recorded as to take only that length of tape which would normally be occupied by six bits. The binary bits are paired, and each pair summed and then converted to analogue representation by recording on magnetic tape as an interval between phase reversals. This also has the advantage that the allocation of a fixed number of digits for a sterling field—which is essential in a word machine—does not lead to the waste of tape that would be expected, because the zeros occupy very little tape. In a character machine, this economy can easily be achieved, but at the cost of inserting a word mark, and possibly of using six bits per digit. Some computers handle figures in binary and thus save 16% in comparison with a fixed-word decimal machine. With the ¼-inch tape, however, there would have been little if any saving achieved by conversion to binary. Any saving in record length on tape had to be derived from the organization of words and fields within the record. Packing of several items into one machine word is the rule rather than the exception (Drummond, 1963).

## Four types of record organization

An inspection of the items in a record showed that they divided the whole record into three main types of section—the name section, the cash section, and the contract section. The contract section might occur several times in one record—where several types of assurance were included within one policy—but the name section necessarily occurred once only. Besides containing the policyholder's address—which varied in length from one policy to another—it contained descriptions which necessarily applied to the whole policy and which were present in almost all records. In contrast to this, the contract section had to make provision for items that were only present in special types of contract, and hence many of these sections would be scattered with void words if any standardization of layout were to be achieved. At the end of the cash section the group of

* *Scottish Amicable Life Assurance Society, Craigforth, Stirling.*

words describing an outstanding premium might require to be repeated.

Thus all together there are four points that require to be provided for in arranging the record—sectionalization, variable length of address, repetition of a group of words a variable number of times, and scattered void words. The first three require either separation by markers or address increments, and the last requires a means of eliminating but recording void words. This can be efficiently achieved by inserting in the first position of the next non-void word a digit specifying the number of preceding void words. Any voids that are not followed by a non-void word within the same section (or record level) do not need to be counted. They are implied by the presence of the next section separator or address increment.

In order to arrange the use of the fast store a figure for maximum block length must be fixed. This should preferably exceed the longest record that can occur in the file, but in this particular case it appeared to be more economic in the use of fast store to break each policy into two records, stored together on the same magnetic tape but treated as separate records for processing purposes. The second record carries information which is not frequently required but may be needed to complete action initiated by the first record. Action must never be initiated through or by the second (trailer) record because the first may be no longer accessible. By this means, a whole policy is permitted to exceed the maximum block length which is already rather long at 200 words of 12 digits each. A typical record of half a policy is quite short, and this ensures that the part of the fast store allocated to a block from tape, but unused because the block does not exactly equal the maximum permissable length, is kept down to 10 words on average—half the length of a typical record. With this block length 15% of the tape is occupied by inter block gaps, markers and redundant data, leaving 85% available for primary data. It is interesting to notice that a doubling of block length to 400 words would increase this tape efficiency factor from 85% to 92% but would require 400 additional words of fast storage adding 5% to the overall hardware cost.

On this machine—the I.C.T. 1300—software is available which permits a simultaneous read/write between tape and fast store to be followed by the examination of the highest key of the block read. If this key indicates that no action is required for any data in the block, a re-entry to the software causes the definitions of the read area and the write area of fast store to be interchanged and another simultaneous read/write to be initiated. This process is very economical in fast-store transfers but has the disadvantage that interblock divisions are perpetuated even if deletions make these blocks unduly small. It is thus necessary to reblock the entire file periodically—perhaps four times a year. Another reason for creating blocks that are below the permissible length of 200 words is that the records have a tendency to "breathe." This effect is caused by that part which records premiums due but unpaid. Adjacent records tend to have similar due dates and thus expand for several days about this time.

**Access to specified words, groups, or records**

Within blocks, software normally permits access to individual records for the purposes of alteration, insertion or deletion. The first of these is the least common and the program must provide for access to individual items within a record—generally for insertion or deletion. This requirement has different effects for each of the methods of organization. If the record is divided by markers into variable-length sections a transfer of the second part of the record to permit the insertion is all that is required. If address increments are used instead of markers then they will require to be updated. Similar considerations apply where repetitive groups of words are involved.

