Reprinted from AMERICAN DOCUMENTATION
Vol. 15, No. 2, April 1964
Printed in U.S.A.

this is a product of
the RCA 1962
project.

COST ANALYSIS AND SIMULATION PROCEDURES FOR THE EVALUATION OF LARGE INFORMATION SYSTEMS

Charles P. Bourne and Donald F. Ford

Cost Analysis and Simulation Procedures for the Evaluation of Large Information Systems

A computer program has been written and used which simulates the several-year operation of an information system and computes estimates of the expected operating costs as well as the amount of equipment and personnel required during that time period. The program has been used for the analysis of several large systems and has proven itself to be a useful research tool for the study of systems with so many components and interrelated operations that an equivalent manual analysis would be extremely cumbersome and time consuming, and perhaps even impractical. This paper describes this program and shows, as an example, some of the results of a simulation of two of several suggested designs for a specific information system.

CHARLES P. BOURNE [1] and DONALD F. FORD [2]

• Introduction

The designers of large information systems have generally used relatively simple and crude procedures to evaluate the merits of their alternative system designs.[3] This may be due in large part to the fact that only in recent years has there been a need to design or evaluate large systems; consequently there was very little opportunity or need to try to use or develop anything more than very elementary techniques. Now that more design work is being done and larger amounts of money are involved, a greater awareness has developed of the need for better tools for design and evaluation.[4] Some work has been done on the cost analysis of various parts of an information system.[5] The authors felt this need over the course of several design and study projects at Stanford Research Institute and gradually modified and used the techniques that are described in this paper. For the most part, these techniques are not new or novel — they might even be considered standard industrial engineering tools — however, they have generally not been applied to the analysis or evaluation of proposed or actual information systems. The intent of this paper is to describe these methods and show how they can be used for this purpose. The description, of course, is too brief to provide all the necessary detail to permit the reader to use this approach without some further study. These methods are used only for the direct economic evaluations, and do not provide a performance evaluation (e.g., relevancy, recall), nor do they include economic values for intangibles such as response time or quality of services provided.

Basic to this procedure is a computer program that accepts detailed descriptions of the operating performance, costs, and interrelationship of all components in a proposed system, and then simulates the operation of that system over a specified time span (chosen as five years in the examples shown) in order to compute estimates of the expected operating costs, amount and type of equipment, and number and classification of personnel required during that time period. The time span is chosen with regard to such factors as the equipment amortization period. The program has been useful for the study of systems with so many components and interrelated operations that an equivalent manual analysis would be extremely cumbersome and time-consuming, and perhaps even impractical. A program such as this would seem to be of most use (1) where a choice must be made between several different alternative systems for implementation, and (2) where an equipment or system designer is interested in determining how sensitive (to costs, etc.) the operation of a system is to changes in the performance or characteristics of different parts of the system (e.g., what would be the overall effect of doubling the speed

[1] Stanford Research Institute, Menlo Park, California.
[2] Thompson Ramo Wooldridge, Inc., Canoga Park, California.
[3] An information system is defined for the purposes of this discussion as a complex of people, equipment, and procedures — working together to perform all the operations needed to provide information to a group of users.
[4] *Information Systems Workshop: The Designer's Responsibility and His Methodology*, 1962. A publication of the American Documentation Institute. Spartan Books, Washington, D.C.
[5] HEILPRIN, L. B. 1962. The Economics of 'On Demand' Library Copying. *Proc. 1962 National Microfilm Association Convention, Vol. 11* National Microfilm Assoc., Annapolis, Maryland, pp. 311–339.

of the searching device and raising its initial cost by 30% while at the same time replacing the input device with another unit with half the speed and one quarter the cost?).

The program was written in a standard ALGOL language and run on the Burroughs 220 computer. The two examples and their illustrations in this paper were obtained during the study of several different ways to mechanize a very large file system. They will be referred to as System 1 and System 2, but will not be described here. Neither of the particular file problems nor the details of the proposed systems affect the methodology of the analysis; they only determine the content.

• The Simulation and Analysis Procedure

STEP ONE: FORMAL DESCRIPTION OF THE CANDIDATE SYSTEMS

Prior to the actual programming and analysis of a specific candidate system, the operation of the candidate must be described and defined in as much detail as possible. This is aided by the construction of block diagrams and tables of operating characteristics of both the equipment and the associated personnel. Flow charts are useful to illustrate all of the equipment used, and all of the operations performed. Fig. 1 is an example of such a diagram. The operating speeds of each piece of equipment, as well as the basic times to perform each of the major manual operations must also be established and recorded. Costs that are common to all of the alternative systems being studied (e.g., acquisition costs of the library collection) are not considered in this analysis procedure.

The following types of input data for each candidate are used by the program:

Time and Cost Data

1. First-, second-, and third-shift wage rates for each of the different labor categories (such as clerk, machine operator, supervisor)
2. Purchase or lease costs for each of the different pieces of equipment used in the system including first-, second-, and third-shift costs where appropriate (up to 40 different types of equipment such as keypunch units, microfilm cameras, motorized file cabinets)
3. Costs for each of the different types of supplies of physical materials used in the system (up to 20 different types of materials such as punched cards, microfilm, paper, magnetic tape)
4. Costs for each of 20 miscellaneous items such as equipment maintenance, postage, transportation, communications.

Statement of Interrelationships

1. Statements that are functions of the file input rate (e.g., X minutes of machine type "A" is required for each input item; Y minutes of a microfilm camera operator's time is required for each input item)
2. Statements that are functions of the volume of search requests (e.g., Y minutes of machine type "A" is required for each search request, and Y minutes of clerk-typist time is required for each search)
3. Miscellaneous statements (e.g., the cards wear out and must be replaced after N searches, and an additional machine type "B" must be installed when the search load exceeds Z searches per month)

Constants

1. Size of the file item (e.g., number of characters, or number of page-size images)
2. Initial file size at beginning of simulation period
3. Amortization period for equipment purchases
4. Rate of return (interest) to be used for amortization calculations
5. Burden and overhead percentages

A summary statement of the input data used for each program run is printed out at the start of the computation (see Fig. 2).

STEP TWO: COMPUTATION OF ESTIMATED MONTHLY OPERATING COSTS

Using the statements about the initial conditions of file size, input rate, and number of search requests, as well as the amount of each type of labor required to process the input items, conduct the searches, and perform all the other necessary tasks, the program determines for a one-month operating period the total number of man-hours required for each category of direct labor. The program then estimates the amount of each type of indirect labor required, such as managers and assistant managers; for example, by adding a manager when the total working staff reaches five persons and adding an assistant manager for each further increment of twenty persons. The salaries of all the direct and indirect labor types are then totaled to determine the basic labor charge. The burden (allowances for vacation sick leave, social security, etc.) and the overhead charges are then used in a standard accounting manner to arrive at the total labor costs.

Next, the required numbers of each type of equipment are determined, based on the capabilities and workloads of each of the individual units of equipment, and taking account of the initial conditions of file size, input rate, and number of search requests. The program can consider each type of equipment on either a lease or purchase basis—as defined by the input statements. Lease charges are considered to be a simple monthly cost. Purchased equipment is amortized over a time interval and at an interest rate specified by the input statements. If the rate of return is set at zero per cent, as done in many crude economy studies, then the cost of the equipment is simply divided equally among the specified time intervals without considering the time value of money. For any other percentage value the correct interest charges are included.

The present program assumes that an integer number of people will be used, so that if the work requires 3.8 clerks, then 4 clerks will be used in the cost estimates. It also assumes that a particular type of laborer (e.g.,

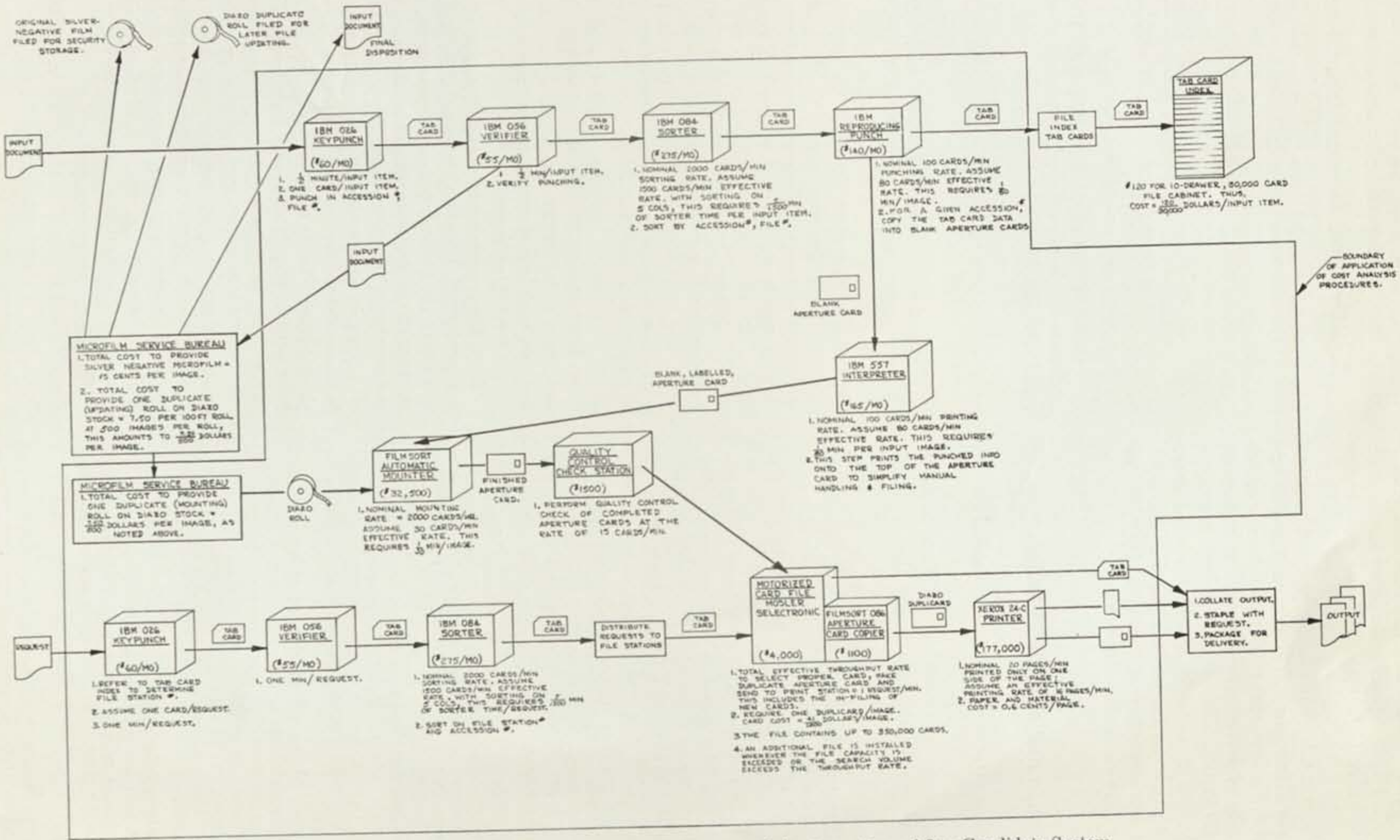


Fig. 1. Representative Flow Chart to Describe Some of the Operations of One Candidate System

GENERAL COST CALCULATIONS—COEFFICIENTS AND INITIAL CONDITIONS

SPECIAL SYSTEM NO. 2		
INITIAL FILE SIZE	0.	ITEMS
RATE OF RETURN	6.0	PERCENT
AMORTIZATION PERIOD	50.0	MONTHS
BURDEN RATE	25.0	PERCENT OF DIRECT LABOR
OVERHEAD RATE	100.0	PERCENT OF (LABOR + BURDEN)
FILE ITEM SIZE	1.0	PAGE

PARAMETER	COST	MINS/ITEM	MINS/SEARCH	MISC.	2ND SHIFT
KEYPUNCH OPERTR	320.00	1.00000	2.00000	.00	368.00
TAB OPERATOR	400.00	.02833	.00333	.00	460.00
MOSLER FILE OPR	300.00	.00000	1.00000	.00	345.00
PRINTER OPERATOR	280.00	.00000	.06250	.00	322.00
CLERK MISC	280.00	.25000	.25000	.00	.00
QUAL CONT INSP	400.00	.06666	.00000	.00	.00
MOUNTER OPER.	280.00	.03333	.00000	.00	.00
MANAGER	750.00	.00000	.00000	1000.00	1250.00
ASST-MANAGER	650.00	.00000	.00000	20.00	.00
IBM 026 KEYPNCH	60.00	.50000	1.00000	.00	36.00
IBM 056 VERIFYR	55.00	.50000	1.00000	.00	33.00
IBM 084 SORTER	275.00	.00333	.00333	.00	165.00
IBM 557 INTERPR	165.00	.01250	.00000	.00	165.00
IBM REPRO-PUNCH	140.00	.01250	.00000	.00	84.00
QUAL CONT STATN	1500.00	.06666	.00000	.00	.00
MOUNTING DEVICE	32500.00	.03333	.00000	.00	.00
XEROX 24-C PRNT	177000.00	.00000	.06250	.00	.00
SELECTOR + COPIER	5500.00	.00000	2.00000	350000.00	.00
DIAZO DUP. ROLL	.00	.01500	.00000	.00	.00
BLANK CARDS	.00	.04000	.00000	.00	.00
INDEX TAB CARDS	.00	.00150	.00150	.00	.00
DUPLICARDS	.00	.00000	.04400	.00	.00
PAPER + TONER	.00	.00000	.00600	.00	.00
MAINTENANCE	.10	.00000	.00000	.00	.00

FIG. 2. Sample Printout of Input Data for System 2

clerk) is free to work on any task (or unit of equipment) that requires that particular labor type, and need not be restricted to a single task. This is the same as assuming a very flexible and mobile staff. Similarly, only integer numbers of equipment units will be used, so that the program will always charge the full cost of each unit of equipment even though the particular unit may be required for only a fraction of the day.

After the material and other miscellaneous costs are determined, a final dollar total is obtained for the operation of this entire system for one month.

With most real information systems, the file continues to grow in size, rather than remain constant. As a general rule, the costs of searching the file also increase as the file gets larger. Consequently, the labor and equipment costs can be expected to increase for each subsequent month's operation. For this reason, and to make the analysis as realistic as possible, costs were computed on a monthly basis for a five-year period, using a continuously increasing file size and computing the costs for each sequential month. If it were not for increasing file size, the problem would be relatively simple since the monthly costs would essentially be constant. An illustration of a five-year series of monthly costs of two candidate systems under two different operating conditions is given in Fig. 3. This figure is a graphic portrayal of the monthly cash flow for each system over a five-year period. The bottom pair of curves of this figure are for two different systems that are operating with an input rate of 500 items per month and a search load of 100 searches per month. Since this work load is not sufficient to saturate the system, the monthly rate is constant. The top pair of curves are for the same two systems with a much larger work load reflecting an input rate of 100,000 items per month and a search load of 100,000 searches per month. In this situation the saturation effect can be seen by the increasing monthly operating cost of System 1.

STEP THREE: REPRESENTATION OF MONTHLY COSTS AS A SINGLE COST FIGURE

In situations such as illustrated by the bottom pair of curves in Fig. 3 where one system is always less expensive than another, it is simple to say which system is economically more attractive. However, in situations such as that illustrated by the top pair of curves in Fig. 3, where one system is less expensive for part of the time and then more expensive for the remainder of the time, the choice is not so simple. In situations where the curves intersect at some time in the future, the choice should consider the time value of money. Two standard and equivalent methods of comparison useful in such situations are the "present worth" and "equivalent annual cost" methods.[6] The use of either method will always result in the same conclusion in the study of two alternatives. The present-worth method determines the present worth of a future time sequence of positive or negative cash flows. In this case, it determines the total amount of money that would have to be put in the bank today at a given interest rate to meet exactly the computed series of monthly costs over the coming five years. The candidate system with the lowest present-worth figure (i.e., the one that initially has to deposit the least amount of

[6] GRANT, E. L. and IRESON, W. G. 1960. *Principles of Engineering Economy*. 4th Ed. Ronald Press, New York.

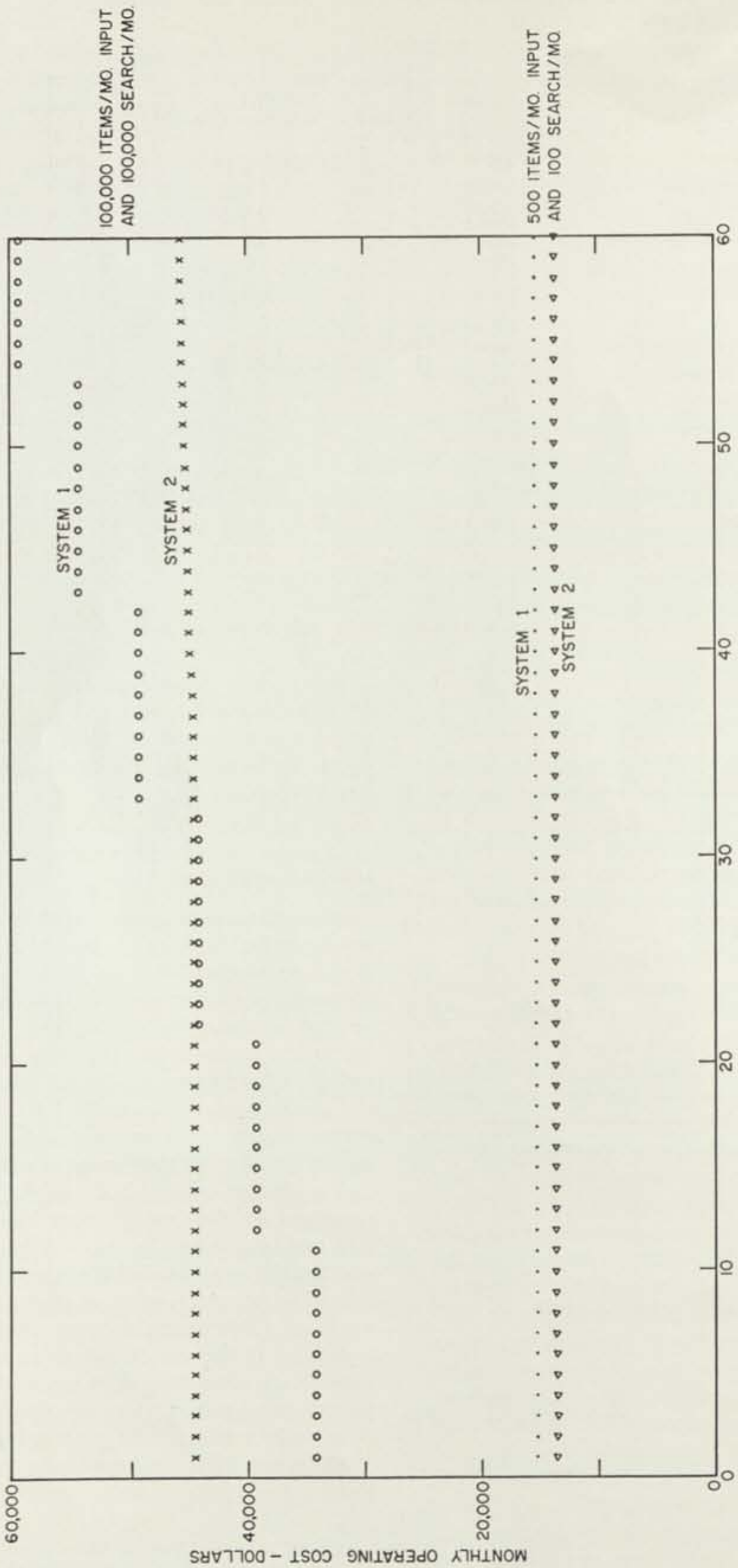


Fig. 3. Estimated Monthly Costs for Two Systems Under Two Different Operating Conditions

money in the bank to meet all the future costs) is obviously the most attractive from an economic standpoint. The second comparison method determines an equivalent annual cost for each candidate over a specified number of years (with the same time period used for each candidate).[7] In our case, a series of unequal monthly costs over a five-year period can be converted to a single number — an equivalent annual cost. Obviously the system with the lowest equivalent annual cost is the most attractive from an economic standpoint.

For each candidate system the program computes the present worth of each of the monthly costs, sums these to obtain a total present worth, and then converts this to an equivalent annual cost for that candidate. The equivalent annual costs are printed out by the program.

Fig. 4 illustrates a portion of the detailed printout of the results of the computation for System 1 and shows that for an input rate of 500 items per month and a search load of 100 searches per month, the system had an equivalent annual cost of $188,514 figured over a five-year period. This particular printout shows how the equivalent costs were divided among labor, equipment, and material. It also provides further detail to show the equivalent cost for each type of labor, equipment, and material. Fig. 5 illustrates a portion of the detailed printout for System 2 and shows that for an input rate of 100,000 items per month and a search load of 100,000 searches per month, the system had an equivalent annual cost of $550,568 figured over a five-year period.

In the examples shown, the cost analysis programs considered only a time series of generally unequal debits (expenses). However, the programs could also accommodate an accompanying time series of generally unequal credits (income) to arrive at a net present worth or equivalent annual cost.

[7] This *equivalent* annual cost should not be confused with the actual annual costs. The equivalent annual cost is obtained by extending for an N-year period, the present worth in equal annual payments, considering some specified interest rate.

STEP FOUR: EXTENSION OF THE ANALYSIS TO MANY DIFFERENT OPERATING CONDITIONS

The simulation and cost analysis described and illustrated above were for a very specific set of operating conditions in terms of input rate and number of search requests. However, the system designer or prospective manager is usually interested in examining the candidate systems under several different operating conditions to determine: (1) under what circumstances one system might be less expensive than the other; and (2) what the range of costs and personnel requirements of a system would be for a range of operating conditions. Given specified ranges and increments of input rates and search loads, this program repeats the simulation and cost analysis procedure for each set of operating conditions and provides a summary printout of the results of all the analyses. Figs. 6 and 7 show such a printout, giving the equivalent annual cost for each candidate for 64 different operating conditions of input rate and search volume, as specified by the row and column headings.

DETAILED BREAKDOWN OF EQUIVALENT AND UNIFORM OPERATING COSTS
(IN DOLLARS PER YEAR)

INPUT RATE	SEARCH LOAD	NAME	EQUIV. ANNUAL COST (DOLLARS)		
500.	100.				
	LABOR				
		KEYPUNCH OPER.	3944.		
		TAB OPERATOR	4930.		
		SPEC. EQUIP. OP.	4930.		
		SPEC. EQUIP. LDR	4930.		
		FILE OPERATOR	4930.		
		CLERK-MISC	3451.		
		QUALITY CONTROL	4930.		
		MANAGER	9244.		
		ASST-MANAGER	0.		
	TOTAL LABOR		41293.		
	TOTAL LABOR AND ADJUSTMENTS			103234.	
	EQUIPMENT				
		IBM 026 KEYPNCH	739.		
		IBM 056 VERIFYR	677.		
		IBM 084 SORTER	3389.		
		IBM 087 COLLATR	3081.		
		IBM REPRO-PUNCH	1725.		
		SPECIAL CAMERA	4453.		
		DEVELOPER	4175.		
		SPECIAL PUNCH	2783.		
		QUAL CONT. STN.	1391.		
		HIGH SPEED FILE	55668.		
		MICS. HAND EQUIP	278.		
	TOTAL EQUIPMENT			78365.	
	MATERIALS AND SERVICES				
		INDEX TAB CARDS	20.		
		SPEC CARD STOCK	17.		
		SPECIAL HOLDER	19.		
		PAPER + TONER	9.		
		MAINTENANCE	6847.		
	TOTAL MATERIALS AND SERVICE			6913.	
	TOTAL COSTS				188514.

Fig. 4. Sample Printout of Detailed Cost Breakdown of System 1

DETAILED BREAKDOWN OF EQUIVALENT AND UNIFORM OPERATING COSTS
(IN DOLLARS PER YEAR)

INPUT RATE	SEARCH LOAD	NAME	EQUIV. ANNUAL COST (DOLLARS)		
100000.	100000.				
	LABOR				
		KEYPUNCH OPERTR	63309.		
		TAB OPERATOR	4930.		
		MOSLER FILE OPR	19599.		
		PRINTER OPERATR	3451.		
		CLERK MISC	10354.		
		QUAL CONT INSP	4930.		
		MOUNTER OPER.	3451.		
		MANAGER	9244.		
		ASST-MANAGER	8012.		
	TOTAL LABOR		127284.		
	TOTAL LABOR AND ADJUSTMENTS			318210.	
	EQUIPMENT				
		IBM 026 KEYPNCH	4733.		
		IBM 056 VERIFYR	4338.		
		IBM 084 SORTER	3389.		
		IBM 557 INTERPR	2033.		
		IBM REPRO-PUNCH	1725.		
		QUAL CONT STATN	417.		
		MOUNTING DEVICE	9046.		
		XEROX 24-C PRNT	49266.		
		SELECTOR + COPIER	16732.		
	TOTAL EQUIPMENT			91684.	
	MATERIALS AND SERVICES				
		DIAZO DUP. ROLL	18989.		
		BLANK CARDS	49306.		
		INDEX TAB CARDS	3697.		
		DUPLICARDS	54236.		
		PAPER + TONER	7395.		
		MAINTENANCE	7546.		
	TOTAL MATERIALS AND SERVICE			140673.	
	TOTAL COSTS				550568.

Fig. 5. Sample Printout of Detailed Cost Breakdown of System 2

EQUIVALENT AND UNIFORM ANNUAL OPERATING COSTS
(IN DOLLARS PER YEAR)

INPUT RATE (ITEMS/MO.)	SEARCH VOLUME (NUMBER OF SEARCHES PER MONTH) 1.	50.	100.	200.	500.	1000.	50000.	100000.
1.	188447.	188453.	188459.	188471.	188506.	188564.	253149.	347374.
50.	188453.	188458.	188464.	188476.	188511.	188570.	253155.	347379.
100.	188458.	188464.	188470.	188481.	188517.	188575.	253160.	347384.
200.	188469.	188475.	188481.	188492.	188528.	188586.	253171.	347395.
500.	188502.	188508.	188514.	188525.	188560.	188619.	253204.	347428.
1000.	188557.	188563.	188569.	188580.	188615.	188674.	253259.	347483.
50000.	270269.	270275.	270280.	270292.	270327.	270386.	346311.	437824.
100000.	379635.	379640.	379646.	379658.	379693.	379752.	484336.	557901.

Fig. 6. Sample Printout of Summary Report of Analysis of System 1

EQUIVALENT AND UNIFORM ANNUAL OPERATING COSTS
(IN DOLLARS PER YEAR)

INPUT RATE (ITEMS/MO.)	SEARCH VOLUME (NUMBER OF SEARCHES PER MONTH) 1.	50.	100.	200.	500.	1000.	50000.	100000.
1.	167613.	167644.	167675.	167739.	167929.	168247.	268953.	413661.
50.	167647.	167678.	167710.	167773.	167963.	168281.	268987.	413695.
100.	167681.	167713.	167744.	167808.	167998.	168316.	269022.	413730.
200.	167751.	167782.	167814.	167877.	168068.	168385.	269091.	413800.
500.	167960.	167991.	168023.	168086.	168277.	168594.	269300.	414009.
1000.	168308.	168339.	168371.	168435.	168625.	168942.	269649.	414357.
50000.	230537.	230568.	230600.	230664.	230854.	231171.	368132.	470535.
100000.	303331.	303363.	303394.	303458.	303648.	303966.	431688.	550568.

Fig. 7. Sample Printout of Summary Report of Analysis of System 2

Note that the annual costs for the specific conditions shown in Figs. 4 and 5 are repeated in Figs. 6 and 7, respectively. However, Figs. 4 and 5 provide much more detailed information. Note also from Figs. 6 and 7 that one candidate is less expensive than the other only in certain well-defined operating regions (e.g., System 2 is less expensive for a condition of 200 items per month input and 50 searches per month, while System 1 is less expensive for a condition of 1,000 items per month input and 50,000 searches per month).

• The Practical Application of this Technique

With this program and method of analysis, relatively accurate cost analyses of proposed systems can be made without actually implementing a full-scale or pilot operation of the proposed system. Consequently, it can be used to test proposed systems that have no counterpart in any existing installation. In addition to being used as an evaluation procedure for proposed retrieval systems, the models can also be used effectively as a research tool to determine the overall effects of varying the characteristics of selected elements in the system. One great advantage of this approach is that once the basic description and characteristics of a proposed system have been assembled into the proper form and analyzed, the program can be quickly re-run, with relatively little additional effort, to study minor variations and changes. In this way a wide range of designs can be studied intensively with a modest amount of effort. It should be noted that the credibility of the analysis depends in large measure on the accuracy and completeness of the analysis and description of the candidate systems, as well as on the accuracy of the basic time and cost data. It is quite possible that a system of predetermined times for standard elemental operations could be used as time estimates for many of the operations of the proposed retrieval systems. The unit times and costs used in our sample evaluations were best guesses and crude estimates.

Much of the program (the method of accounting for overhead charges and computing present worths and equivalent annual costs) is applicable to any system and forms the framework for the basic program. It would be relatively simple to modify or change these procedures if necessary. Some parts of the program must be written for each specific candidate system to describe its performance and the interrelationships between the various components of the system. Consequently, some custom and detailed programming is required for the analysis of each candidate system, and the program as it currently exists cannot be used without some specific re-programming for each system to be studied. The analysis and re-programming effort for a candidate system may take weeks or months, depending upon the complexity of the system and the ready availability of the time and cost data and detailed system description. The cost of the necessary analysis and programming effort, as well as the computer time, is small in relation to the cost of errors in judgment for organizations planning to develop or install any complex and expensive information systems.

STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA

1962 Project date April 9 – Oct 9

Project ES-4058

Engineering Sciences Division

CLIENT CONFIDENTIAL

Proposal for Research ES 62-31

22 February 1962

TECHNICAL ASSISTANCE IN THE EVALUATION OF RCA EQUIPMENT

Colins Flint
George Arnovick
John Murphy

Mr. B. B. Saffe, Buyer
Data Systems Division
Radio Corporation of America
8500 Balboa Boulevard
Van Nuys, California

Dear Mr. Saffe:

This letter proposal has been prepared as a result of recent conversations between RCA and Stanford Research Institute personnel concerning possible assistance that the Institute might provide in the information retrieval program of RCA. In particular, this document is intended to provide the basis for a research agreement between RCA and the Institute--the outline of which is described below.

RCA is planning a wide range of activities in the information retrieval field, and presently intends to develop and market several different types of equipment for information retrieval. Preliminary studies conducted by RCA have indicated that a substantial market for such information retrieval equipment exists. At this point the company is primarily interested in evaluating the relative merits of its own and its competitors' equipment in specific application areas, and formulating plans for future research and development efforts. A special full-time RCA committee has already been established to perform this function; this committee would like to receive some additional assistance from Stanford Research Institute.

The Institute is prepared to offer its services in a study to help provide supporting information to the RCA committee. The principal steps in the proposed approach are outlined below:

1. Familiarity with RCA products and capabilities will be achieved by SRI through a review of pertinent literature and attendance at briefings conducted by RCA at Van Nuys. It is expected that travel will be restricted to a maximum of six trips to the Los Angeles area during the project.

2. Working together with the RCA Committee, SRI would devote its efforts to the following tasks:

   a. A review and evaluation of RCA's advanced research and development (A R & D) program in information retrieval. The A R & D program is the research effort supported by RCA funds, and currently consists of three major efforts: a study of descriptors and list structures for library classification systems; the development of a mass memory--a magnetic card system with capability for storing one million bits per card with digital, video, or microfilm on each card; the development of an automatic retrieval comparator (ARCOM).

   b. A formulation of A R & D plans for the next five years.

   c. A preliminary definition of users' requirements for information retrieval equipments. RCA's present equipment will be reviewed and comments made on its applicability to various user problem areas.

3. A simulation will be done of the operation of four information retrieval systems selected by the Committee for further study. The simulation effort would use the programs and techniques developed by SRI on a recent National Science Foundation project, and would provide estimates of the operating costs of each system over wide ranges in operating parameters.

The foregoing effort would require approximately 3 man-months of engineering effort, divided between a systems engineer and a programmer. The estimated cost is $12,000. Approximately $1,400 of the total project funds would be required for computer running time.

The studies will be performed by Mr. Charles Bourne and Mr. Don Ford of the Engineering Sciences Division; biographies of both investigators are attached. Informal letter reports would be prepared at the conclusion of each major task, and an oral presentation will be provided at the conclusion of the project. Work on this project could begin on or about March 15, 1962 and would be conducted over a six-month period.

Stanford Research Institute is a not-for-profit, non-endowed organization. In view of this status, the Institute requests each research client to make an advance deposit to cover working capital requirements for the project being undertaken. The Institute's invoicing policy is described in its Standard Research Agreement, two copies of which are enclosed, and which would be considered a part of this contract.

The advance deposit on this project is $4,000. Upon completion of the project, any unexpended project funds on hand are returned to the client.

It is hoped that the foregoing outline provides sufficient preliminary information concerning our possible assistance to RCA. Any changes can be made that you desire, provided that they are agreeable to both parties. If RCA elects to proceed with the study, this proposal can be established as a project on the basis of a letter of acceptance, or by returning two signed copies of the enclosed Research Agreement which is considered a part of the proposal. The proposal will remain in effect until 15 April 1962.

We trust that this letter will provide the basis of a mutually rewarding association. If you have any questions concerning the program outlined above, please do not hesitate to get in touch with us.

Sincerely,

Chas Bourne

Charles P. Bourne
Research Engineer

Approved:

R. C. Amara

R. C. Amara, Manager
Systems Engineering Department

for J. D. Noe, Director
Engineering Sciences Division

Bourne, Charles P. - Research Engineer, Systems Engineering Department

Mr. Bourne received a B.S. degree in Electrical Engineering from the University of California in 1957. He is working on an M.S. degree in Industrial Engineering at Stanford University, specializing in data-processing and operations research.

He served in the U.S. Marine Corps during 1950-1951. In 1952-1953 he was an Instructor in Guided Missile Instrumentation, Maintenance, Operation, and Telemetry at Convair Guided Missile Division. While he was a university student he did summer work as an Engineering Aide at Stanford Research Institute on design studies for the ERMA computer system.

In June 1957 Mr. Bourne became a Research Engineer on the staff of the Institute. He has participated in a government project to investigate storage, retrieval, and reproduction techniques for a file of several million engineering drawings; engineering and operational evaluations of several new general-purpose digital computer systems for various computer manufacturers; technical planning for digital computer system installations; and a government project to design a comprehensive mechanized system for accumulating, reviewing, disseminating, storing, and retrieving abstracts of European technical literature. He also supervised the operation and programming effort and conducted the systems studies that determined programming requirements, choice of computer, and expansion capabilities for a large digital computer system currently used as part of a military reconnaissance system. He has provided product planning assistance in the design of magnetic tape systems for commercial data processing equipment and market research and product planning for commercial information retrieval equipment, and has conducted system studies for the design of a very large memory for information retrieval problems. He has provided technical assistance in the design of the data processing and display portions of a world-wide bomb damage assessment center for the joint military services, and has also assisted in the evaluation of the data processing and display portions of a large air defense system. He is currently working on pattern recognition and the processing of graphic information with digital computer techniques, as well as methods for automatically abbreviating and coding English text material.

Mr. Bourne has written several articles for technical journals, dealing with information retrieval and technical information problems, and with studies in non-linear mechanisms. He is a member of the Institute of Radio Engineers, the American Documentation Institute, the National Microfilm Association, and the Association for Computing Machinery, and was Chairman of the sessions on Information Retrieval and Machine Translation at the 1959 Western Joint Computer Conference and at the 1960 Annual Conference of the American Documentation Institute. He is currently the Chairman of the San Francisco Bay Region Chapter of the American Documentation Institute.

Ford, Donald F. - Research Engineer, Mathematical Sciences Department

Mr. Ford's work at the Institute has included responsibility for the use scheduling, and maintenance of an extensive digital computer installation for a military reconnaissance system. He has also supervised a programming group for the implementation of the production programs. The final results of this project included a set of routines that amounted to some 80,000 program steps. Complete documentation was accomplished in the form of ten operating manuals and two volumes of system description.

He has cooperated with the Economics Division at Stanford Research Institute in the evaluation of two foreign computers with several American computers for business applications.

Recently, Mr. Ford has been participating in an effort to automate SRI's own accounting and project reporting system. This involves standard magnetic-tape file manipulation and extensive efforts with input-output. He has designed an extremely versatile tape record system and is designing a generalized reporting routine to fit that system.

Before joining the staff of the Institute in 1958, Mr. Ford was with the ElectroData Division, Burroughs Corporation in Pasadena, where he was engaged in programming and engineering work on computers. During the period 1953-1954, he was employed by the Phaostron Electronics Company in South Pasadena.

Mr. Ford studied engineering at Pasadena City College for two years and holds a B.S. degree in Agriculture from Arizona State College. He is attending Stanford University part time, with the object of acquiring a Masters Degree in Operations Research in June 1962.

Publications co-jointly with C. P. Bourne of SRI include: "The Historical Development and Predicted State of the Art of the General Purpose Digital Computer," Proceedings of the 1960 Western Joint Computer Conference, San Francisco, May 1960 (this was an opening paper given at that conference); "A Study of the Statistics of Letters in English Words," Information and Control (March 1961); "A Study of Methods for Systematically Abbreviating English Words and Names," Journal of the ACM (in press).

CB

The Record cc: J. Norton (2) 11 July 1962

Contracts - P. L. Browning

ES-4058

During the absence of the undersigned on 29 June 1962, it was disclosed that the above referenced contract had been terminated by RCA. This termination notice was informally told to C. P. Bourne by the technical people at RCA. Mr. Bourne advised them that the Institute would stop work on this project immediately and no further charges would be accumulated.

Inasmuch as RCA has accepted the Institute's standard research agreement which contains a 30 day termination notice, this office has requested Jim Norton to find out what late costs if any have been recorded and what they are for, so that a satisfactory termination dettlement may be worked out with RCA.

As of the date of this memo no formal notification of the termination of the agreement has been received by this office, however the undersigned has contacted Mr. B. B. Saffe of RCA that in fact the agreement has been terminated and that he was in the process of forwarding a termination notice.

This office will keep the Division advised of any future developments.

05

STANFORD RESEARCH INSTITUTE

PROJECT AUTHORIZATION

PROJECT NO.
BB-4058
B1(643)-4058 MASTER & SUBS

~~[illegible]~~ STOP WORK MASTER & SUBS 6-29-62

Client: Radio Corporation of America
Classification: Unclassified
Client Authorization: P.O. 112068-4001-82-L17
Project Supervisor: R. C. Amara; D. Ford -2
Project Title: Technical Assitance in the Evaluation of of Equipment
Project Leader: C. P. Bourne -1
Starting Date: 4-9-62
Proposal No.: BB 62-31
Termination Date: 10-9-62
Project Funds: $12,000.00
Issue No.: ~~1~~ ~~2~~ 3 Date: ~~[illegible]~~ ~~[illegible]~~ 7-20-62

SCOPE: Sub numbers have been established as follows:

B1(643)-4058-1	10,600.
B1(643)-4058-2	1,400.

FISCAL INFORMATION

Overhead	110%
Development	15%

CLIENT CONTACT AND ADDRESS

Radio Corporation of America
Data Systems Division
8500 Balboa Blvd
Van Nuys, California

BILLING INSTRUCTIONS

Send invoices to above address via J. C. Norton.

