X-563-67-82

NASA TM X-55714

THE GSFC SCIENTIFIC DATA STORAGE PROBLEM

FACILITY FORM 602

N67-19861 (ACCESSION NUMBER)

(THRU) 1

16 (PAGES)

(CODE)

TMX-55714 (NASA CR OR TMX OR AD NUMBER)

07 (CATEGORY)

BY

RICHARD G. HOLMES

WILLIAM B. POLAND

ALLAN M. DEMMERLE

JANUARY 1967

NASA

GODDARD SPACE FLIGHT CENTER

GREENBELT, MARYLAND

X-563-67-82

THE GSFC SCIENTIFIC DATA STORAGE PROBLEM

Richard G. Holmes and William B. Poland
Goddard Space Flight Center
Greenbelt, Maryland

Allan M. Demmerle
National Institute of Health
Bethesda, Maryland

January 1967

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

CONTENTS

THE GSFC SCIENTIFIC DATA STORAGE PROBLEM

INTRODUCTION

Since the beginning of the United States Space Program, about eight years ago, the Goddard Space Flight Center has experienced a constantly increasing flow of telemetry data. In contrast to the experience of military missile ranges, the Earth Orbiting Scientific Satellites for which Goddard Space Flight Center is responsible tend to radiate telemetry data continuously at very high rates for long periods of time (six months to several years).[1] Large quantities of data resulting from these transmissions are recorded at receiving stations all over the world and mailed to the Goddard Space Flight Center where the data are processed (see Figure 1). Individual experiment raw data, along with such other useful information as satellite attitude and orbit data, are given to the experimenters whose spacecraft equipment generated the data originally.

The original data recordings are stored at the Goddard Space Flight Center. Over 140,000 reels of magnetic tape (90,000 digital and 50,000 analog tapes) are presently stored. Data are presently stored at an average approximate rate of 35,000 digital tapes per year (see Figure 2). As the number of satellites and the data rates increase, the storage of these data becomes a very significant problem.

Originally, both the analog tapes, containing the signals as received at the tracking stations, and the digital tapes, containing the raw data in digitized, slightly edited form, were stored. There is some prospect that the policy with regard to storing both of these types of tapes containing essentially the same data will be changed to require the storage of digital tapes only for archival purposes.

The reasons why these reformatted original digital data recordings are stored are varied. Most importantly, perhaps, the data must be stored to allow for reinterpretation at some future date. It is expected that improved analysis techniques may produce results at a future time which were not achievable initially. In addition, it may be desirable to conduct a concentrated study of portions which were shown by other sources to be of interest. It is conceivable that interest in these original data recordings would be shown at some date well into the future for presently unpredictable reasons. Furthermore, these original recordings are the most fundamental identifiable product of extremely expensive satellite programs. A conclusion to be drawn from these considerations is that, although one can expect change in the policy toward retention of some of the data having questionable value, it seems likely that there will be a continuing requirement for retention of at least a large portion of the data which has been or will be received.

Some of the more important problems associated with long term storage of data on magnetic tape are:[2,3] (1) cost of the data stored by virtue of the cost of material, (2) print-through of data from one layer to the next, (3) blocking (layer to layer adhesion), (4) mechanical distortion of the tape, and hence, difficulty in reading the data back from the tape, all as a result of stresses in winding, (5) erasures due to stray magnetic field, (6) sensitivity to the environment. The last two of these can be minimized by storage in a controlled environment; however, the others are common for magnetic tape on a reel. Some relief for the problems of print-through, blocking, and reeling stresses can be had by periodic controlled rewinding of the tape. Such an operation is rather costly for an archieve of tens of thousands of reels of tape since the first such pass effectively doubles the operational load, and the second pass triples it and so on. This has dubious value as a long period solution.

The greatest drawback, however, to a magnetic tape archive is the cost of the material. At the present cost per reel of tape of about $15, roughly three quarters of one million dollars worth of digital magnetic tape is now invested in long term data storage.