The description "repetitive" has been found to have two sub-divisions. In the first of them alterations may be required to a specified group out of several successive groups forming a repetition. The specification should be such that the correct group is found whichever position it occupies in the repetition. It must, therefore, be specified by an item or items within its contents. This sub-division attaches no special importance to the first or last items in the repetition—the order is fortuitous. The second use of the word repetitive applies to groups where the order is all important and an alteration to a specified group is not required. In this case insertions are made at one end of the repetition and deletions at the other. A list of outstanding premiums in order of due date is the typical example. Payments delete the oldest items from one end while new items are periodically inserted at the other end. Reminder action is normally governed by the oldest outstanding premium.

It appears to be normally the case that whenever action is required on a record, access is required to several places in that record. This is often caused by the need to recalculate the implicit key date. Thus rather than employ a complex subroutine, to make access across packed voids several times, it is probably more economical to unpack the record before processing and repack it afterwards. This permits subsequent direct addressing of the items required by presenting optional details in fixed locations within their record section.

However, a major point of importance is that access must be rapid to those items of record that alter frequently, and they must be positioned so as to reduce the requirement to move other items to permit their insertion or deletion. This shows that such items are best located at the end of the record—not the beginning since this is the natural position for record keys and for address increments permitting rapid access to the end. The locations of these keys can thus remain fixed. The centre of the record is then filled by those items required for reference but which seldom alter. An example of a first half record is given in **Table 1**.

**Table 1**

**An example of the first half of the record of a policy**

| *Contents of each 12 digit word* | *Description of each part of each word* |
|---|---|
| *First section* | |
| 025/070463/100 | Length of record of half policy.<br>Date and type of last explicit alteration. |
| 00241165/16/00 | Length of record to end of addresses.<br>Date and type of next implicit alteration. |
| 00264713/3600 | Policy number and statistical codes. |
| 600000000/700 | Assignment or Scheme number, clauses on policy. |
| 010463/0700/10 | Date assurance commenced.<br>Prospectus and circumstances of sale. |
| 140/003/0216/02 | Agency or Scheme. |
| 160/140/025711 | Statistical codes. |
| *Second section* | |
| 1/251125/0/3/05/0 | Sex and Date of Birth.<br>Restrictions on use of this address.<br>Words in name, in address, and in Life Assured (if different). |
| J. B. * Smith Esq., | Contributor six characters per word, asterisk showing position for alphabetical sorting. |
| 47 Main St.,<br>Halifax, Yorkshire | Address using commas as line separators. |

If the next word commences with non-zero the format of the second section is repeated. The entire record may cease at this point.

| | |
|---|---|
| *Third section* | |
| 000/220665/000 | Length to a repeat of premium details (if any).<br>Date of last premium payment. |

Up to eight words describing the periodic premium and the method of collection. Individual words may be omitted by adding one to the first position of the next non-void word (if any) within these eight. This decimal position is normally zero.

*Fourth section*

Groups of three words, each describing one outstanding or prepaid premium, the first being the oldest and the last the most recent. Access to the latest is rapidly achieved using the length of record as a modifier.

**Mapping of blocks on input**

So far, it has been assumed that reading from tape presents a block of records in fast store but provides no information as to its internal construction. If parallel programming is permissible, this may have enabled some other program to make progress provided the computer is large enough to permit efficient parallel running. In many cases time sharing within the same program is all that is available. It is difficult to employ this facility if the new block is considered not yet available and the old block is being written to tape and cannot now be altered. There seems, however, to be no logical reason why a program should not inspect the new block word by word as it comes in from tape, and extract key words and separator addresses. These can then be inserted into standard fast-store locations, or into modifying registers if sufficient are available, so that rapid access may subsequently be made to individual record items. This would provide a steady occupation for the program controller during reading of magnetic tape, and would reduce the time required for altering records. The same program could insert voids that had been packed before writing to tape.

**Implicit and explicit key words**

All files are naturally stored in order, and this is often in numerical order of account number, or policy number in insurance files. Thus alterations to records, principally premiums received, will require a comparison with the policy number of every file record. When equality is found this explicit action can be processed. However, there are other functions required of the file which must be initiated internally on a specified date. These implicit actions include production of premium renewal notices and notices of benefits becoming payable. For this reason every record must carry the date of the next implicit action expected to be required. The time required to perform the action can be reduced by attaching a "type of action" coding to the date. On every updating run, today's date must be compared with this implicit key word, and whenever a record is altered for any reason, this key word may need to be recalculated.