GENERAL COST CALCULATIONS (IN DOLLARS PER YEAR)

RACE SYSTEM NO. 1 (ON-LINE)

INITIAL FILE SIZE	0.	
RATE OF RETURN	6.0	PERCENT
AMORTIZATION PERIOD	50.0	MONTHS
BURDEN RATE	25.0	PERCENT OF DIRECT LABOR
OVERHEAD RATE	100.0	PERCENT OF LABOR+BURDEN

ONE PAGE PER FILE ITEM

ROW	~~NAME~~ COST ELEMENTS	COST	MINS/(input) ITEM	MINS/SEARCH	MISC.	2ND SHIFT
1	KEYPUNCH OPER.	320.00	1.00000	2.00000	.00000	368.00
2	TAB OPERATOR	400.00	.02083	.03330	.00000	460.00
3	CAMERA OPERATOR	400.00	.07812	.00000	.00000	460.00
4	DEVELOPER OPER.	400.00	.00187	.00000	.00000	460.00
5	CARD-PUNCH OPR.	400.00	.00013	.00000	.00000	460.00
6	MAGAZINE LOADER	400.00	.00015	.00000	.00000	460.00
7	RACE-FILE OPER.	400.00	.00000	.50000	.00000	460.00
8	CLERK-MISC	280.00	.20000	.20000	.00000	.00
9	QUALITY CONTROL	400.00	.00781	.00000	.00000	.00
11	MANAGER	750.00	.00000	.00000	1000.00000	1250.00
12	ASST-MANAGER	650.00	.00000	.00000	20.00000	.00
21	IBM 026 KEYPNCH	60.00	.50000	1.00000	.00000	36.00
22	IBM 056 VERIFYR	55.00	.50000	1.00000	.00000	33.00
23	IBM 084 SORTER	275.00	.00330	.00330	.00000	165.00
24	IBM 087 COLLATR	250.00	.00500	.00000	.00000	150.00
25	IBM REPRO-PUNCH	140.00	.01250	.00000	.00000	84.00
26	CARD CAMERA	1600.00	.07812	.00000	.00000	.00
27	DEVELOPER	1500.00	.00187	.00000	.00000	.00
28	CARD PUNCH	1000.00	.00013	.00000	.00000	.00
29	QUAL CONT. STN.	500.00	.00781	.00000	.00000	.00
35	32 MAGAZNE RACE	200000.00	.00000	.00833	*	.00
36	MISC.HAND EQUIP	1000.00	.00000	.00000	.00000	.00
43	INDEX TAB CARDS	.00	.00300	.00150	.00000	.00
45	RACE CARD STOCK	.00	.00280	.00000	.00000	.00
46	RACE MAGAZINE	.00	.00310	.00000	.00000	.00
47	PAPER + TONER	.00	.00000	.00800	.00000	.00
48	MAINTENANCE	.10	.00000	.00000	.00000	.00

GENERAL COST CALCULATIONS (IN DOLLARS PER YEAR)

RACE SYSTEM NO. 1 (ON-LINE)
INITIAL FILE SIZE 0.
RATE OF RETURN 6.0 PERCENT
AMORTIZATION PERIOD 50.0 MONTHS
BURDEN RATE 25.0 PERCENT OF DIRECT LABOR
OVERHEAD RATE 100.0 PERCENT OF LABOR+BURDEN
ONE PAGE PER FILE ITEM

ROW	NAME	COST	MINS/ITEM	MINS/SEARCH	MISC.	2ND SHIFT
1	KEYPUNCH OPER.	320.00	1.00000	2.00000	.00000	368.00
2	TAB OPERATOR	400.00	.02083	.03330	.00000	460.00
3	CAMERA OPERATOR	400.00	.07812	.00000	.00000	460.00
4	DEVELOPER OPER.	400.00	.00187	.00000	.00000	460.00
5	CARD-PUNCH OPR.	400.00	.00013	.00000	.00000	460.00
6	MAGAZINE LOADER	400.00	.00015	.00000	.00000	460.00
7	RACE-FILE OPER.	400.00	.00000	.50000	.00000	460.00
8	CLERK-MISC	280.00	.20000	.20000	.00000	.00
9	QUALITY CONTROL	400.00	.00781	.00000	.00000	.00
11	MANAGER	750.00	.00000	.00000	1000.00000	1250.00
12	ASST-MANAGER	650.00	.00000	.00000	20.00000	.00
21	IBM 026 KEYPNCH	60.00	.50000	1.00000	.00000	36.00
22	IBM 056 VERIFYR	55.00	.50000	1.00000	.00000	33.00
23	IBM 084 SORTER	275.00	.00330	.00330	.00000	165.00
24	IBM 087 COLLATR	250.00	.00500	.00000	.00000	150.00
25	IBM REPRO-PUNCH	140.00	.01250	.00000	.00000	84.00
26	CARD CAMERA	1600.00	.07812	.00000	.00000	.00
27	DEVELOPER	1500.00	.00187	.00000	.00000	.00
28	CARD PUNCH	1000.00	.00013	.00000	.00000	.00
29	QUAL CONT. STN.	500.00	.00781	.00000	.00000	.00
35	32 MAGAZNE RACE	200000.00	.00000	.00833	*	.00
36	MISC.HAND EQUIP	1000.00	.00000	.00000	.00000	.00
43	INDEX TAB CARDS	.00	.00300	.00150	.00000	.00
45	RACE CARD STOCK	.00	.00280	.00000	.00000	.00
46	RACE MAGAZINE	.00	.00310	.00000	.00000	.00
47	PAPER + TONER	.00	.00000	.00800	.00000	.00
48	MAINTENANCE	.10	.00000	.00000	.00000	.00

INPUT RATE	SEARCH LOAD	NAME	EQUIV. ANNUAL COST
500.	1000.		
	LABOR		
		KEYPUNCH OPER.	3929.
		TAB OPERATOR	4911.
		CAMERA OPERATOR	4911.
		DEVELOPER OPER.	4911.
		CARD-PUNCH OPR.	4911.
		MAGAZINE LOADER	4911.
		RACE-FILE OPER.	4911.
		CLERK-MISC	3438.
		QUALITY CONTROL	4911.
		MANAGER	9209.
		ASST-MANAGER	0.
	TOTAL LABOR		50959.
	TOTAL LABOR AND ADJUSTMENTS		127399.
	EQUIPMENT		
		IBM 026 KEYPNCH	736.
		IBM 056 VERIFYR	675.
		IBM 084 SORTER	3376.
		IBM 087 COLLATR	3069.
		IBM REPRO-PUNCH	1719.
		CARD CAMERA	445.
		DEVELOPER	417.
		CARD PUNCH	278.
		QUAL CONT. STN.	139.
		32 MAGAZNE RACE	55635.
		MISC.HAND EQUIP	278.
	TOTAL EQUIPMENT		66770.
	MATERIALS AND SERVICES		
		INDEX TAB CARDS	36.
		RACE CARD STOCK	17.
		RACE MAGAZINE	19.
		PAPER + TONER	98.
		MAINTENANCE	5691.
	TOTAL MATERIALS AND SERVICE		5862.
	TOTAL COSTS		200033.

INPUT RATE	SEARCH LOAD	NAME	EQUIV. ANNUAL COST
1.	100000000.		
	LABOR		
		KEYPUNCH OPER.	40002544.
		TAB OPERATOR	834266.
		CAMERA OPERATOR	4911.
		DEVELOPER OPER.	4911.
		CARD-PUNCH OPR.	4911.
		MAGAZINE LOADER	4911.
		RACE-FILE OPER.	12503439.
		CLERK-MISC	3256021.
		QUALITY CONTROL	4911.
		MANAGER	15349.
		ASST-MANAGER	5164125.
	TOTAL LABOR		61800299.
	TOTAL LABOR AND ADJUSTMENTS		154500740.
	EQUIPMENT		
		IBM 026 KEYPNCH	2791023.
		IBM 056 VERIFYR	2558438.
		IBM 084 SORTER	43223.
		IBM 087 COLLATR	3069.
		IBM REPRO-PUNCH	1719.
		CARD CAMERA	445.
		DEVELOPER	417.
		CARD PUNCH	278.
		QUAL CONT. STN.	139.
		32 MAGAZNE RACE	2169772.
		MISC.HAND EQUIP	278.
	TOTAL EQUIPMENT		7568804.
	MATERIALS AND SERVICES		
		INDEX TAB CARDS	1841919.
		RACE CARD STOCK	0.
		RACE MAGAZINE	0.
		PAPER + TONER	9823569.
		MAINTENANCE	217105.
	TOTAL MATERIALS AND SERVICE		11882594.
	TOTAL COSTS		173952130.

INPUT RATE	SEARCH LOAD	NAME	EQUIV. ANNUAL COST
1.	1.		
	LABOR		
		KEYPUNCH OPER.	3929.
		TAB OPERATOR	4911.
		CAMERA OPERATOR	4911.
		DEVELOPER OPER.	4911.
		CARD-PUNCH OPR.	4911.
		MAGAZINE LOADER	4911.
		RACE-FILE OPER.	4911.
		CLERK-MISC	3438.
		QUALITY CONTROL	4911.
		MANAGER	9209.
		ASST-MANAGER	0.
	TOTAL LABOR		50959.
	TOTAL LABOR AND ADJUSTMENTS		127399.
	EQUIPMENT		
		IBM 026 KEYPNCH	736.
		IBM 056 VERIFYR	675.
		IBM 084 SORTER	3376.
		IBM 087 COLLATR	3069.
		IBM REPRO-PUNCH	1719.
		CARD CAMERA	445.
		DEVELOPER	417.
		CARD PUNCH	278.
		QUAL CONT. STN.	139.
		32 MAGAZNE RACE	55635.
		MISC.HAND EQUIP	278.
	TOTAL EQUIPMENT		66770.
	MATERIALS AND SERVICES		
		INDEX TAB CARDS	0.
		RACE CARD STOCK	0.
		RACE MAGAZINE	0.
		PAPER + TONER	0.
		MAINTENANCE	5691.
	TOTAL MATERIALS AND SERVICE		5691.
	TOTAL COSTS		199861.

INPUT RATE	SEARCH LOAD		NAME	EQUIV. ANNUAL COST
1000.	100000000.			
		LABOR		
			KEYPUNCH OPER.	40002544.
			TAB OPERATOR	834266.
			CAMERA OPERATOR	4911.
			DEVELOPER OPER.	4911.
			CARD-PUNCH OPR.	4911.
			MAGAZINE LOADER	4911.
			RACE-FILE OPER.	12503439.
			CLERK-MISC	3256021.
			QUALITY CONTROL	4911.
			MANAGER	15349.
			ASST-MANAGER	5164125.
		TOTAL LABOR		61800299.
		TOTAL LABOR AND ADJUSTMENTS		154500740.
		EQUIPMENT		
			IBM 026 KEYPNCH	2791023.
			IBM 056 VERIFYR	2558438.
			IBM 084 SORTER	43223.
			IBM 087 COLLATR	3069.
			IBM REPRO-PUNCH	1719.
			CARD CAMERA	445.
			DEVELOPER	417.
			CARD PUNCH	278.
			QUAL CONT. STN.	139.
			32 MAGAZNE RACE	2169772.
			MISC.HAND EQUIP	278.
		TOTAL EQUIPMENT		7568804.
		MATERIALS AND SERVICES		
			INDEX TAB CARDS	1841956.
			RACE CARD STOCK	34.
			RACE MAGAZINE	38.
			PAPER + TONER	9823569.
			MAINTENANCE	217105.
		TOTAL MATERIALS AND SERVICE		11882703.
		TOTAL COSTS		173952240.

8 June 1962

RACE-1 (ON-LINE)

HEADING INFO FOR PRINTOUT

Estimates of Equivalent Annual Operating Costs (in dollars per year)

Race-1 on-line system:
Initial File Size: Zero
Rate of Return: 6 percent
Amortization Period: 50 ~~60~~ months
Burden: 25 percent of direct labor
Overhead: 100 percent of loaded direct labor
Number of pages per file item: 1

Range of Variables for Initial Run

Input Rate (items/month)	Request Volume (no.reports/mo.)--limited to 11
1	1
100	100
500	1000
1000	~~5000~~
5000	10000
10000	50000
20000	100000
50000	500000
70000	1000000
100000	10000000
200000	50000000
1500000	100,000,000
700000	
1000000	
2000000	
5000000	
7000000	
10000000	
20000000	
50000000	
100,000,000	

General Assumptions for cost analysis procedure:

1. None of the equipment or people operate for more than 2 shifts (16 hours) per day, 22 days per month.

2. If more machine time is needed, that machine is used on the 2nd shift. If still more machine time is needed, then add another machine on 1st shift, and so on.

3. Add a 15% salary increase for 2nd shift workers.

4. For a given shift, and within each labor type, personnel are free to work on any machine that requires their services. That is, a tab operator is not particularly slaved to a particular machine.

5. People and equipment are obtained and charged for in integer amounts. That is, the system cannot use "1/2 time" people or equipment.

6. The system is to operate essentially in real time, permitting no extensive backlog to develop. That is, the operating procedures must be such as to answer search requests within two working days.

7. Unless otherwise noted, annual equipment maintenance cost for purchased equipment = 10% of capital equipment cost. This includes labor and necessary replacement parts. The cost of maintaining leased equipment is included in the lease price (e.g. as with the IBM equipment).

8. Equipment amortization to be done in 50 months at 6% rate of return.

9. Labor Burden (vacation, sick leave, retirement, etc.) = 25% of direct labor cost.

10. Overhead (power, insurance, office supplies and furniture, telephone, etc.) = 100% of loaded direct labor.

11. The equipment cost excludes excise and sales taxes (10-15% extra cost in some cases).

12. The requests are uniformly distributed throughout the file collection.

13. The system is to provide high quality image storage (e.g. 120 lines/mm resolution) suitable for storing maps, drawings, and other detailed images.

14. General cost factors not specifically considered in this analysis: floor space, file conversion, parallel operation with present system, file corrections or modifications to data in the file.

15. If used in this analysis, floor space cost = \$5 per square foot per year.

    RACE - 32 magazine module = 2 x 12 = 24 ft^2

    Printer = 24 ft^2

    Cameras and Processors (each) = 16 ft^2

    } not including aisle or working space

16. Each request will result in a single copy printout. Requests for multiple copies of the same image will be treated in this analysis as separate requests.

RACE-1 (with 1 page per file item)

EQUIPMENT REQUIREMENTS:

Input:	No. min./input item	($) Purchase	Monthly Rental 1st Shift	Monthly Rental 2nd Shift(60%)
1. IBM 026 Keypunch	0.50000	--	60.	36.
2. IBM 056 Verifier	0.50000	--	55.	33.
3. IBM 084 Sorter 1/300	.00333	--	275.	165.
4. IBM 087 Alpha Collator 1/200	.00500	--	250.	150.
5. IBM Reproducing Punch 1/80	.01250	--	140.	84.
6. Race Card Camera-Punch Unit 10/128	.07812	16,000	--	--
7. Race Card Developer $\frac{60}{250(128)}$	.00187	15,000	--	--
8. Race Card Punch $\frac{1}{60(128)}$	.00013	10,000	--	--
9. Race Quality Control Station $\frac{1}{128}$	.00781	5,000	--	--

Request:	No. minutes/request			
1. IBM 026 Keypunch	1.0	--	60.	36.
2. IBM 056 Verifier	1.0	--	55.	33.
3. IBM 084 Sorter 1/300	.00333	--	275.	165.
4. Race with 32-mag. raceway(print-limited)	.50000	200,000	--	--

An additional Race will be required whenever the file exceeds (32)(256)(128) = 1,048,576 items, or the printing load exceeds 2(60)(60)(16)=115,200 pages per day.

Other:

1. Fixed initial cost for each installation for a hand viewer, notcher, and retrieval jig = $1,000.
2. Annual maintenance cost of RACE equipment = 10% of the capital equipment cost. This results in a fixed monthly maintenance cost of 0.833 percent of the capital equipment cost.

RACE-1 (with 1 page per file item)

LABOR REQUIREMENTS

Input:	No. minutes/input item		1st Shift Monthly Salary	(15% extra) 2nd Shift Monthly Salary
1. Keypunch & Verifier operator $\frac{1}{2}+\frac{1}{2}$	1.0		320	368
2. Tab operator 1/300+1/200+ 1/80 = 1/48	.02083		400	460
3. Race operator-camera 10/128	.07812		400	460
4. Race operator-developer 60/(250)(128)	.00187		400	460
5. Race operator-card punch 1/(60)(128)	.00013	.08027	400	460
6. Race operator-load mag. and files (5 min/mag)=5/(256)(128)	.00015		400	460
7. Clerk-misc. handling	.2		280	322
8. Quality control inspector 1/128	.00781		400	460

Output:	No. Minutes/output request	1st Shift Monthly Salary	2nd Shift Monthly Salary
1. Keypunch and verifier operator 1+1	2.0	320	368
2. Tab operator 1/300	.00333	400	460
3. Race operator-file (print-limited)	.5	400	460
4. clerk-misc. handling	.2	280	322

Other:

1. One Manager ($9,000 per year) needed full time as soon as the system requires more than 5 people. As soon as the system requires over 40 people, this manager will require a salary of $12,000 per year. As soon as the system requires over 100 people, the manager will require a salary of $15,000 per year.

2. One assistant manager ($8,000 per year) is needed when the staff reaches a total of 20 people. One additional assistant manager is required for each additional group of 20 people.

3. Common to both the RACE and the aperture card system:
   clerk-file index tab cards = 0.5 min/input item
   clerk--collate output w/requests = 0.2 min / request

MATERIAL COSTS

Input:			cost per input item (dollars)
1.	index tab cards	1.50/10000+1.50/1000	.0030
2.	RACE card stock	0.35/128	.00274
3.	RACE magazine	100/(256)(128)	.00305

Request:		cost per request (dollars)
1.	tab cards 1.50/1000	.00150
2.	paper and toner 0.8cents/page	.00800

Other: (Common to RACE and aperture card system):		cost per input item(dollars)
1.	microfilming to obtain 35-mm silver negative	.15000 per input image
2.	provide duplicate diazo roll film 7.50/500	.01500 per input image
3.	index tab card file 120/30,000	.004000 "

manual
Quality Control - should it be included

— done after the race card punch. (resolution, density, spot checks, check that images are in the original frame.

1 min / card inspection time
5K equip.
salary? 400/mo.

no conversion costs figured — no corrections or modifications

Better Q.C.
material on-line RACE } requirements for later
floor space requirements

cost of duplicate files when additional parallel files are required
price books for materials, etc.

changes on RACE drawing

check 8 cents/image microfilm cost

~~Verify & keypunch input = 1/2 min each.~~

~~Race card camera $10,000 (instead of 25) and 10 min/card~~

~~128 images / card~~

~~256 cards/mag.~~

~~Race card punch $10,000 and 60 cards/min.~~

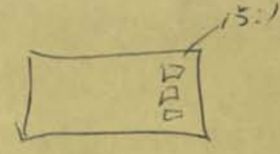
~~follow Race card punch w/ Q.C.~~ ←

~~Remove collation of output from analysis.~~

16 May 62

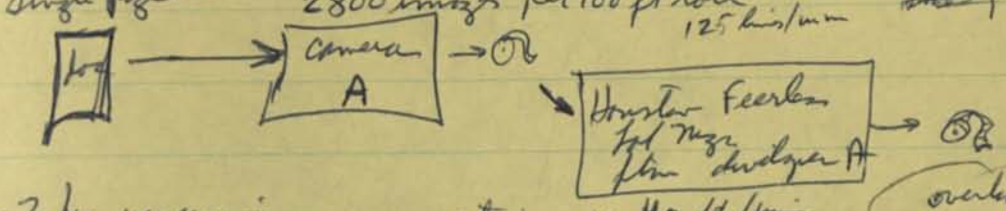
meeting at RCA Van Nuys

Gerry Richards
Hans Ullman
Charles Fenwick



output = single copy of hard copy. (single side printing) shooting for 12 lines/mm

Recordak flow camera $3000 ~~2895~~) stands, etc., speed etc.
average sustained rate = 30 exposures/min.
2800 images per 100 ft roll
125 lines/mm ~~shooting~~



continuous → 10-40 ft./min
$5000. cost of chemicals, etc?
does this include drying?

overlooking quality control & checking.

assume 2 hrs. pre-processing & cleaning for every 1 hrs. of filming.

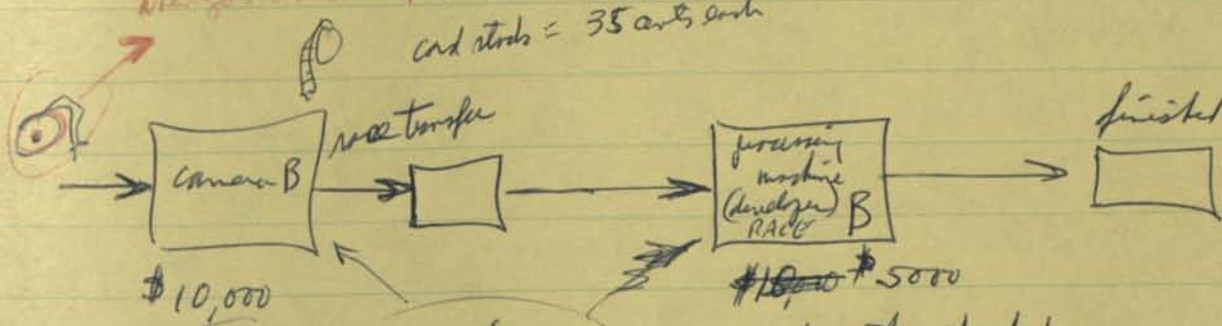
doc sent somewhere else out of the system

→

figure file conversion w/ 1-2 yr. amortization ? time to convert ?

then goes to archive (some files in OK) overlook this for now.

card stock = 35 cents each



$10,000

2 min/card exposure time

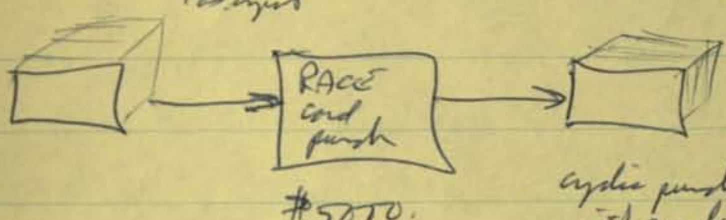
~~$10,000~~ $5000
cards go thru the short way at 40 ft/min.
200 cards/min.
chemicals, etc.?

personnel analysis -

Would like AR&D recommendations by mid-July for RCA Committee report

— Zero code is not used.

batches of 256 cards (as required) can punch in selected nos. by a dial (for updating)



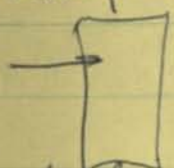
#5010.
120 cards/min.

cyclic punching
with numbers 0-256

→ load into magazines
(must have completely full magazines – with dummies or extra cards, otherwise the groups don't work)

time to load a magazine from punch & put it into RACE: say 5 min / magazine.

missing doc.



[10" per minute with [illegible]]

each item stamped w/ accession # (manually). multiple page items have only the 1st page stamped

1 tab card/item is prepared at this time.

cols 1-7 = accession #
8-9 = group # (group = 1-32 magazines)
10-11 = magazine # (within a group)
12-14 = RACE card #
16-18 = aperture # (for 0-125) of starting page
19-23 = no. of apertures for this item.
24-25 if you have to go to another card, this no. tells you how many cards you have to expect
26 no. of magazines to expect.

can get this

at time of stamping, this info is taken off.
you don't know what card you're on or what page you're on.

need more clarification?

update procedure

simple: remove all index traces of where this card is (to keep it from being referenced) & add new entry. then periodically pull out the bad images & clear the old address for re-use. & prepare a new RACE card from archive storage.

update cost = (cost we have this to put an item in the file) + (effort need provide this to change the index)

OUTPUT

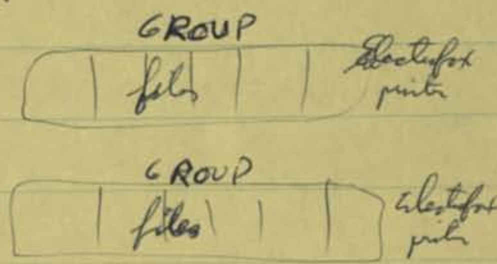
~~punch~~

1 tab card / ~~request~~ has the accession # on it & all the necessary address data.

Cards are sorted & batched into order by group (& magazine?) ~~accession~~

if more than 1 group is used

i.e. if more than 1 printer is required.



need a cardreader for each ~~group~~ (station) ~~stack~~

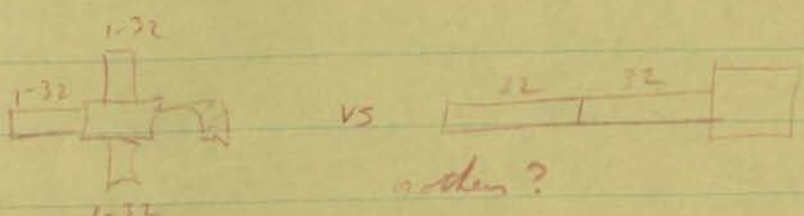
IBM Mod 523 — or some cheap card reader 100 c/min is faster than the printer can take.

$100^K for 32 magazines frame (air supply, gates, electronics, cabinets)

add $100 for each magazine

printer cost $100^K cut sheet printer, not a roll printer.

add another line of 32 ~~station~~ when 1st 32 is full. (at cost of magazines & cabinets) + 30K?



access time always ≤ 180ms because of card buffer to be used.

$100K

Printer — cut sheet fed in asynchronous ∴ little waste paper.

2 photo stations (½ sec in on each side) —

~~2~~ pages/sec (printed on both sides) — 4 pages/sec (printed on single sides)

3 pages/sec effective rate ~~? check this.~~

1 cent/page for Electrofax paper

1 qt toner / 200,000 pages = ? } say 1.2 cents/page unless told otherwise

~~Eight hours / 200,~~

assemble, staple, & group items w/ requests, & postage.

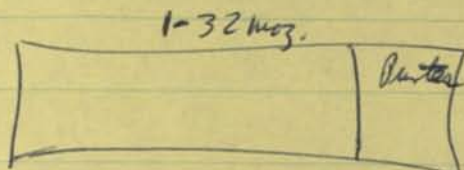
effort ?

maintenance cost ? say ~~[illegible]~~ 15% of capital cost.

~~replacement of insurance~~ preventive maint. - 4 hr. / day.

30 minutes to replace 32 magazines w/ 32 mag. from a file cabinet.

access



This is the basic unit we will use.

there will be a hold station s.t. the access time to get a card for printing, is negligible.

use manual magazine storage (eg Patent office)

use non linear (exponential) relationships.

space — $5 per sq. ft. of floor space.

Space requirement

32 mag module = 12×2 = 24 sq. ft.

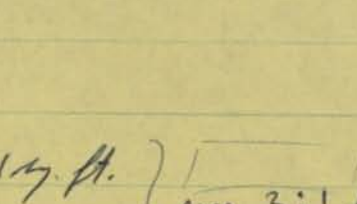
printer = 24 sq ft.

} use 3:1 for determining aisle space, etc.

Camera = 16 sq ft each

processor = " "

storage racks for magazines



? 2?

(10) magazines per running ft. of 5' shelving

STANFORD RESEARCH INSTITUTE

To:

Date:

From:

Location:

Subject:

Answering:

---

assume

1. uniform dist. of strikes against the files
2. print only on one side of paper @ 2 pp/sec.
3. Hold station on each RACE
4. full magazines in system only
5. Batching of requests – 1 per day

RACE costs.

	#
1. Raceway & Holding Sta. } 2. Power & air Supply.	150 K
3. Control for 32 mag. } 4. Printer & Read Sta.	50 K
5. Magazines	100
6. Cards	35¢
7. Files	$60 /48 mags.
8.	
9. Hand viewer	$500 ?
10. Hand notcher	$50
11. " decoder	$100

SRI-255

STANFORD RESEARCH INSTITUTE

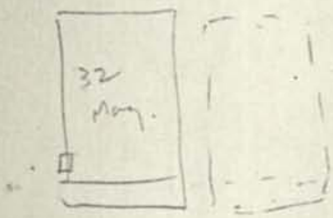
To: Date:

From: Location:

Subject: Answering:

---

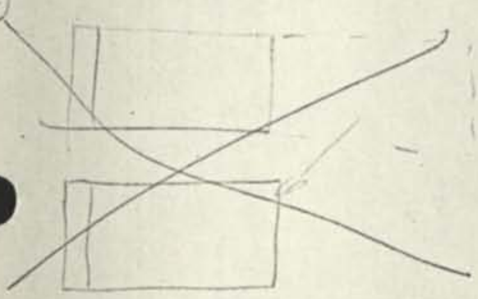
1)



add identical modules
if print capacity or file
cap. is inadequate

---

2)

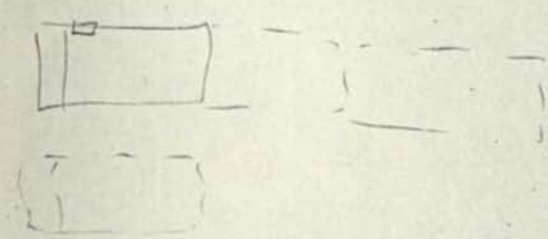


a) add storage modules in series - not to exceed N modules, if storage capac. is inadequate (how does access time vary w/ words dist)

b) in printer saturates: add printer (& redistribute the file?)

---

3)

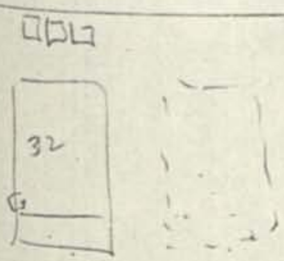


same as 2.

½ sec access time from any module

---

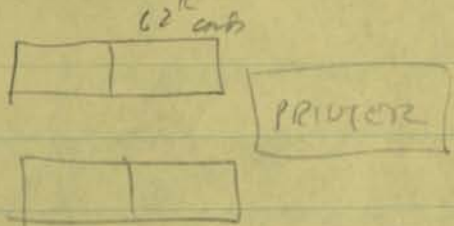
4)



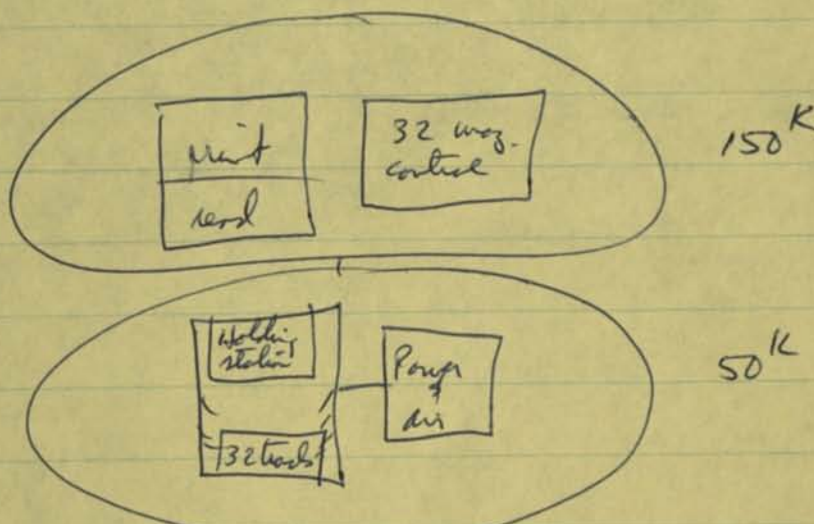
1. all magazines stored on shelf & loaded at search time
2. put mags on shelf when storage capac. is exceeded
3. add modules as print capacity is exceeded

RCA visit – J.C. Flint 5 June 62
at SRI

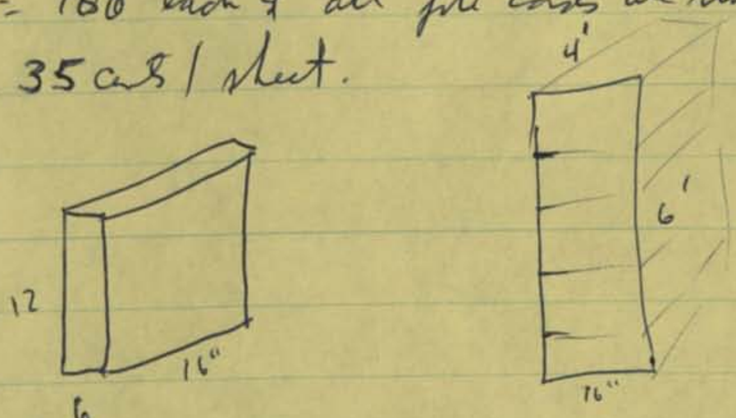
Matching – assume 1 calendar day response to search requests
use 1 page/item printout as 1 test case, then maybe 10 or 100.
Prices – 4 RACE + 1 comprehensive Filmsort



4 cabinets to meet storage requirement, but 45 gets to handle print vol. – How to resolve conflict?



Magazines = $100 each & all file cards are stored in magazines
film stock ≈ 35 cents / sheet.



= $60.

hand viewer $200 – 2000.
hand notcher $150.
hand retrieval jig $250.

~~NO [illegible]~~

input = $8\frac{1}{2} \times 11$" items

file conversion to be included? over what period of time? at what amortization?
NO — maybe model test and competition

what are the GS rates?

~~How long is roll of printer paper?~~ assume no lost time to paper change ~~How long to change a roll?~~

microfilm filming cost? ~~planetary vs flow?~~

GS-3 page, 1.75 – 2.00 hr ± for RACE loader

say 6-8 ~~5~~ cents/input item for single sheets for file conversion — large batch, pause breaks
(2- 3×5 cards positioned on a page may double)

~~8 [illegible]~~

8 cents/input item for regular input

page 245
pt 2

Aperture Card System - 1 (with 1 page per file item)

EQUIPMENT REQUIREMENTS

Input:	No.min/input item	Purchase	Monthly Rental($) 1st Shift	Monthly Rental($) 2nd Shift
1. IBM 026 Keypunch	0.5	--	60.	36.
2. IBM 056 Verifier	0.5	--	55.	33.
3. IBM 084 Sorter 1/300	.00333	--	275.	165.
4. IBM Reproducing Punch 1/80	.01250	--	140.	84.
5. IBM 557 Interpreter 1/80	.01250	--	165.	99.
6. Quality Control Station 1/15	.06666	1500	--	--
7. Filmsort Automatic ~~Moniter~~ Mounter 1/30	.03333	32,500	--	--
Request:	**No. min/request**			
1. IBM 026 Keypunch	1.0	--	60.	36.
2. IBM 056 Verifier	1.0	--	55.	33.
3. IBM 084 Sorter 1/300	.00333	--	275.	165.
4. Mosler Selectronic tab card file & Filmsort 086 card copier	1.0	4000+1100	--	--
5. Xerox 24-C Printer 1/16	.06250	177,000	--	--

An additional Mosler Selectronic file will be required for each additional 350,000 file items, and for each additional 1(60)(16) = 960 requests per day.

An additional Xerox 24-C Printer will be required for each additional 16(60)(16) = 15,360 requests per day.

An additional Filmsort automatic mounter will be required for each additional 30(60)(16) = 28,800 input items per day.

Other:

1. Maintenance cost of Xerox 24-C printers = $50/month.

Annual maintenance cost of Filmsort Automatic Mounter = 10% of purchase price.

Aperture Card System - 1 (with 1 page per file item)

LABOR REQUIREMENTS

monthly salary ($)

Input:	No.min/input item	1st shift	2nd shift(15%extra
1. Keypunch operator $\frac{1}{2}+\frac{1}{2}$	1.0	320	368
2. Tab Operator 1/300 + 1/80 + 1/80	.02833	400	460
3. Clerk - misc. handling	.25	280	322
4. Quality Control Inspector 1/15	.06666	400	460
5. Filmsort Automatic Mounter Operator 1/30	.03333	280	322

Request:	No.min/output request		
1. Keypunch Operator 1+1	2.0	320	368
2. Tab Operator 1/300	.00333	400	460
3. Mosler tab file operator	1.0	300	345
4. Printer operator 1/16	.06250	280	322
5. Clerk - misc. handling	.25	280	322

Other:

1. Manager and assistant manager as defined for RACE - 1
2. Common clerks for RACE and aperture card systems as defined for RACE-1

Aperture Card System - 1 (with 1 page per file item)

MATERIAL REQUIREMENTS

Input:	Cost/input item (dollars)
1. Diazo duplicate (mounting) roll 7.50/500	.01500
2. Index tab cards 1.50/1000	.00150
3. Flank aperture cards 40/1000	.04000

Request:	Cost/Request (dollars)
1. Tab cards 1.50/1000	.00150
2. Duplicards 44/1000	.04400
3. Paper, toner, and all other expendable supplies for the Xerox 24-C printer	.00600

Other: (Common to RACE and aperture card systems)	Cost per input item (dollars)
1. Microfilming to obtain 35-mm silver negative	.15000 per input image
2. Provide duplicate (updating) diazo roll film 7.50/500	.01500 per input image
3. Index tab card file 120/30,000	.00400 per input image

6/26/62

Notes on a telephone conversation with Bill Whitney, Bay Microfilm, Palo Alto
(DA 6-1812)

cb

----

1. They have seen the Filmsort Uniprinter 041 (card-to-card printer) in operation at Sandia. There are only 4-6 of these machines in the U.S. today. The rated speed is 2000 cards/hour, but there are still enough bugs in the equipment to reduce the effective rate to 1000-1200 cards/hour.

2. For their own card copy work, they use a Technifax card copier ($700) rather than a Filmsort unit since it gives them better copies.

3. For their own aperture card mounting, they use a Filmsort Semi-Automatic Mounter ($4070) that provides a mounting jig as well as a viewing screen for inspection purposes. It mounts a 100' roll of 35-mm film, chops, and mounts the 2" frames in a standard Mil-D card. The equipment is rated at 400 cards/hour, but the effective rate, as used at Bay Microfilm is 175-200 cards/hour. The aperture cards cost about $40/thousand in small lots.

4. Engineering drawings are filmed at Bay Microfilm to exact DOD specs (120 lines/mm resolution, etc.) at a cost of 24 cents per image for sheet drawings and 38 cents per image for roll drawings--regardless of drawing size (from A to E size). For this price, they will take stacks of drawings, pre-sorted by size, and provide the images in 100' rolls of silver negative 35-mm microfilm. If the specs can be relaxed to omit some of the special DOD requirements, then the work can be done at a cost of 15 and 25 cents per image for sheet and roll drawings, respectively.

5. Aperture card mounting is done at a price of 2 7/8 - 3 cents per card.

6. If business records, engineering charge orders, or other images are to be recorded that do not require as much resolution as the engineering drawings, then they can be filmed ~~and provide~~ 4 to a frame (i.e. 4 page images to each 35-mm aperture) at a cost of about 40 cents per aperture card--assuming that the records are pre-sorted by color and size.

~~Mr. Lowell Ross~~ (MMM - Filmsort) Millbrae ~~PL 6-0800~~
761 - 1155
Film .041 - 25^k 2000/hr.—

Bill Gannon 3MM ←

Filmsort auto Mounter : ~~rate?~~ only 1 few built.
purchase: 32,500
rental :
maint.: ≈ 10% /yr. 65/yr. for 600 hours

Filmsort 086 card copier : rate? 300/hr. — effective
purchase : 1100 1021.
rental : 1 yr lease : 85/mo.
10/mo for 2nd yr.
maint. included

flash op. card cost } ? mil-D 5-100^k 39.43/thous.
Dryplasiod " " } over 100^k - 250^k 36.50/thous
5-100^k 44.11/thous.
41.34 "

Literature & prices?

How many images on a 100' reel of 35-mm film? Mil spec ≈ 500

(# 4071)

Filmsort Semi-Auto Mounter 100' roll 2" pull down Mil D ap. card

cut & mount 400 cards / hr.

(175-200) " (effective)

must ~~remove~~ install aperture apertures - this takes

viewing screen allows check of drawing

this is OK ~~[illegible]~~ $40 / thous. for small lots

RCA
16 May

Aperture Card System

2860 ap. cards / drawer
22 drawers / standard cabinet } ≈63,000 / module

consider each gal can handle 4 of these cabinets

doc.

use service bureau costs if available.

pre-process to remove staples, etc. (same as RACE)

→ Camera A (same as RACE) → 35mm (same as for RACE) → Develop & same as RACE →

roll-to archive

→ Roll-to-Roll Contact Printer → diazo 35mm

effective rate? use service bureau rate: 6.00 / roll for processing + labor.

→ Filmcard Mounter (086) 1000–2000 cards/hr. → mil D aperture 35mm

use service bureau costs.

doc.

tab card prepared for every input item & interpreted

item: accession #, no. of pages in the item, station (= block of 220,000 cards)

→ reproduce blank aperture cards, interpret ap. cards,

→ maint. cost of tab equip.

Is there to be a file conversion? Over what period? How much?

• [box] → goes to manual file drawer

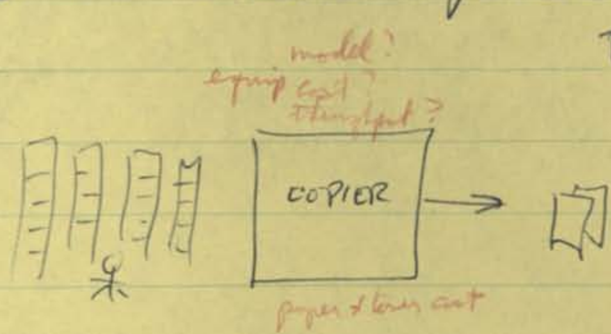
---

request – prepare request as tab card. – Sort by accession #
distribute to stations

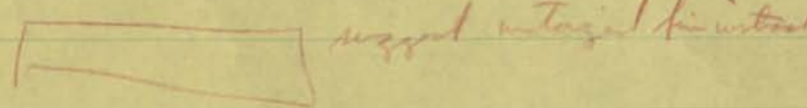
search & retrieve by batch – can pull 1 card in 15 sec. for random request
(regardless of ~~length~~ no. of cards pulled in response)

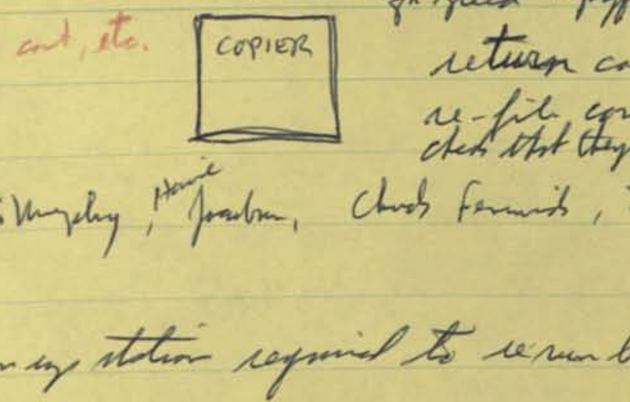
hard copy output (start with copier at each station, & then go to Copyflow)

Then manually re-file the cards
collate & staple output, &
collate w/ order card, &
put in "out" basket.



vs.





high speed Copyflow from ap. cards. (single side printing)
return cards to stations.
re-file cards (twice as much effort say/min.)
check that they're all there, & insert back in file.

IS+R Committee J. Mingsley, Howie Josephson, Chuck Fenwick, Sid Kaplan (DSC), Bill Woods

Exceptions – clean up station required to rerun bad images, etc.