BACKGROUND

Thus, there is a need for an archival system which could handle the large volumes of data at a substantial saving in cost and with improved performance when compared to standard reels of magnetic tape. With consideration of this type in mind, work was begun in 1964 to develop an archival system to meet Goddard Space Flight Center's needs. As a first step, studies were performed to establish achievable levels of performance which could be expected in a hypothetical archive system, and to evaluate the benefits that could be derived from an operational system meeting these performance levels.

As a result of the external studies and the protracted program at GSFC, two procurement actions for an archival system were initiated. The first resulted in no proposals, and the second, based on revised specifications, produced six proposals. Systems involving photographic techniques with wet processing, magnetic tape techniques, and other more unconventional techniques were proposed. However, this second procurement was finally cancelled by the Government, primarily because of unexpectedly high initial cost of the systems and the feeling of Goddard management that the technical and operational requirements involved a substantial risk for still higher costs and possibly for lack of fully satisfactory operation of the system. A new attempt to obtain an acceptable system will be made in the future. It is one purpose of this presentation to summarize the experience and planning which have developed so far in the hope that the notions we have arrived at will be of some value to others concerned with our type of systems.

THE STUDY PHASE

The initial studies were intended to establish levels of performance for an archive system; however alternate solutions of the problem were also considered. Among these was the concept of data compaction. It is obvious to anyone who studies the experiments carried by the different spacecraft that much of the data generated is of little value, and preservation of all these data is not warranted. For example, an instrument designed to study a discrete occasional event, such as a solar flare, frequently generates telemetry data at the maximum required rate continuously rather than only as needed. In the past, inefficiencies of this type have been an inevitable part of the telemetry data systems. Much of this inefficiency can be removed with better instrumentation. As spacecraft become more sophisticated, various sorts of data compression are being adopted to improve the efficiency of use of the telemetry channel. It is doubtful, however, that such compression techniques will be effective in substantially reducing data flow in the near future because they require prejudgment of the nature of the data to a degree the experimenter may not accept. Whatever benefits are derived from data compression may not relieve the archive work load because while it may relieve channel capacity that capacity would probably be used up by the addition of more experiments. We would then have the same total data, but more of it would be useful.

The studies further indicated that data recording techniques are available which would provide for a decrease in physical size of a hundred to one when compared with magnetic tape. Additionally, these techniques seem to assure the required degree of longevity and reliability and would provide a system with a very low operating cost when compared to tape. Some desirable by-products would also be available. Since the system would be interfaced with a digital computer, indexing, bookkeeping, and performance evaluation could thus be automated. Also, it would appear feasible to have a fair degree of flexibility for on-line data retrieval (equivalent to several hundred or more reels of magnetic tape units on line) giving the system some attributes of a mass on-line storage as well as those of an archive. Based on these results, a plan was developed for setting up an archival system for telemetry data at the Goddard Space Flight Center.

SYSTEM DESCRIPTION

The basic plan was to transcribe the data found on magnetic tape on to a different medium which is much less expensive than tape, thereby allowing the magnetic tape to be returned for reuse many times. A large reduction in material cost is necessary so that the savings found in reusing the tape will amply cover the additional operating expense of the new system (see Figure 3).

When evaluating the desirability of different systems, the cost of the material appears to be the most important factor; however, it is not the only factor. Number of people required as system operators, computer time for index generation and other functions, record keeping, and expendable supplies (chemicals, etc.) all contribute to operating expenses. In the remainder of this article, estimated expenses shown will be per 10^8 bits of data, i.e., the approximate equivalent of one fully packed reel of digital tape, and are compared to the Goddard Space Flight Center initial cost per reel of tape of about $15.

The new archive medium will provide the greatest potential for saving. Materials with the cost per equivalent reel of tape of approximately two dollars are practicable. This is a reduction over magnetic tape of about seven to one. Systems have been proposed with a much smaller material cost approaching a few cents for a reel of tape.

Efforts to reduce the cost of material below the neighborhood of $1 per reel have little value in the overall economics of a system to meet Goddards needs because the other economic factors discussed become so much more important.