These two keys may be summarized at the beginning of a block of data by recording there both the highest policy number and the earliest date which occurs in any record within that block. It is then possible to avoid individual examination of the records within the block whenever no action is required on any of them. This can eliminate the need for internal transfers of that block and save considerable time if file activity is low. There may be occasions, however, when action is required on all records of a given subsidiary key. At these times individual examination cannot be avoided. If the requirement is to produce a statistical report on the file, a recent past-generation copy file can be processed as a separate job from the normal updating routine.

The various types of record organization described above occur at various levels within each other—

though no more than three levels within a record were apparent in the files being studied. As a result, the methods of accessing and keying the records can be quite complex, but the advantage of doing so can be judged by the fact that in practice one policy record can be more than ten times as long as another. This variation is unusual but must be provided for. Fixed-length records would thus waste about 90% of the magnetic tape available, and increase overall processing time to an unreasonable extent.

### Conclusion

Financial records, which form a considerable part of the general data-processing problem, can be separated by normal double-entry book-keeping into individual accounts, each for a specific asset, customer, or supplier, etc. Each of these accounts can be allotted a serial number, and will generally also contain the next date on which an action must be taken if the account does not alter, subsidiary keys such as the agency involved, the occupation or industry of the account holder, etc., the name and address for correspondence, and details both of the items forming the outstanding balance of account and of the contract underlying the account, or each item of it.

The general layout of a record of an account should be arranged with the fixed-length items at the beginning and the items subject to frequent change at the end, as follows:

1. Record separator or length of this record.
2. Principal keys, the account number and the earliest implicit alteration date.
3. Subsidiary keys and statistical keys.
4. Variable-length items subject to occasional alteration, such as name and address and possibly contract details. Void items may be fairly common in the latter.
5. Repetitive items subject to frequent change normally by addition or deletion. The items forming the balance of the account together with their description and due dates are the principal category under this heading.

### References

CORMACK, A. S. (1965). "Interaction between user's needs and language—compiler—computer systems," *The Computer Journal* Vol. 8, p. 8.

DRUMMOND, J. (1963). "Some aspects of recording Graduated National Insurance contributions," *The Computer Journal* Vol. 6, p. 1.

GILES, P. (1964). "Problems of Life Assurance data processing," *The Computer Journal* Vol. 7, p. 95.

## Book Review

*Modern Control Theory*, by J. T. Tou, 1964; 427 pages. (Maidenhead: *McGraw-Hill Publishing Company Ltd.*, 112s. 6d.).

A decade ago, ideas in control theory appeared to be in the doldrums. Linear techniques for design and analysis, firmly based on the Laplace and Fourier Transforms, were highly developed and provided adequate cut and try methods for control engineering design. Attempts to apply transform techniques to non-linear and multivariable systems had been pushed as far as they would go and an apparent impasse had been reached.

It was no doubt the growing ease of numerical computation which sparked off the revolution which has since taken place in control theory. Going back to the original differential equations of the system and facing the difficulties they presented led to the rapid development of techniques based on the variational principle of mechanics, which provided a well developed mathematical basis for the design of optimal control systems. This led to the state space concept and to the principles of optimality as expounded by Bellman and Pontryagin.

This book is one of the first texts dealing solely with these new methods of design and analysis. It is written for the engineer rather than for the mathematician, and contains a chapter on the necessary mathematical foundations of the subject, *e.g.* set theory, vector spaces, and vector matrix differential equations. The book covers the topics of state space and state transition methods, variational calculus, dynamic programming and the maximum principle. A final chapter is entitled "Computer Control Theory". This is not quite so comprehensive as it may sound, being concerned mainly with the application of dynamic programming techniques to linear systems, but including the case of inaccessible state variables. Such practical topics as computer storage requirements are not discussed.

Many examples are worked throughout the text and this certainly adds much to the value.

As one of the few books on the modern theory, this volume will inevitably attract the attention of the practising control engineer who is trying his best to bridge the much discussed gap between control theory and its industrial application. He will not be disappointed, and, as a bridge across the gap, the book must be given full marks.

P. H. HAMMOND