(RCA) VIDEO FILE SYSTEM

DESCRIPTION

The video recorder is used to store, on magnetic tape, document images produced by a camera or transferred from another recorder. The recorder is also used to reproduce, in video form, any given stored image according to a coded index number recorded on digital tracks along the tape edge. The recorder is capable of automatically searching for a video frame, given an index number or a serial number, and then to read out, or erase and write over any such addressed frame. The equipment has two modes of operation:

(1) A high speed search mode with tape movement in either direction.

(2) A slower video mode used for recording or playback with tape movement in the forward direction only.

Sub-Units

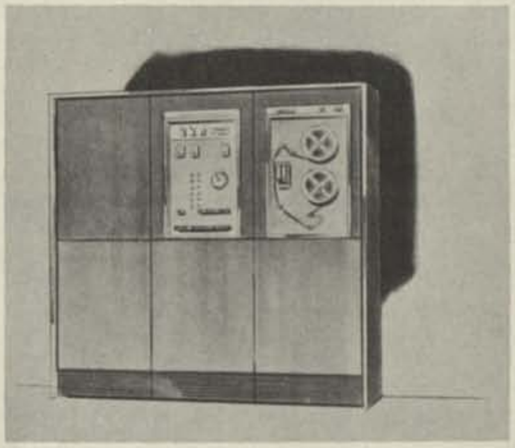
Tape Transport-Contains all components for reading, writing, and erasing digital and video signals on magnetic tape. Included is a tape transport mechanism, servos, read and write amplifiers, modulator and demodulator, and a service control panel.

Electronic Unit - Contains all electronic switching and addressing circuits necessary for controlling all recorder functions.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

In the high-speed search mode, control is exercised by sensing the digitally coded sequential index number from a control track at the side of the tape. The index number, representing the position of the magnetic tape, is compared with the index number of the required document. While sensing the digital data track, the tape is moved at speeds up to 300 in/sec in the correct direction.



RECORDER

As the tape approaches the desired frame (within a distance depending upon the tape speed) a slow-down-to-stop program stops the tape at a fixed number of frames before the desired position. The recorder is subsequently switched to the video mode and the video head wheel is brought to bear upon the magnetic tape; the tape is automatically brought into synchronism with the master clock (external to the recorder) by frame pulse, and the video-control clock track by line pulse. The tape head wheel then starts to scan successive images under internal servo control until the addressed frame is reached. The video signal is used to drive either a printer, or another recorder.

Alternatively, a serial number, which is not sequential, may be written in the digital tracks and may be addressed. The equipment will then find the corresponding sequential index number and use it in the same manner as if the index number had been originally inserted.

While recording a document, the recorder operates in synchronism with the master clock to control the camera operation.

SPECIFICATIONS

• Video Recording	Transverse Scan
• Recording Medium	Magnetic tape 2 inches wide
• Tracks	1 video, 4 digital
• Tape Reels	14-inch reels (7200 feet)
• Scan Rate	2 frames per second
• Transverse Tracks	160 per frame
• Transverse Tracks	4 per headwheel revolution
• Horizontal Lines	10 per track
• Track Pitch	15 mils
• Track Width	10 mils
• Video Response	1.5 mc at 90%
• Linear Tape Search Speed	300 inches per second
• Record Speed	4.8 inches per second
• Electronic Unit	All electronic switching and addressing circuits necessary for controlling all recorder functions required by the record, printout and merge tapes operations.

PHYSICAL CHARACTERISTICS

Overall dimensions (approximate) - Height 72 inches, width 47 inches, depth 22 inches. Weight (approximate) - 1400 pounds.

OPERATING REQUIREMENTS

Power (approximate) - 3.5 KW at 120 volts, ±10%, 60 cycles, single phase; 4.4 KVA. Average dissipated heat 12,200 BTU/hr.

Cooling - This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than ±2°F from a reference point over a 6-hour period. The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6-hour period, the relative humidity should not vary more than ±5% from a nominal operating point.

4 Dec   Video-file changes

old	new
1. 70,000 datasets/roll	36,000 data./roll
2. tape cost $100./roll	$425/roll.
3. tape replaced after 10,000 passes	replaced after 20,000 passes
4. output material = $1.06/search	out. mtl. = $1.20/search
5. video file equip = $1. mill.	= $1.35 mill.

6. search time:

1 strike/pass

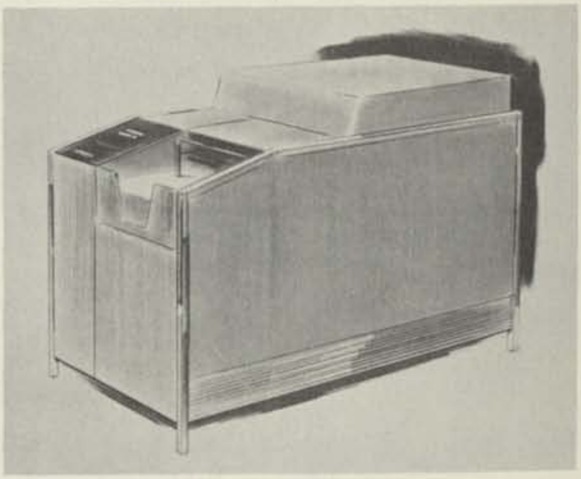
10 strikes/pass

50

100

RCA VIDEO FILE SYSTEM

CAMERA



DESCRIPTION

The camera unit is used to convert images contained on documents to composite video signals capable of being stored on magnetic tape by the Video File Recorder. In addition, a unique index number is stamped on the document and is recorded on the digital control track at the same time the composite video is recorded. This index number is used for subsequent identification of the video frame.

In operation, stacks of single sheet documents feed at a uniform rate past the optics of an electro optical scanner and are discharged into a bin. During the time the document passes through the equipment one or both sides can be scanned.

Sub-Units

Document Feed Unit - Contains all paper handling mechanisms necessary for accepting stacks of paper documents; feeding the documents one sheet at a time past the electro optical scanners and discharging the sheets into a bin. This unit also contains the indexable document stamp.

Scanner - Contains all components necessary for the conversion of images contained on the documents to composite video signal.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

The feed mechanism consists of a hopper and a rotating vacuum drum which align the sheets for scanning. The drum alternately sucks the page to the curved surface and then blows it off after the scanning mechanism is passed. When two sides of a document are to be scanned, a secone drum accepts the page from the first drum, and then presents the opposite side of the page to a second scanning mechanism. The rotating drums are synchronized to the system master clock (external to the camera), and one video frame is produced for each side of a document scanned. The movement of the sheets is perpendicular to the plane containing the image. Sheet movement is at constant velocity and distance from the optical scanner. The paper scan start and end positions are sensed by a photocell whose output signal is used to initiate the video recording.

The electro optical scanner consists of a kinescope which generates a circular trace. The face of the tube is in contact with a fiber optic bundle the other end of which illuminates the line of paper being scanned. Photo multiplier tubes detect the reflected light and, after amplification, generate the video signals for recording. The document indexing stamp applies a seven digit decimal number to the document prior to recording. This number is automatically incremented after printing (starting from an initial manual setting) and consists of a five digit frame or index number plus a two digit tape reel number.

SPECIFICATIONS

• Document Size*	10 by 12 inches maximum 7-3/8 by 3-1/4 inches minimum
• Paper Type	16-1b bond or card stock
• Document Rate	1 sheet per second when recording both sides, or 2 sheets per second when recording one side.
• Scanning Line Pitch	125 lines per inch
• Line Scan Rate	3200 lines per second
• Video Bandwidth	1.5 mc
• Signal to Noise Ratio	30 db

Note: Documents are not acceptable if stapled, folded, wrinkled, or abraded.

* When various sizes of documents are to be recorded, they must be placed in the hopper only one size at a time.

PHYSICAL CHARACTERISTICS

Overall dimensions (approximate) - Height 60-1/2 inches, width 30 inches, and depth 72 inches. Weight (approximate) - 800 pounds.

OPERATING REQUIREMENTS

Power (approximate) 1.5 KW at 120 volts, ±10%, 60 cps, single phase; 2 KVA. Average dissipated heat 5,300 Btu/hr.

Cooling – This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than ±2°F from a reference point over a 6 hour period.

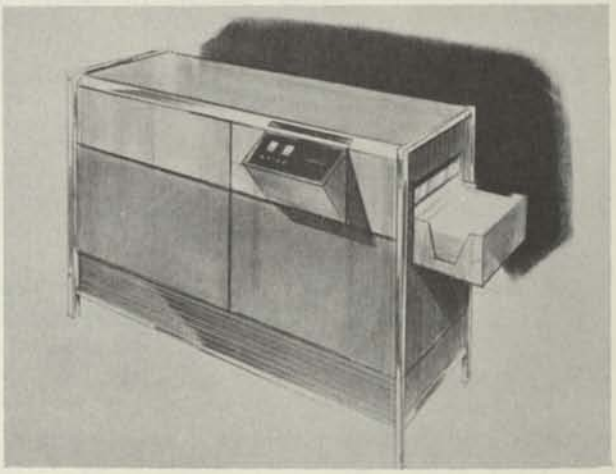
The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6 hour period the relative humidity should not vary more than ±5% from a nominal operating point.

RCA VIDEO FILE SYSTEM

DESCRIPTION

The printer unit is used to obtain hard copy printouts of images stored on video magnetic tape. The unit employs the electrofax system for video reproduction, and delivers either a continuous roll or a succession of sheets. Regardless of the size of the original document, single copies are produced in size 10 by 12.8 inches.

In addition to reproducing the exact recorded image each printed document may be stamped with a three digit destination number.



PRINTER

After the document is reproduced a three digit decimal number may be stamped on the document to indicate the desired recipient. This number is supplied from the digital tracks on the video tape via the recorder.

Sub-Units

Paper Feed Unit - Contains all paper handling mechanisms necessary for accepting blank paper roll, controlling the paper flow through the printer, cutting the paper when sheet printouts are required, and temporarily storing the output reproductions.

Reproducer - Contains all components necessary for the conversion of composite video signals back to the original images.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

Reconstituted video signals from the recorder are applied to the electro optical scanner within the printer. The electro optical scanner consists of a fiber optic bundle and kinescope which generates a circular trace. The face of the tube is in contact with the fiber optic bundle whose other end forms a straight line of equal width as that of the maximum width paper. The illumination from the kinescope is transferred through the fiber optic bundle to the charged paper surface upon which the image is to be reproduced. The action of the light is to discharge the paper so that a developer can be applied to render the image visible.

SPECIFICATIONS

• Paper Flow Rate	25.6 inches per second
• Document Size	10 by 12.8 inches
• Scanning Line Pitch	125 lines per inch
• Line Scan Rate	3200 lines per second
• Horizontal Resolution	Sine wave response of 70% at 900 TV lines in 10 inches
• Fiber Optic Linear Deviation	±0.04 inch
• Scanning Linearity	±0.5%

PHYSICAL CHARACTERISTICS

Overall dimensions (approximate) · Height 60 inches, width 30 inches, and depth 96-1/2 inches. Weight (approximate) - 1000 pounds.

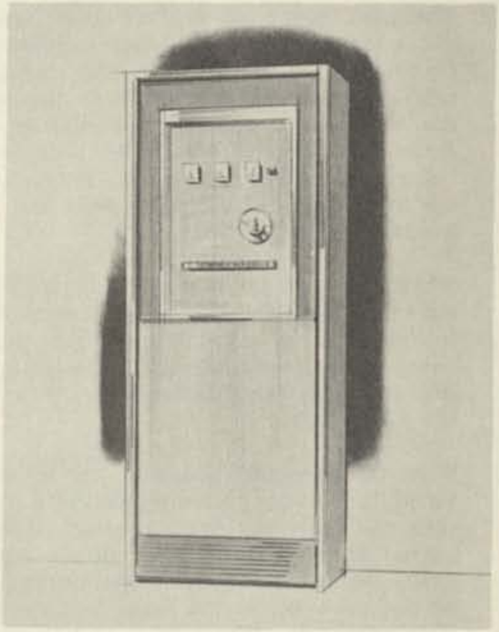
OPERATING REQUIREMENTS

Power (approximate) 2 KW at 120 volts, ±10%, 60 cps, single phase; 2.5 KVA. Average dissipated heat 7,000 Btu/hr.

Cooling - This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than ±2°F from a reference point over a 6-hour period.

The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6-hour period, the relative humidity should not vary more than ±5% from a nominal operating point.

VIDEO FILE SYSTEM



MERGE CONTROL

DESCRIPTION

The merge control unit is used to simultaneously address and control two recorder units during the merging video tapes operation. This equipment functions to:

(1) Update a master file by erasing and replacing out-of-date documents by their newer counterpart.

(2) Update a master file by adding additional documents to the video tape's empty locations.

(3) Stack or batch requested document images from the master file on an output video tape along with a three digit decimal number indicating origin of the request.

(4) Sort and merge the master file video tapes.

(5) Insert digital information, such as serial number, upon the control digital track of the master file tapes.

Sub-Units

Paper Tape Reader- Reads information from seven-level punched paper tape. Includes unwind and take-up facilities for handling approximately 400 feet of tape; it also handles short lengths of tape.

Electronic Unit - Contains all electronic switching and addressing circuits necessary for controlling the paper tape reader and two recorders.

Power Supply–Furnishes the necessary power to all components at the required voltage levels.

OPERATION

The operator makes the system ready for use by (1) placing video tapes in the tape transport unit of each recorder and making the necessary adjustments for satisfactory recording, and (2) placing punched paper tape previously prepared by a computer on the paper tape reader contained in the merge control unit.

After the system is set up, the START button is pressed, and the paper tape is incremented to the first character, which is the operand digit. The operand is decoded in the merge unit for one of four possible operations. These are:

(1) Update the master file by erasing and replacing a video frame on the master tape by a video frame on the input tape.

(2) Update the master file by transferring a video frame from the input tape to an empty location on the master tape. Simultaneously, digital information from the paper tape (such as serial number) can be inserted on the control digital track of the master tape.

(3) Stack or batch requested documents by transferring a video frame from the master tape to an empty location on the input tape. Simultaneously, a three-digit decimal number from the paper tape representing the request originator is inserted on the control digital track of the input tape.

(4) Insert only digital information such as serial number upon the control digital track of a master file tape.

After the operand is decoded and the operation determined, a switch is selected according to the required direction of video signal flow. In the case of digital transfer alone, none of the video switches are selected.

The next seven characters read from the paper tape indicate the index number of the frame on the master file tape to be acted upon. This index number is transferred to the recorder containing the master tape. After the transfer has been completed, a "start-searching" signal is transmitted to the recorder and a high-speed search for the frame commences.

While the recorder is performing its search, the following eight characters are read from the paper tape. At this time, the operation being performed determines the interpretation of these control characters read from the paper tape.

When the operation concerns the transfer of video information, seven of the eight characters designate the index number of the required frame on the input tape. For this case, the index number is transferred to the second recorder in the same manner as the first recorder. Similarly, a "start-searching" signal is also transmitted to the second recorder which then also commences search at high speed. As the search control logic in each recorder approaches the desired index number to within a predetermined number of frames, it initiates a slow-down-to-stop program. When a recorder has stopped at a predetermined number of frames ahead of the desired frame, a "recorder-ready" signal is transmitted to the merge control unit.

While both recorders are searching for the ready position, the paper tape is again advanced by an additional eight characters. The characters read at this time represent digital information, if any, to be inserted on the video tape digital track along with the transferred video. This information represents a serial number when a new video frame is to be transferred to an empty location on a master tape, or it may identify the request originator if a video frame is to be transferred from the master tape to an empty location on the input tape.

When the merge control unit receives both "recorder-ready" signals, it transfers the digital information described above to the proper recorder, and simultaneously starts both recorder units in synchronism. Both recorders initiate a start program to bring the video tapes to record speed in servo lock and in frame synchronization. While the video tape is still ahead of the required frame, the recorder control system performs a serial subtraction between the required index and the present index number. The difference between the index numbers represents "the number of frames to go" until the required frame is reached. The merge control unit receives this difference information from both recorders and performs a comparison to determine if the recorders are in synchronism. When the recorders are not in synchronism, one of the "number of frames to go" will be smaller, and that recorder is commanded to slip one frame by the merge control unit.

The merge control unit also decodes the difference information for "one frame to go". If both recorders simultaneously attain "one frame to go", the merge control unit allows the transfer of the video information during the next frame. If one recorder attains "one frame to go" before the other, the video transfer is inhibited and the merge control unit reinitiates the "start-searching" signal and the foregoing process is repeated. After two unsuccessful attempts to complete the operation, an alarm is set.

While the video and digital information are being transferred, the paper tape in the merge unit is again advanced to read the operand and the master file index number. Upon completion of the preceding operation, the system is ready to start the next operation.

When the operation concerns only the addition of digital information with no video transfer, the processing is significantly simplified. All operations, as mentioned previously, are considered to be the same for the first seven paper tape characters following the operand character. These characters indicate the index number of the frame on the master tape recorder along with the "start-searching" signal. While the recorder is performing its high-speed search, the following eight characters, representing a serial number, are read from the paper tape. When recorder operation is completed and the recorder stopped, the "recorder-ready" signal is transmitted to the merge control unit and the serial number is transmitted to the recorder. The recorder is then started and the serial number is recorded upon the control digital track adjoining the required frame. While the digital information is being inserted, the paper tape is again advanced to read the operand and the master file index number. Upon completion of the preceding operation, the system is ready to start the next operation.

SPECIFICATIONS

PHYSICAL CHARACTERISTICS

Overall Dimensions (approximate) - Height 72 inches, width 19 inches, and depth 22 inches. Weight (approximate) - 450 pounds.

OPERATING REQUIREMENTS

Power (approximate) - 1 KW at 120 volts, ±10%, 60 cps, single phase; 1.25 KVA. Average dissipated heat 5, 500 Btu/hr.

Cooling - This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than ±2°F from a reference point over a 6 hour period.

The relative humidity must not drop below 20% nor exceed 65% within a maximum dew point at 58°F. Within a 6 hour period, the relative humidity should not vary more than ±5% from a nominal operating point.

RCA VIDEO FILE SYSTEM

STATUS AND SWITCHING PANEL

DESCRIPTION

The status and switching panel is used to provide a centralized panel whereby the condition and usage of each equipment in the system is indicated. In addition, this equipment provides a central interconnection point for all system equipments. Provision is also made for mounting the system master clock and a spare master clock in this unit housing. These clocks, however, are not part of the status and switching panel.

Sub-Units

Indicator Panel - Contains all indicators required to display the status of each unit in the system.

Patch Panel - Contains terminations for all system equipments to be cabled to this panel.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

During operation, all system equipments are interconnected at the status and switching panel. Whenever an equipment fails, the subsystem can be maintained operationally by patching in an identical available unit in place of the failed unit. This allows complete interchangeability of all similar system equipments.

The status panel indicates in one centralized location, the condition and interconnection of each equipment. No control functions are performed at this panel.

SPECIFICATIONS

Status Display per Equipment	(1) Available
	(2) In use
	(3) Out of service
	(4) Alarm

PHYSICAL CHARACTERISTICS

Overall dimensions (approximate)- Height 72 inches, width 48 inches, and depth 28 inches. Weight (approximate) - 1000 pounds.

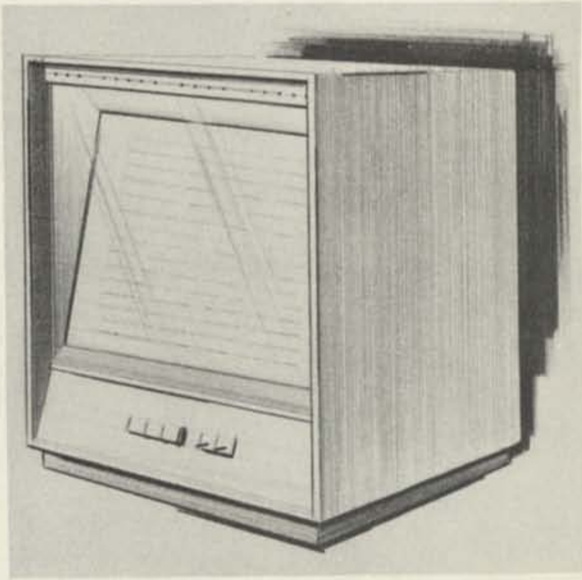
OPERATING REQUIREMENTS

Power (approximate) - 0.5 KW at 120 volts, ±10%, 60 cps, single phase, 0.65 KVA. Average dissipated heat 1,800 Btu/hr.

Cooling - This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F the temperature should not be allowed to vary more than ±2°F from a reference point over a 6 hour period.

The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6 hour period, the relative humidity should not vary more than ±5% from a nominal operating point.

RCA VIDEO FILE SYSTEM



SOFT COPY READER

DESCRIPTION

The soft copy reader is used to display any document obtained in a roll output from the Video File System printer unit. The reader is capable of being incremented by one document, in the forward or backward direction, in response to pushbutton control. In addition, the unit is capable of moving the document roll at high speed in either direction in response to pushbutton control.

The reader is used to display one complete 10 by 12.8 inch frame. The entire mechanism is of desk top size.

SPOOLER

The spooler unit provides a high speed, off line video tape rewind mechanism.

VIDEO FILE SYSTEM

DESCRIPTION

The monitor unit is used to observe the picture quality and evaluate the waveform display of the frequency components of a video signal. The monitor can be plugged into any video signal contained within the Video File System. This unit is portable with a self-contained power supply.

Sub-Units

Kinescope and Associated Electronics - Contain all components necessary for viewing the composite video signals.

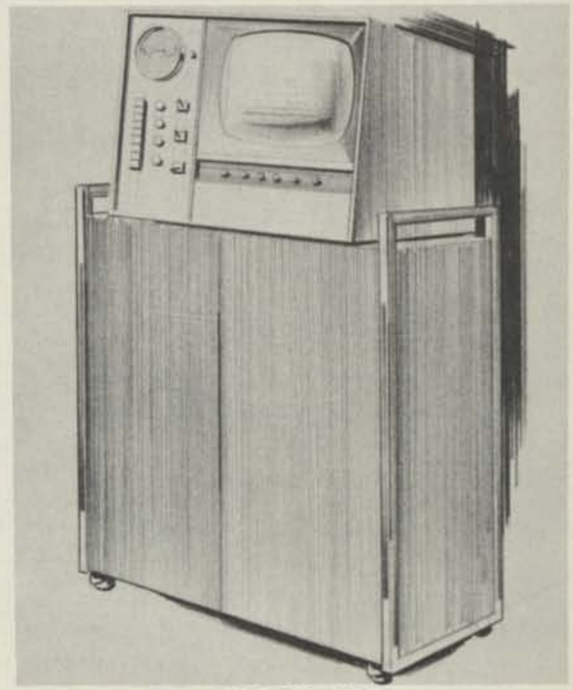
CRO and Associated Electronics - Contains all components necessary for displaying input signal waveforms.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

The monitor unit contains a rectangular kinescope, which utilizes low voltage electrostatic focusing and a 90-degree deflection system, to display a picture 6-3/4 by 4-5/8 inches.

The 3-inch CRO tube has a flat faceplate to minimize parallax between lines. An edge-lighted scale, calibrated for sync or composite signals, is located over the face of the CRO for evaluating the frequency components of the input signal waveform. The CRO deflection is at half horizontal or half vertical scan frequency, permitting two cycles of the waveform to be displayed. All the operating controls normally used are located on the front panel.



MONITOR

SPECIFICATIONS

• Composite Video	0.3 to 2 volts p-p
• Sync	4 volts nominal p-p
• Picture Size	7-3/4-inch diagonal
• Brightness	200 foot lamberts
• Bandwidth	2 mc
• Linearity	±2%
• Display	3-inch flat face with green filter
• Waveform	1/2 horizontal or 1/2 vertical sweep rate

PHYSICAL CHARACTERISTICS

Overall dimensions (approx) Height 15-7/8 inches, width 8-1/2 inches, and depth 20 inches. Weight (approximate) - 49 pounds.

OPERATING REQUIREMENTS

Power (approximate) 0.2 KW at 120 volts, ±10%, 60 cps, single phase; 0.25 KVA. Average dissipated heat 700 Btu/hr.

Cooling- This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than + 2°F from a reference point over a 6 hour period.

The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6 hour period, the relative humidity should not vary more than + 5% from a nominal operating point.

VIDEO FILE SYSTEM

DESCRIPTION

The master clock is used to provide the horizontal and vertical synchronizing signals to all units of the Video File System consistent with the frame and line rate. In addition, this unit provides a central timing source for all digital controls in the Video File System.

The unit is completely self-contained in a independent housing, but provision is made for storing two master clocks in the status and switching panel.

Sub-Units

Electronics Unit - Contains all counters, pulse shapers and drivers required to obtain the output synchronizing signals.

Power Supply - Furnishes the necessary power to all components at the required voltage levels.

OPERATION

The master clock consists of a crystal controlled oscillator for the primary standard frequency source. This primary frequency is shaped and counted down to obtain the required synchronizing frequencies. The output from the clock are pulses, from low impedance drivers, at the frequency and duration as required by the Video File System.

SPECIFICATIONS

• Composite Sync	Modulated 3.2 KC
• Horizontal Drive	3.2 KC
• Traverse Trigger	320 cps
• Head-wheel Trigger	80 cps
• Vertical Drive	2 cps
• Vertical Blanking	2 cps
• Digital Clock	500 cps
• Grating Signal	Test pattern

PHYSICAL CHARACTERISTICS

Overall dimensions (approximate)- Height 8 inches, width 20 inches, and depth 12 inches. Weight (approximate) - 49 pounds.

OPERATING REQUIREMENTS

Power (approximate) 0.5 KW at 120 volts, ±10%, 60 cps, single phase; 0.65 KVA. Average dissipated heat 1,800 Btu/hr.

Cooling- This equipment operates satisfactorily in a temperature range from 65°F to 85°F. Below 70°F and above 80°F, the temperature should not be allowed to vary more than ±2°F from a reference point over a 6 hour period.

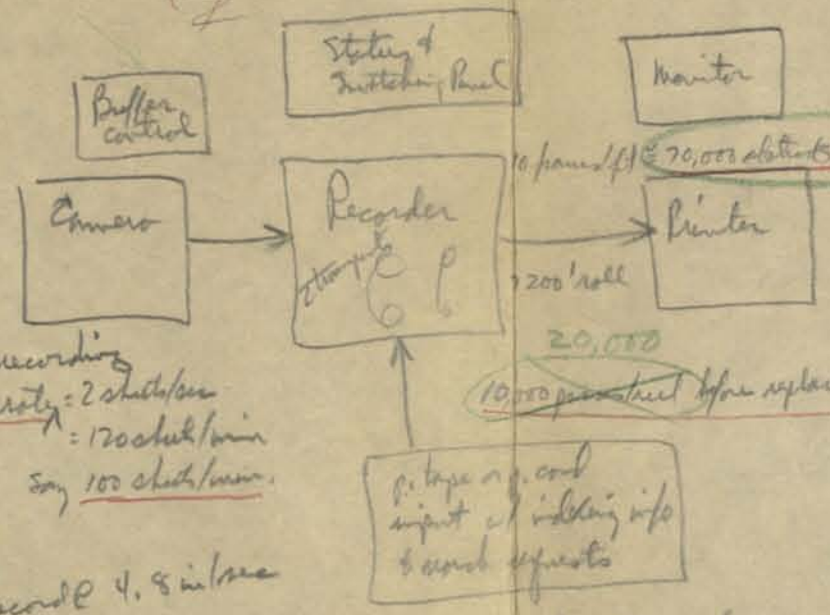
The relative humidity must not drop below 20% nor exceed 65% with a maximum dew point at 58°F. Within a 6 hour period, the relative humidity should not vary more than ±5% from a nominal operating point.

Video File A

assume 1 search at a time.
assume each search produces 60 abstracts

No displays consoles or remote

video file total of equip. cost = $1.35 mill.
Maint, power, insurance, etc = $1000/mo.
keypunch & verifier



36,000
~~34,000~~/roll
10 frames/ft ~~70,000 abstracts/roll~~
1200'/roll
output rate = 25 in/sec
1-2 sheets/sec
maybe 4 sheets/sec
20,000
~~10,000 passes~~ before replacement
c 1500 in/sec

recording input rate = 2 sheets/sec
= 120 sheets/min
say 100 sheets/min.

record @ 4.8 in/sec

high speed search = 300 in/sec
4.8 min to pass a roll or rewind

tape change time = 6 min.
rewind = 5 min.

1 cent/sheet for paper
1/2 " for toner
waste, etc.
say 2 cents/sheet

System to record complete abstracts (one abstract/sheet)

for 60 abstracts: $1.20/search for materials

FIXED
machine operator: 1 min. of labor D for each min. of machine time

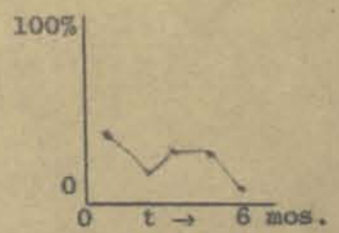
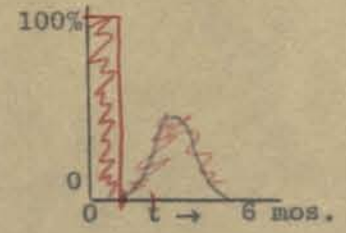
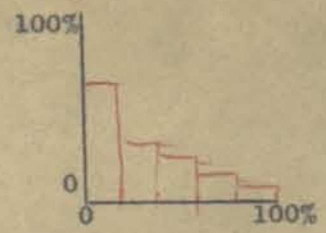
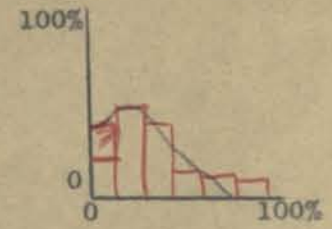
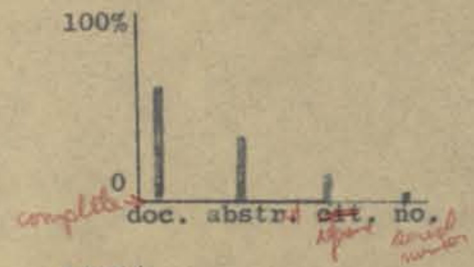
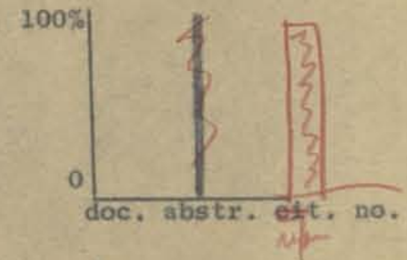
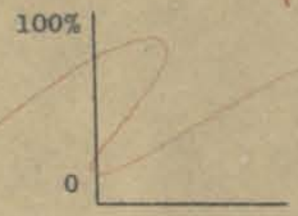
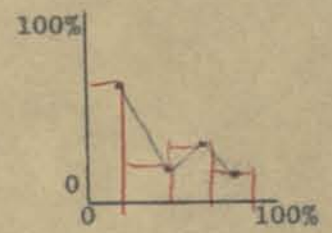
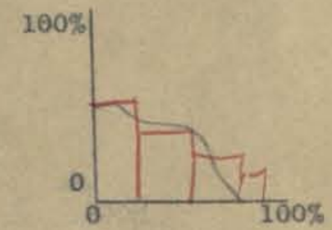
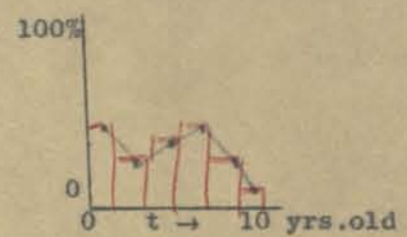
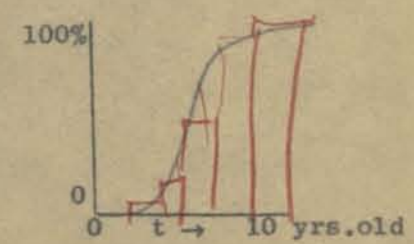
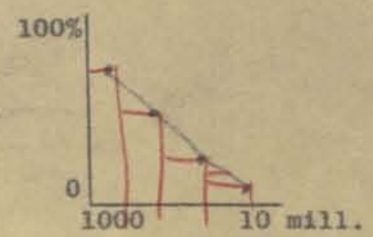
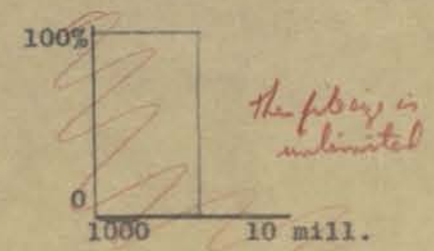
INPUT
indexing: 21 min. of labor C for each input item.
typing of abstracts: 10 min. of labor D for each input item
sorting, batching, & filing of paper abstracts & others of original source mat. 1 min of labor D for each input item
keypunching of index or serial no. = 1 min of key punch & verifier operator time/input item (labor D)
keypunch & verifier = 1 min/input item
tape machine time = 5 min. setup + .01 min. per input item
assume 1 batch/mo. for input
tape material = 1 reel of video tape for each ~~70,000~~ 36,000 input items @ $100. ~~250~~ 425.
p. card for indexing info. = 1 card for each input item

SEARCH
accept inquiry & code in machine language = 20 min. of labor for labor C for each search
punch search info into machine language = 1 min of key punch & verifier time/search (labor D)
package output for delivery = 10 min/search of labor D

keypunch & verifier time = 1 min/search
tape machine time = 5 min. setup + 4.8 min. tape time + 0.5 min scan time = 10.3 min per search of the file in less than ~~70,000~~ 36,000 abstracts.
add 5 min tape change time for each additional 70,000 file items.
tape regeneration time = 10 min per reel of ~~70,000 abstracts~~ 36,000. The tape file is to be replaced after 10,000 searches at a material cost of $100. ~~250~~ 425 /roll
printer paper = $1.00 per search

Total
1. Labor: 1 min. of labor D for each min. of machine time
21 min of labor C and 13 min of labor D for each input item
20 min of labor C and 12 min of labor D " " search
2. Equip.: keypunch: $— 1 min. per input item and 1 min per search
Verifier: $— " " " "
Videofile complex: $3 mill. 5 min + .01 min per input item per month
10.3 min. per search of the file in ≤ 70,000 items and 10 min/search for each additional 70,000 items
10 min per reel for tape regeneration (after 10,000 searches)
3. Material: $0.05 per input item
~~1.00 per search~~ 1.20.
4. Maint, power, insurance, etc. = $1000/mo.

Sample Evaluation of the Performance of Edge-notched Card System A

REQUIREMENT	MEASURED USER REQUIREMENTS	MEASURED SYSTEM PERFORMANCE	MEASURE OF AGREEMENT BETWEEN PERFORMANCE AND REQUIREMENT	RELATIVE IMPORTANCE OF THIS REQUIREMENT	SCORE FOR THIS REQUIREMENT
Response time for major group of relevant references					
Tolerable fraction of relevant material overlooked					
Desired form of response					
Tolerable amount of effort required to communicate with the system					
Tolerable fraction of irrelevant material					
Minimum age of file contents desired					
File size					

TOTAL SCORE ______

Note: (1) The vertical scale on the plots of user requirements refer to the fraction of users with that equipment. requirement.

(2) The vertical scale on the plots of system performance refer to the fraction of searches with that performance.

INTERNAL CORRESPONDENCE

DATE: May 25, 1962

TO: G. Arnovick

FROM: J. C. Flint

SUBJECT: SRI Contract P.O. 112068-4001-82 (S.O. 219710)

REFERENCE: Machine Organization and Applications Study For SRI (S.O. 219710)

Enclosed are the results of my preliminary study, requested by C.P. Bourne, to aid SRI to apply our data to their computer model.

The range of applications being considered by DSD is too large to be successfully applied to the SRI three-dimensional model. By carefully selecting certain applicational areas within the range, we can still compare operational costs vs years, for the various pieces of equipment.

I am continuing my portion of the study in the application areas, including optimization of the machine organization of the various equipments to be compared. From this we will be able to determine the sub-areas of applications that will give a fair and honest comparison between equipments.

J. C. Flint

JCF:mw

Encl.

Machine Organization and Applications Study

For SRI (S.O. 219710)

May 25, 1962
J. C. Flint

I. APPLICATION RANGES

In order to test the capability of various equipments to perform in the IS&R and printing field, retrieval volumes in pages per day and storage volumes in pages, were chosen as shown in Figure 1. The diagonals through each square are percent-of-retrieval lines. Percentage-of-retrieval refers to the number of pages output per day divided by the storage volume X 100. On the right hand side of the page are the search-retrieve-print times per page, for each of the output levels chosen.

II. IS&R EQUIPMENT

Originally the study tentatively called for the comparison of a RACE film machine against some sort of manual film machine.

The basic storage unit of the RACE machine is one magazine containing 256 cards, 128 frames per card. This amounts to about 32,000 pages per magazine module. This is also approximately equal to one-7,200 foot video file reel or one-1,000 foot Recordak 16 millimeter film reel. This makes for easy comparison of these three equipments in terms of their basic storage module. It was decided to use an Electrofax printer (of at least 2 pages per second output) with the RACE.

The manual film system chosen was Filmsort (aperture card). The printer with this equipment would be a XEROX machine with 3 seconds per page output rate.

III. MECHANIZED SYSTEMS QUANTITIES (See Figure 1)

Since seven different retrieval volumes and eight different storage volumes are compared in the block chart, there are 56 applications represented.

The ISR machine organization chart is useful to determine the minimum number of printers and storage modules required. The number of printers is determined by dividing the printing capability of the printer, in the pages per day, into the daily requirements for the application. The number of storage modules is determined by dividing the module capability into the application storage requirements. M and P are minimal and usually are not the final figures for each application. A careful applications engineering study including optimization of the machine organization is required before any final and reliable quantities of the two equipments can be known. This is necessary before any comparison of equipment efficiency can be made.

Figure 2 gives the number of printers (P) and the number of storage modules (M) required for either RACE, video file, or the Recordak reels.

Application R2-S4 is the minimal job on the chart. The .0017 printer means that the printer only operates .17% of the day. The .312 module means that the module is only 1/3 full. The other 55 applications increase the equipment numbers, by a factor of 10 in either direction, or both, as indicated.

Applications similar to ASTIA or the Patent Office are in the range R4-R5-R6 and S7-S8. Note that the quantity of equipment grows rapidly.

IV. MANUAL FILM EQUIPMENT

The manual equipment chosen was patterned after the 3-M Filmsort equipment using aperture cards with one-35 mm frame per card. The storage module chosen was the standard 62,000 card cabinet. The printer chosen was the automatic XEROX equipment which has an output rate of 3 seconds per page.

ISR MACHINE ORGANIZATION CHART.

J.C. FLINT | 5-29-62 | ①

FIG. 1

J.C. FLINT | 5/23/62 | (1)

STORAGE & RETRIEVAL CHART

STORAGE: ONE RACE MAGAZINE ≈ ONE 7200 FOOT VIDEO FILE REEL ≈ ONE 1000 FOOT RECORDAK OPTICAL REEL,
STORAGE UNIT = M = 32,000 PAGES.
PRINTERS: ONE PRINTER = 2 PAGES PER SECOND = 57,600 PAGES PER FULL 8 HOUR DAY
PRINTER UNIT = P

PAGES OUTPUT PER DAY ↑ / PAGES STORED →	S-4 (10,000; 10^4)	S-5 (100,000; 10^5)	S-6 (1,000,000; 10^6)	S-7 (10,000,000; 10^7)	S-8 (100,000,000; 10^8)	S-9 (1,000,000,000; 10^9)	S-10 (10,000,000,000; 10^{10})	S-11 (100,000,000,000; 10^{11})	
		X=1000	X=100	X=10	100%	10%	1%		.1%
R-8 (100,000,000; 10^8)	1700-P / .312-M	1700-P / 3.12-M	1700-P / 31.2 M	1700-P / 312-M	1700-P / 3,120-M	1700-P / 31,200-M	1700-P / 312,000-M	1700-P / 3,120,000-M	.0003 SEC/PAGE; .01%
R-7 (10,000,000; 10^7)	170-P / .312-M	170-P / 3.12-M	170-P / 31.2-M	170-P / 312-M	170-P / 3,120-M	170-P / 31,200-M	170-P / 312,000-M	170-P / 3,120,000-M	.003 SEC/PAGE; .001%
R-6 (1,000,000; 10^6)	17-P / .312-M	17-P / 3.12-M	17-P / 31.2-M	17-P / 312-M	17-P / 3,120-M	17-P / 31,200-M	17-P / 312,000-M	17-P / 3,120,000-M	.030 SEC/PAGE; .0001% (ONE MILLI-PERCENT)
R-5 (100,000; 10^5)	1.7-P / .312-M	1.7-P / 3.12-M	1.7-P / 31.2-M	1.7-P / 312-M	1.7-P / 3,120-M	1.7-P / 31,200-M	1.7-P / 312,000-M	1.7-P / 3,120,000-M	.3 SEC/PAGE; .00001%
R-4 (10,000; 10^4)	.17-P / .312-M	.17-P / 3.12-M	.17-P / 31.2-M	.17-P / 312-M	.17-P / 3120-M	.17-P / 31,200-M	.17-P / 312,000-M	.17-P / 3,120,000-M	3 SEC/PAGE; .000001%
R-3 (1,000; 10^3)	.017-P / .312-M	.017-P / 3.12-M	.017-P / 31.2-M	.017-P / 312-M	.017-P / 3,120-M	.017-P / 31,200-M	.017-P / 312,000-M	.017-P / 3,120,000-M	30 SEC/PAGE; .0000001% (ONE MICRO-PERCENT)
R-2 (100; 10^2)	.0017-P / .312-M	.0017-P / 3.12-M	.0017-P / 31.2-M	.0017-P / 312-M	.0017-P / 3,120-M	.0017-P / 31,200-M	.0017-P / 312,000-M	.0017-P / 3,120,000-M	5 MINUTES/PAGE.