Another major expense for a system is the man power required. Four men, including operators, supervisors, and warehousemen, and not including maintenance people or any allowance for the usual estimating contingencies, seem to be required for most systems of the size necessary to transcribe about 40,000 tapes per year. This work load can be handled on one shift. Such a staff results in a cost of about one dollar and fifty cents per reel of tape not including contingencies. This amount is subtracted from savings per tape accrued by transcribing on less expensive medium. Minimizing the number of operators required by automating or eliminating manual operations is thus desirable. For instance, a system using a photographic medium might have an automatic developing process. For this and other reasons, a system which requires no additional processing after the transcribing operation has much appeal.

Also, among the economic factors is the possible use of a computer for indexing, record keeping, etc. for these large amounts of data as well as other possible data manipulations. A minimum requirement is for a computer to index by accession number (AN) each reel of magnetic tape to be archived. The computer would be provided with identifying descriptors, such as satellite number, orbit number, date, etc. The computer would provide the AN and would maintain the cross index to use for data retrieval. When retrieving data, the computer would be supplied with the descriptor information of interest from which all the pertinent AN's would be retrieved. Estimates have been made that 15%, or more, of the available time of a 360/40 computer would be required to perform the necessary indexing of retrieval functions for a 40,000 tape/year work load. This expense amounts to approximately fifty cents per reel of tape. The use of large computers in the archive writing process for such functions as data buffering, error code generation, etc. should be avoided on the basis of economics.

Volumetric compression is a desirable feature of an archival system, but is not very important from an economic point of view at the Goddard Space Flight Center where storage space is fairly inexpensive. The total annual cost to store a reel of magnetic tape at a Goddard Space Flight Center warehouse has been computed to be approximately five cents. However, in other geographical areas, the storage space may be much more expensive, in which case volume compression is an important consideration. In any case, even at Goddard Space Flight Center, for operational reasons one would prefer that the data not be stored in a large warehouse. To store it elsewhere requires some compression.

Another important class of characteristics for archival equipment is that which affects its operational efficiency and utility. Since the archival equipment would be operated by relatively unskilled people, it is important that its operating requirements be kept as simple as possible. For example, a system which requires wet processing of the archival medium, even though the process be automated, would clearly tend to be less desirable in this respect than some form of magnetic tape recording, which requires no additional processing at all. A system requirement for operation in a clean room would be viewed less favorably from an operational standpoint than one which had no such requirement. In some instances the specialized environment, for example the clean room, was proposed as an internal environment created within the equipment. This effectively relieves the undesirable facility requirement, but creates potential difficulty in loading and unloading. Complex operational requirements or special facilities by their very existence tend to compromise reliability. For these reasons, the designers of an archival facility must weigh the operational requirements as heavily as the purely technical ones. In short, any special requirements for environment, chemical processing, or other functions, although not disqualifying the system of themselves, may well outweigh purely technical characteristics in determining a selection.

The physical size of the minimum retrievable sub-unit of the archive (analogous to a single book in a library) is also an important consideration in the design of an archive system. It is conceivable that units could be manufactured which are capable of holding the data for many thousands of reels of magnetic tape, or which are capable of holding only a single record. Intuitively, the ideal would seem to be somewhere between these extremes. Our thinking points toward a unit holding at most the equivalent a few reels of tape. Fast turn-around time (the time elapsed between writing data onto the archive unit and verifying that the data were accurately written) is desirable, and, usually, small archive sub-units facilitate this. If problems in writing the data on the archive material are detected quickly, a minimum of material is discarded or wasted, fewer magnetic tapes require temporary storage, and traffic and scheduling problems are diminished. Small archive sub-units also lead to less wear and abrasion to data which are not needed in a given retrieval request. If a large archive sub-unit is used, for example a large roll of film, the data from hundreds of reels of magnetic tape will pass over the reading mechanism to retrieve any single file contained on the film roll. Also, probably a shorter search time is obtained with a small unit. Lastly, there is the matter of up-dating, or, in the extreme, purging the archive. If selected files are to be up-dated, or removed from the archive, the job is facilitated with a small address-able data unit. Less copying of data from the obsolete to the up-dated unit is then necessary, and less material is destroyed.