FIG. 2

J.C. FLINT | 5-24-62

BASIC FILM SORT - EQUIPMENT

STORAGE = M ≈ ONE CARD FILE CABINET = 62,000 CARDS
PRINTER = P = ONE XEROX @ 3 SEC./PAGE.

PAGES OUT PER DAY ↑ / PAGES STORED →	10,000x	1000x	100x	10x	100%	10%	1%	.1%
R-8 100,000,000 10^8	10,000 M P / .161	10,000 / 1.61	10,000 / 16.1	10,000 / 161	10,000 / 1,610	10,000 / 16,100	10,000 / 161,000	10,000 M P / 1,610,000
								.01%
R-7 10,000,000 10^7	1000 / .161	1000 / 1.61	1000 / 16.1	1000 / 161	1000 / 1,610	1000 / 16,100	1000 / 161,000	1000 / 1,610,000
								.001%
R-6 1,000,000 10^6	100 / .161	100 / 1.61	100 / 16.1	100 / 161	100 / 1,610	100 / 16,100	100 / 161,000	100 / 1,610,000
								.0001%
R-5 100,000 10^5	10 / .161	10 / 1.61	10 / 16.1	10 / 161	10 / 1,610	10 / 16,100	10 / 161,000	10 / 1,610,000
								.00001%
R-4 10,000 10^4	1.0 / .161	1.0 / 1.61	1.0 / 16.1	1.0 / 161	1.0 / 1,610	1.0 / 16,100	1.0 / 161,000	1.0 / 1,610,000
								} 3 SEC/PAGE .000001%
R-3 1,000 10^3	.1 / .161	.1 / 1.61	.1 / 16.1	.1 / 161	.1 / 1,610	.1 / 16,100	.1 / 161,000	.1 / 1,610,000
								.0000001%
R-2 100 10^2	.01 M P / .161	.01 / 1.61	.01 / 16.1	.01 / 161	.01 / 1,610	.01 / 16,100	.01 / 161,000	.01 M P / 1,610,000
	10,000 10^4	100,000 10^5	1,000,000 10^6	10,000,000 10^7	100,000,000 10^8	1,000,000,000 10^9	10,000,000,000 10^{10}	100,000,000,000 10^{11}
	S-4	S-5	S-6	S-7	S-8	S-9	S-10	S-11

FIG. 3

Figure 3 is the corresponding equipment chart for this equipment. Again the general ASTIA - Patent Office areas covers 10^4 to 10^6 pages out per day, and 10^7 to 10^8 pages stored.

V. RELATIVE PRICE COMPARISONS

A rough idea of the magnitude of dollars involved in each of the 56 applications can be obtained by determining the cost of the printer, the cost of the basic storage unit, and the cost of storage cards. It must be kept in mind that these are by no means the total costs, and accurate costs vs performance comparisons cannot be made from the following charts.

Figure 4 refers to RACE. Figure 5 refers to the manual system.

Note that because both equipments are print-limited, the costs go up rapidly above 100,000 pages per day retrieved on both sets of equipment. While the Electrofax is 6 times as fast as XEROX, this still does not improve the cost significantly in retrieval rates above 1 million a day.

Note also that storage volumes above 10 million, on both equipments, require from 1 to 4 million dollars in printers, storage modules, and storage cards alone. The cost of peripheral equipment, labor, paper, indexing equipment, sorting equipment, ETC are still to be added to this figure.

VI. MACHINE ORGANIZATION

Referring to figures 2 and 3, it is quite evident in the areas of R6-S5 and R4-S9 for instance, that additional storage modules and printers, respectively, are required to make a "possible" system. Therefore the cost will increase accordingly. More important, the determination of a "possible" equipment, let alone an optimal equipment special machine organization will have to be determined for each square.

This means that it is highly misleading and irrelevant to conduct the SRI 3-dimensional model tests as originally planned.

VII. CONCLUSIONS & RECOMMENDATIONS

1. The relative comparison of RACE and manual film systems can only be made over a small range of printing and storage parameters.

2. Optimization of both equipments with thorough system and machine organization investigations are absolutely necessary to insure any valuable information from the SRI results.

J.C. FLINT | 5/23/62

STORAGE & RETRIEVAL COSTS (PARTIAL)

BASED ONLY ON: ① ONE BASIC RACE ≈ \$100,000, ② ONE PRINTER ≈ \$100,000, ③ ONE MAGAZINE ≈ \$250.

	10,000X	1000X	100X	1000% 10X	100%	10%	1%	.1%	
100,000,000 R-8 10^8	\$189,070,250	\$189,070,780	\$189,077,800	\$189,148,000	\$189,850,000	\$196,070,000	\$258,737,000	\$888,870,000	.0003 SEC/PAGE .01%
10,000,000 R-7 10^7	\$19,070,250	\$19,070,780	\$19,077,800	\$19,148,000	\$19,850,000	\$26,070,000	\$88,737,000	\$718,870,000	.003 SEC/PAGE .001%
1,000,000 R-6 10^6	\$2,070,250	\$2,070,780	\$2,077,800	\$2,148,000	\$2,850,000	\$9,070,000	\$71,737,000	\$701,870,000	.03 SEC/PAGE .0001%
100,000 R-5 10^5	\$370,250	\$370,780	\$377,800	\$448,000	\$1,150,000	\$7,370,000	\$70,037,000	\$700,170,000	.3 SEC/PAGE .000 01%
10,000 R-4 10^4	↑	↑	↑	↑	↑	↑	↑	↑	3 SEC/PAGE .000 001%
1,000 R-3 10^3									30 SEC/PAGE .000 000 1%
100 R-2 10^2	\$200,250	\$200,780	\$207,800	\$278,000	\$980,000	\$7,200,000	\$70,200,000	\$700,000,000	5 MIN/PAGE
PAGES OUT PER DAY ↑	10,000 10^4	100,000 10^5	1,000,000 10^6	10,000,000 10^7	100,000,000 10^8	1,000,000,000 10^9	10,000,000,000 10^{10}	100,000,000,000 10^{11}	
PAGES STORED →	S-4	S-5	S-6	S-7	S-8	S-9	S-10	S-11	

FIG. 4

J. C. FLINT | 5-24-62

FILMSORT (PARTIAL PRICES)

BASE ONLY ON: ① XEROX PRINTER @ $30,000
② APERTURE CARDS @ $36.00/1000.
③ CARD CABINETS @ $100/CAB.

Diagonal labels: 10,000x, 1000x, 100x, 10x, 100%, 10%, 1%, .1%, .01%, .001%, .0001%, .00001%, .000001%, .0000001%

Pages out	S-4: 10,000 (10^4)	S-5: 100,000 (10^5)	S-6: 1,000,000 (10^6)	S-7: 10,000,000 (10^7)	S-8: 100,000,000 (10^8)	S-9: 1,000,000,000 (10^9)	S-10: 10,000,000,000 (10^{10})	S-11: 100,000,000,000 (10^{11})
R-8: 100,000,000 (10^8)	$300,000,460	300,003,800	300,037,700	300,376,200	303,376,200	337,620,000	676,200,000	$4,062,000,000
R-7: 10,000,000 (10^7)	30,000,460	30,003,800	30,037,700	30,376,200	33,762,000	67,620,000	406,200,000	3,792,000,000
R-6: 1,000,000 (10^6)	3,000,460	3,003,800	3,037,700	3,376,200	6,762,000	40,620,000	377,200,000	3,765,000,000
R-5: 100,000 (10^5)	300,460	303,800	337,706	676,200	4,062,000	37,920,000	376,500,000	3,762,300,000
R-4: 10,000 (10^4)	↑	↑	↑	↑	↑	↑	↑	↑
R-3: 1000 (10^3)	|	|	|	|	|	|	|	|
R-2: 100 (10^2)	$30,460	33,800	67,700	406,200	3,792,000	37,650,000	376,230,000	$3,762,030,000

PAGES OUT ↑
PAGES STORED →

FIG. 5

4 April 1962
RCA visit at SRI

rather have 3 mos. instead of 6.

May | June | July

Colins Flint – prin. contact – Boston DSD/AR&D coord.

time schedule

picked 2 of 4 systems – decide what the other 2 should be.

get AR&D Quart. reports for briefing – DSD/AR&D supplied by RCA

[illegible] Project Mgr. — [illegible] co. [illegible] in 184 R

[illegible] with the money

[illegible] in S. [illegible] Group — support to

full time on this proj. — Collins Flint. work for Group —
{Ray Clawson – Data Systems Center, [illegible], [illegible] – Systems and programming capability [illegible]
{Sid Kaplan

interested in joint with will provide support & data as requested

[illegible] Data [illegible] Systems [illegible] software org. – [illegible] data proc. field.

1st 2 systems firmly selected.

RACE (Random Access Card System)

4" x 16" edge notched card in magazine of 256 cards

2 sided magnetic card: 350 bits/in, 96 tracks total (48 on each side + 2 clock)

selected card is transported 400 in/sec past a read head

transfer rate 140 KC for 48 bit word

1–32 magazines in any transport. – 7.7 bill. bits of storage

180 msec access time to 1st ~~card in~~ magazine

450 msec

32

they have re-circulate [illegible]
180–450 to return the card | 180–450

They have a feasibility model
hope to have microfilm mag. storage capability by end of year. –

5 lb magazine ($500 each)

≤ 100K w/o aux. equip.

Image RACE

Microfilm – same card size 256 images – 16mm – 25:1 reduction resolution 12 lines/mm full size. Mylar – 500 lines/mm for film image (film ≈ 1000, optics limit to 500) silver halide film. – No indexing data – just address data – (at least for now) can split the cards. or have combination of image magazines & digital. – 2 reading stations allow Electrofax to print both sides of the paper in continuous form.

Want to compare this image storage system w/ Filmsort

store cards in magazine ($500) or in plastic bins ($5?)

no scratch problem w/ cards. they are separate in the magazines.

Digital cards make at least 5000 passes w/o damage.

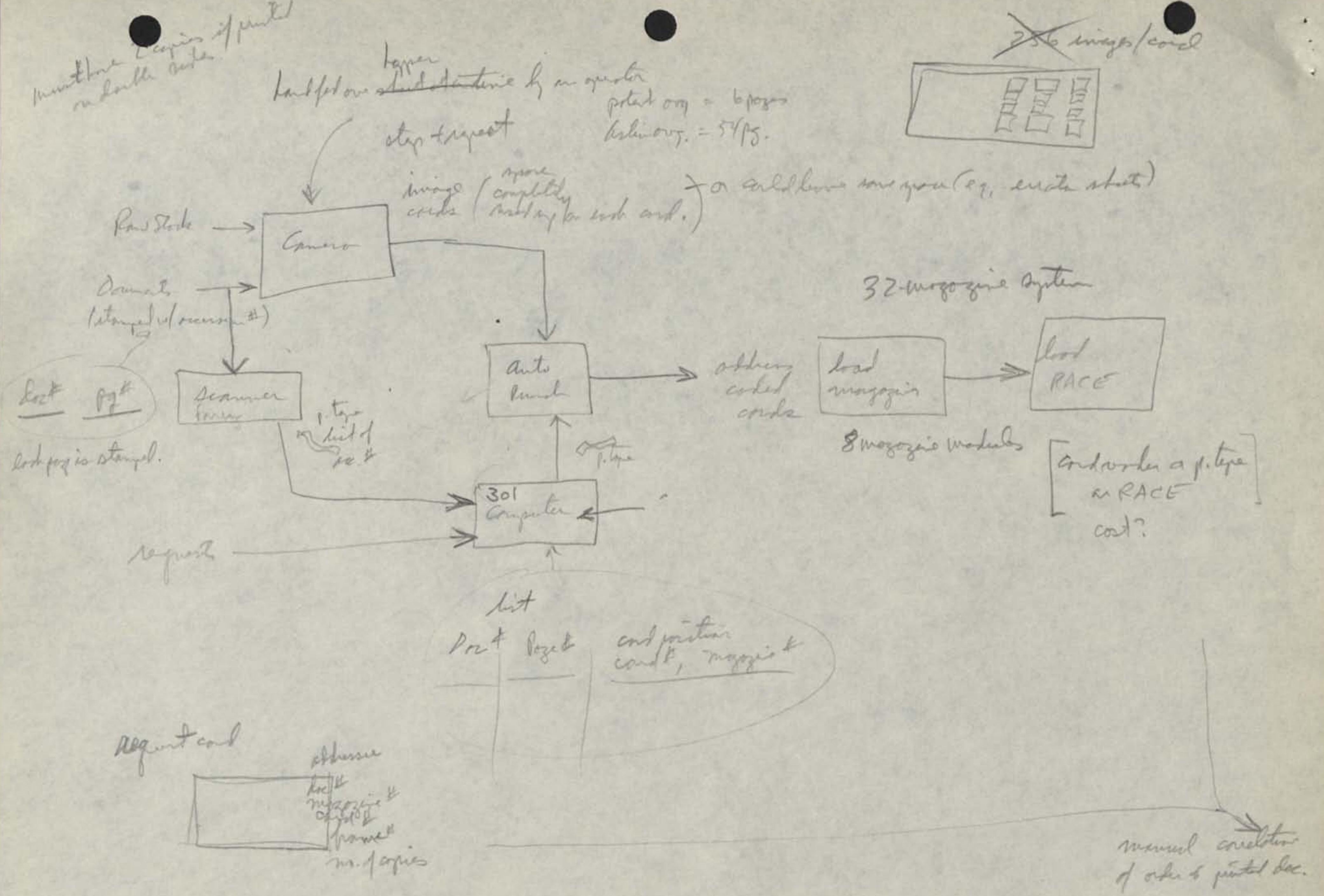
Program budget – what should ~~be done~~?

Other 2 systems — RCA internally split on RCA Video file – it may be explored by RACE. Is there any point in having both? they may be competitive. – May not be able to get video recording on a RACE card. ~~[illegible]~~ For places that need video – RACE out quit it. Video should be advantageous for ~~problem~~ application w/ file updating.

Video file may be a candidate for simulation.

1. camera speeds
2. clerical effort required to unbind, label, collate 2 copies, ^punch & rebind
3. unit card cost? (25 cents ea)
4 scanner limit?
5 and punch speeds?
time for paper change?
equip costs — Maybe by May 25
→ unit costs? ←
amortization 3 or 5 yr., maybe both rate?
301 configuration? cost?

Sep. cost
1. language processing — auto. syntactic analysis report (recent AD paper)
2. programming languages. —

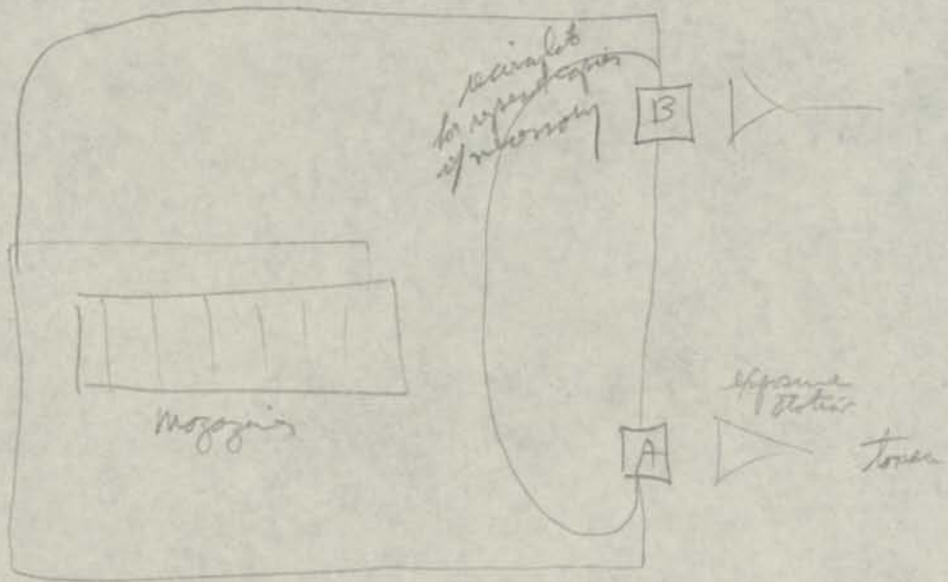

must have 2 copies if printed on double sides.
hand fed over hopper by an operator
potent avg = 6 pages
Ashim avg. = 5½ pg.
step + repeat
image cards (space completely used up on each card.) or could leave some space (e.g. errata sheets)
256 images/card
Raw Stock
Camera
Documents (stamped w/ accession #)
Doc #
Pg #
each page is stamped.
Scanner
p. tape list of pg #
Auto Punch
p. tape
address coded cards
load magazine
load RACE
32-magazine system
8 magazine modules
Card reader or p. tape on RACE cost?
301 Computer
requests
list
Doc #
Page #
card position card #, magazine #
request card
address
doc #
magazine #
card #
frame #
no. of copies
manual correlation of order & printed doc.

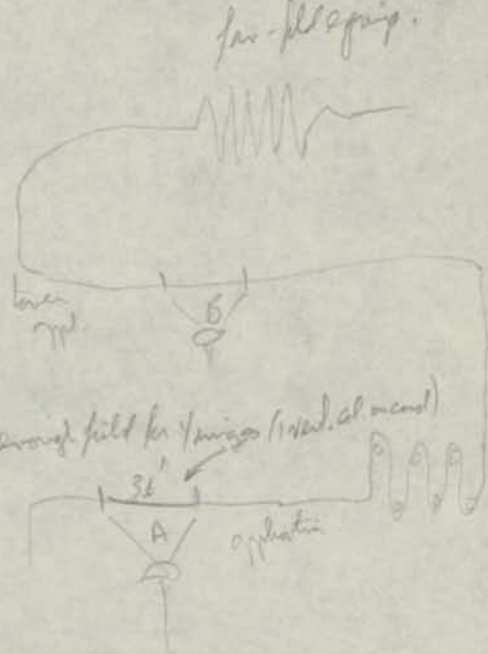
suggest we use 125 % O.H. (incl. burden) —
13%

m

$ 4000 for each module of 8 magazines.

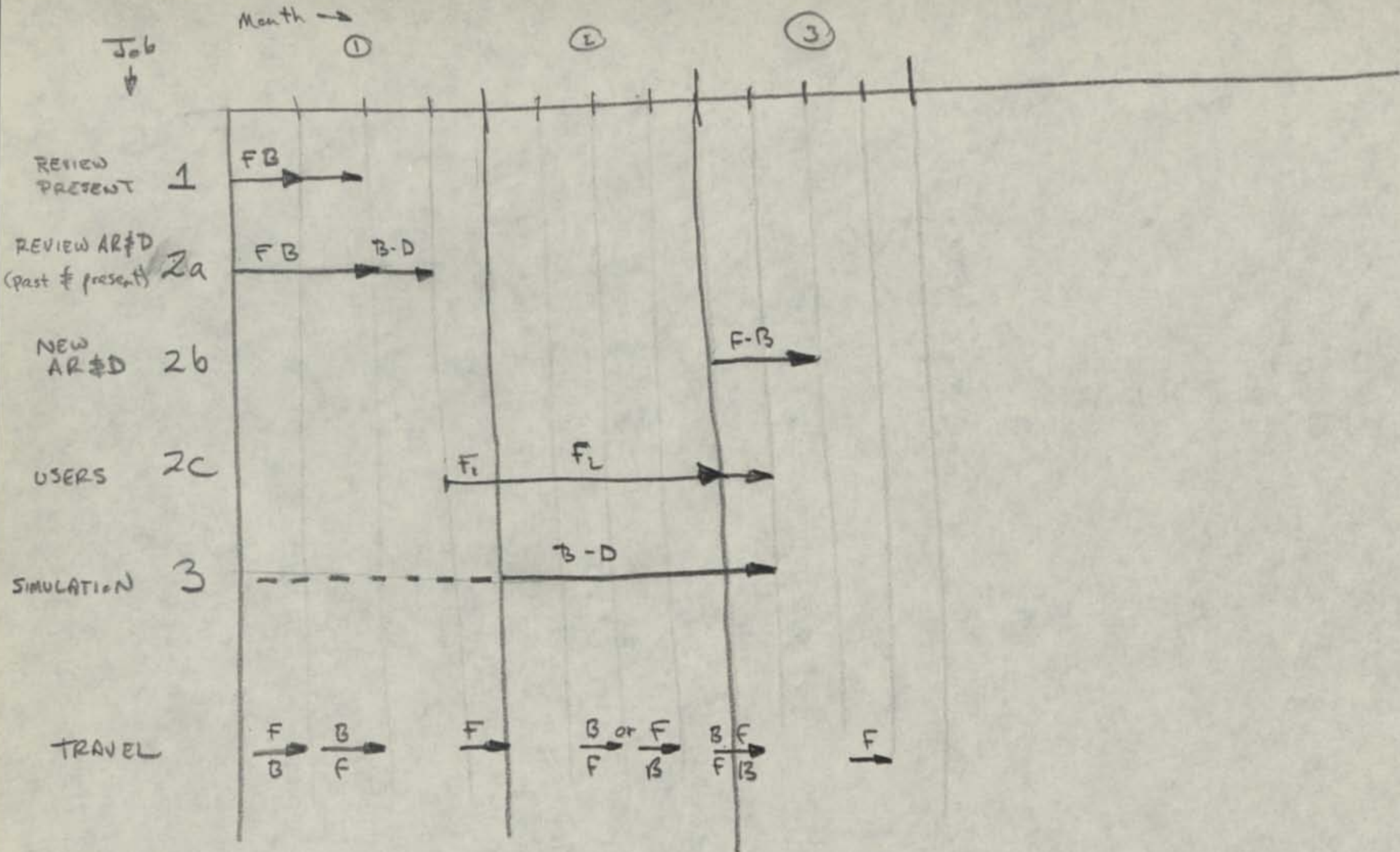
8½ x 11 reading
~~10 x 12~~
print 8 pages every 3 sec.





paper cost .75 cents / sq. ft.
for both side

toner?
Total unit cost / page.



B - Bourne

F - Flint

D - Don Ford

RCA Simulations

Cost Parameters

1. Wages: Use civil service wages & job descriptions for standards *where are they?*
2. Equipment amortization: Use ~~5 year period~~ 50 months   Rate ? 6%
3. Burden (25%) & O.H. (100%) ~~[illegible]~~ ?
4. RACE Card cost?
5. RACE Equipment Cost (for each module)?

   magazine
   8 magazine module
   printing station
   camera
   card punch
   operator console (inquiry station)
   output folding & stapling unit

6. RACE complete maintenance costs (for each module) parts & labor ?

   8-magazine module
   printing station
   camera
   card punch
   operator console (inquiry station)

7. RACE film processing cost?
8. printing costs (paper & toner) per pages:

INTERNAL CORRESPONDENCE

DATE: May 8, 1962

TO: Distribution

FROM: J. J. Murphy

SUBJECT: IS&R AD HOC COMMITTEE REPORT

This is the first of two reports which will issued by the Ad Hoc IS&R Committee. The purpose of this committee is to evaluate the requirements of the IS&R field as they reflect on present RCA DSD capabilities and to determine what direction our marketing and engineering efforts should take to optimize our position to capture business over the interval of the immediate future through the next five years.

It is the opinion of the committee that if we are to penetrate this field then all of our limited resources should be directed toward a definite end. This means that we should, for instance, plan our AR&D program for the next five years and modify it only if significant developments have eliminated the necessity to pursue further development in a given area.

In order to accomplish such planning it is necessary to know the following:

1. What are present customer requirements?

2. What are future customer requirements?

3. What does RCA have now that will meet "1." above?

4. What should RCA have to meet "2." above?

This report deals primarily with "1." and "3.". Although the treatment is somewhat incomplete, this report is issued primarily to elicit comments from those who read it. Such comments should be directed to J. J. Murphy, Ext. 3415, RCA, Van Nuys.

This report was prepared by J. J. Murphy, I. Sperling, and G. Richards.

J. J. Murphy

J. J. Murphy
Bldg. 307/2
Van Nuys

JJM:dp

FORM DEL 1241

Distribution:

Committee Members

G. Arnovick
D. Clemison (DSC)
C. Fanwick
R. Geiger
S. Kaplan (DSC)
E. Moulton
J. Murphy

Others:

D. Beaumariage
A. Bos
G. Breitwieser
G. Decker
H. Lesser
J. Minker
G. Murray
H. Nordyke
P. Scholz
H. Wege

TABLE OF CONTENTS

I. SUMMARY

A. Conclusions

1. Video File is most applicable to Print Factory applications.
2. RACE represents a mass storage capability which is orders of magnitude better than anything which will be available in the period 1964-1966.
3. RACE should eventually be a part of the 4100 series product line as a mass memory.
4. A Microfilm RACE should be developed before (and in preference to) a Video system.
5. For the period covering the next 25 years the majority of information storage and retrieval operations will be conducted using general purpose computers and standard peripheral lines with some additional special purpose peripheral equipment to enhance the capability of the general purpose machines. The availability of Superdensity Tape Stations by mid 1963 will reinforce this tendency.
6. The Intelligence Data Processing market is undoubtedly very large and has a great variety of problems. Hardware sales in this area will probably occur only when a product has been developed and offered on the market. Apparently, there are development funds available but DSD (Van Nuys) does not understand the problem requirements and does not know and understand the customer or his plans. It is not known how well this problem is understood by DSC nor what specific plans have been adopted by DSC due to complications introduced by security restrictions.

I. SUMMARY (continued)

B. Recommendations

1. Approach Signal Corps for development support for a digital RACE. There is a definite requirement for such a mass memory. (SIGC Prob. #14 - Problems Guide Volume VI, 1961)

2. Devote the balance of this year's AR&D to development of a Microfilm RACE.

3. Postpone attempts to market a document retrieval system to such agencies as SSA and ARC until "2." above has been accomplished.

4. Pursue applications of the "Print Factory" incorporating Video File and Microfilm RACE when each is most applicable. Market Video File for the small to medium file size applications with moderate dynamic loading where updating or communication is a requirement.

5. Increase efforts to coordinate DSC and Van Nuys to attempt to capture additional business. It is suggested that "working level" coordinators be set up at each location to enhance information exchange and contribute to cooperative efforts.

II. AREA OF INTEREST

Almost every aspect of data processing involves storing and retrieval of information, requiring both serial (tapes) or random (cores) access to data stored. Furthermore, there is a great wealth of technical art dealing with address coding, file structuring, storage allocation, and indexing and programming which permits handling information once it has been reduced to digital form. There is no operation with such data that cannot be

II. AREA OF INTEREST (continued)

implemented with existing data processing hardware, regardless of whether we are speaking of natural languages or artificial, and involving syntactical or semantic operations.

We choose in this analysis to eliminate consideration of general purpose computer systems which may be programmed in a variety of ways using standard peripheral equipment to accomplish storage and retrieval operations on data which has been reduced to digital form.

We will consider special purpose devices to be used in conjunction with computers to accomplish these operations more efficiently; however, the greatest attention will be directed toward operations involving data which is largely, and must remain, in analog form such as documents, maps, radar data and other electromagnetic signals, and imagery.

The purpose of this exclusion is simply that the field of "document" storage and retrieval holds the greatest promise for development and sales of hardware which would not already be covered by RCA existing EDP product lines. It is this area where we need to develop new engineering insight and adjust our marketing techniques to suit the demands of the field.

III. MARKET CHARACTERISTICS

A. Where is the market?

At various times enthusiastic market predictions made by "experts" in the field have assessed the annual value of the "IS&R" market as approaching the value of the national debt. The Five Year Data Systems Program for DEP (1962 - 1966) delineates a much more modest market value and a correspondingly modest capture ratio for RCA. It is felt that these latter predictions are much more accurate than some figures that have been disclosed by other organizations. It is true that if all of the libraries, technical institutes, government agencies and commercial concerns were to automate their present information storage and retrieval systems as they are now organized and operated, the capital investment and operating costs would approach astronomical proportions; however, the present trend seems to indicate that this will never come about for the following reasons:

1. There will be a concentration of information into "Information Centers" because individual users simply cannot afford the capital investment required to convert to and operate automated IS&R systems, and this concentration will mean small quantity "high cost" installations.

2. Certain IS&R activities will never in the foreseeable future (10 - 20 years) convert from manual operations because concentration will limit their independence, and the cost of efficient automated devices will remain out of their reach. It is most likely that these activities will incorporate semi-automatic or partial systems such as Jonkers Termatrix or RAP 600, or manual card search.

The attempts toward centralization have already begun their inexorable trend. Examples of this are seen in the case of:

1. ASTIA which will probably combine and control Cambridge Air Force Research Center Library, IDEP, ONR Library, ARPA Library.

III. MARKET CHARACTERISTICS

A. (continued)

2. Army Record Center providing combined record facilities for all services and the Veterans' Administration for both active and inactive service records.

3. A centralized record facility built around the Social Security Administration for all government civilian activities.

4. Defense Supply Agency will combine the parts cataloging and instruction book keeping for all Services.

5. Concentration of medical records with the Department of Health, Education and Welfare (NLM).

6. Centralized Weather Data Center (Weather Bureau).

7. Centralized Finance and Accounting for Services.

(A centralized Intelligence Record Center appears unlikely.)

Such a trend dictates that we recognize who these agencies will be and concentrate our marketing efforts toward those whose requirements we can satisfy. At the present time after limited examination of the market, the agencies that we can profitably pursue are probably limited to a few due to the status of our product line.

The most typical candidates at the present time appear to be the following:

1. ASTIA
2. Patent Office
3. Medlars (NLM)
4. Air Force Satellite Weather Data (433L)

III. MARKET CHARACTERISTICS

A. (continued)

These agencies either have money and a well-established requirement (433L), or a well-established requirement and an excellent chance of obtaining money (ASTIA, Patent Office, Medlars).

The Air Force Finance Center to whom we hope to sell our first system is a bootstrap operation. They do not have a clearly defined requirement nor firm commitment of funds, but are well supplied with enthusiasm and desire to sell the Video File system to their cognizant superiors.

In the near future other customers may come in reach as our product line develops. Typical examples of such might be the following:

1. Social Security Administration
2. Army Record Center
3. FBI
4. CIA
5. DSA
6. Weather Bureau
7. NAFC, AAFC
8. State Motor Vehicle Agencies
9. Commercial concerns desiring to handle drawings, handbooks and reports.

There is undoubtedly an extremely large market which could be tapped immediately for development funds and hardware procurement if we understood the requirements. This market is the Intelligence Data Processing field inherent in security classification and record keeping, signature analysis, Reconnaisance Intelligence Census, Foreign Technology, etc.

III. MARKET CHARACTERISTICS

A. (continued)

Some examples of these problems appear in Appendix A; however, it is not known of which agency or department these examples are typical nor are the problems understood well enough to offer solutions.

B. What do they want?

Most applications discussed above fall into two general categories:

1. Print Factories

   a. Monotonically increasing files.

   b. Little or no updating. *

   c. High volume, multiple hard copy output.

   d. No quantity soft copy requirements.

   e. Batched requests to file.

2. Record Files

   a. Monotonically increasing or static file size.

   b. Updating requirements.

   c. Quantity soft copy requirements.

   d. Random unbatched requests or batched serial requests.

   e. Low quantity hard copy.

Some applications such as the Patent Office have combined requirements for record files (Patent search) and print factories (Patent Copy Service).

It is a safe conclusion to draw that no one system will meet requirements of both categories. For the sake of efficiency the Print Factories will require a serial processing of continuous medium particularly where collated output is desired. If a random access system is

* Updating defined as "erase and replace".

III. MARKET CHARACTERISTICS

B. (continued)

employed then the problem of unit record accounting is encountered . Appendix B contains some examples of Print Factories and Record Files. In general, the Print Factory problems are amenable to Video File applications but the Record Files are not. The Print Factories are interested in systems that will:

1. Represent the least effort to convert data base.
2. Give maximum flexibility of format and font.
3. Eliminate back ordering and advance ordering, and provide on demand service.
4. Reduce their file volume and floor space.
5. Provide modularity for expansion.
6. Provide high quality printed product.
7. Provide non-volatile storage.
8. Automatically collate.
9. Have capability for multiple copies.

The Record File applications requirements are typical of the traditional storage and retrieval problems.

1. Archival storage. (Non-volatile)
2. Easy and rapid input.
3. Random access.
4. Flexibility of file structure and ability to rearrange at will.
5. Minimum search time (usually minimum access time).
6. Facility for logical operations with data.
7. Soft and hard copy.

III. MARKET CHARACTERISTICS

B. (continued)

8. Some requirements for communication.

9. Some requirements for relatively widely located segments of file.

10. File integrity.

11. Unit record file capability.

The Print Factory applications can usually afford a sophisticated automatic system because of the sometimes prodigious cost of printing by customary techniques and the poor service obtainable.

The Record File problem is often a different proposition because of the relatively low operating costs of maintaining a manual (or semi-automatic) file which does not permit amortizing capital equipment. Many agencies are able to operate surprisingly efficient files with low cost manual labor and still provide good service. Problems arise in these cases with labor turnover and questionable file integrity. There is a breakpoint when sheer volume of activity causes these systems to founder primarily due to the long file updating time. Some files that have been studied have up to one year backlog of updating and many have a two to three month backlog.

C. What are they buying?

The answer to this question is briefly "very little". There seems to be a conservative "go easy" attitude on the part of agencies in a position to buy storage and retrieval systems. By far the greater emphasis in the field has been toward utilizing general purpose computers in conjunction with mass memories to accomplish search and retrieval of information that can be reduced to digital form. The appearance in the near future (five years) of workable page readers should increase the

C. (continued)

utilization of computers for IS&R operations on documents of certain types. This will increase the demand for high density fast access mass memory systems. This trend will take place even in the face of the inefficiency of general purpose machines for this type of problem because of the availability of standardized product lines including software and peripheral gear, and the flexibility derived from them. This trend is reinforced by the availability of a wide variety of competitive lines and rental arrangements which avoid the necessity for large capital investments and the possibility of obsolescence of same.

In the document storage and retrieval field the trend is overwhelmingly toward microfilm mainly because there isn't anything else that is practical at this time. The weakness of existing microfilm systems is usually at the output end due to the slow and frequently messy film processing. Microfilm has the disadvantage that it is difficult to search, update and print out but it provides very cheap soft copy random access (aperture cards) and high storage density at lowest cost.

The following is a partial list of agencies with document systems in existence:

CIA	Walnut (microfilm) Aperture Cards Strip Film Paper
Patent Office	Aperture Cards Paper
ASTIA	Strip Film
ARC	Paper
AFC	Paper
SSA	Paper Microfilm Reels Aperture Cards

III. MARKET CHARACTERISTICS

C. (continued)

FBI	Aperture Cards Strip film
Weather Bureau	Paper Aperture Cards
Navy Printing Office	Paper
Medlars	Paper
BUSANDA (DSA)	Paper
Library of Congress	Paper Aperture Cards
OACSI	Film (slides)
Army Engineers	Film (slides)
Navy Security	Paper Aperture Cards Strip film

Some agencies, notably ASTIA and Patent Office, because of either management desire or necessity, are considering the further automation of their systems and are leaning toward the improvement of their existing microfilm systems by converting their output techniques and employing more efficient printing techniques.

IV. DISCUSSION OF RCA COMPETITIVE STATUS

The only practical way to compete in an open market is to have wares to sell. RCA DSD is in the very awkward status of being without wares to sell. The best that can be said in this regard is that DSD has great promise of having a product line in two to three years.

There are many established document storage and retrieval systems in existence all of which employ microfilm or photographic processing in some form. Most of these systems and devices are described in Appendix C. There are far fewer special purpose IS&R systems that are designed to work with digital data. The GE 250 Search Comparator, the Heatwohl H-44, CRAM and Finda-fact are examples of this type of system. GE has produced about seven for use internally to GE. CRAM is sold exclusively with NCR 315 computer and is employed for business data processing. The quantity sales of the H-44 are unknown but are estimated to be in the order of five to ten at most. Finda-fact has had one sale and no others as far as we have been able to determine.

Of the document systems the Walnut installation is the only one in existence and IBM is not offering it as a standard product line. The following table lists the probable quantity of equipments sold or in existence to date:

ARTOC	1
VERAC	1
FLIP	2
RAPIDAC	2 - 4
File Search	2
GE -250	7
H-44	5 - 10
CRIS	1
Walnut	1
Termatrix	50 - 60 (rough estimate)
RD-900	1
Magnacard	3
Magnavue	1
Filmsort	500 (rough estimate)
Rapid Selector	1

CRAM	200 (rough estimate)
Finda-fact	1
Houston-Fearless	1

Two observations may be drawn from this data:

1. Having a working model of a system with many desirable features does not insure sales success.

2. To date the systems enjoying greatest use are those which are:
    a. Semi-automatic or manual,
    b. Require minimum capital investment,
    c. Offer great flexibility of file organization.

The one exception to this is CRAM which is enjoying singular success due to its relatively low cost and high storage capability and its tie-in as part of a computer system with the NCR 315.

The lack of success of many of these systems may be due in part not to the disadvantages of the devices but to the deficiencies of the concerns manufacturing them. The lack of a comprehensive product line, a well-established service arrangement and insufficient marketing organization are contributing to the lethargy of sales for many of these devices. It is here that RCA can capitalize on its organization to its advantage once a product line has been established.

At the present time the area which appears most fruitful for RCA is the "Print Factory". The reason we can be competitive in this area is simply that there is nothing on the market at present nor for the next two years that can compete with Electrofax printing from the standpoint of speed, cost, flexibility, or quality. Whether the printer is driven from a film medium or video medium, performance is still superior to anything else available. The capability for improvement provides considerable margin of safety to maintain this lead in performance. Secondly, we can devise a storage system which employs serial search techniques using video tapes which will satisfactorily meet speed requirements. The cost of the video storage medium incurs a 10 to 1 disadvantage over microfilm and also means less storage per reel, requiring a greater number of reel changes. It has advantages, however, in terms of higher search speeds, editing and rearrangement, auto-collation and multiple copies by means of buffer loop, flexibility of format and positive or reversal processing without change in toner or paper by inverting video (simultaneous positive or negative processing employing successive applications of positive or reversal toner is possible with film). The competition in this area may come from A. B. Dick with their Videograph with the system employing the ~~Printa-Pix~~ Videograph tube.

V. WHAT RCA SHOULD DO

A. Recommendations for Marketing

At the present time it appears that we will be capable of producing a working Video File system (prototype) within one year and delivering quantity items within two years.

The prospect for a random access system is dependent upon the development of the RACE equipment. If we assume that a digital RACE is developed first, a prototype would be available in 24 months from January 1, 1962. Assuming a parallel development is undertaken then either a microfilm or video RACE would be available at the same time if the effort were started

V. WHAT RCA SHOULD DO

A. (continued)

on or before July 1st. If a parallel effort cannot be mounted then an additional 8 to 10 months would be required to produce a prototype of either a video or microfilm system.

As a consequence of this schedule, it is obvious that we should not pursue business that requires a random access mass data storage, i.e., Social Security Files. Rather, we should concentrate on seeking support for the development of the RACE. In this regard it is more likely that we will receive such support if we emphasize the document (microfilm, video) storage capability with concurrent digital storage. In other words, the customer most likely to pay for development is one who has a document storage problem. It should be recognized that the user expects the computer industry to develop mass digital stores with their own funds; however, this emotional feeling has not as yet been extended to image storage systems primarily because no concern has as yet achieved well-recognized success and the field is still considered to be a specialty market.

A timetable is reproduced below which indicates the probable availability of various mass storage systems which shows not only the competition but indicates the breadth of market which it will stimulate.

Data for this table has been obtained primarily from the 1960 Report of the Senate Committee on Government Operations Document #113 and from other sources such as Datamation and upon predictions based on knowledge of current research programs at IBM and RCA.

V. WHAT RCA SHOULD DO

A. (continued)

One conclusion which can be drawn from the table is that all possible effort should be made to get support for RACE and attempt to develop all three capabilities simultaneously and as early as possible.

It should be noted that the Video File will be more or less concurrent with RACE and if much further delay is encountered in getting started in Video File, it may not be available until after a RACE with document storage capability has been designed.

One way to insure the development of Video File is to attempt to market it with customers who are already working with video imagery data. Both NASA and the Weather Bureau have expressed interest in video storage of Weather Satellite data; however, neither agency has a clear cut plan as yet for the ground handling and storage. The Air Force 433L Program is at present most active in this area and has at least two active R & D programs to develop a "real time" processing system for ground handling of video satellite data. It is believed that they would be amenable to another approach and engineering is preparing material for a presentation of such an approach.

B. Research and Development

Comment on this area is withheld until results of the SRI study are available.

V. WHAT RCA SHOULD DO (continued)

TABLE I

(1) IBM Photoscopic Discs

(2) AVCO VERAC

(3) RCA RACE (digital)

(4) RCA Micro

(5) RCA Video

(6) RCA Video File

(7) Magnavox Magnacard (Kerr)

(8) Magnavox Magnavue (Kerr)

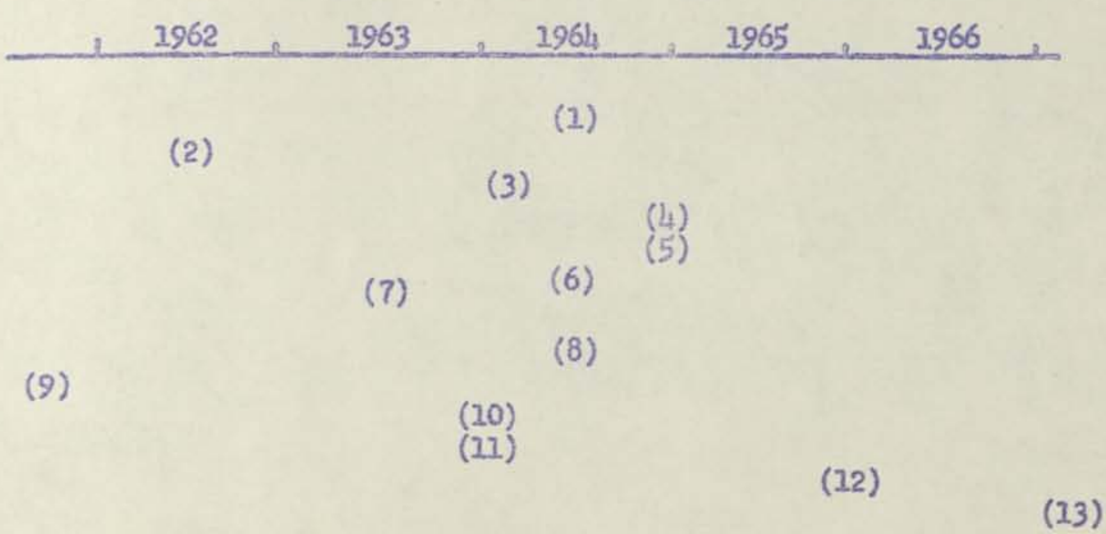
(9) NCR CRAM

(10) Bryant Disc Improved

(11) Telex Disc Improved

(12) Thin Film Memories

(13) Cryogenics



VI. DISCUSSION OF STORAGE & RETRIEVAL SYSTEMS

A. ARCOM Study

I. OBJECT:

This application study was primarily intended to study ARCOM (Automatic Retrieval COMparator) in various system environments in an attempt to evaluate its performance in information storage and retrieval systems. In addition, the application study was intended to develop sets of system specifications and rough costs which would describe to a prospective customer equipment which would meet some of their literature searching requirements.

II. CONCLUSIONS

An analysis of the solutions to the hypothetical retrieval problems studied indicates:

A. ARCOM is more suited to high retrieval rate problems than to low retrieval rate problems.

B. One set of equipments cannot be specified, on the basis of economics, as the universal ARCOM system.

C. ARCOM is not a speed limiting device with respect to peripheral attachments. It reaches a limitation on the basis of the number of search terms which must be compared in a limited time against a given file.

D. In order to ascertain the usefulness of the devised ARCOM systems, similar systems must be devised using other types of comparators. The various solutions to the same problem can then be compared against one another on the basis of 1. economics, 2. flexibility, and 3. expandability.

III. THEORY OF OPERATION OF ARCOM

The following chapter is extracted from an Engineering Memorandum, "ISAR - TECHNICAL CONCEPT ANALYSIS" by Thomas I. Ress, the inventor of ARCOM. In this memorandum Mr. Ress uses the term "Automatic Bibliographer" instead of the acronym ARCOM.

The Automatic Retrieval COMparator (ARCOM) consists of a cathode ray tube and a planar optical system as shown in Figure 1. In operation, a user inserts a card bearing the interrogating index terms, in the form of punched hole patterns, to be compared with the index blocks connected to the document identification stored on a magnetic tape file. The cathode ray tube presents the file's index terms in binary form through a sequence of vertical beam displacement, as shown in Figure 2, which is then optically transmitted to the card bearing the interrogating terms. For every bit position a luminous dot appears either in the '0' slot or the '1' slot. The optical system is used to transform the binary light signals on the face of the cathode ray tube to the columnar bit positions on the card.

A second optical system, horizontally oriented and located behind the card, is used to collect all the light passing through a row on the card and conducts this light to a photocell. If the user punches the interrogating card in such a manner that every row contains an index term, and one hole, either the '0' or the '1', is always punched for every bit position, then whenever a bit of the interrogating term and the index term match, no light is transmitted through the card to the photocell. Conversely, whenever a bit of the interrogating term and the index term do not match, light is transmitted through the card to the photocell. Therefore, a match or coincidence is detected by a null condition for the photocell corresponding to a row on the card.

Punching a hole for every bit position allows a "don't care" situation in which neither hole is punched. If ARCOM were to be used as a "natural language" searching comparator then the "don't care" situation can be used to take care of variable suffix situations where a root of a word or words appearing within at least one individual index item satisfied the request criteria.

IV. POTENTIAL APPLICATIONS

The most appropriate applications for ARCOM appear to be

A. Scanning of full text tapes to appropriately select index terms for future retrieval purposes.

B. Scanning full text tapes to retrieve documents containing desired descriptors with or without consideration of their relative order or frequency of occurrence.

C. Scanning (for retrieval) tapes containing appropriately selected index terms and either the document number or a complete bibliographic citation for each document indexed.

Unless we assume a completely fixed vocabulary the first application to automatic indexing of documents from full text tapes implies the need to include, as they are encountered, significant terms which are new to the system. Such inclusion must be governed by human intervention in such cases or by elaborate evaluation and "self-learning" systems. The inherent complexity of such a system would involve studies considered to be beyond the scope of the present evaluation.

The scanning of full text tapes for document retrieval is inherently impractical due to excessive storage requirements and search time. Also, since the document must be searched before a decision can be made to copy the document, buffering or back up problems would add additional complications, unless the user is to be satisfied with only a reference or citation.

It should be noted that the third application implies that the logical consideration or order or frequency of occurrence of the indexing terms has been performed, hence the logic connected with the searching selector would be simpler in the third application than in the second application.

Due to the above reasoning, the major emphasis of this study was directed toward the third application.

V. GENERAL METHOD OF EMPLOYING ARCOM

In order to retrieve information from a collection of stored material, a selection and correlation process must be performed which will inspect all the stored index information, decide whether or not it is related to the information required, and, if it is, provide a means for locating the original documents. The selection and correlation process, then, consists of scanning the encoded index blocks, matching the various index entries against a list of criteria (which characterizes the information requirement for which the search is being conducted), and combining the index entries thus matched to decide whether or not the proper logical combination of criteria is satisfied. Last of all, if the document represented by such inspection of an index block results in positive identification of a document, then it must be identified appropriately, by an accession or serial number, in order that the original document be found in the storage file.

VI. HYPOTHETICAL RETRIEVAL PROBLEMS

Several operating parameters were chosen to indicate the range of possible application of this device.

Assume that:

A. Customers already have libraries of stored documents with the following characteristics:

		Class I	Class II
Total Number of Documents in Library		10^6	10^5
Average Number of Index Terms/Document		40	40
Maximum Number of Characters/Index Term		20	20
Average Number of Documents Retrieved/Request		20	20
Request/Day	High Rate	10^4	10^4
	Low Rate	100	100

B. All work must be accomplished during an eight hour (28,800 seconds) working day.

VII. FILE STORAGE REQUIREMENTS

General:

Two requirements which are necessary, though not sufficient, to determine a satisfactory file storage mechanism are:

A. Required Capacity

B. Required Processing Time

Required Capacity:

Without a knowledge of the information format, the minimum possible storage requirement is -

Min. Storage = Index Term Storage + Document Identification Storage

The total storage requirement is -

Total Storage = Minimum Storage + Marker Codes

Therefore,

Total Storage $\geq$ Minimum Storage

A. Natural Language (alpha-numeric six bit code)

Assume:

1. Index terms are expressed in "natural language".
2. Average length of index terms is 10 characters.
3. Documents are identified by an accession number in straight binary coding.

Index Term Storage:

Class I

10^6 doc. x 40 terms/doc. x 10 char./term x 6 bits/char.

= 2,400 x 10^6 bits.

Class II

10^5 doc. x 40 terms/doc. x 10 char/term x 6 bits/char.

= 240 x 10^6 bits.

Document Identification Storage:

Class I

20 bits/doc. x 10^6 doc. = 20×10^6 bits

Class II

17 bits/doc. x 10^5 doc. = 1.7×10^6 bits

Therefore, for Class I

Total Storage Required $\geqq 2.42 \times 10^9$ bits,

and for Class II

Total Storage Required $\geqq 241.7 \times 10^6$ bits.

B. Artificial Numeric Language

Assume:

1. Index terms are expressed in a straight binary numeric code.
2. Total number of different index terms is approximately 32,500 and each term is expressed as a 15 bit binary number.
3. Documents are again identified by an accession number in straight binary coding.

Index Term Storage:

Class I

10^6 doc. x 40 terms/doc. x 15 bits/term = 600×10^6 bits.

Class II

10^5 doc. x 40 terms/doc. x 15 bits/term = 60×10^6 bits.

Document Identification Storage:

Class I

20 bits/doc. x 10^6 doc. = 20×10^6 bits.

Class II

17 bits/doc. x 10^5 doc. = 1.7×10^6 bits.

Therefore, for Class I

Total Storage Required $\geq 620 \times 10^6$ bits,

and for Class II

Total Storage Required $\geq 61.7 \times 10^6$ bits.

Although the structural differences between "natural language" and formalized synthetic languages or codes seem quite obvious at first glance, there are surprising similarities in the procedures necessary in applying either form of language to automatic data processing equipment.

The first factor to consider in this "paradox" is that neither type of language is indeed "natural" to the machine complex. Characters alone do not make a language but only meaningful combinations of characters, known as words, permit all the desired meanings and variations in an extensive vocabulary. Regardless of the type of language used, the machine makes sense only of combinations of its own "native language" representation of the characters of whatever type of language may be imposed upon its information store.

The basic difference, then, that exists between "natural" language and formal codes is one of organization of the basic characters of that language which is used, and generally favors the formal language for machine use in reduction of redundancy, hence of space requirements, in its words. This makes for greater efficiency in storage space and decreases the time for digital sequential input or output of such words.

At some compromise to the redundancy factor, a formal code may incorporate hierarchical relationships between various words in its vocabulary. This relieves to a considerable extent the problems of synonyms and related words that accompany the use of natural language. Because of the large number of synonyms and non-explicit relationships in natural language we must consult some form of table of meanings or thesaurus for use words just as we must do in translation between natural language and formal codes.

In summary, so far as the machine is concerned either language requires table look up, and only the redundancy or space factor is really significant. Therefore, at this time it was decided to consider only the artificial numeric language in the machine complex.

Required Processing Time:

The required processing time can be determined as follows:

$$\text{Processing Time} = R \times \frac{28.8 \times 10^3 \text{ sec./day}}{\text{request/day}}$$

where: R = the number of requests simultaneously processed.

At this time an estimate must be made regarding the number of requests that can be simultaneously processed by one ARCOM unit. Extracting again from "ISAR - Technical Concept Analysis" by Thomas I. Ress, " . . . It is anticipated that a single interrogation card could carry up to 100 descriptor words . . ." If we assume a maximum of 10 index terms per request, then 10 requests can be simultaneously processed by one ARCOM unit and the processing time can now be defined as a function of the number of ARCOM units employed.

R = 10 x (No. of ARCOMS employed)

R = 10 x A

Therefore, for the high request rate

$$\text{Processing Time} = \frac{28.8 \times 10^3}{10^4} \times 10 \times A$$

$$= 28.8 \times A \text{ (sec.)}$$

and for the low request rate

$$\text{Processing Time} = \frac{28.8 \times 10^3}{10^2} \times 10 \times A$$

$$= 2,880 \times A \text{ (sec.)}$$

VIII. AVAILABLE FILE STORAGE MECHANISMS

The following table shows the relative characteristics of RCA digital storage media investigated during the course of this study. The capacities indicated are the maximum usable information storage per unit tape reel. In the case of the 381 tape station the capacity listed is for six tape reels or one tape reel per read/write deck.

TABLE I. Digital Storage Media Characteristics

	TRADEX	HSTL	381	582	681
TAPE SPEED (in./sec.)	1,000	1,000	30	100	150
RECORDING DENSITY (bits/in.)	200*	250	250	667	800
TAPE LENGTH (feet)	20,000	2,000 in.	1,200	2,400	2,400
CAPACITY (10^6 bits)	720	48	151**	288	345
READ TIME (sec.)	240	2	2,880	288	192
READING RATE (10^6 bits/sec.)	3	24	0.053	1.0	1.8
START TIME (millisec.)	1 Min.		20	3.5	6.0
STOP TIME (millisec.)	1 Min.				
USABLE TRACKS	15*	96	7	15	15
TAPE WIDTH (in.)		2	½	3/4	3/4
TAPE REELS (in.)	36		8	10½	10½
SALE PRICE ($\$10^3$)	100	50	74.9	43.26	56.3
COST/BIT/SEC.	0.1	0.002	1.4	0.043	0.031

* The Tradex tape station presently has only two digital tracks, used for control data, with a recording density of 100 bits/in. For the study it was assumed that with slight modification the number of digital tracks could be raised to 15 with a recording density of 200 bits/in.

** The 381 tape station combines six read/write decks in a single unit.

RACE

RACE, the acronym for random access card equipment, presently under investigation at RCA Van Nuys has the following preliminary specifications:

Cards/Magazine	256
Channels/Card	96
Recording Density	350 bits/in.
Bits/Card	
a. Single Side	470×10^3
b. Both Sides	940×10^3
Access Time	
1st Magazine	0.18 sec.
32nd Magazine	0.45 sec.
Card Reading Speed	400 in./sec.
Read Time/Card	
a. Single Side Recording	0.1 sec.
b. Recording on Both Sides	0.2 sec.
Capacity (32 Magazines)	
a. Single Side Recording	3.85×10^9 bits
b. Recording on Both Sides	7.7×10^9 bits
Estimated Cost	$\$200 \times 10^3$
Reading Rate	4.7×10^6 bits/sec.
Cost/Bit/Sec.	
a. Single Side Recording	0.04
b. Recording on Both Sides	0.04

361 Data Record File

Records/File	128
Usable Sides	2
Bands/Side	2

Cells/Band	10
Bits/Cell	5.4×10^3
Capacity	27.6×10^6 bits
Reading Rate	15×10^3 bits/sec.
Read Time	1,850 sec.
Cost	21.15 K *
Cost/Bit/Sec.	1.4

* This includes the cost of the Data Record File Control 317-1.

IX. CHOICE OF FILE STORAGE MECHANISM

General:

The file storage requirements determined in Section A is duplicated here for convenience and is shown in Table II.

Table II. File Storage Requirements

	Capacity (10^6 bits)	Processing Time (sec./ARCOM)
Class I (a)	620	28.8
(b)	620	2,880
Class II (a)	61.7	28.8
(b)	61.7	2,880

The criterion used in the selection of the file storage mechanism is one of economics. For each file storage mechanism tabulated in Section VIII the number of ARCOMS required to meet or exceed the desired processing times will be calculated and tabulated. Also, the number of tapes, records, magazines and/or mechanisms required to meet or exceed the desired capacity will be calculated and tabulated. The total cost of the ARCOMS and the file storage mechanism will determine which of the file storage mechanisms shall be chosen.

The cost of one ARCOM unit, complete with the criteria evaluation logic, has been roughly estimated to be about $20,000.

Class I (a):

	Tapes, etc. Required	ARCOM Required	Total Cost ($\$10^3$)
TRADEX	1	11*	320
HSTL	13	1	670
385	25	419	8,455

	Tapes, etc. Required	ARCOM Required	Total Cost ($\$10^3$)
582	3	30	643
681	2	14	336
RACE	{5.2 / 2.6}	5	300
361	2,880	1,440	28,881

* Includes 60 seconds start time.

On the basis of economics TRADEX with 11 ARCOMS, the 681 with 14 ARCOMS and RACE with 5 ARCOMS are all within the same approximate cost range. Therefore, a finer evaluation must be made.

It has been shown that all three equipments satisfy or exceed the criteria presented. It remains yet to determine the operating limits for the three equipments.

	Capacity (10^6 bits)	Processing Time (sec.)
TRADEX	720	27.3
681	690	27.4**
RACE	722**	30.7

* 6 magazines of single side recording or 3 magazines of recording on both sides.

** The 681 equipments processing time should actually be higher than the 27.4 seconds indicated due to the assumption of zero time for tape changes.

The most logical choice of the three equipments is RACE because of the future expansion capability in storage capacity and the number of requests to be handled per day. If the storage requirement remains the same, the addition of one ARCOM allows an increase of 2,000 in the number of requests that can be handled per day. This is an increase of 20% for an increased cost of less than 6.7% of the original capital expenditure. If the storage requirements increase while the request rate remains constant, then an additional ARCOM is again required as well as additional RACE magazines.

In summary, the use of RACE and ARCOMS together allows for modular system expansion at very reasonable additional cost.

Class I (b):

	Tapes, etc. Required	ARCOM Required	Total Cost ($$10^3$)
TRADEX	1	1*	120
HSTL	13	1	670
381	25	5	175
582	3	1	63
681	2	1	76
RACE	{5.2 / 2.6}	1	220
361	2,880	15	321

* Includes 60 seconds start time

Again, on the basis of economics alone, the 582 with one ARCOM and the 681 with one ARCOM are within the same approximate cost range and satisfy or exceed the criteria presented. If we assume a tape change time of 30 seconds and that the format is such that the tapes can be searched in both directions (zero rewind time), then the operating limits for these equipments are:

	Capacity (10^6 bits)	Processing Time (sec.)
582	864	824
681	690	414

Since the allowable processing time is 2,880 seconds and both tape mechanisms are well below this figure, then the 582 tape station with three tapes will more than satisfy the requirements with a savings of $13,000 compared with the 681 tape station with two tapes. Thus, the equipment selected will allow reasonable expansion of file storage as well as the number of requests to be handled per day at no additional cost.

Class II (a):

	Tapes, etc. Required	ARCOM Required	Total Cost ($\$10^3$)
TRADEX	1	1*	120
HSTL	2	1	120
381	3	5	175
582	1	3	103
681	1	2	96
RACE	0.26	1	220
361	288	144	2,901

* Includes 60 seconds start time.

For this problem the 681 tape station, utilizing approximately 20% of its tapes capacity, and two ARCOMS will more than satisfy the requirements with a savings of $7,000 compared with 582 tape stations and three ARCOMS.

Class II (b):

	Tapes, etc. Required	ARCOM Required	Total Cost ($\$10^3$)
TRADEX	1	1*	120
HSTL	2	1	120
381	3	1	95
582	1	1	63
681	1	1	76
RACE	0.26	1	220
361	288	2	61

* Includes 60 seconds start time.

If the 361 (Data Record File) were chosen as the file storage mechanism for this problem then 160 record discs would have to be removed and replaced "on line" during each search operation. The 361 can handle only 128 records simultaneously. For an additional 3-1/3% capital investment the 582 tape station and one ARCOM provides an automatic tape search operation. Also, since only approximately 25% of a 582 tape is utilized the selection of this equipment complex will allow a great deal of expansion in file storage and the number of requests that could be handled per day at no additional cost.

BUFFER STORAGE AND PRINTOUT MECHANISM

The two remaining items required to complete the searching selector system are the printout mechanism and a buffer storage medium.

Printout Mechanism:

The printout mechanism must be capable of printing out all the retrieved document's accession numbers along with the searchers identification number during a single searching cycle. The entire printout may occur during the following search cycle when processing time requirements does not allow the printout during the same search cycle. This requires twice the buffer storage that would be required if the printout occurred during the same search cycle.

It has previously been assumed that the accession number is in straight binary coding; therefore, the printout will be assumed to be in octal code. The printer to be selected must be capable of the minimum requirements outlined below:

Required Printer Characteristics

	Class I		Class II	
	(a)	(b)	(a)	(b)
Accession Number (char./doc.)	7	7	6	6
Requests Processed/Cycle	50	10	20	10
Searchers I.D. (Char./Requests)	2	2	2	2
Spacer Char.	1	1	1	1
Average No. Doc/Request	20	20	20	20
Total Octal Char.*	10^4	2×10^3	3.6×10^3	1.8×10^3
Processing Time (sec.)	144	2,880	57.6	2,880
Printout Time (sec.)	144**	2,000	57.6**	2,500
Printout Rate (Char./sec.)	70	1	63	0.72

* Assumes worst case where all retrieved documents are different.

** Printout lags search by one search cycle.

A relatively cheap printer which meets the requirements of Class I (b) and Class II (b) is the Hewlett-Packard 562A digital recorder. The purchase price of this printer is approximately $1600.

The most economical approach to meet the requirements of Class I (a) and Class II (a) is to use two Hewlett-Packard 562A digital recorders simultaneously in parallel. This scheme gives an effective printout rate of 110 characters per second.

Buffer Storage:

The buffer storage must be capable of storing the retrieved document accession number along with the searchers identification number at a compatible rate with the reading of the file storage mechanism. Interleaved with this operation, for the case of Class I (a) and Class II (a), the buffer storage must be capable of feeding the printout mechanism.

For all four cases the most universal buffer storage mechanism is the coincident current core memory. A reasonable estimate of the cost of these memories is based on an average figure of 20 cents per core. This figure includes all the required control electronics.

Required Buffer Storage

	Capacity (10^3 bits)	Approx. Cost ($)
Class I (a)	60	12,000
(b)	6	1,200
Class II (a)	22	4,400
(b)	5	1,000

GLOSSARY

Document	-	Any publication considered as a single unit for indexing purposes.
Index Block	-	All the index entries pertaining to a single document.
Index Entry	-	A component unit of an index; may be either an index term or a group of index terms considered as an integral unit.
Index Term	-	A single word or phrase similar to a subject heading, used to designate a single concept.

VI. DISCUSSION OF STORAGE & RETRIEVAL SYSTEMS (continued)

B. Document Storage Systems

1. Video File

Since the Video File has been discussed and described in some detail previously, it is the purpose of this report to deal with certain isolated aspects of the system and other competitive techniques. Further analysis is planned on Video File beyond that attempted in this report and the results of this analysis will be reported at a later date.

The Video File is a segmental multiple record, serially accessed, unordered, sequentially indexed file. It is segmented because the file is distributed over several reels of tape according to any arbitrary scheme. It is multiple record file because any one segment (reel) of the file contains many unrelated records. It is serially accessed because each segment of the file must be searched serially record by record from some arbitrary initial position to the position desired. It is unordered because the sequence of records in any segment of the file need bear no relationship whatsoever to the adjacent records. It is sequentially indexed because the location number of each succeeding frame (record) is incremented by one, and every location in the segment is identified by a unique continuous sequence.

The above definitions delineate the characteristics of Video File and from these characteristics certain conclusions can be drawn concerning the applications of Video File.

VI. DISCUSSION OF STORAGE & RETRIEVAL SYSTEMS

B. Document Storage Systems

1. Video File (continued)

a. Large files mean many segments. The feasibility of a multi-segmented file is a question of economics and dynamic load. If the dynamic request load on the file demands that a certain portion of the file be on-line over a given interval, then this establishes the minimum hardware which must be provided to accommodate the segments of the file in use. Whether this is feasible depends upon the unit cost of the file mechanisms. In Video File this refers to tape transports. At approximately $70,000 per transport, high dynamic loads with large files rapidly become too expensive.

b. Serial access does not permit most efficient search for records which are not retrieved in sequence.

If the requests to the file cannot be so arranged that random searching of the file is minimized or eliminated, then the delay incurred by traversing unwanted records (even at accelerated speeds) becomes burdensome (See Fig. 3). It is this simple fact which reduces the number of applications for Video File because for certain applications it is not convenient for the requests to be batched and sequentially ordered.

If efficient use of the file dictates that requests be batched, the question arises "Why bother with soft copy?"

The primary motivation for soft copy is to permit inspection of the contents of any location in the file without the necessity

AVERAGE VIDEO FILE RETRIEVAL RATE
(ASSUMING BATCH-SORTED REQUESTS)

E. W. MOULTON
APR 17 1962

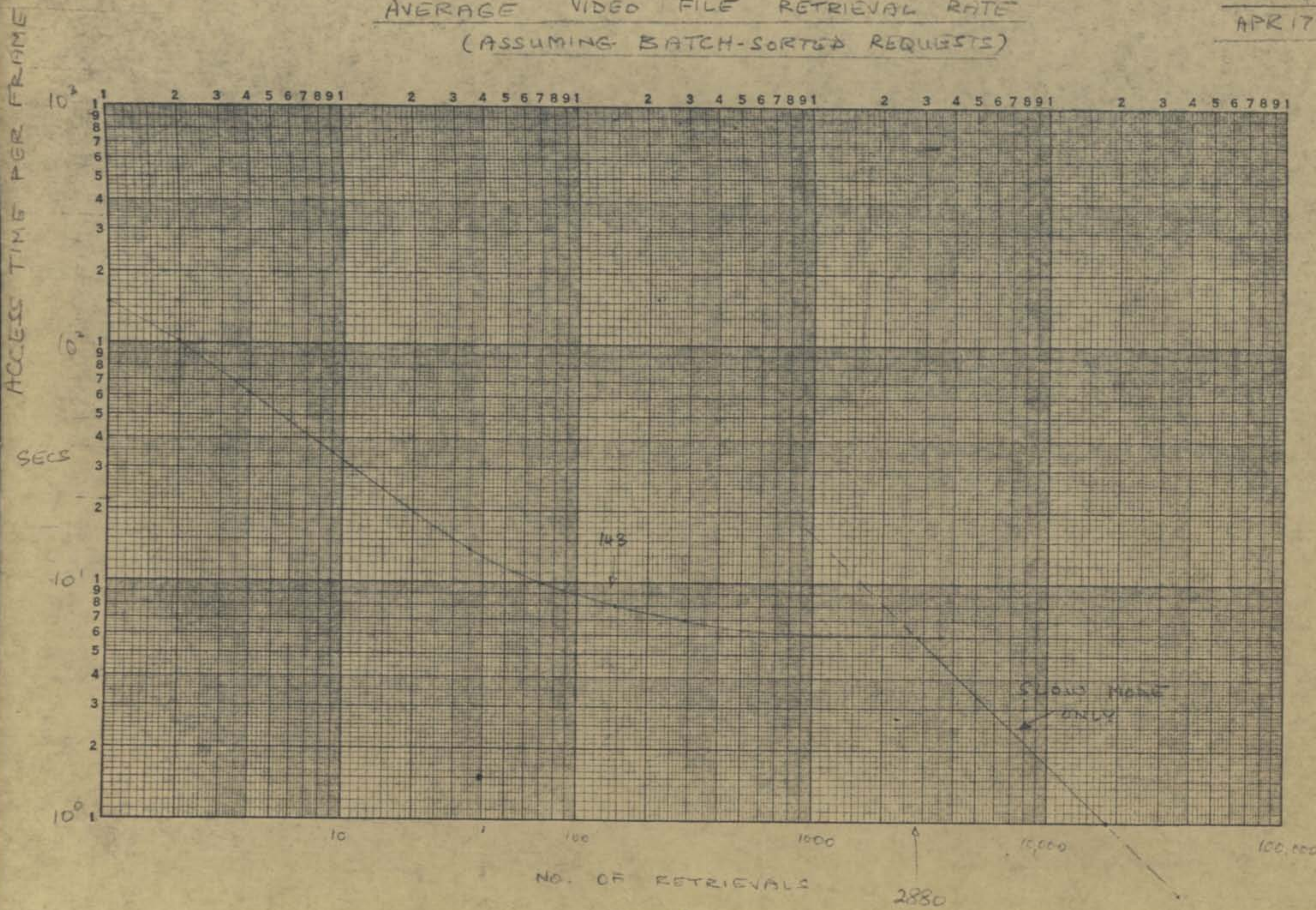


Figure 3.

VI. DISCUSSION OF STORAGE & RETRIEVAL SYSTEMS

B. Document Storage Systems

1. Video File

b. (continued)

of a tangible output (hard copy). An observer may very well wish to see the contents of any arbitrary location in a serial file but it is a very inefficient use of the files. One observer can tie up the file so that production work cannot be accomplished. A buffer system does not actually relieve the situation unless the observer is willing to relinquish the segment of the file on time so that it may be replaced with another segment desired by another observer or required for production purposes. If an observer or group of observers are willing to organize their requests to the file so that the utilization of the file is more efficient, then why not produce hard copy and look at that? The cost of the paper (1¢ per sheet) is inconsequential compared to the cost of inefficient use of the file. As a parallel, consider the likelihood of a large scale computer installation costing $600/hr. being tied up while GS-5 clerks observe a monitor displaying alpha-numeric records which could just as well be printed out at much faster rates and distributed to each clerk.

Nevertheless, some soft copy capability will undoubtedly be required for some applications. The important consideration here is that high quantities will <u>not</u> be required and the viewer can be a relatively high cost item and still represent only a small increment of the total system cost. The question arises, "What should a Soft Copy Viewer cost?" It would seem reasonable that 5% of system cost could be allocated to soft copy viewers.

VI. DISCUSSION OF STORAGE & RETRIEVAL SYSTEMS

B. Document Storage Systems

1. Video File

b. (continued)

In the case of AFAFC the hardware cost is roughly 1.5 million dollars at manufacturing cost level. This would permit $75,000 to be allocated to Soft Copy Viewers. Two (2) such viewers should satisfy requirements which would put the allocable cost for SCV at $37,500 at manufacturing cost level which seems well above the figure which has been determined by engineering cost estimates at 17 - 20K.

The Soft Copy Viewers for a random access system such as RACE are quite another consideration and will be discussed at a later date.

APPENDIX A

APPLICATIONS OF IS&R SYSTEMS FOR INTELLIGENCE PROBLEMS

1. Fingerprinting - A device for identifying and classifying ten fingers of up to 50×10^6 individuals, grouped into classes of whorls and similar identifying means for rapid access.

2. Sonar Noises - Detecting and classifying of signatures relating to living organisms and known types of water-borne craft. Access to groups of originators by noise spectra.

3. Cloud Cover and Aerial Reconnaisance - By means of catalogued types of cloud formations and known geographical outlines. Accessible by picture from Nimbus, Tiros or high-flying craft.

4. Census Figures - Including DODAC (53 categories) with Target and Capability analysis for strategic intelligence.

5. Various Classified Problems of Similar Nature to above

An optimum IS&R system retrieves 0.1 to 1.0% of its storage per day (C. Bourne, SRI), and on these terms the Video File is competitive.

PROGRAM	AGENCY	PROBLEM	COMPETITION
1.	FBI	Indexing to permit search on partial file.	Paper Cards
4.	Industry Biographical CIA FTD	Indexing, abstracting, translation of languages.	Printed Index of Abstracts
5.	Problem #1	No unretrieved data. File size - 7×10^6 Records. 1 Record = 200 character average. 10,000 characters maximum/retrieval. 15 minutes Access Time (Random order) Multiple Access required of 75 - 100 records/day with 3 minutes average Access Time.	

PROGRAM	AGENCY	PROBLEM	COMPETITION
5.	Problem #2	Auto descriptions as input. Owners of autos are output. Insurance companies could use this system.	
	Problem #3	File Size - 2 X 10^6 Documents of 2 pages each. 150 Retrievals/week. Net growth 1000 documents/week.	
	Problem #4	File Size - 250,000 documents of 2 - 3 pages each. Updated without additions. Immediate access required.	
	Problem #5	File Size - 50 X 10^6 Cards to form a multiple-ordered file. Reproduction of Abstracts of 20 minutes to 2 day access time. 100 - 200 retrievals/day.	
	Problem #6	10/12 X 10^6 Security PSQ File. 1000 Requests/day. File integrity is of utmost importance. Present system is a repository of punched cards.	

APPENDIX B

ASTIA PRINT FACTORY

Fiscal Year	1962	1963	1964	1965	1966
Documents in File	262,000	400,000	500,000	600,000	700,000
Average No. Pages/Document	50	50	50	50	50
New Documents/Year	30,000	100,000	100,000	100,000	100,000
Requests/Year	1,400,000	2,400,000	3,900,000	5,400,000	7,200,000

ASTIA RECORD FILE

Items in File	660,000	760,000	860,000	960,000	1,060,000
No. ASTIA Users	4,500	5,900	7,200	8,600	10,000
Descrip./Search	15	15	15	15	15
Descrip./Document	12	12	12	12	12
*RDTE Update/Year	15,000	80,000	84,000	88,000	92,000
*RDTE Req./Year	4,500	27,000	27,000	27,000	27,000
Bib. Req./Year (Priority)	4,500	27,000	27,000	27,000	27,000
(Routine)	500	3,000	3,000	3,000	3,000

* Research, Development, Test & Evaluation

PATENT OFFICE PRINT FACTORY

Documents in File	3,000,000+
Average No. Pages/Doc.	6 (640 max.)
New Documents/Year	250,000
Req./Year	3,800,000
Average No. Copies/Req.	2
Copies/Year	6,575,000

PATENT RECORD FILE (Patent Examiners)

	(1962)	
Items in File		
U. S. Patents	3,000,000	
Foreign Patents	3,500,000	
Non-Pat. Literature	1,000,000	
No. Classes	350	
No. Sub-Classes	55,000	
No. Users	1,500	
Update/Year	0	
New Additions/Year	1,200,000	
No. of Descrip./Document	?	(25 Experimental)
No. of Descrip./Search	?	

ARMY RECORD CENTER FILE

Documents in File	16,000,000	
Pages/Document	16	
New Documents/Year	400,000	(jackets)
Requests/Year	41,600	

NAVY PUBLICATIONS & PRINTING SERVICE

Drawings (Current) (To be located solely at Philadelphia)	25,000,000	(Single pages, various sizes)
Format: 35mm	500,000	
105mm	?	(Construction Drawings)

NAVY PUBLICATIONS & PRINTING SERVICE (continued)

(1962)

New Drawings/Year	1,000,000 (estimate)

MISA (Military Industrial Supply Agency)

File Size	750,000 items 25,000 pages
Copies Printed Each Run	40,000

MGSA (Military General Supply Agency)

File Size	120,000 items 5,000 pages
Copies Printed Each Run	40,000

APPENDIX C

INFORMATION PROCESSING SYSTEMS COMPARATIVE CHARACTERISTICS

I. GENERAL

To assess the capabilities of Information Processing Systems with respect to the field of Information Storage and Retrieval is a difficult and unrewarding task. The "difficult" aspect relates primarily to the problem of accomplishing a system analysis without a specific application in mind. Since the specifications of competitive systems are drawn mainly from publicity releases, they can be considered suspect; comparisons based on these specifications are very misleading and are therefore "unrewarding".

An analysis of comparative characteristics does provide certain useful bits of information, however. The general role of a particular system or component can be defined for IS&R applications. The scope of activity and direction of many manufacturers can be assessed. Approximate system performance with regard to storage capacities and output rates can be stated. Certain internal rates can be given but are somewhat meaningless unless applied to a specified activity.

The following lists of Information Processing Systems represent a compilation of proposed and existing systems slanted toward IS&R applications. True IS&R components are relatively scarce; most components with electronic search, storage, or retrieval functions are offshoots from business data processing systems. Systems designed for vast document storage are primarily of the microfilm variety. This fact may be of a transient nature reflecting the dependence on microfilm for reducing storage requirements in the past.

It should also be mentioned that the use of the term "Systems" as applied to the following listings is perhaps a misnomer. The comparative recency of IS&R as a field of major interest has resulted in tremendous activity by many manufacturers to market services and equipment in this field. Much of the equipment represents only a "partial" system; only selected aspects of the IS&R field, such as search, storage, or display are affected by this equipment. No Information Processing System can be described as "all-purpose" in the field of IS&R; the applications are much too complex for the current state-of-the-art. Information Processing Systems are usually designed for a specific application, or group of applications, and extension of these systems to other applications normally entails a compromise between optimum efficiency, and economy or timeliness.

II. SYSTEM CATEGORIES

The types of Information Processing Systems for IS&R can be categorized in several ways. One might make an initial division as "military" or "commercial". Although this division might be worthwhile on the basis of current use, it is often found that, after a period of time, a military system is followed by a commercial equivalent. Another division might relate to the role of the system in an IS&R application, i.e., whether the system is designed to accomplish complete or partial information processing. A third division could be based on the storage medium, e.g., microfilm or magnetic tape.

Since there is such a disparity among the many types of Information Processing Systems, it is questionable whether any division is meaningful. This doubt is increased by the fact that few of these systems are operational in the sense of having a record of proven worth behind them. Several of the systems are still in the developmental phase; the indicated roles of these systems in the IS&R picture may not necessarily coincide with their ultimate place in the field.

The breakdown of equipment characteristics utilized in this report avoids any attempt at arbitrary classification of systems. Instead, the objective has been to categorize the operations within the IS&R field and provide system information where applicable. It is felt that, by this method, the reader can define the system application and capability most relevant to his own objectives.

III. EXPLANATION OF TABULATION CATEGORIES

The following lists of Information Processing Systems occur under five major headings: General, Input, Storage, Retrieval, and Summary.

A. General

Manufacturers, system designations, and a brief description of system operations are given.

B. Input

Coding systems for later retrieval are indicated, techniques for entry of the material into the system are described, and the rate of input is given.

C. Storage

The type and size of the unit storage medium is given, the capacity of the unit record is listed, and pertinent data is stated to describe system capacity.

D. Retrieval

This section describes how, by what means, and how quickly material is accessed from storage.

E. Components

The various components which constitute the system are listed, sizes and costs are indicated where known, and an availability statement is made.

F. Summary

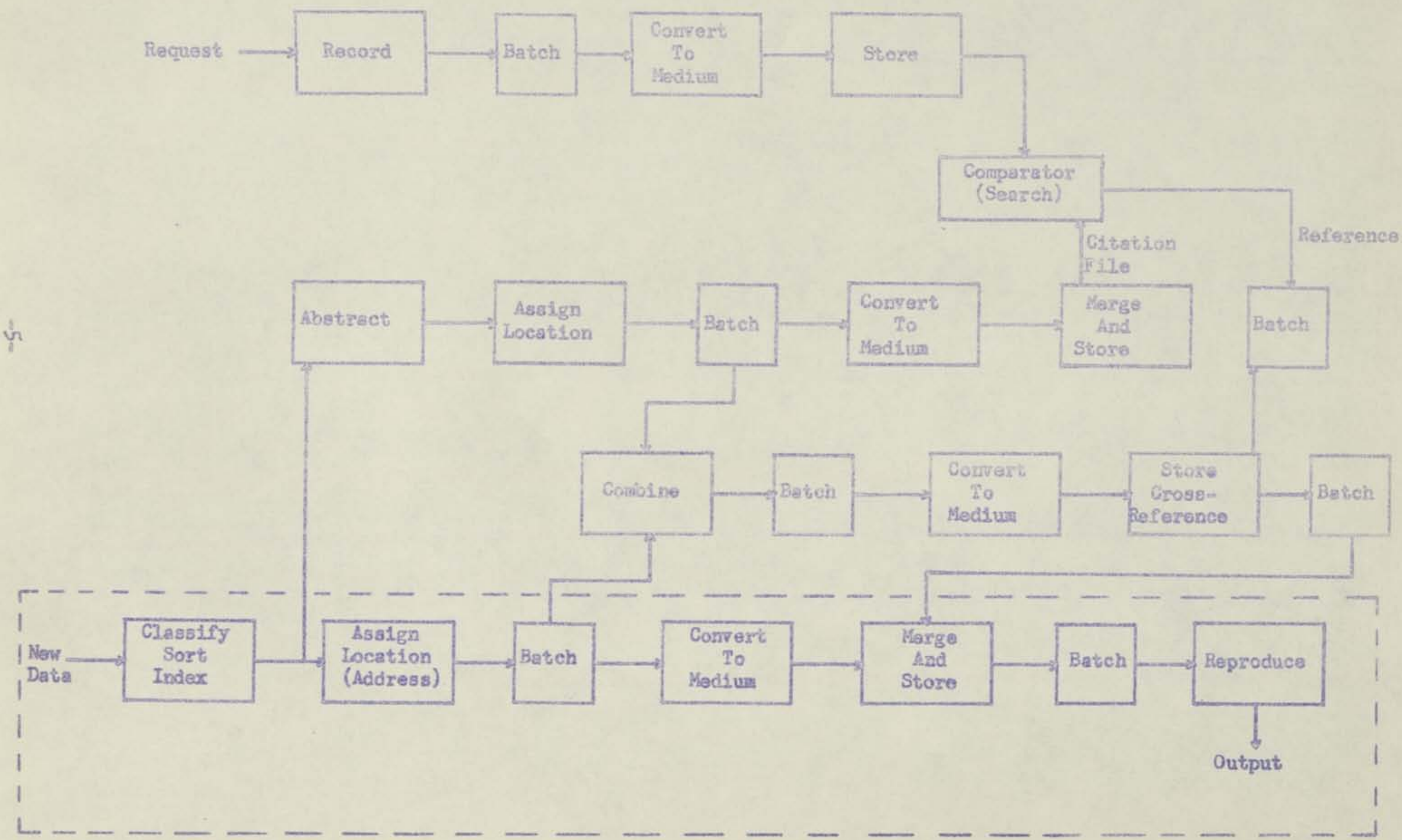
Pertinent characteristics of each system are summarized with respect to the IS&R function.