A rapid simple verification scheme is required for any data archive to provide assurance that the data are recorded correctly before disposing of the original record. A rapid scheme is necessary to minimize the operation time, and a simple scheme is necessary to minimize operator efforts and set up time, and to assure reliability. The same verification subsystem might be used for testing the endurance of data in the archive. This operation could be done by sampling data from the store on a regular basis. The technique used is dictated by system consideration. A magnetic recording system allows for a simple, highly reliable, immediate bit by bit verification by read-after-write. Operationally, this scheme probably could not be improved upon greatly. Systems which record on a photographic medium must delay verification until after the developing process, after which bit-by-bit verification is much less practical. However, in such a case, error correcting codes can be added to the data before recording on the archive material and used in verification.

The functional configuration of the data archive is shown in Figure 4. Two basic processes are represented in the diagram: the introduction of raw data into the archive and the withdrawal of data in response to retrieval requests. Data enters the system from the region indicated in Figure 1 through the block labeled "Data Tape Sources." The data first undergoes an input writing and verification process, and is then placed in the Store. After the data on the archive unit is verified to be correct the data tapes undergo a rehabilitation process and are returned to the data Sources for reuse. Upon receipt of a data request the computer, which maintains all indexing and housekeeping functions, translates the request information received in the form of such descriptors as time, satellite number, station number, etc., into specific accession information and initiates the retrieval process. Retrieval requests are constrained to a basic set of descriptors; sophisticated retrieval processes involving associative searches or other complex operations are not contemplated for this system. The verification function encompasses not only rapid validation of the data written into the archive, but also periodic endurance testing of the data which resides in the archive. The Store function is that part of the archive in which data is retained. It is contemplated that a portion of this data will be accessable automatically by computer. In addition the Store function will provide supplementary storage of containers of the archival medium off-line. The output of the archive from the reading system will be in the form of digital computer tapes. In some implementations, the reading, writing, and verification functions may be physically combined or closely associated.

CONCLUSIONS

The Goddard Space Flight Center is carrying out a continuing program in the field of archiving for telemetry data. Although the basic characteristics of the equipment being developed are determined by the telemetry archival requirements, a study of the problem has indicated that any practical system should incorporate properties that would make it useful in other applications as well: specifically, it appears desirable to increase the data accessing capability beyond that strictly necessary for the purely archival requirements. This greatly enhances the operational properties of the equipment and makes it useful for a much wider variety of applications than would otherwise be the case. It also seems clear that, for facilities having a large data flow, the reduction in physical volume available in most archival systems is a far less important property per se than cost per bit for the storage medium or low cost digital computer processing requirements. These two factors tend to be of roughly comparable importance, and decreasing the cost of one far below that of the other probably would not lead to an advantageous compromise. Two other properties which materially affect the desirability of a system are verification and endurance testing. What industry may be asked to provide in future is a small system having a modest initial cost with ability to be incremented as data volume requires, a data rate commensurate with that of advanced magnetic tape units, demonstratable archival properties, and an intermediate level of data accessability to facilitate at least block searches of the store and to permit efficient verification and endurance testing.

These major considerations and the others already discussed will be applied in another attempt to develop a satisfactory archival system which will meet not only the technical but also the budgetary constraints with which we are confronted. The basic requirements which have motivated the development thus far, a very high input data rate and a rapidly mounting accumulation of telemetry tapes, seem certain to make archiving a necessity.