FLOW DIAGRAM

OF A

SIMPLE DOCUMENT RETRIEVAL SYSTEM

SIMPLE DOCUMENT RETRIEVAL SYSTEM (FUNCTIONAL)



-5-

COMPARATIVE CHARACTERISTICS

OF

INFORMATION PROCESSING SYSTEMS

TABLE I

GENERAL INFORMATION

MANUFACTURER	SYSTEM	DESCRIPTION
Aeronutronic Division Ford Motor Company	ARTOC DATAVIEW (AN/MSQ-19)	ARTOC (Army Tactical Operations Central), and DATAVIEW, the display system, being developed for USSC. Graphic information received, electronically processed and displayed. Hard copy provided.
Avco Corporation	VERAC	Storage and retrieval system for graphic information. Stored image can be displayed, transmitted by TV, or printed.
Benson-Lehner Corp.	FLIP	Console storage and retrieval unit for display of recorded information.
Eastman Kodak, or Recordak	DACOM (DAtascope Computer Output Microfilmer)	Converts computer machine language recorded on magnetic tape to human language on microfilm. Simultaneously combines output image with background format. Utilizes flying spot character generator.
	MINICARD	Complete ISR System incorporating analysis, coding, recording, storage, and retrieval of graphic information. Provides duplicate record, viewing, or enlarged copy.
Ferranti-Packard	RAPID ACCESS LOOK-UP SYSTEM (RAPIDAC)	Storage and retrieval system for graphic information. Provides display capability.
FMA, Inc.	FILE SEARCH	Storage and retrieval system for graphic information. Console provides display, hard copy, and duplicate record.
General Electric	SEARCH COMPARATOR	A search system designed to furnish address of stored information.
Heatwole Associates	H-44	Console stores coded descriptive information and retrieves a list or count of documents.

MANUFACTURER	SYSTEM	DESCRIPTION
Information Retrieval Corp. Information for Industry, Inc.	CRIS (Command Retrieval Information System)	A mass image S&R unit, designed for compatability with external coding systems, which provides hard or soft copy.
International Business Machines	WALNUT	Prototype S&R system for CIA with storage and retrieval functions separated. Stored document duplicated on aperture card for further use.
Jonker Business Machines	TERMATREX	A manual S&R system, utilizing a special drilled card for each term, which furnishes item location.
Laboratory For Electronics	RD-900 RASTAD	Provides electronic display of alphanumeric and line drawings stored in digital form.
Magnavox Corp.	MAGNACARD	Digital storage on magnetic cards. Provides hard copy printout.
	MAGNAVUE	Intrudes film aperture to magnetic card, and provides soft copy option.
	MEDIA	Appears to be "poor man's MINICARD". Storage and retrieval of graphic information.
Minn. Mining	FILMSORT	Records graphic information on punched aperture card. Provides display and duplication components.
National Bureau of Standards	RAPID SELECTOR	Storage of graphic information in conjunction with hard copy retrieval.
	MICROCITE	Mass image storage medium; projected image determined by illuminated card coordinates.
National Cash Register	CRAM	Stores digital information on magnetic cards to furnish a reference list for computer use.

MANUFACTURER	SYSTEM	DESCRIPTION
Radio Corporation of America	ARCOM	Automatic retrieval - comparator system to search index tapes and furnish reference locations.
	RACE	Random access card system for storage and retrieval of alphanumeric and pictorial information.
	VIDEO FILE	Utilizes video techniques for storage and retrieval of documentary information, and furnishes rapid hard copy printout.
Ramo Wooldridge	RW-400 AN/FSQ-27 (Polymorphic)	Large scale data processing system that is self-organizing, handles simultaneous requests, has module arrangement, is flexible & expandable, and has variety of uses.
Rese Engineering, Inc.	FINDAFACT 2510	Transcribes punch card data to magnetic tape, retrieves information upon request, and maintains magnetic tape files.
System Development Corp.	RAP 600	Stores graphic information in slide form, and provides viewing capability. Used on teaching machine.
Dr. Jacques Samain, Paris	FILMOREX	A film clip search, storage, and retrieval system; output used for viewing, enlarged print, or duplicate.
Houston-Fearless Corp.	AUTOMATIC IMAGE RETRIEVER	An electronically controlled random access photographic storage and retrieval system with viewing station and enlarged printout.

TABLE II

INPUT DATA

SYSTEM	CODING	INPUT METHOD	INPUT RATE
ARTOC	14-bit edge notched holder	CRT, camera, processor, and slide	6 second film processing, 1 second Kalvar print
VERAC	Location specified by row, column, & plane 7 digit address	By optical reduction 140:1 and 70:1	Camera can be fed at 15 pages per minute
FLIP	Location indicated by binary-coded bits	Optical reduction 48:1	
DACOM	n.a.	Magnetic tape to character generator to microfilm	14,000 characters/second
MINICARD	Alphanumeric photographic 294-2730 bits	Optical reduction: text 60:1; maps 39:1; photos 20:1. Code: paper tape to camera. Automatic selecting & sorting	Code: 35 characters/second information: 500 pages/hour
RAPID ACCESS LOOK-UP SYSTEM	3 character, 18 bit digital code	Photographic	440 pages/hour
FILE SEARCH	Maximum 392 code bits; photographic	Code: punched card to camera information: 25:1 optical reduction	500 pages/hour
SEARCH COMPARATOR	Digital	Normal input equipment	200 bits/inch 100 inches/second
H-44	Numerical	Victor portable keyboard	
CRIS	Location numbers	Photographic: master negative to Kalvar positive	

SYSTEM	CODING	INPUT METHOD	INPUT RATE
WALNUT	Address only; search via mag. tape for abstract list in RAMAC	Optical reduction: first to microfilm, then to Kalvar. Final record 35:1	1500 frames /hour
TERMATREX	Card = term Hole = item (inverted system)	Holes drilled in term cards to indicate items - manual	Depends on speed of operator
RD-900	Digital		
MAGNACARD	6-bit alphanumeric max. = 126 char./card	Via Magnacard Reading Station, paper tape reader, punched cards, or magnetic tape.	Read Station: 5400 cards/minute
MAGNAVUE	Digital, maximum character 756	Same as above, and optical reduction	Same for magnetic data
MEDIA	Binary coded decimal form	Optical reduction; manual sorting and filing	
FILMSORT	Standard punched card coding 55-68 character	Optical reduction 30:1	
RAPID SELECTOR	Binary dots on film 6 bits = machine inform.	Optical reduction 10:1 to 25:1	
MICROCITE	Address only; coincides with card coordinates	Optical reduction onto transparent cylinder	
CRAM	8-bit binary (edge notches)	From computer via tapes or cards	80 K char/second
ARCOM	Binary coded input compared to binary vocab.	Index tape digitally displayed to shutter matrix for word comparisons	Limited only by speed information can be presented to comparator

SYSTEM	CODING	INPUT METHOD	INPUT RATE
RACE	14 bit binary, edge notching	Normal digital or video techniques	Digital - 140K bits/second Video - 2.7 documents/second
VIDEO FILE	Digital coding along tape edge (4 tracks)	Camera unit converts document images to video signals for tape storage; index number recorded on digital track at same time.	1 sheet/second (both sides) 2 sheets/second (one side) Read-in: 4.8 inches/second
RW-400			
FINDAFACT 2510	Numeric	Magnetic or paper tape, punched cards	
RAP 600	10 bit address	Optical reduction variable	
FILMOREX	Digital coding by photographic means	Optical reduction on 30 meter roll of 70 mm film	
AUTOMATIC IMAGE RETRIEVER	7-digit alphanumeric code	Optical reduction, 40:1 or 80:1, address on mag. drum	

TABLE III

STORAGE DATA

SYSTEM	SIZE AND TYPE	UNIT CAPACITY	SYSTEM DATA
ARTOC	Film slide Kalvar	One image/slide	1000 slides for group display, 750 slides for console display
VERAC	High resolution film sheet or plate 8" - 12" sq.	Matrix = 100 x 100 images 10K legal size pages/plate	Basic capacity = 1K plates = 10M legal size pages
FLIP	16 mm - 1200' microfilm reels	72,000 frames (1 image/frame)	Unlimited number of reels; manual change
DACOM	16mm microfilm 100' magazine	160K lines of data	Unlimited number of magazines; manual change
MINICARD	16 x 32 mm high res. film chip	Max. 12 legal size pages/chip	2000 chips/file stick 800 K chips/file cabinet number of cabinets unlimited
RAPID ACCESS LOOK-UP SYSTEM	16 mm film loop	1 image/frame 880-1000 pages	
FILE SEARCH	35 mm - 1000' microfilm reel	30K pages/reel	Unlimited - Manual reel change
SEARCH COMPARATOR	2400' reel of magnetic tape	900K 7-character words	Unlimited - Manual reel change
H-44	¼" audio magnetic tape; 10½" reel	5-subject codes for 30K documents	Unlimited - Manual reel change
CRIS	Kalvar film scroll 2' x 400'	500K 8½" x 11" pages	Unlimited number of scrolls

STORAGE

SYSTEM	SIZE AND TYPE	UNIT CAPACITY	SYSTEM DATA
WALNUT	0.9" x 15.5" Kalvar film strip	99 images/strip	990 K images/bin potential = 100 M pages
TERMATREX	9 5/8" x 11½" (1 size) term card	10 K-40 K items depending on card size	Term capacity unlimited item capacity limited by hole capacity of card
RD-900	High density magnetic file drum	1.5 M+ alphanumeric characters	33 file drums = 50 M+ alphanumeric characters
MAGNACARD	1" x 3" magnetic card, mylar base	756 characters	3000 cards/magazine system capacity = 50 M characters
MAGNAVUE	1" x 3" magnetic card, incl. film clip; 1" reserved for image	450 characters and 1 image	Two 10-magazine file blocks = 27 M alphanumeric characters and 60 K images
MEDIA	16 x 32 mm film clips	2 document pages and digital data	200 clips/capsule, 100 capsules/drawer, 20 drawers/file } 400 K clips
FILMSORT	Punched card with microfilm window	1 image plus code	Unlimited
RAPID SELECTOR	3000 feet 35mm reel of microfilm	24,000 frames	Unlimited reels, manual change
MICROCITE	15" long x 4.75" diam. film cylinder	18 K abstract images	Unlimited
CRAM	3¼" x 14" x .005" magnetic card	21,700 alphanumeric 35,500 decimal	255 cards/magazine 16 magazines/computer system additional magazines require manual change

SYSTEM	SIZE AND TYPE	UNIT CAPACITY	SYSTEM DATA
ARCOM	Perforated card	Up to 100 terms	Unlimited
RACE	Magnetic card, or microfilm 16" x 4.5"	940 K bits digital 16 pages video	Based on 128 magazine system: 30.8 x 10^9 bits digital, or 524 K pages video, or combinations
VIDEO FILE	Mag. tape on 14" reel, 2" x 7200'	33,320 pages	Unlimited
RW-400	Tape Module: 2500' reel magnetic tape, 1" wide; Metricard film record	200 13-bit words/inch	
FINDAFACT	Magnetic tape reel	20, 80, or 960 characters	Unlimited
RAP 600	Standard 2" x 2" 35 mm slide	1 image/slide	600 slides
FILMOREX	Film clip 70 mm x 45 mm	2 document pages plus code; max. format of text = 30 x 45 cm.	Unlimited number of clips; manual feed to selector
AUTOMATIC IMAGE RETRIEVER	Film clip 63 x 87 mm	850 pages/clip	Basic module stores 10,240 clips, or 8,704,000 pages

TABLE IV

RETRIEVAL DATA

RETRIEVAL

SYSTEM	INDEXING	CONTROL	RETRIEVAL RATES
ARTOC	Localized random access	MOBIDIC computer and symbol generator	1 image out of 1000 = .05 min. microfilm copy at 2 frames/sec.
VERAC	External alphabetical or magnetic tape index searched for locations; then location dialed	Manul or paper tape	1 image out of 1M = 1 sec.
FLIP	External index; user requests location number	Keyboard	60 inches/sec., or 300 frames/sec.
DACOM	Code format between film images	Plugboard	1 page out of 180 K = 15 sec.
MINICARD	Multiple entries; duplicate clip for each level of indexing to reduce search time	Plugboard: Selector-Sorter	1200 clips/minute
RAPID ACCESS LOOKUP SYSTEM	Address only; IBM card punch supplies 3 character request	Keyboard	1 of 880 pages = 3 sec.
FILE SEARCH	56 characters/frame; can be serial or by descriptors	Punched card via Flexowriter	Average 4.5 minutes for 1000' reel
SEARCH COMPARATOR	Digitally coded search words	Keyboard, card reader, paper tape, or magnetic tape	900 K words in less than 5 minutes
H-44	Up to 5-subject codes	Keyboard	Search/count = 1500 5-subject documents/minute Search/print = 1000 ...
CRIS	External index; user requests location number	Keyboard	Estimates are from 7-20 seconds average
WALNUT	Location punched on card, transferred to magnetic index with keywords	Keywords punched on paper tape which initiates mag. index search; RAMAC	Index search 500 records per second; 6 sec. average access to document

SYSTEM	INDEXING	CONTROL	RETRIEVAL RATES
TERMATREX	External listing for each hole	Coincidence viewing on light table	Manual listings; speed depends on ability and quantity
RD-900	10 character word identifies drum, track, starting sector, number of sectors to read	Keyboard	Seconds from request to display
MAGNACARD	Table lookup for magazine, then automatic	CDC-160	90 cards/second 500 lines/minute
MAGNAVUE	Same as MAGNACARD	Same	Average retrieval 17 seconds
MEDIA			
FILMSORT	Standard punched card	Sorter or collator	Card retrieval rate
RAPID SELECTOR	1-6 word machine "code" for document information	Bank of photocells scan code	300 feet/minute 2400-3000 pages/minute
MICROCITE	Address only; keyed to inverted card system	Mechanical	
CRAM	Digital	NCR 315	Card selected and read - 1/6 sec or 170 milliseconds. Reaccess 1 m.s.
ARCOM	Binary card perforations		Manual card change for term comparison
RACE	Digital identification	Computer or special control unit	Access to retrieval information From 1st mag. = .18 sec. From 32 mags. = .45 sec. Read: 400 inches/second
VIDEO FILE	Sequential index number	Paper tape input to merge control unit	Search: 300 inches/second Readout: 4.8 inches/second Print: 2 pages second

SYSTEM	INDEXING	CONTROL	RETRIEVAL RATES
RW-400	Computer Module: 38 basic instructions and full indexing capability	Display Analysis Console Flexowriter	Analex high-speed printer 900 lines/minute
FINDAFACT		Plugboard	1000 lookups in 1 hour in file of 1M 80-char. records
RAP 600	Address stored in 6-15 computer	Keyboard	0.12 minutes for standard load (average)
FILMOREX	Coding capability for 20 descriptors or cross-references	Manual	Selector handles 600 cards/minute
AUTOMATIC IMAGE RETRIEVER	Address only	Keyboard accepts 1-64 address requests simultaneously	Single unit average - 4.8 sec. group of 64 - 2.5 min.

TABLE V

COMPONENTS

COMPONENTS

SYSTEM	COMPONENTS	SIZE	COST	AVAILABILITY
ARTOC	MOBIDIC computer, symbol generator, camera, processor, console display, projectors		275 K	Developed for Signal Corps
VERAC	Step-repeat camera; Storage and retrieval unit; Optical or television unit; Head-copy reproduction machine	Library of 70K ft/sq into 1K ft/sq		No prices until market study made
FLIP	Self-contained console	56" deep x 31½" wide x 53½" high 900 lbs.	50K	Operational
DACOM	Module containing character generator, display tube, 16 mm camera, & logic and control subsystem			
MINICARD	Typewriter - Tape punch, Copy cameras, Film processor, Film lubricator, Film cutter, Viewers & printers, Selector-Sorter, & Computer-Duplicator		2.5 - 3.5 M	4 installations in Gov't, one at Kodak
RAPID ACCESS LOOK-UP SYSTEM	Filming table, Look-up and display unit, Interrogator		34 K	
FILE SEARCH	Recording Unit Retrieval Unit	71" x 55" x 50" (Retr. unit)	115 K	First commercial installation at Bureau of Ships, 1961
SEARCH COMPARATOR	Console, Optional alphanumeric printer	H-72", D-30" W-40"		Prototype displayed
H-44	Console, keyboard, and printer	67" x 22" x 43"	9.5 K	Developed and tested
CRIS	Basic unit is console with film control, projector and display, and keyboard; other features optional	Console is desk-size		Available Spring 1962; Mfg. by Litton Systems, Inc.

SYSTEM	COMPONENTS	SIZE	COST	AVAILABILITY
WALNUT	Microfilm unit, card punch, tape punch, image converter, tape reader, RAMAC computer			No plans for marketing
TERMATREX			1-6 K	In production
RD-900	File drums, symbol generator, viewing console, control units			
MAGNACARD	Card handler, control unit, Printer, CDC-160 computer		Component price list available	12 months from order date
MAGNAVUE	Same, plus display station		250-450 K	
MEDIA			Supposed to be "low-cost"	Under development
FILMSORT	Filmac viewer and printer, standard commercial equipment		10-25 K	Off-the-shelf
RAPID SELECTOR	Camera & encoder, microfilm copier			An N.B.S. development
MICROCITE	Uniterm cards, camera, machine with viewer and printer			An N.B.S. development
CRAM			Rent $38 K/ month	Early 1962; with NCR 315 system
ARCOM	Control mechanism, tape input, ARCOM unit, printer			Under development
RACE	Control mechanism, track and magazines, hard and soft copy equipment			Under development
VIDEO FILE	Merge control, Printer, Monitor, Soft copy reader, Status and switching panel, Video recorder	Recorder: 72" x 47" x 22"; Merge Control: 72" x 19" x 22"		Under development

SYSTEM	COMPONENTS	SIZE	COST	AVAILABILITY
RW-400	Computer mod., Buffer mod., Central exchange, Drum mod., Tape mod., Tape adapter, Peripheral buffer, Display buffer, Display analysis console, Computer Comm. console, Printer, Plotter, Paper tape reader, User mod.			Presently available components listed
FINDAFACT	Control Unit, Tape Stations, Input-Output			
RAP 600	G-15 computer, random access projector, standard photo equipment		8 K	Presently available
FILMOREX	Filmorex camera, Filmorex selector	Selector: 75 x 50 x 40 cm.		Operation; 2 French product
AUTOMATIC IMAGE RETRIEVER	Camera, keyboard, selector, viewer, printer		250 K	Under development for USAF: delivery scheduled April 1962

TABLE VI

SUMMARY STATEMENTS

SUMMARY

SYSTEM	COMMENTS
ARTOC	Designed specifically for military field use as a display system; limited storage capabilities.
VERAC	Primarily a storage system; requires separate search system and is virtually impossible to update.
FLIP	Display console utilizing microfilm reel storage; requires separate search system; serial access and manual reel change; no hard copy provided.
DACOM	Designed for business applications and may not be suited for IS&R; fills component role of system which requires microfilm output from computer.
MINICARD	Large storage capacity and random access filing; very expensive, separate clip required for each indexing level and cross-reference; clip combines index with storage; manual selection of file sticks.
RAPID ACCESS LOOK-UP SYSTEM	Temporary individual display of limited library requiring little updating; requires separate search system.
FILE SEARCH	Combines search and storage function in console utilizing microfilm reels; manual reel change; updating possible by splicing; slow retrieval.
SEARCH COMPARATOR	Search system only utilizing magnetic tape storage; may contain abstracts; could be used in conjunction with separate document storage.
H-44	Search system only utilizing magnetic tape storage; limited storage; printout capability.
CRIS	Storage system only; updating virtually impossible; manual scroll change; requires external search system.
WALNUT	Complete IS&R system with human interrupt between search output and retrieval input; very expensive; image storage accessed by address randomly.
TERMATREX	Manual search system only by coincident viewing; inverted card files with each card a term and each hole an item.

SYSTEM	COMMENTS
RD-900	Display system utilizing drum storage and symbol generator; generates line drawings from digital storage.
MAGNACARD	Computer-controlled, random access card file; stores digital information only; table look-up for magazine reference.
MAGNAVUE	MAGNACARD with film aperture; includes display stations.
MEDIA	Film clip storage system; information of characteristics lacking; clip similar to MINICARD with less unit storage.
FILMSORT	Standard punched cards with film aperture. Primarily manual operation with automatic card sorting.
RAPID SELECTOR	Modified version of an old microfilm storage system. Large reel storage capacity; microfilm printout; updating only by splicing.
MICROCITE	Manual-mechanical system primarily for browsing through abstract collection on film cylinder.
CRAM	Random access magnetic card storage for NCR 315 system. Designed for business applications.
ARCOM	Search comparator suited to high retrieval rate problems. Scans tapes for comparison with interrogation terms. May also compile index term lists.
RACE	High-speed random access card storage system; operation with or without computer; storage in digital, video, or microfilm form; hard and soft copy output.
VIDEO FILE	Mass document storage in video form; computer controlled; high speed serial access; updating, transmission, and reorganization capability; hard and soft copy.
RW-400	Computer controlled module system; film, tape, and other storage; expandable with varied components; not primarily IS&R but adaptable.
FINDAFACT	Component in system for manipulating magnetic tape files; limited IS&R application.

SYSTEM	COMMENTS
RAP 600	Temporary group display of limited library; teaching machine application; computer controlled random access slide file for projection.
FILMOREX	French film clip system available through Benson-Lehner; not known to be in use in U.S.A.; similar to MINICARD but 70 mm. clips.
AUTOMATIC IMAGE RETRIEVER	Computer controlled film slide storage system; being developed for RADC; random access, large storage, fast retrieval.

INTERNAL CORRESPONDENCE

DATE: 8 March 1962

TO: G. Arnovick

FROM: P. Tanner

SUBJECT: Information on Indexing Procedures

Introduction

Volumes #6 thru #8 of, "Current Research and Development in Scientific Documentation," were examined for information on indexing procedures. In the majority of procedures cited, the indexing methods are detailed. In the remaining cases, the index procedure is at least named and the operations or uses are indicated. The same or similar information appearing in more than one volume was duplicated.

Volume #6

Page 26, The Chemical Abstracts Service: G. M. Dyson

The material of the Chemical Abstracts formula indexes is being classified in a preliminary edge-notched card form to provide significant and useful subdivisions of the whole material. Experimental runs on a variety of methods of storage - aperture cards, microcards, Minicards - are being set up to ascertain the most suitable means of building a document storage file of abstracts (about 2 to 2.5×10^6 are involved).

Page 38, Lockheed Aircraft Corporation: R. P. Mitchell

An algebraic representation of syntax has been found which covers a large subclass of English sentences. It may then be possible by a computer, to perform operations such as collating, indexing, abstracting, and making bibliographies.

Page 43, Planning Research Corporation: H. P. Edmundson and W. V. Owens

A full-scale experiment in indexing and abstracting by means of the "multiterm" technique was carried out. For each document, significant word-groups, or "multiterms", were determined by frequency criteria, and sentences with highest number of multiterms were selected as "topic sentences". Multiterms of highest significance serve as an index to the document; and the topic sentences of highest significance serve as an abstract.

Page 45, Ramo-Wooldridge: J. Kuhns, etc.

A probabilistic model for the assignment of index tags to documents has been hypothesized and partially tested. When a subject heading or other index tag is assigned to the document, a weight factor is also assigned which measures the degree of relevance or appropriateness of that tag or heading to the document being indexed.

Page 3, American Institute of Physics: R. E. Maisell

The permuted title index system, an index produced by the Listomatic System, and a simple listing of title pages, are being studied. Using the Listomatic System, preliminary findings do not show any economic advantage of the high speed filming of punched cards over monotype composition and letterpress printing. More economic means for cumulative index preparation may include the storage and cumulation of type by the printer or the reassembly and photographing of printed index entries previously cut into strips.

Page 6, The Chemical Abstracts Service: G. M. Dyson

A series of experiments on the use of Varitype setting with Fotolister camera collation has been started. C A formula indexing has been demonstrated using 4,000 cards.

Page 30, General Electric Co.: L. E. Saline

The Enriched Coordinate Indexing (ECI) concept is used. It has been designed for use in systems requiring machine storage and retrieval of large quantities of diverse information. ECI combines the techniques of classification and coordinate indexing. Class indicators can be added so that the machine can distinguish among descriptors of identical spelling but different meanings in context. Syntax is preserved by role indicators and descriptor links. Generic terms, implied but not in a document, may be added to the index record.

<u>Volume #7</u>

#2.19, Page 30, IBM: T. R. Savage, H. P. Luhn

An auto-indexing method, produces for a set of documents a "Keyword-in-Context Index". Index is derived from texts available in machine-readable form by establishing a list of keywords, and by extracting portions of titles, abstracts, or texts containing such keywords. These portions are then arranged in alphabetic order of the keywords to form the index. IBM 704 Programs for this may be available through SHARE.

#1.2, Page 3, American Institute of Physics: H. C. Wolfe

A compilation is made of major subject headings and cross-references. Many journals use different terms for the same concept, therefore feasibility of some form of coordination is being studied. Index density (number of subject entries per article) is related to ease of retrieval.

#2.8, Page 21, Communanute Europeenne De L'Energie Atomique (Euratom): P. Braffort and A..Leroy

Indexing for document retrieval is based on the use of an artificial documentation language as the internal language of a computer. The artificial language can be represented by diagrams whose knots would be "notions" and the braches "relations" between the notions. Efforts deal with creation of a detailed vocabulary and grammar of the documentation language and with the operations of automatic translation between natural languages and the documentation languages.

There is an automatic analysis of documents. This operation consists of establishing semantic bonds between certain elements of a stemma or between elements of several stemmas (parts of a diagram). When the operation is completed by determining solutions to the problems of synonymy and polysemy, it makes possible the desired diagram representation. This in turn leads to completion of the general diagram, which will eventually include the entire contents of published scientific tests arranged so that many "semantic hierarchies" are apparent.

#2.20, Page 32, Itek Corporation: J. W. Kuipers, etc.

The grammar of the predicate calculus in symbolic logic forms the pattern for the representation of subject matter, with the object of reducing and controlling the expressional variety of input tests without altering significantly their informative meaning. This method applies also to the generation of index data.

In operation, for each incoming document a single "bibliographic card" is typed in an interpretable serial format on a Flexowriter. The format includes several variable-length fields. A computer has been programmed to extract and rearrange the punched type data to produce accession lists, author lists, subject lists, title concordances, and records which have been arranged in format to meet the needs of a special processing machine, such as the Minicard search selector.

#2.14, Page 26, Gmelin Institute Documentation Center: E. H. Pietsch and K. Schneider

EAM equipment is used to produce indexes of bibliographic summaries. Plans are for using this equipment for preparing more detailed subject indexes making extensive use of descriptors and, beyond that, of key phrases extracted from the original publications. The equipment will make it possible to obtain, by automatic means, index texts which can be arranged by one or a combination of several criteria.

One specific project relates to the formulation of subject headings, descriptors, key terms, or information units of maximum "statement density"; correlation of these terms with subject headings; determination of their frequency of use; the establishment of semantic relationships between terms; and allied questions.

#2.16, Page 28, Herner and Company; S. Herner, etc.

An index for the unclassified reports of the AEC is being prepared. This index, which gives the class notation of every term and concept in the classification schedule (already complete), serves both as a key to the schedule and as a code dictionary for the application of the scheme in mechanical storage and retrieval systems. Since the code is based on a multilevel classification, it can be used to facilitate generic as well as specific searches.

Volume #8

#1.12, Page 10, Institute for Scientific Information: Eugene Garfield

The first cumulative molecular formula index of 43,000 new chemical compounds was sorted and printed by punchcard equipment. An author index of 11,000 entries was similarly prepared. A new generic formula index was prepared to enable one to find, in one place, all compounds containing the less frequently occurring elements. A basic recognition grammar of organic chemistry was prepared enabling computer conversion of chemical names to chemical formula.

#1.13, Page 12, National Institutes of Health: L. L. Cahoon

A modified Uniterm system of indexing was used to prepare a scientific subject-matter index on 12,500 research projects. An average of 12 main terms, each with approximately 5 modifiers, were extracted from summary statements. IBM cards are used and sorted alphabetically by main terms.

#2.12, Page 30, General Electrical Company; L. E. Saline

The system's current design provides for the sequential updating and searching of a serial file. The file consists of unit records composed of "terms". A "term" is defined as any combination of alpha-numeric characters. Terms can be grouped or associated at four levels, i.e., as associated words, phrases, sentences, or the complete unit record. Two modifiers may be associated with each term; the modifiers are used to reduce problems of polysemy and syntax often assocated with English word indexes.

#2.39, Page 49, U. S. Army Chemical Corps R and D) Command: C. J. Maloney and H. W. Batchelor

Indexing is done by a series of lattice or two-dimensional tables. The subject heading and its code are in the upper left-hand corner of the lattice. An array of codes for eight categories, into which a main subject heading may be divided, is formed at line intersections along the left-hand margin of the lattice. A similar array of eight codes for subcategories or descriptors is formed across the top margin by using 8 subcodes, into which each primary category may be divided. The terms for a total of eight categories and 64 subcategories may then be written near the intersections of the respective row and column lines.

Another indexing method uses two matched forms for indexing documents. The first form contains a series of subject headings and their descriptors. The second form contains code numbers that correspond, in context and exact position on the page, with their respective terms on the first form. At the bottom of each form are blanks for writing the document title, number, author, and other pertinent information. The forms are used together, with carbon paper insert. The indexer checks pertinent subject headings or descriptors for a given document and writes in any other terms and information needed to index properly and to identify the document.

#2.40, Page 50, U. S. Patent Office: D. D. Andrews

A coordinate index file for miscellaneous transistor circuits, consisting of 500 original and 800 cross-reference patents, is being established. The file emphasizes the depth (an average of over 100 descriptors per document) required for patent searching, and an extensive group of topological terms. Peek-a-boo cards are being punched and display means, such as the Microcite Reader, are being considered.

P E Tanner

P. E. Tanner
Systems Engineering

VM

D. Ballard 3/1/62

HIERARCHIAL INTELLIGENCE RECORDING & ASSOCIATION MODEL (HIRAM)

A basic requirement for an intelligence storage system is the ability to draw together all relevant pieces of information in answer to interrogations which may be made at almost any hierarchical level of relationship. Systems employing simple descriptors and low level associations between those factors allow relatively simple computer processing, but place upon the analyst the burden of determining the relevancy of the retrieved information, much of which will not necessarily be pertinent.

The multiplets of the ACSI-MATIC system provide a logically manipulatable information unit, but appear to require linking two or more such multiplets to convey the equivalent meaning of a simple grammatidal statement. The multiplet does not appear to be sufficient for an automatic logical consistency check as to the completeness, to say nothing of the reasonableness, of an incoming statement. This appears to limit the possibilities of utilizing automatic input processing instead of human analysts to encode incoming information.

The techniques of machine language translation (MLT) have been quite well explored, and offer means for automatically analysing the syntactical structure of sentences. The semantic content of a sentence is dependent both upon the words used and upon their relative order of use. Automatic syntactical analysis and automatic reference to vocabulary lists (formally equated to hierarchical code lists) governed by a formal set of rules may be used to construct operating formal statements of the form shown in equation (1)

$$\left\{ I \left[t(S \cdot O \cdot A_m) \rightarrow P \right] \right\}_1 , \cdots \left\{ I \left[t(S \cdot O \cdot A_m) \rightarrow P \right] \right\}_n \rightarrow R \quad (1)$$

where

$\{\ \}$ ≈ Operating level formal statement,

R ≈ Total stored intelligence item (gives location of storage and act as a link between statements included in a particular item),

I ≈ Field of Interest (e.g. Economics, Law, Military strategy, etc),

t ≈ Time of statement or origination of subject or object (A_m may modify),

S ≈ Subject taking the action,

O ≈ Object acted upon or co-subject of intransitive actions,

A ≈ Action,

m ≈ Modifier of action,

P ≈ Product or Result, and

$\rightarrow$ ≈ Leads to.

A symbol outside of a bracket may modify the heirachical structure of the code elements within that bracket, e.g. I modifies S, O, A_m and P. t may modify S, O and A_m but is unlikely to modify P as this should be chosen to include time-stable terminology. For example, S and O might contain names of countries or cities whose names may be subject to change with time. t itself may express either relative time, as dates, or absolute time relationships such as elaped time, velocity, rate, etc.

It goes without saying that rules are made only to be broken. Even though we may establish a means for automatic encoding of intelligence data, we must realize that some data have a priori more intrinsic worth than others, and intelligence analysts may prefer to short-cut the machine procedure for input.

Figure 1 takes into account the above document evaluation and shows a suggested flow of data and the processing sequence for input operations.

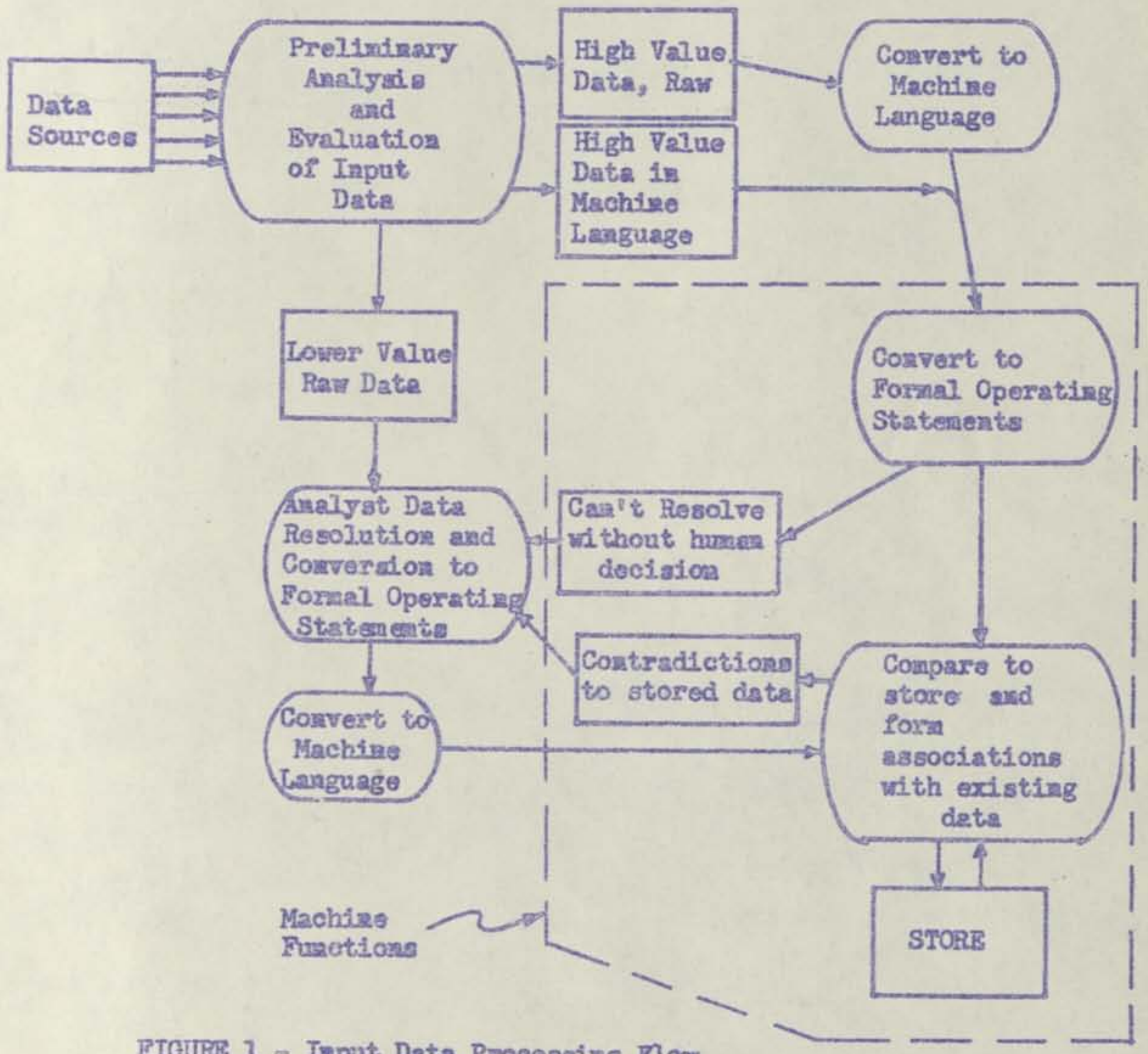


FIGURE 1 - Input Data Processing Flow

Various index files of the formal statements will be derived from the combination of input data and logical decisions applicable to those data by the system or by the human analysts.

A file of logical statements will be created to serve as a check upon the reasonableness of incoming statements. For example, an input statement regarding the movement of the troops of one nation through the territory of another nation cannot be considered as reasonable unless (a) these two nations have some treaty or agreement regarding such movements or (b) these two nations are at war with each other or (c) one of these nations is in a critical geographical location with respect to some aggressor nation. Such statements may themselves be derived from verified input data.

Any contradictions to the stored logical rules of reasonableness, any lack of completeness or other inherent defects of the statement would be sensed automatically, and cause the statement to be transmitted to the analyst for further investigation.

DEFENSE ELECTRONIC PRODUCTS

DATA SYSTEMS DIVISION

APPLIED RESEARCH PROPOSAL

TITLE: A Study of Algorithms and the Determination of Descriptor Characteristics for Self-Organizing Indexing and Searching Systems

Statement of Problem:

Considerable effort has been expended by others to establish the mathematical basis for various indexing and searching methods, however those few cases where actual micro-library searches have been used fail to establish a yardstick for comparison of a variety of approaches. Numerous methods of indexing and searching have been explored as isolated cases, but little can be deduced as to their relative merits. Comparison of various approaches upon a uniform set of subjects is proposed.

Objectives:

Derivation of effective procedures and algorithms for indexing and searching by comparative experiments performed upon representatives samples of collections of documents covering rather narrow fields of technical specialization. These techniques in turn would lead to better functional specifications for efficient search and retrieval systems applicable to large scale information collections.

Approaches:

1) Establishment of suitable document collections, to be referred to as micro-libraries.

2) Establishment of a "yardstick" for measurement of the results of all subsequent experiments.

3) Formulation of computer programs for indexing and searching experimentation, including initial concordance runs and evaluations.

Work Proposed:

1) Establish a micro-library sampling a field of narrow technical specialization. The micro-library would contain at least 100 articles.

2) Produce a concordance of words found in these articles, and listings for every article carrying all words by frequency.

3) A group of persons representing the users of this micro-library would formulate typical research questions in the chosen field. Later this group would select from the micro-library all articles which appear most pertinent to answer these questions.

4) The results of this direct selection process would be used as a yardstick to evaluate rules and algorithms for automatic indexing and search procedures, to test them by computer searches, and to alter them until their effectiveness approaches or exceeds the results of the direct search. Algorithms for association factor indexing, probabalistic indexing, Uniterm descriptor indexing, Western Reserve University logical statements and other methods will be treated as components of self indexing schema.

5) As background for experimentation, a bibliography on various analytical and practical approaches to these problems will be compiled.

6) Time and money permitting additional micro-libraries will be selected and subjected to the same procedures as used upon the original experiments. This will permit verification of the validity of the conclusions drawn from a limited sample as compared against a wider spectrum of knowledge.

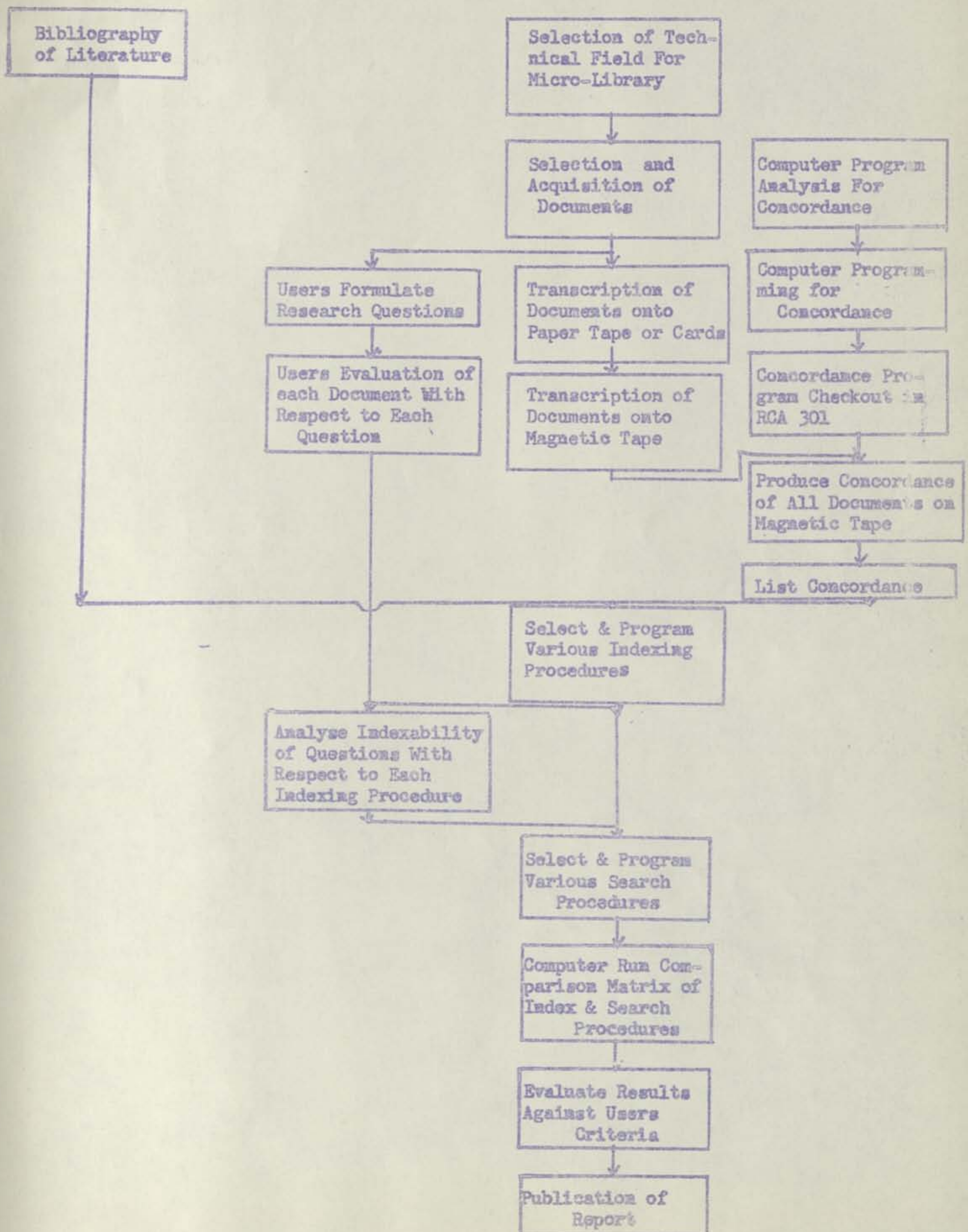
Probability of Success:

Excellent. Past attempts in this area have shown very promising results. RCA's Rapid Information Search and Self-Indexing Systems would benefit immeasurably by this study.

Profit Potential:

This study would enchance RCA's position in the fast growing field of information search and retrieval by raising the processing capabilities of related machine systems. In addition the study would assist in obtaining systems contracts from Government and Military Agencies.

DIAGRAM OF TASKS RELATIONSHIPS



PROPOSED WORK SCHEDULE

CALENDAR MONTHS

TASKS	1	2	3	4	5	6	7	8	9	10	11	12	MAN-MONTHS ENG.	CLERICAL
Bibliography of Literature													1	
Computer Program Concordance													2	
Selection of Technical Field for Micro-Library													½	
Selection & Acquisition of Documents													1	1
Transcription of Documents														6
Users Formulate Research Questions													1	
Users Evaluation of Documents With Respect to Questions													3	
Select and Program Indexing Procedures													7½	
Analyse Indexability of Questions With Respect to Indexing Procedures													1½	
Select And Program Search Procedures													7½	
Computer Run Index and Search Procedures													2	
Evaluate Results													2	
Total													29	7

INTERNAL CORRESPONDENCE

RCA Confidential

DATE: 20 March 1962

TO: G. Arnovick

FROM: P. Tanner

SUBJECT: Automatic Analysis and System Synthesis for IS & R Applications

I. Statement of Problem

When presented with a specific IS & R application, the problem arises as to what equipment would best be suited for the job. This presents a voluminous problem since many IS & R Systems should be considered. Moreover, EAM Equipment and certain computers should be investigated, especially in the cases where a storage and retrieval system (per se) does not meet all of the application requirements. In other words it would require an exhaustive analysis of all practical combinations of equipment in these three areas in order to determine what system, or collection of systems and separate components, would best fit the problem.

To do such an analysis manually would take more manpower than most installations have, or are willing to put on such a project. To do the job properly would also require a complete file of detailed information on each component of equipment that can be used for information storage and retrieval. Examples of required system characteristics are cost, availability, and capabilities (functions that can be performed). Certain application requirements with their degree of importance are also needed. These would include allowable cost, when equipment is needed, and the functions that are to be performed. The application parameters must then be matched against each system's corresponding characteristics to determine if a match can be attained. The process becomes more complex when a system fails to meet the requirements at a certain point. Then other equipment must be analysed for the possibility of interfacing and subsequent integration into the overall system. As a result many practical systems would be generated.

The subsequent problem would then be how to choose the generated system which best fits the application requirements. This would be difficult to do manually (if at all possible) due to the large matrix of requirements versus generated systems and their characteristics.

Another major problem is involved in the approach to the analysis of the application requirements and their bearing on the synthesis of the system components. What application parameters does one consider and what method is to be used in comparing them with the systems' characteristics? Should we examine one system at a time with respect to all of the application requirements, should all systems be analysed with respect to one requirement or function at a time, or is there a better method?

FORM DEL 1241

It should be clear that any manual procedure for selecting an IS & R System, given a specific application, is either incomplete or too time consuming. The scope of the required analysis and synthesis indicates the use of automatic equipment, i.e., a computer with an internally stored compiler-generated program.

II. Research Objectives and Method of Solution

The overall objective of research in the area of IS & R application analysis is to evolve a practical and automatic method for system synthesis. This requires that given any IS & R application, one or more compatible systems can be specified with all of their quantitative and qualitative characteristics.

To solve this problem in an effective and efficient manner, a computer must be used. It should be chosen with the following job requirements in mind: primary and secondary storage, input-output equipment, processing speed, and automatic programming facilities.

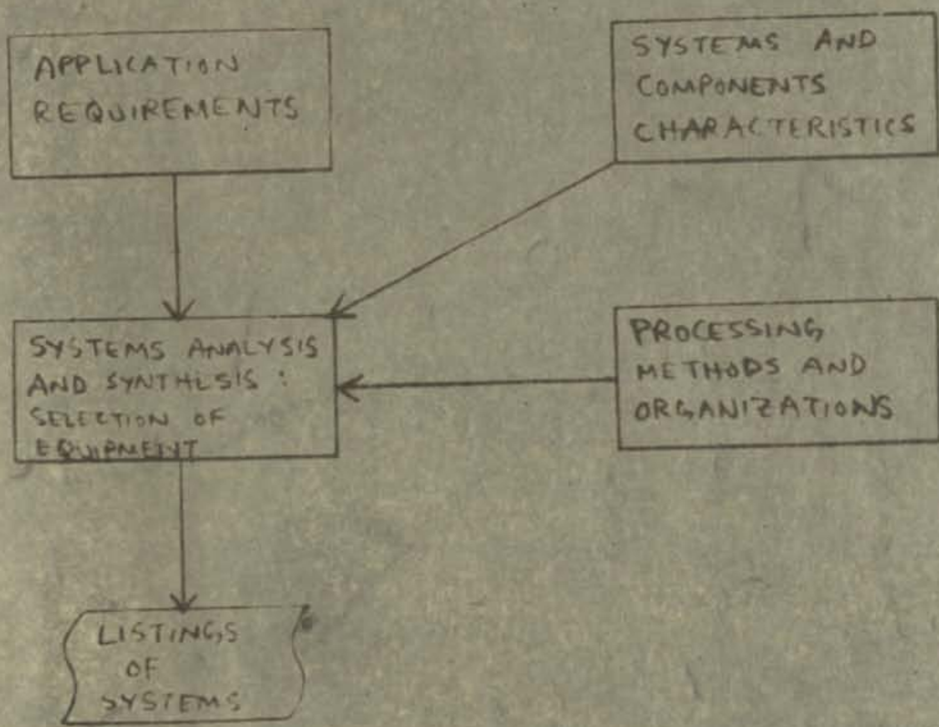
The first task to be accomplished is to generate a complete list of system component characteristics and ancillary information. The parameters would be divided into four categories: input, processing, output, and ancillary. The principal input-output items would include: media, coding, rates, capacity, and unit of information (as character or word). Processing characteristics would include those mentioned for input-output, as applied to internal storage units, and capabilities, i.e., functions that can be performed. Among the ancillary parameters are cost, availability, and size of components.

At this point the appropriate parameter information must be gathered for each IS & R System, EAM device, and certain computers. A file can now be generated by, and stored in, the computer to be used for the application analysis and system synthesis.

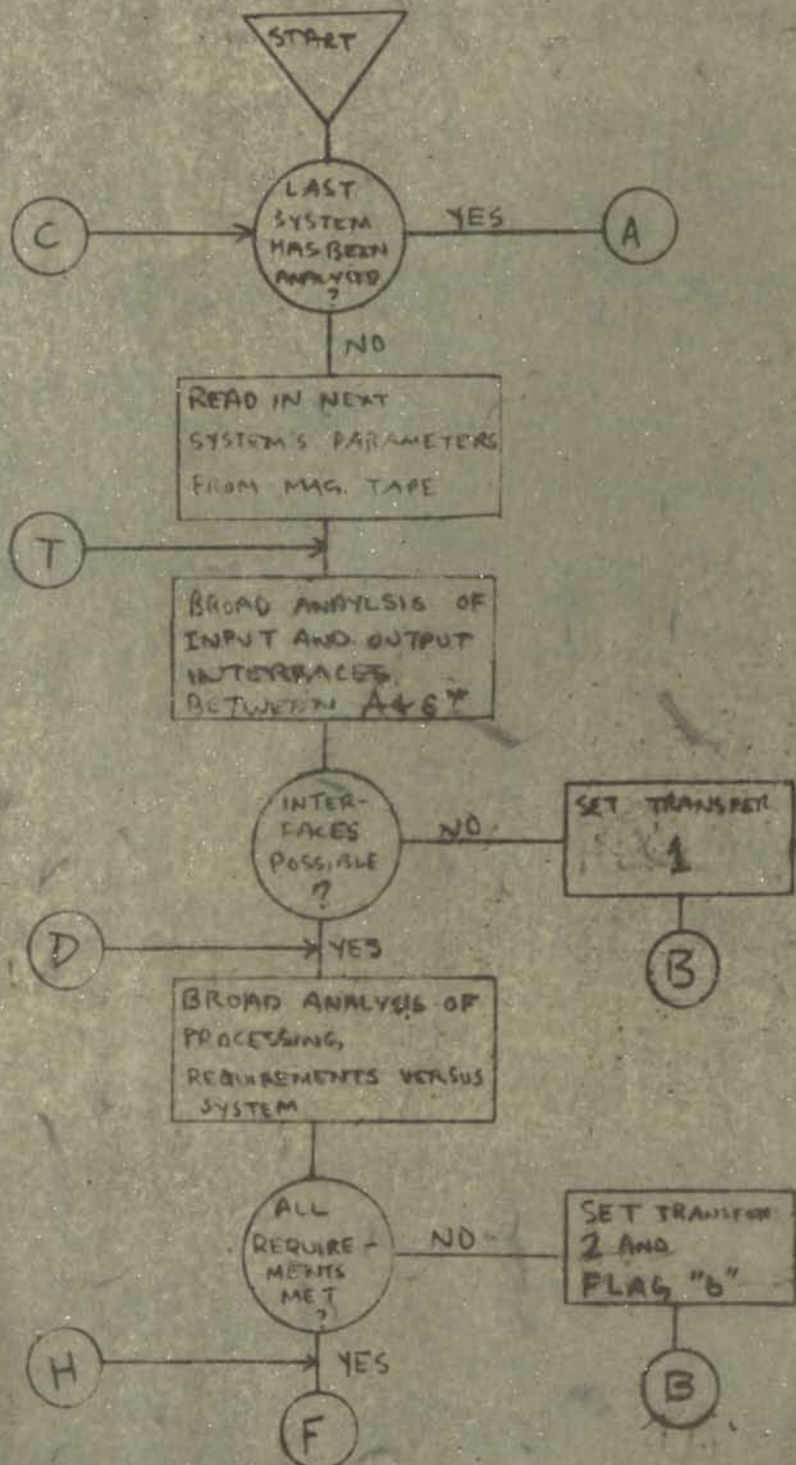
The next step is to form a complete set of application requirements and characteristics. These would include: formats, volume and type of data, information flow mode (as continuous or intermittent), processes to be performed, and storage requirements. Some of the parameters listed above for the IS & R systems would also be germane to the application characteristics. Examples of this are cost, availability, and rates of input.

The most difficult problem to be solved is how to process the requirements for a given application, of course, they must be compared with the characteristics of various system components, but the system synthesis is more involved than this. What must be generated is a framework or network for a generalized information processing system. This would allow the step by step analysis of a given application from initial input to final output. Major items in the network would include file generation, updating, and requests. Subdivisions would include sorting, merging, and code conversion.

DIAGRAM OF METHOD OF APPROACH TO AUTOMATIC COMPILER GENERATED APPLICATION ANALYSIS FOR IS&R SYSTEMS



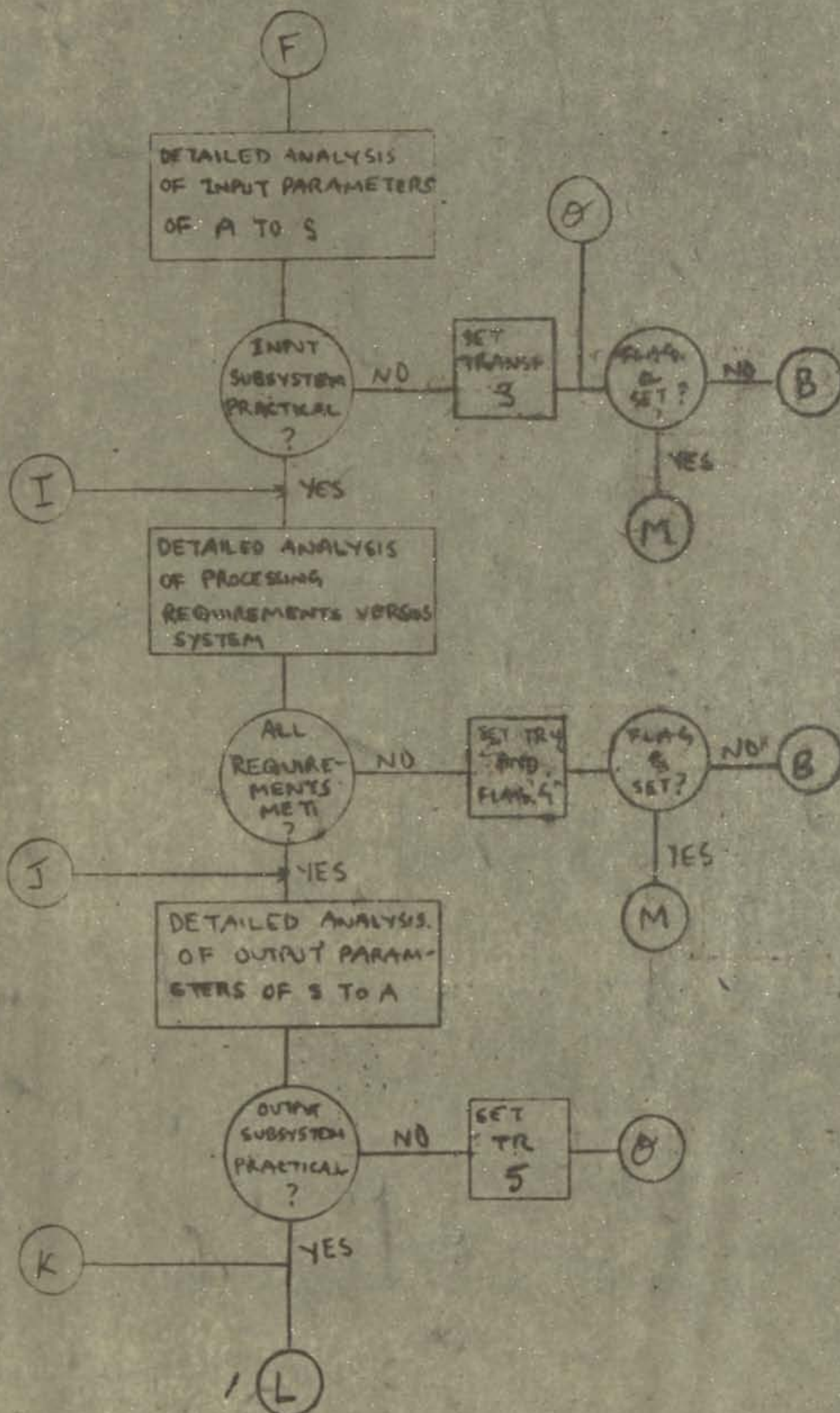
BROAD FLOW DIAGRAM OF AUTOMATIC COMPILER GENERATED APPLICATION ANALYSIS FOR IS4R SYSTEMS

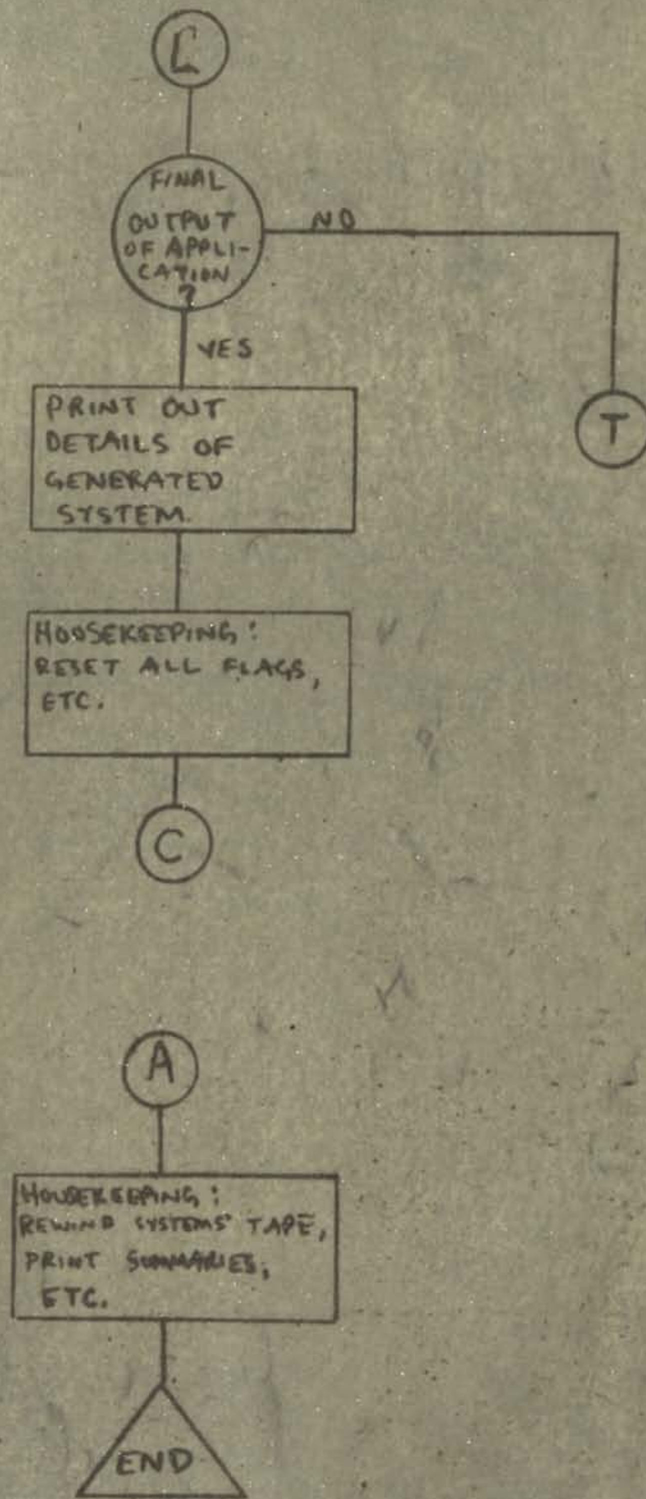


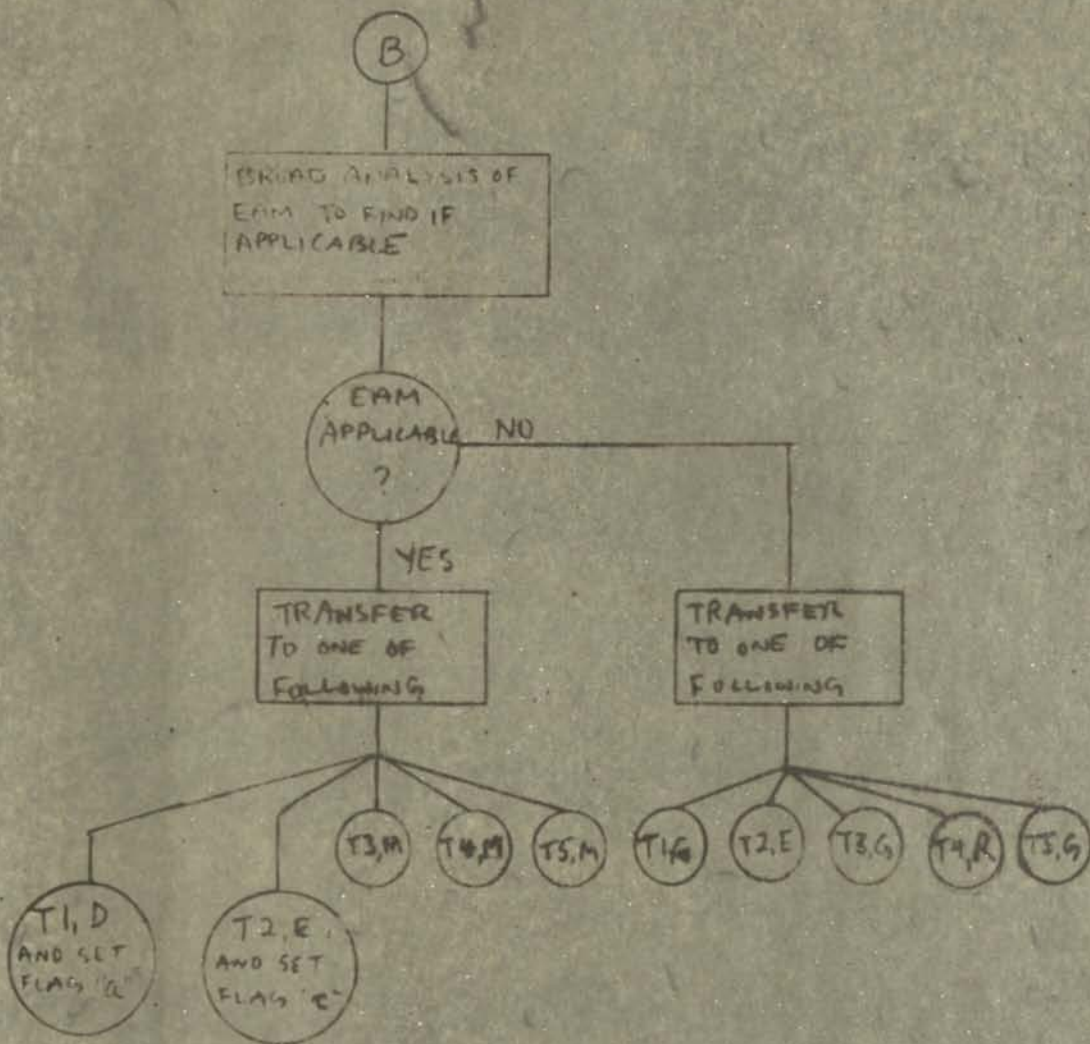
*A = APPLICATION
S = SYSTEM

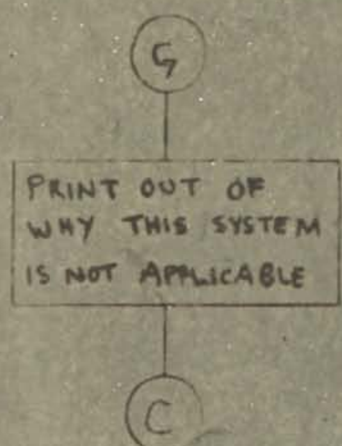
3/9/62

P. TANNER

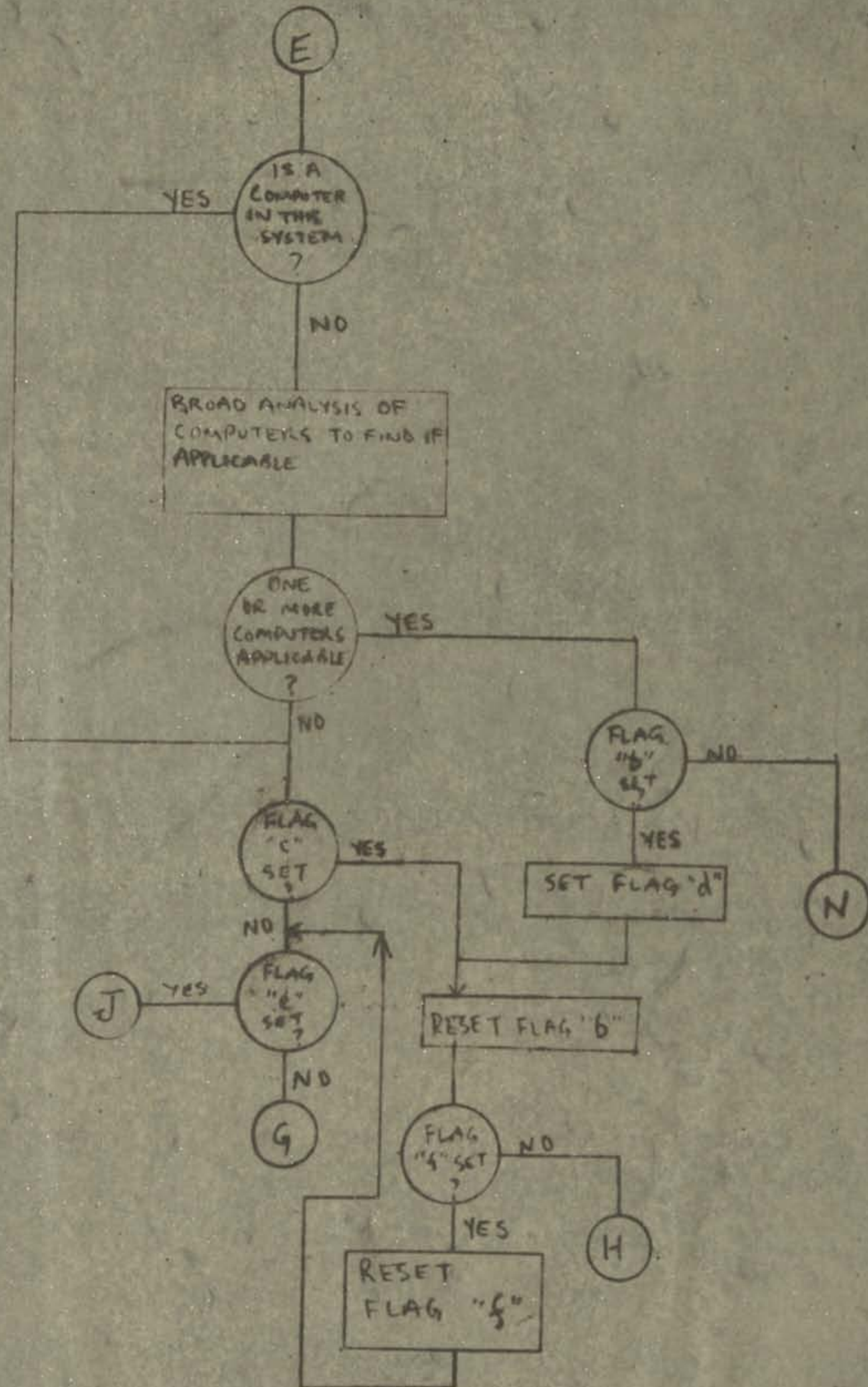

F
DETAILED ANALYSIS OF INPUT PARAMETERS OF A TO S
INPUT SUBSYSTEM PRACTICAL ?
NO
YES
I
B
M
DETAILED ANALYSIS OF PROCESSING REQUIREMENTS VERSUS SYSTEM
ALL REQUIREMENTS MET ?
NO
YES
J
FLAG & SET?
B
M
DETAILED ANALYSIS OF OUTPUT PARAMETERS OF S TO A
OUTPUT SUBSYSTEM PRACTICAL ?
NO
SET TR 5
YES
K
L

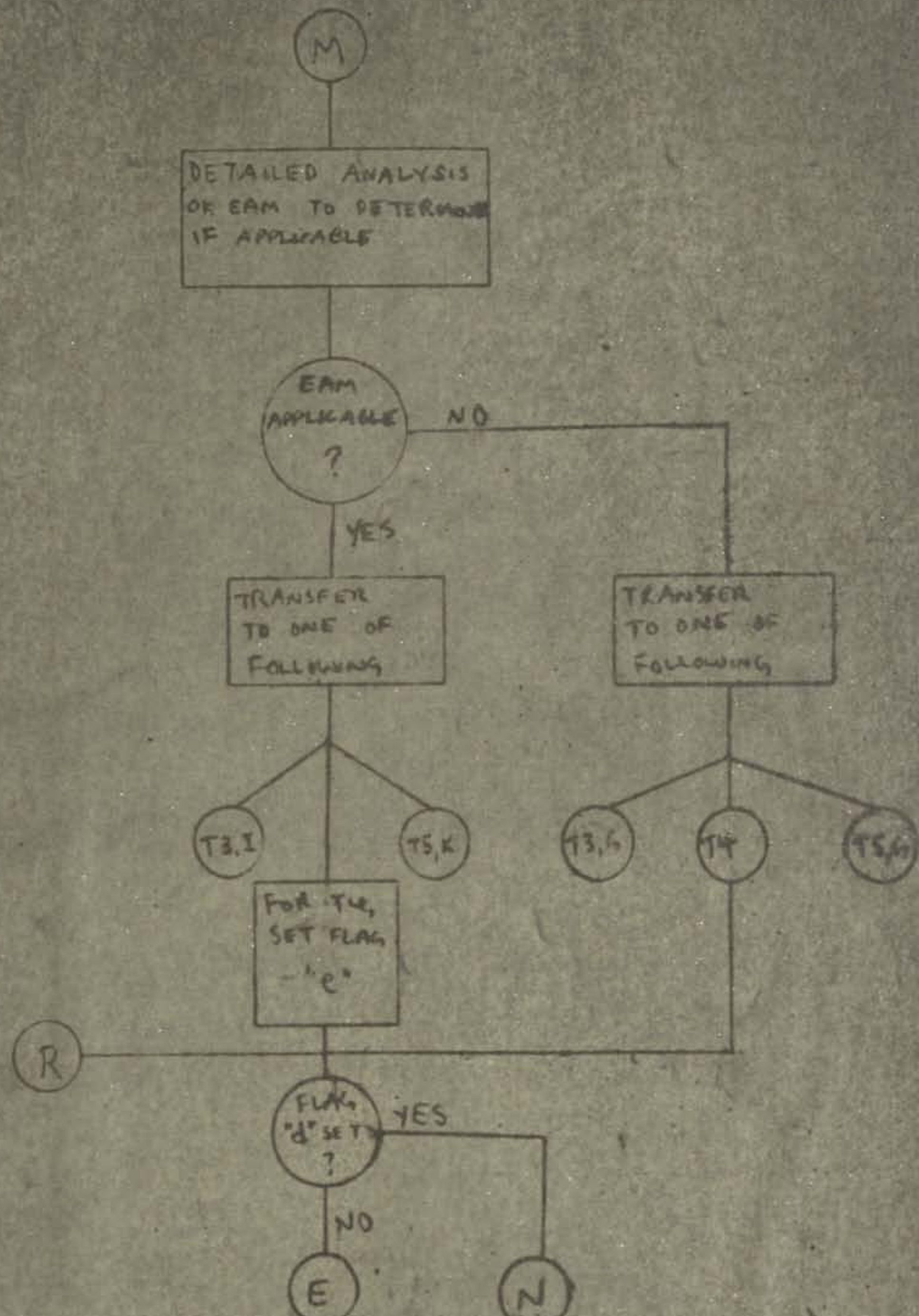

L
FINAL OUTPUT OF APPLICATION ?
NO
YES
PRINT OUT DETAILS OF GENERATED SYSTEM.
T
HOUSEKEEPING: RESET ALL FLAGS, ETC.
C
A
HOUSEKEEPING; REWIND SYSTEMS' TAPE, PRINT SUMMARIES, ETC.
END

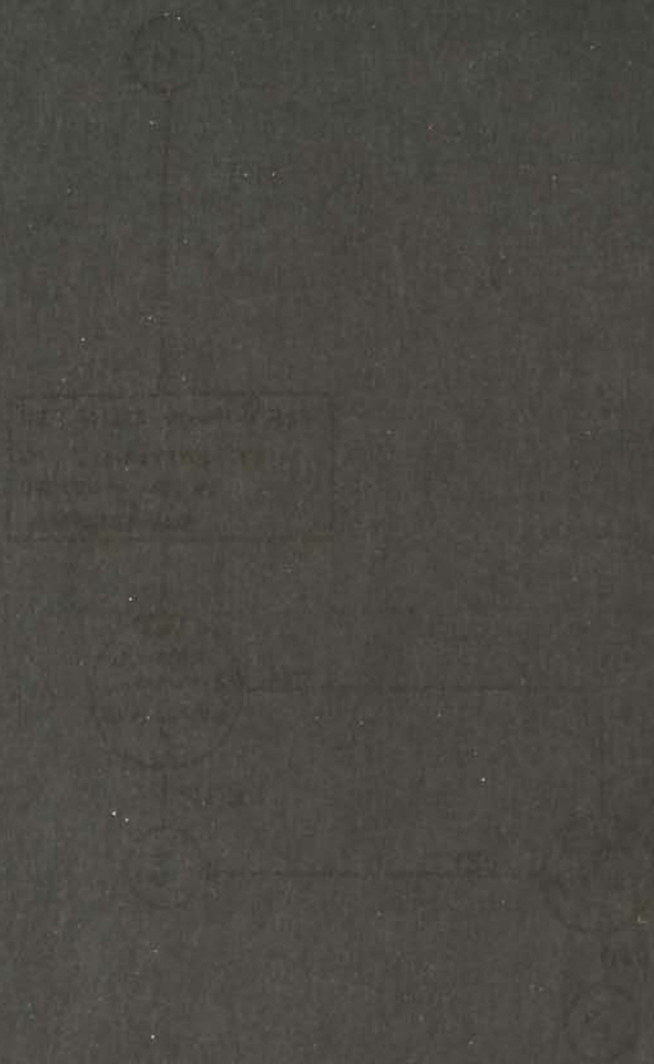

B
BROAD ANALYSIS OF EAM TO FIND IF APPLICABLE
EAM APPLICABLE ?
NO
YES
TRANSFER TO ONE OF FOLLOWING
TRANSFER TO ONE OF FOLLOWING
T1, D AND SET FLAG
T2, E AND SET FLAG
T3,M
T4,M
T5,M
T1,G
T2,E
T3,G
T4,R
T5,G


G
PRINT OUT OF WHY THIS SYSTEM IS NOT APPLICABLE
C

⑤


E
IS A COMPUTER IN THIS SYSTEM ?
YES
NO
BROAD ANALYSIS OF COMPUTERS TO FIND IF APPLICABLE
ONE OR MORE COMPUTERS APPLICABLE ?
YES
NO
FLAG "b" SET
NO
YES
SET FLAG "d"
N
FLAG "C" SET ?
YES
NO
FLAG "e" SET ?
YES
J
NO
G
RESET FLAG "b"
FLAG "f" SET ?
NO
H
YES
RESET FLAG "f"


M
DETAILED ANALYSIS OF EAM TO DETERMINE IF APPLICABLE
EAM APPLICABLE ?
NO
YES
TRANSFER TO ONE OF FOLLOWING
TRANSFER TO ONE OF FOLLOWING
T3,I
T5,K
T3,G
T4
T5,G
FOR T4, SET FLAG - "e"
R
FLAG "d" SET ?
YES
NO
E
N



FLAGS

FLAG	SET IF
a	Broad analysis of EAM examined and found O.K. for I/O interfaces
b	Transfer from "Set TR 2" (Reset prior to "F")
c	Broad analysis of EAM examined and found O.K. for broad processing requirements
d	Broad analysis of computers examined and found O.K. for broad processing requirements
e	Detailed analysis of EAM examined and found O.K. for detailed processing requirements
f	Transfer from "Set TR 4" (Reset prior to "J")

PRELIMINARY LIST OF PARAMETERS FOR SYSTEMS AND APPLICATION

A. INPUT PARAMETERS

1. MEDIA (System-S, Application-A)

   1.1 Tape: paper, magnetic, video, digital-video (see note at end of this Sect.)
   1.2 Cards: punched, notched, aperture, mag. (digital or video), graphic, drilled.
   1.3 Paper: documents, print, written
   1.4 Voice
   1.5 Magnetic: core, disc, drum
   1.6 Camera: video, optical
   1.7 Keyboard
   1.8 Photo
   1.9 Film chip
   1.10 Microfilm
   1.11 Console

2. CODING (S,A)

   2.1 Alphanumeric
   2.2 Binary (include type-as excess 3)
   2.3 Binary coded Decimal
   2.4 Hollerith
   2.5 Digital
   2.6 Graphic

3. RATE PER (S, Equipment Info. Unit; A, Input Item-In Units of #8)

   3.1 Second
   3.2 Minute
   3.3 Hour
   3.4 Day

4. TIME (A, for Total of Input Items)

   4.1 Seconds
   4.2 Minutes
   4.3 Hours
   4.4 Days

5. STORAGE CAPACITY (S, Equipment - In LCD)

   5.1 Per Unit
   5.2 Total

6. FORMAT (A)

   6.1 Variable field (limits, if any)
   6.2 Fixed field (length)

7. <u>VOLUME (TOTAL)</u> (A) = Storage Requirement

7.1 Documents
7.2 Pages
7.3 Items
7.4 Words
7.5 Characters
7.6 Bits

8. <u>LENGTH (A)</u>

8.1 Document (# of pages)
8.2 Pages (# of words)
8.3 Item (# of characters)
8.4 Word (# of characters)
8.5 Character (# of bits)

9. <u>INFORMATION FLOW MODE</u> (A, For Total Input)

9.1 Continuous
9.2 Intermittent and uniform
9.3 Random

10. <u>INFORMATION BASIC UNIT</u> (S,A)

10.1 Page
10.2 Item
10.3 Word
10.4 Character
10.5 Bit

NOTE: The items in each subclass would also be numbered, for example:

1.1 Tape: (1) paper, (2) magnetic, (3) video, etc.
This numbering system will be used for the identification, computer input, and use of system characteristics and application requirements. An example appears at the end of this paper.

B. <u>PROCESSING PARAMETERS</u>

1. <u>FUNCTIONS</u> * (A; S, All Those Applicable)

1.1 Character code translation (see A2 and C2)
1.2 Word code translation (identifier ⇄ word)

*Include whether manual or automatic

1.3 Indexing (include characteristics as # of levels, etc.)
1.4 Sort: class, numeric
1.5 Merge: class, numeric
1.6 Store: record, file, insert, buffer
1.7 Search: select, batch, compare, scan
1.8 Select: abstract, erase, sample
1.9 Batch
1.10 Examine (for decisions, checking)
1.11 Optical reduction
1.12 Exchange*
1.13 Duplicate
1.14 Collect
1.15 Distribute

2. STORAGE MEDIA (A, used during process, other than I-O-see A1 & C1;
(S, All Those Applicable)

2.1 Magnetic: core, tape, disc, cards, drum
2.2 Photoscopic
2.3 Register

3. STORAGE CAPACITY (S, of subsystem - in its units - See #6)

3.1 Per Unit
3.2 Total

4. STORAGE REQUIREMENTS (A, for process, in units of Subsystem, See #6)

4.1 Data
4.2 Program Instructions
4.3 Working Area

5. TIMING OF PROCESS (See B1, B2) DO DURING ANALYSIS

5.1 Seconds
5.2 Minutes

NOTE: Compare I-O rates where applicable in function or subfunction to determine if buffering is required.

6. STORAGE INFO. UNIT (S, include length in a lower term**)

6.1 Page
6.2 Item
6.3 Word
6.4 Character
6.5 Bit

* As tape reels off and on machine (manual)

** Example: 64, 65,6 = 6 bit char. : for RCA 301 HSM

C. OUTPUT PARAMETERS

1. MEDIA (S,A)

   1.1 Tape: paper, magnetic, video, digital-video
   1.2 Cards: punched, notched, mag. (digital or video), graphic, aperture, drilled
   1.3 Paper: print, written
   1.4 Microfilm
   1.5 Magnetic: core, disc, drum
   1.6 Display: tube, board, screen
   1.7 Camera
   1.8 Photo
   1.9 Film chip
   1.10 Console
   1.11 Plotter

2. CODING (S,A)

   2.1 Alphanumeric
   2.2 Binary
   2.3 BCD
   2.4 Hollerith
   2.5 Digital
   2.6 Graphic

3. RATE (S, Equipment Info. Unit; A, Output Item-in Units of #8)

   3.1 Second
   3.2 Minute
   3.3 Hour
   3.4 Day

4. TIME (A, for Total of Output Items)

   4.1 Seconds
   4.2 Minutes
   4.3 Hours
   4.4 Days

5. STORAGE CAPACITY (S, Equipment-in LCD)

   5.1 Per Unit
   5.2 Total

6. FORMAT (A)

   6.1 Variable field (limits, if any)
   6.2 Fixed field (length)

C. <u>OUTPUT PARAMETERS</u> (Cont.)

7. <u>VOLUME (TOTAL)</u> (A) = Storage Requirement

7.1
7.2 Pages
7.3 Items
7.4 Words
7.5 Characters
7.6 Bits

8. <u>Length (A)</u>

8.1 Report (# of pages)
8.2 Page (# of words)
8.3 Item (# of chars.)
8.4 Word (# of chars.)
8.5 Character (# of bits)

9. <u>INFORMATION FLOW MODE</u> (A, for Total Output)

9.1 Continuous
9.2 Intermittent and uniform
9.3 Random

10. <u>INFORMATION BASIC UNIT</u> (S, A)

10.1 Page
10.2 Item
10.3 Word
10.4 Character
10.5 Bit

D. <u>ANCILLARY PARAMETERS</u>

1. <u>Cost</u>

1.1 System (to buy and/or rent)
1.2 To run application (based on 1.1 and/or buying time)

2. <u>Availability</u>

2.1 If system bought or leased
2.2 If time bought on another's system

3. <u>Size of system Components</u>

NOTE: Required only if 1.1 and 2.1 apply

Example of how the input parameters would be written for a specific application:

(I 112; 221; 31,750; 41,0.1; 62; 73,75; 83,10; 91; 104)

This code represents the following information in serial order;

Start, input parameters, magnetic tape, binary excess 3 code, 750 items per second (input rate), 0.1 second to read input, input in fixed field format, total volume is 75 items, length of item is ten characters, data flow is continuous, the basic unit is a character.