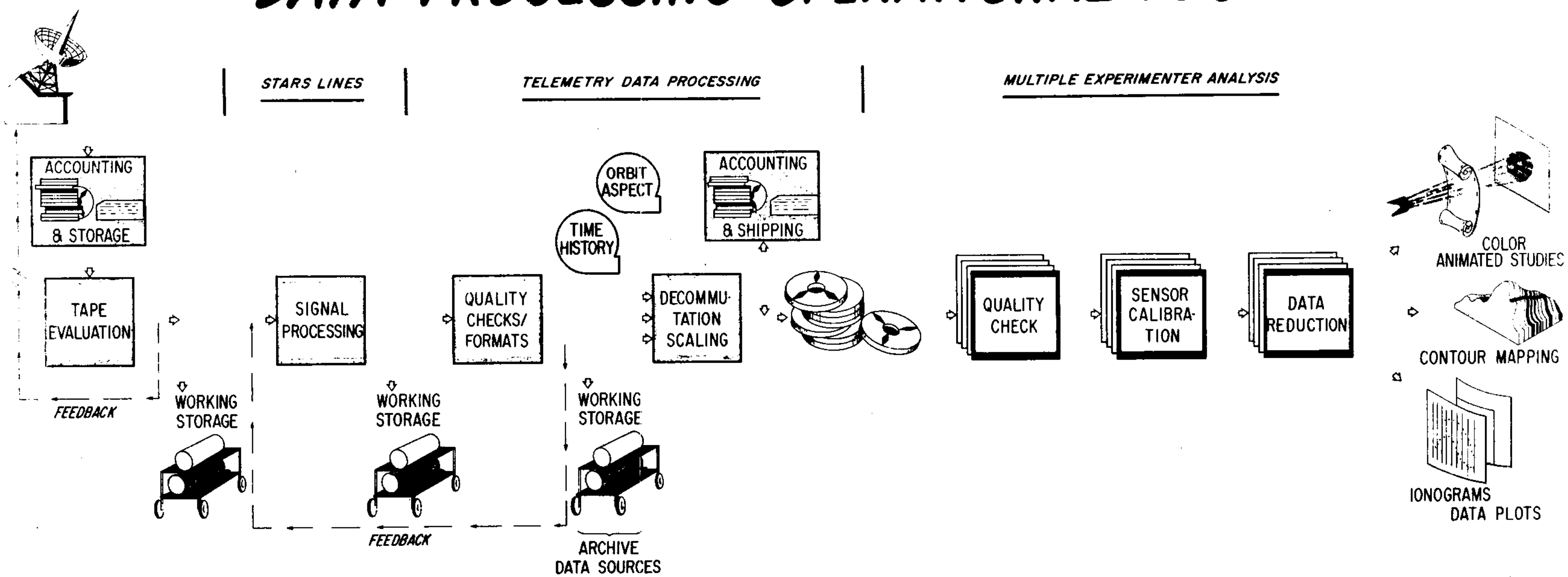
REFERENCES

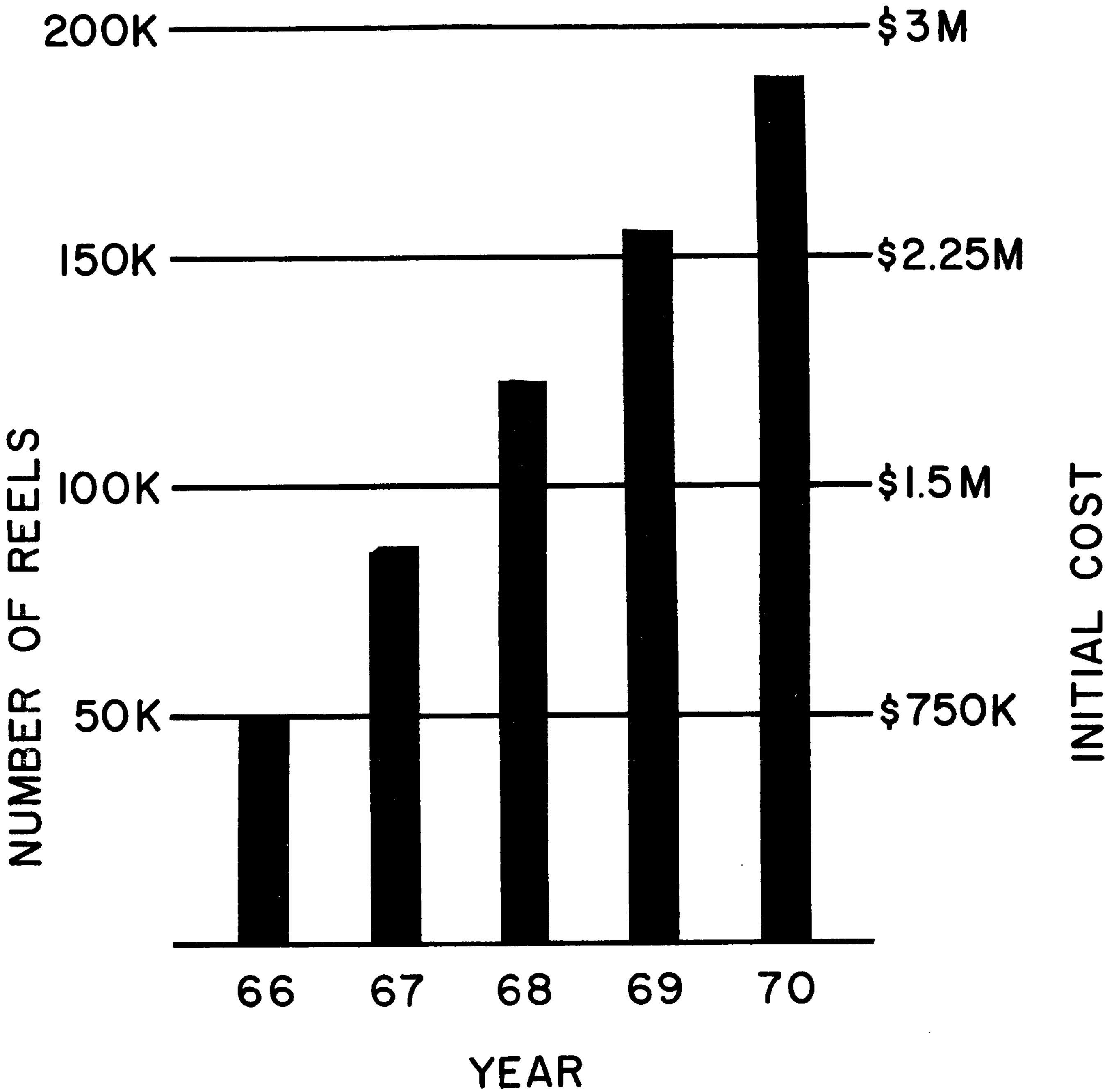
1. Habib, E. J., Telemetry Processing for NASA Scientific Satellites Proceedings of the 20th Annual ISA Conference and Exhibit October 1965.

2. Westmijze, W. K., Studies on Magnetic Recording, Phillips Research Report, 1953.

3. Radocy, Frank, Tape Storage Problems, Journ. of the Audio Engineering Soc., January 1957.

DATA PROCESSING OPERATIONAL FLOW





PROJECTED STORAGE REQUIREMENT FOR DIGITAL DATA (35K REELS PER YEAR AT $15.00 PER REEL)

FIG. 2

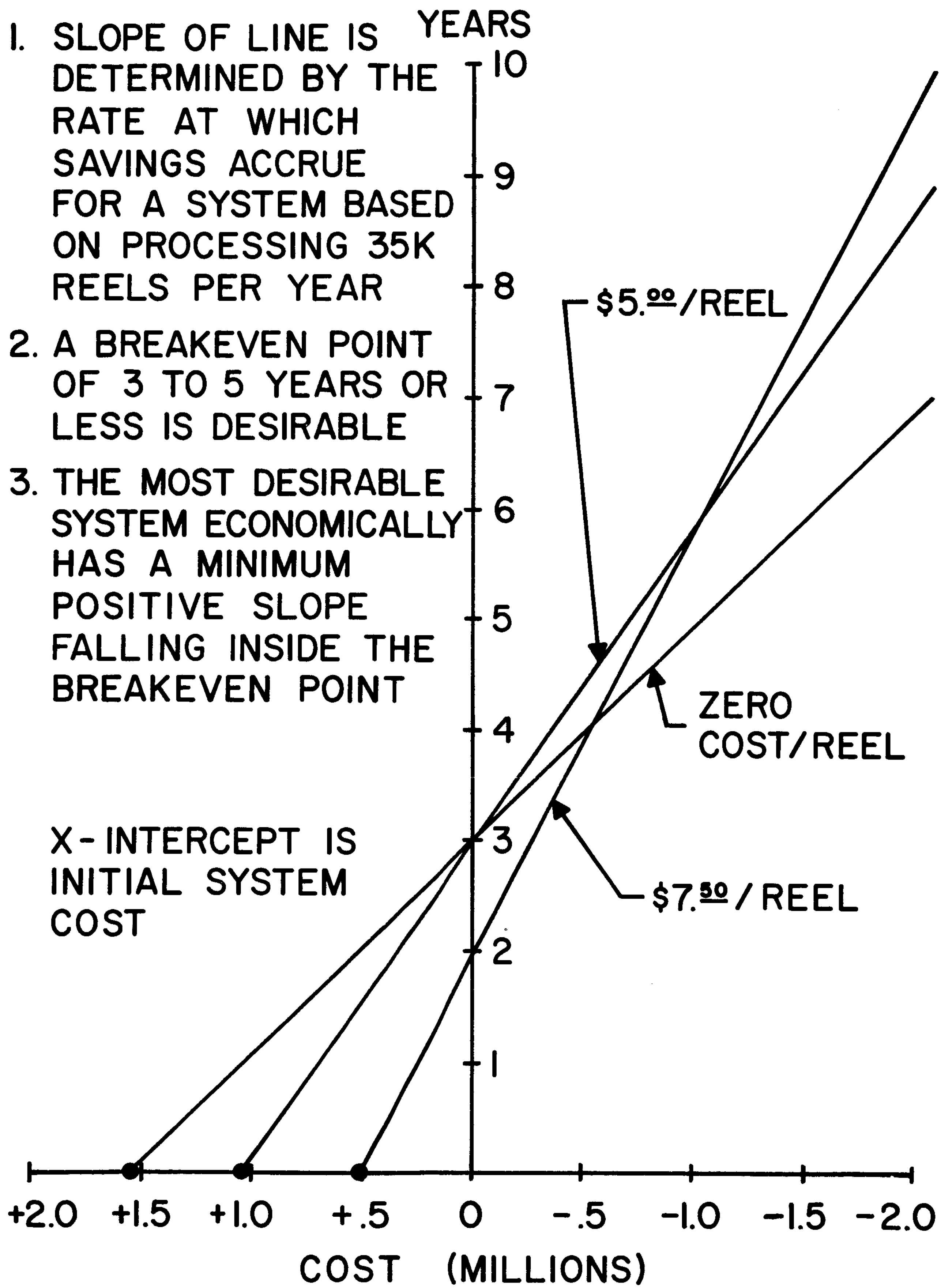
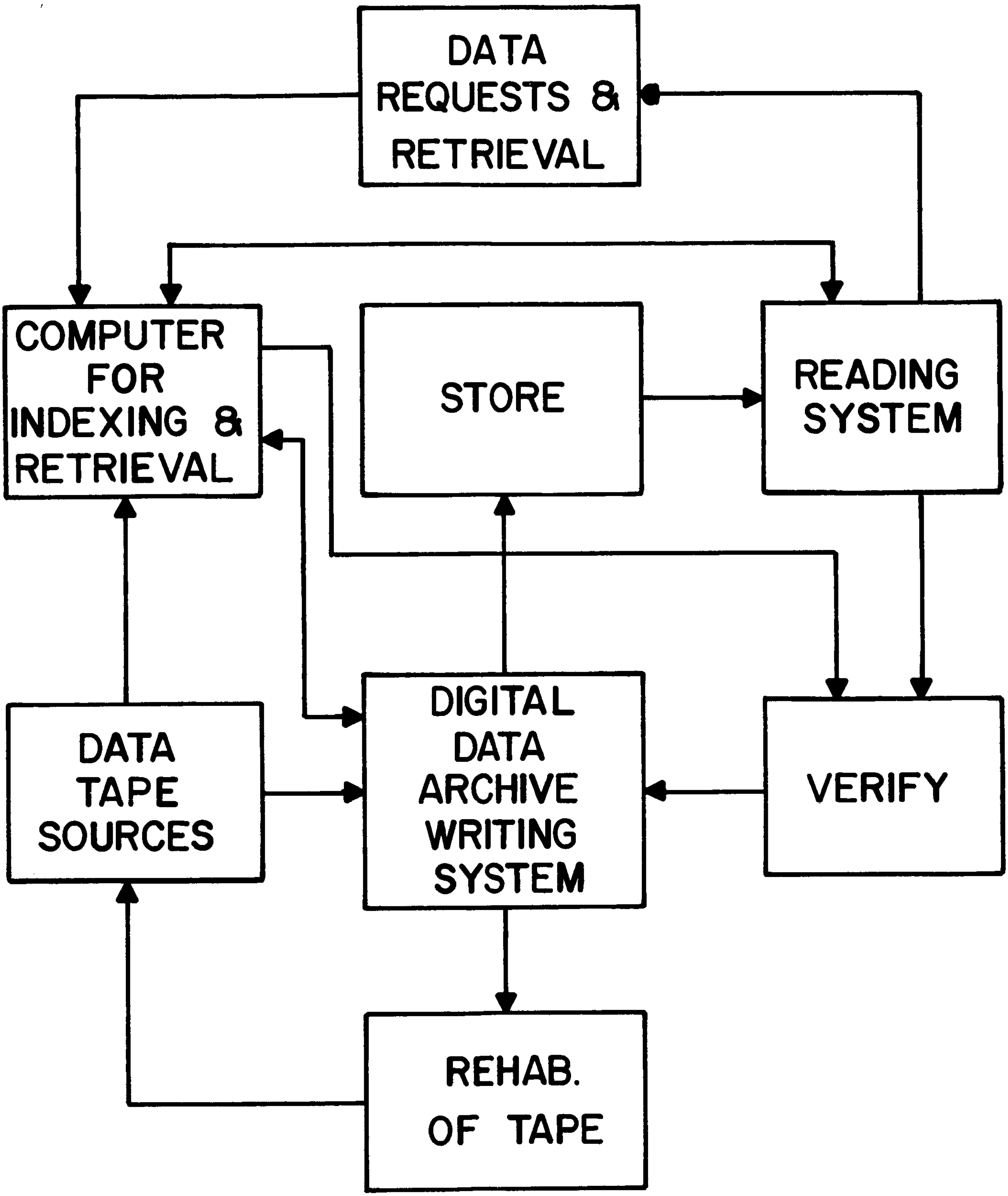

1. SLOPE OF LINE IS DETERMINED BY THE RATE AT WHICH SAVINGS ACCRUE FOR A SYSTEM BASED ON PROCESSING 35K REELS PER YEAR
2. A BREAKEVEN POINT OF 3 TO 5 YEARS OR LESS IS DESIRABLE
3. THE MOST DESIRABLE SYSTEM ECONOMICALLY HAS A MINIMUM POSITIVE SLOPE FALLING INSIDE THE BREAKEVEN POINT
X-INTERCEPT IS INITIAL SYSTEM COST
YEARS
10
9
8
7
6
5
4
3
2
1
$5.00/REEL
ZERO COST/REEL
$7.50/REEL
+2.0
+1.5
+1.0
+.5
0
-.5
-1.0
-1.5
-2.0
COST (MILLIONS)

FIG. 3



DATA ARCHIVE FUNCTIONAL DIAGRAM
FIG. 4