ANALYSIS OF MANPOWER FOR APPLICATION ANALYSIS

FOR IS&R SYSTEMS

MAN MONTHS E, C*	AREAS OF WORK (Not all items are in chronological order)
	A. Systems
1.0, 0.2	1. Selection of systems and equipment (IS&R, EAM & Computers)
0.7, 0.1	2. Generation and structuring of system parameters.
3.0, 0.5	3. Gathering of parameter information for systems and equipment(see 1)
1.0, 0.5	4. Preparation of systems' parameters for computer input
5.7, 1.3	
	B. Applications
0.5, 0.1	1. Select types of applications (& a few test cases)
0.5, 0.1	2. Generation & Structuring of application parameters
0.5, 0.1	3. Gathering of parameter information for application test cases (See 1)
0.2, 0.1	4. Preparation of application parameters for computer input.
1.7, 0.4	
	C. Analysis (include flow diagrams in all cases)
4.5, 0.5	1. Determine method of application analysis
0.5, 0.1	2. Computer functional analysis (equipment)
4.3, 0.5	3. Computer program analysis (file generation, I-O, processing)
9.3, 1.1	D. Computer Program
4.0, 0.5	1. Preparation (in problem oriented language)
2.5, 0.2	2. Input, compilation, and debugging.
0.3, 0.1	3. Generation of systems' file in storage
3.5, 0.5	4. Run test applications and modify analysis and program (system checkout)
10.3, 1.3	
1.0, 0.2	E. Write Report
2.0, 0.7	F. Contingent Manpower
30.0, 5.0	TOTAL MAN-MONTHS
15 MONTHS	MINIMUM ELAPSED TIME
Months	G. Equipment Rental
1.00	1. EAM (various components)
0.25	2. Computer

* E = System's Engineer
C = Typist, EAM and Computer Operators

RCA Strengths

1. Equipment

RADE – [Jan '64 avail prototype] image, digital, video (should receive greatest emphasis) [is this really true? / nothing else competitive right now.]

ARCOM – [Jan '63 avail prototype]

Video-File (Primarily for Print Factories) – should try to get some operating experience with this system.

[what experience or background does RCA have in talking to printers?]

Content-addressed memories (MIRF)

2. Resources

Marketing [(card handling, microfilm, imagery, data proc)]

Engineering & Equip. Development

Software & problem orientation [(ASTIA employees]

3. Markets

Intelligence (estimated to be large, but undefined)

Info Centers

~~ASTIA~~
Army Records Center
Social Security Admin.
Defense Supply Agency (Bureau)
HEW - medical records
Weather Data Center

ASTIA
Patent Office
Medlars (NLM)
A.F. Satellite Weather Data (433L)
A.F Finance Center (best prospect for Video File)
[most typical candidates]

FBI
CIA

NAFC, AAFC
State Motor vehicle agencies
Commercial Concerns
Navy Printing Office
Library of Congress
OACSI
Army Engineers
Navy Security

all (what jobs? →
pre-emption of large systems.
do you know enough about each of these?
what are the stumbling blocks to implementation?

TO: The File January 19, 1961

FROM: Charles Bourne

SUBJECT: Notes of (3 January) Meeting (at SRI) with Representatives of RCA-Van Nuys

---

Mr. John J. Murphy, Manager of Systems Projects, and Mr. Irving Sperling, Senior Member of the Engineering Staff (both of RCA Data Systems Division, Van Nuys, California) talked with C. Bourne and R. Amara today about RCA activities and plans in the field of information retrieval. Discussions were also held recently with Mr. George Arnovick, Leader, Systems Engineering Staff (also from RCA-Van Nuys) to exchange information about our respective projects in this field. RCA is becoming increasingly active in this field and has plans for an appropriate line of products and services.

Mr. Murphy has requested consulting assistance from SRI to work with an RCA ad hoc committee on information retrieval to help establish some goals and guidelines for RCA's work in this field. The following preliminary work statement was developed during our discussions.

<u>Preliminary Work Statement--RCA-SRI Effort</u>

The SRI representatives will be given the following major tasks:

1. Become familiar with RCA products and capabilities pertinent to information retrieval. (RCA will provide the necessary briefings and documentation. Travel will probably be restricted to the Los Angeles area.)
2. Review and evaluate RCA's advance research and development (AR&D) program. (AR&D) is the internal research effort supported by RCA funds and currently consists of three major efforts--(a) a study of descriptors and list

structures for library classification systems; (b) the development of a mass memory--a magnetic card system with capability for storing 1 million bits per card with digital, video, or microfilm on each card; (c) the development of an automatic retrieval comparator ARCOM.

3. Participate in the formulation of AR&D plans for the next five years. (This will be done primarily by group discussions with the committee members.)
4. Assist in the preliminary definition of users' requirements for a number of problem areas. Review present RCA equipment and note its applicability to the various problem areas.
5. Using the analysis programs developed on SRI's recent NSF project, simulate the performance of about 5 or 6 of the most promising systems suggested by the committee studies, to determine their operating costs.

Informal written reports are to be prepared for each of the major tasks. The effort would start as soon as possible, preferably with 20 to 30 percent of C. Bourne's time over a six calendar month period. A possible alternative is a two-month effort at 90%, followed by four months at 10%.

One of the most attractive features of this potential work is the opportunity to extend and further test some of the programs and simulation procedures developed during the NSF project. This work for RCA would have a useful research content that would be useful in subsequent dealings with NSF or other organizations.

Mr. Murphy expects to return to his office in about a week, at which time he will send us a more complete work statement to be used as a basis for our proposal preparation.

CB/rt

cc: Noe/Wing
D. Cone
R. Amara
C. Bourne ✓
D. Ford

TP 1125

J. C. FLINT

RACE PRESENTATION

April 1962

RCA
Data Systems Division
Defense Electronic Products
RADIO CORPORATION OF AMERICA
Van Nuys, California

TP 1125

RACE PRESENTATION

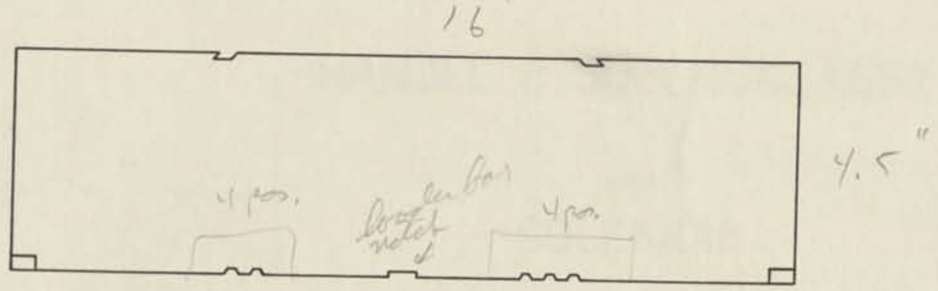
April 1962

Data Systems Division
Defense Electronic Products
RADIO CORPORATION OF AMERICA
Van Nuys, California

BASIC MEMORY COMPONENTS

- CARDS
- MAGAZINE (REPLACEABLE)
- CARD SELECTION MECHANISM
- TRANSPORT & GATES
- READ/WRITE STATION (DIGITAL)
- READ/WRITE STATION (VIDEO)
- RE-INSERTION MECHANISM
- CONTROLS

INFORMATION STORAGE CARD



DIGITAL

TRACKS-96

CHARACTER-6 BITS

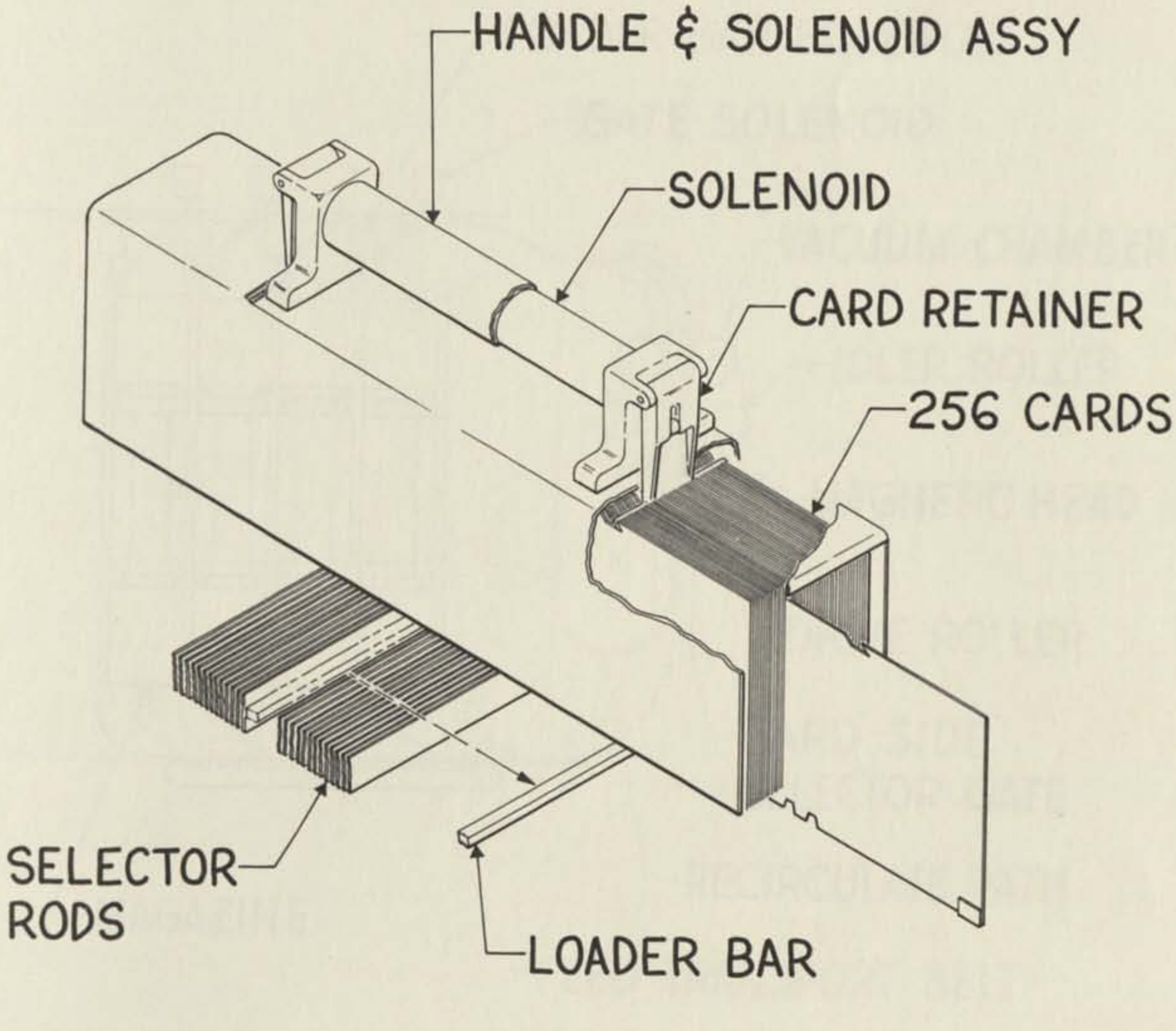
BIT DENSITY-350 PER INCH

CAPACITY-940 X 10^3 BITS

VIDEO

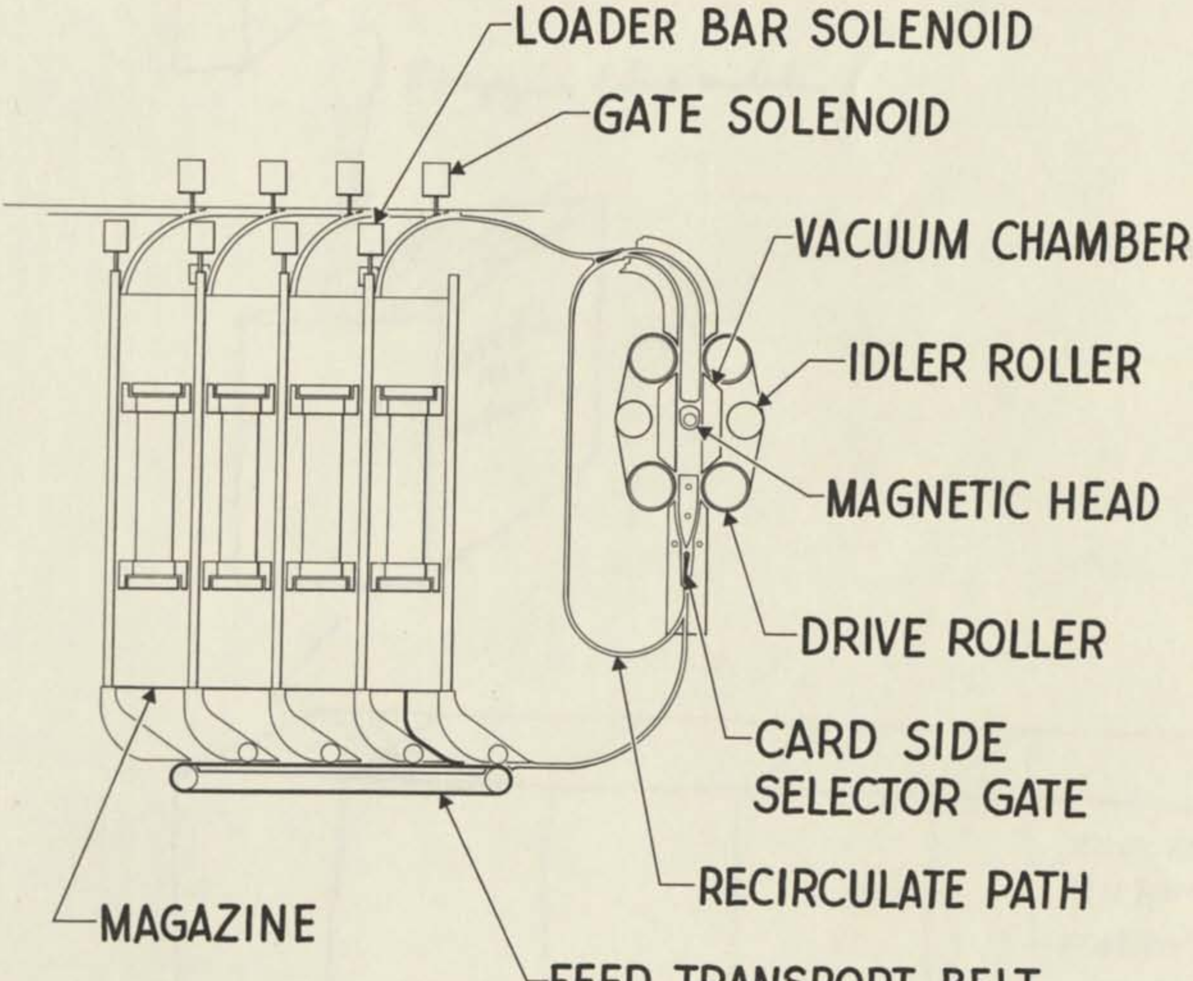
DOCUMENTS-32 (8½" X 11") PER CARD

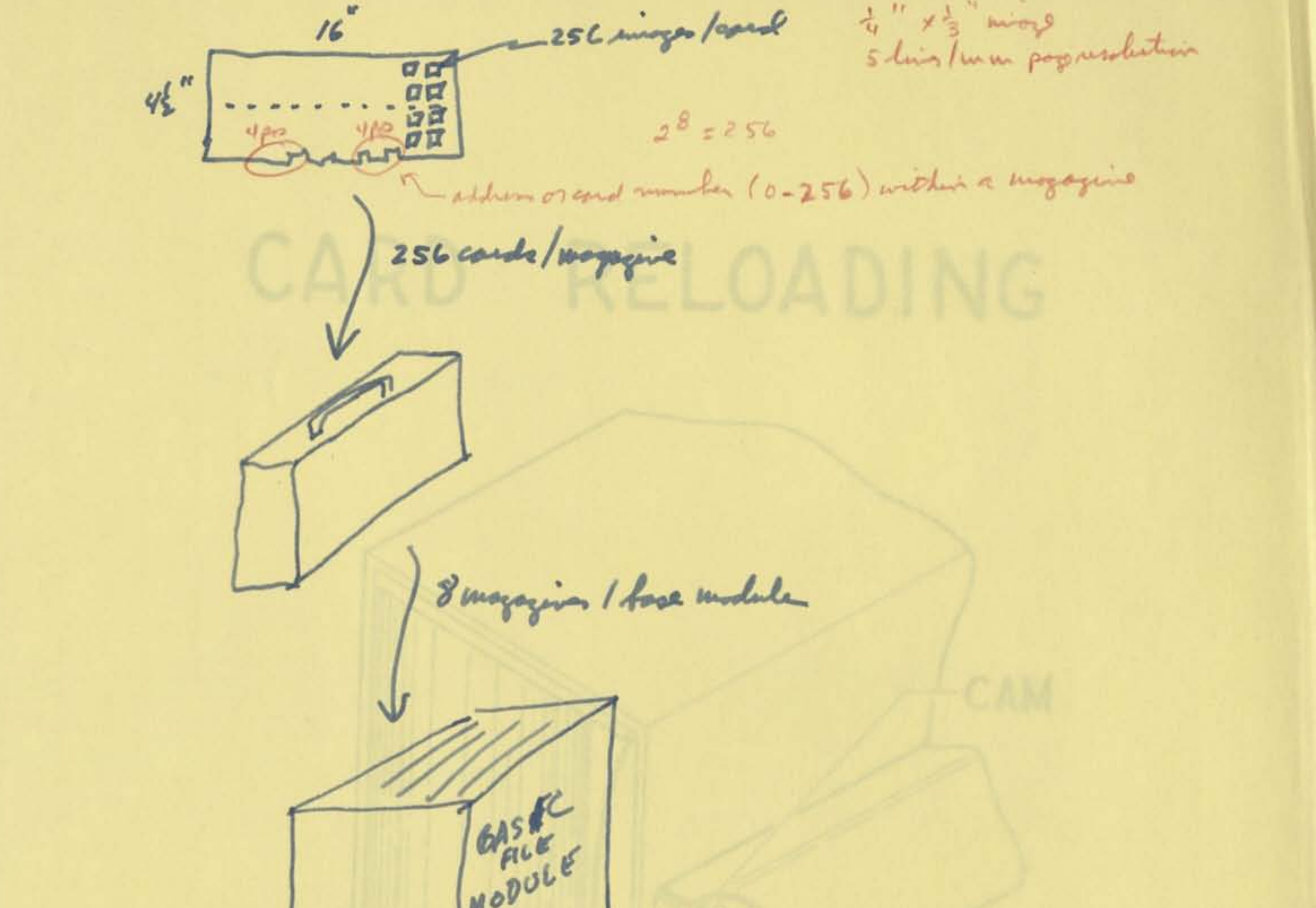
MAGAZINE


HANDLE & SOLENOID ASSY
SOLENOID
CARD RETAINER
256 CARDS
SELECTOR RODS
LOADER BAR

RANDOM ACCESS MASS MEMORY

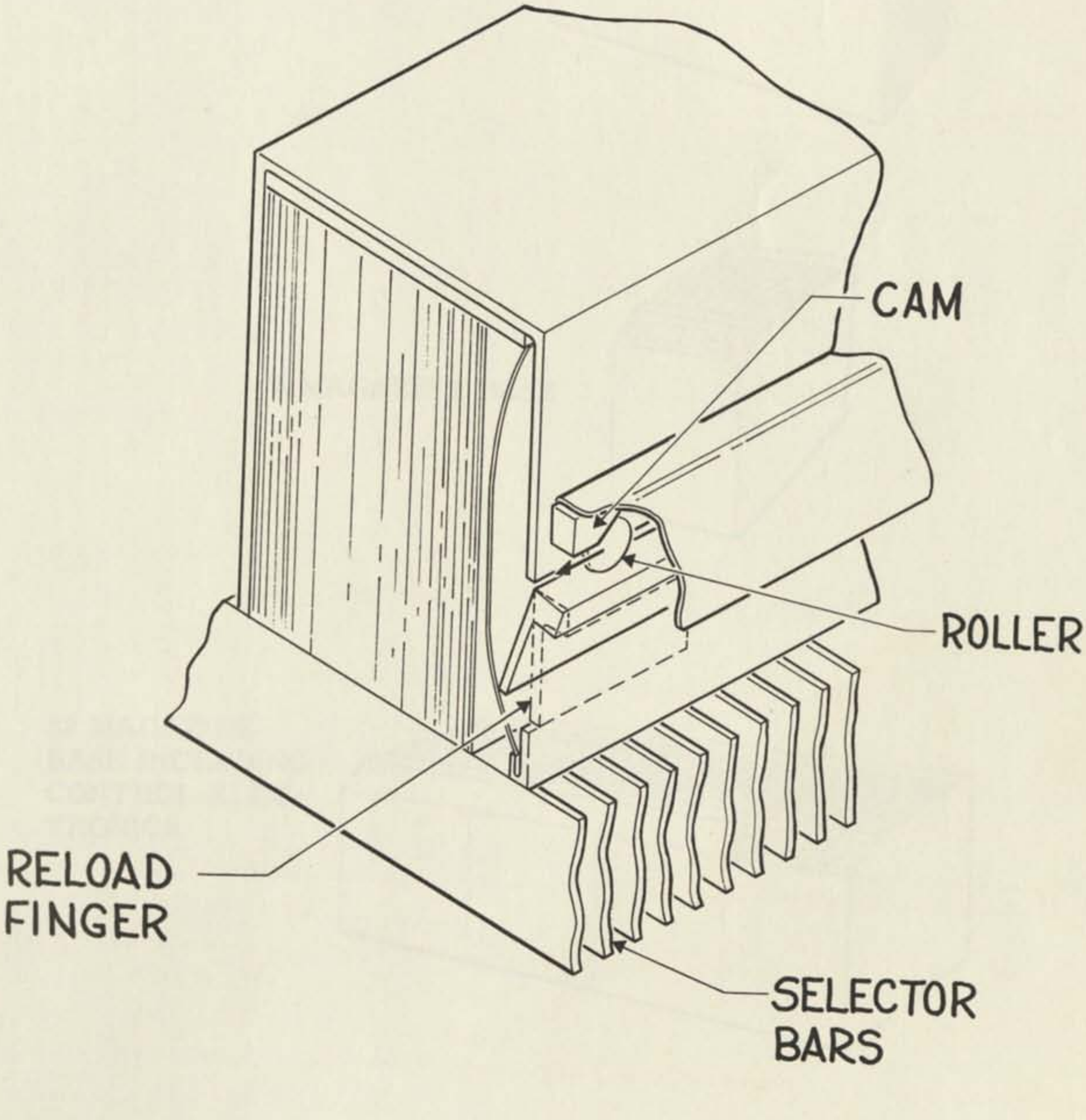
PLAN VIEW

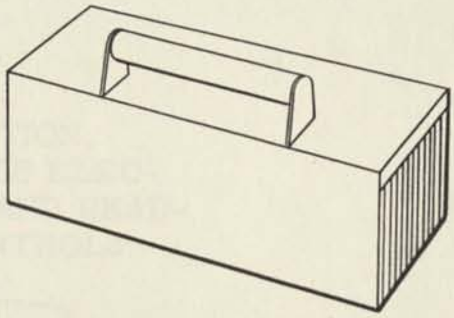




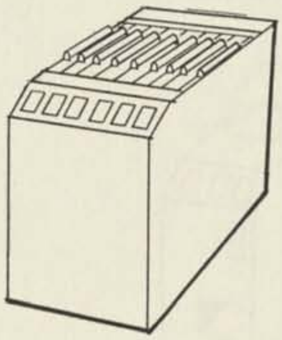
32-MAGAZINE FILE
(may have up to 128 magazines)

CARD RELOADING

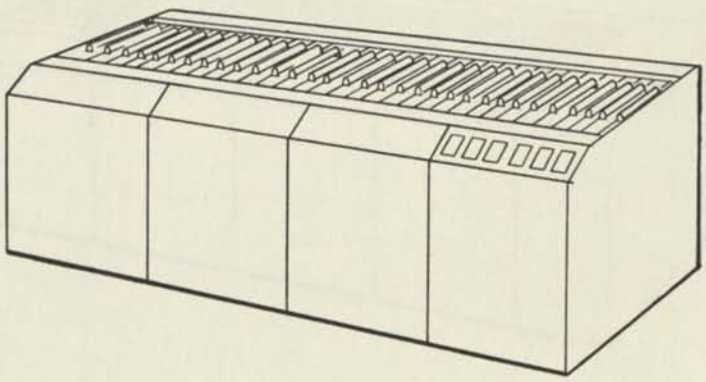




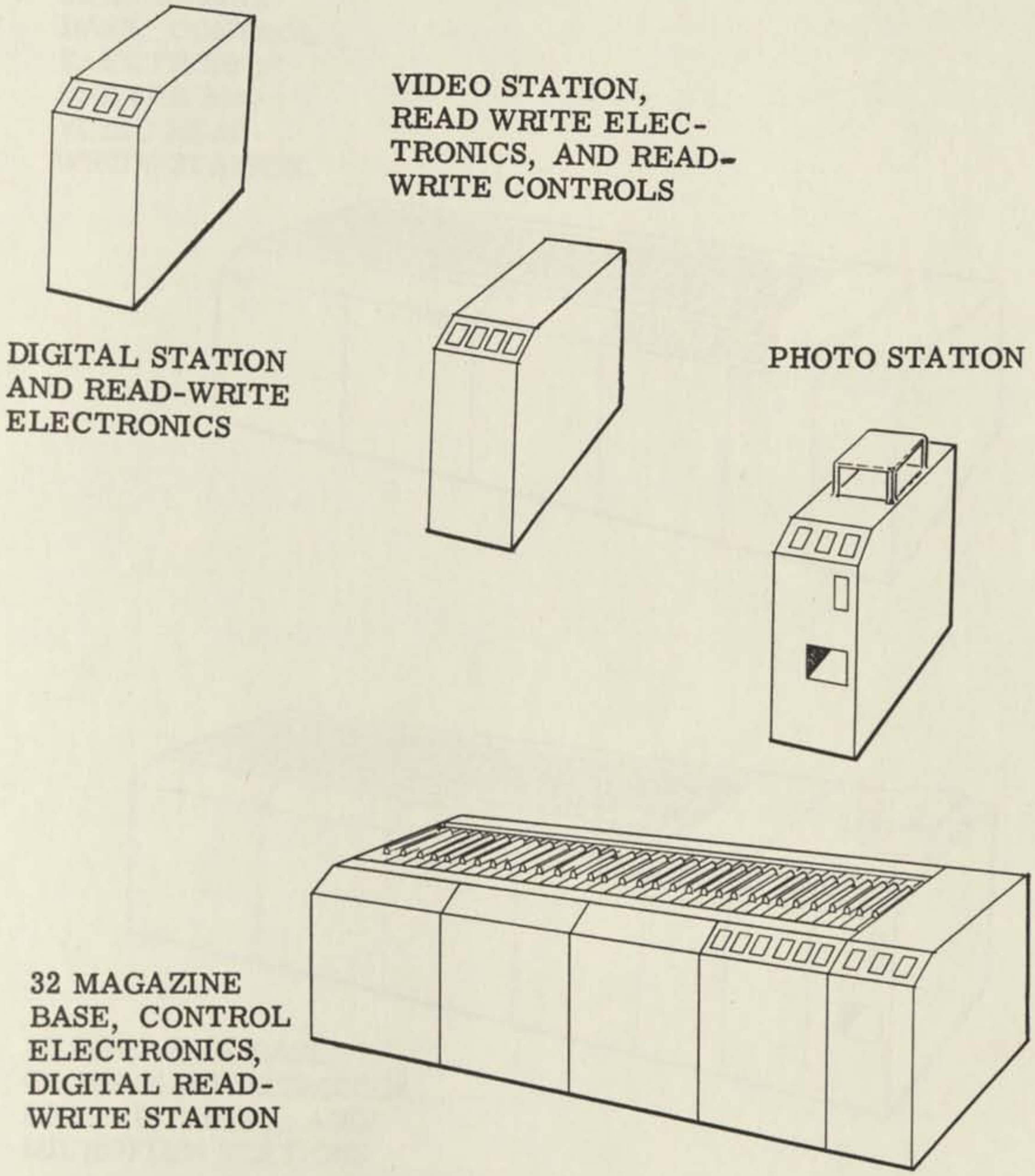
EXTRA MAGAZINE
WITH PRECODED CARDS



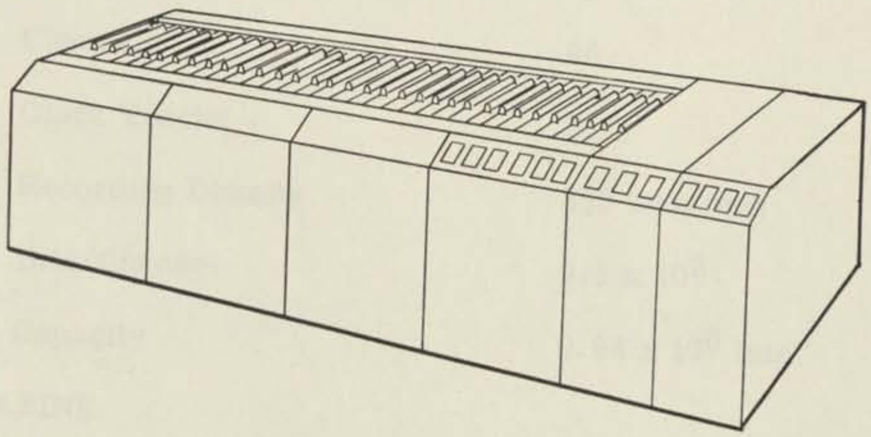
8 MAGAZINE BASE

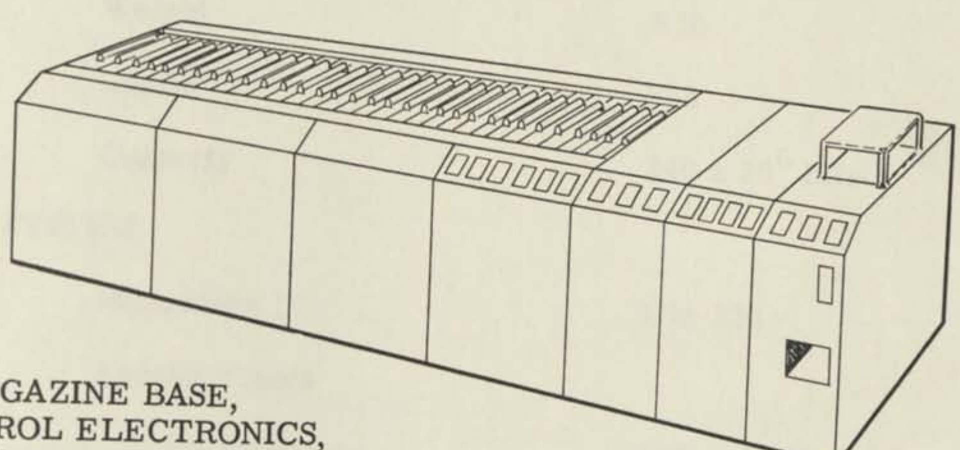


32 MAGAZINE
BASE INCLUDING
CONTROL ELEC-
TRONICS


VIDEO STATION, READ WRITE ELEC- TRONICS, AND READ- WRITE CONTROLS
DIGITAL STATION AND READ-WRITE ELECTRONICS
PHOTO STATION
32 MAGAZINE BASE, CONTROL ELECTRONICS, DIGITAL READ- WRITE STATION

32 MAGAZINE BASE, CONTROL ELECTRONICS, DIGITAL AND VIDEO READ-WRITE STATION





32 MAGAZINE BASE, CONTROL ELECTRONICS, DIGITAL, VIDEO, AND MICROFILM STATIONS

RACE CHARACTERISTICS

DIGITAL	
CARD	
Size	4-1/2" x 16"
Channels	96
Clock Tracks	2
Recording Density	350 bits/inch
Bits/Channel	9.8×10^3
Capacity	0.94×10^6 bits
MAGAZINE	
Size	16" x 5" x 2-1/4"
Weight	5 lb
Cards $2^6=64$ $2^7=128$	256 $= 2^8$
Capacity	240×10^6 bits
SYSTEM	
Magazines	1 to 128
Access Times	
1st Magazine	180 msec
32nd Magazine	450 msec
Read/Write Rate	140 kc

from where to where? from magazine 128 to the print station?

RACE CHARACTERISTICS (Cont)

32 MAGAZINE SYSTEM

Cards	8,192
Capacity	7.7 x 10^9 bits
Size	96" x 36" x 26"
Weight	1,250 lb

128 MAGAZINE SYSTEM

Cards	32,768
Capacity	30.8 x 10^9 bits
Size	192" x 36" x 52"
Weight	4,250 lb

VIDEO

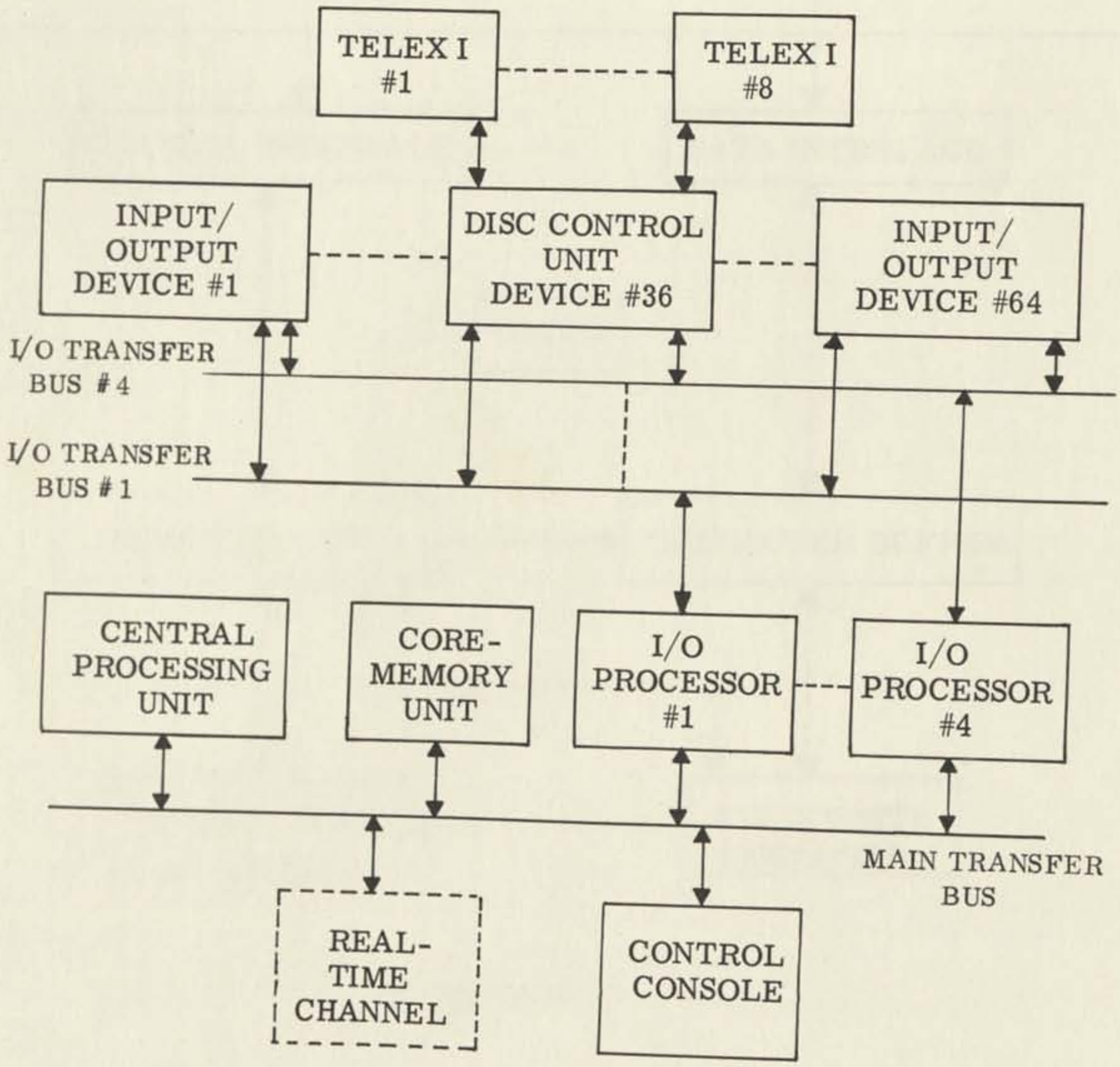
DOCUMENT SIZE	8-1/2" x 11" sheet
Per Card	32 documents
Per Magazine	8,192 documents
Per 32 Magazine System	262,144 documents
Per 128 Magazine System	1,048,576 documents
VIDEO RATE	2.7 documents/sec
RESOLUTION	125 TV lines/inch

RACE CHARACTERISTICS (Cont)

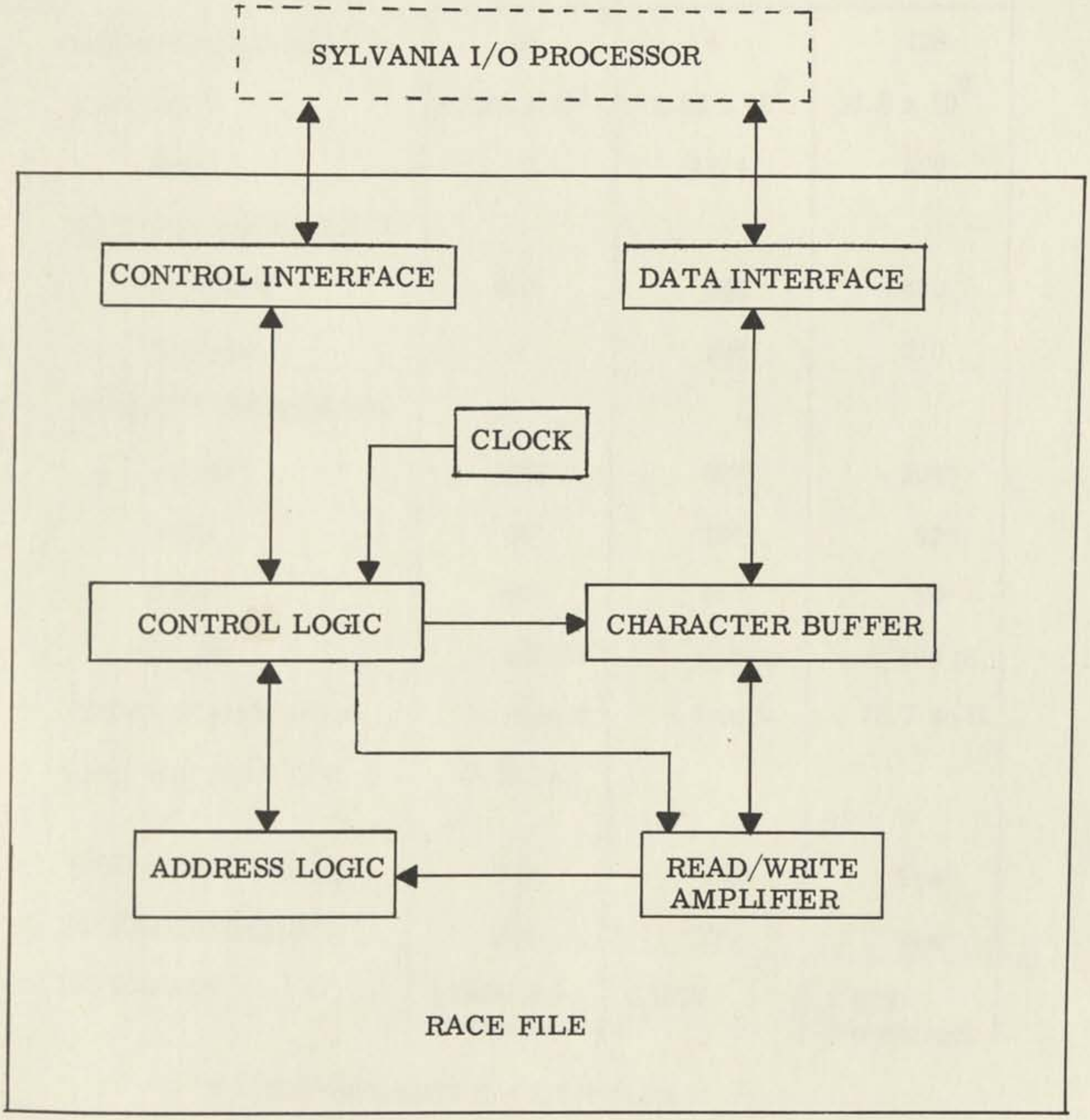
MICROFILM

IMAGES/CARD	350 maximum 250 suggested
DOCUMENT SIZE	8-1/2" x 11"
RESOLUTION (FULL PAGE DISPLAY)	5 lines/millimeter
FRAMES/MAGAZINE	64×10^3
FRAMES/128 MAGAZINE SYSTEM	8.192×10^6
IMAGE SIZE ON CARD	1/4" x 1/3"

SYLVANIA DATA PROCESSING SYSTEM BLOCK DIAGRAM



RACE BLOCK DIAGRAM



COMPARISON OF TELEX I AND RACE

	TELEX I	RACE	
DISCS/MAGAZINES	16	8	128
CAPACITY	0.155×10^9	*1.92×10^9	30.8×10^9
Ratio	1	12.4	200
ACCESS TIME (MSEC) #			
Maximum	400	240	450
Average	?	206	310
PHYSICAL DIMENSIONS			
Length	84"	36"	204"
Width	36"	26"	52"
Height	60"	36"	36"
Weight	2,500 lb	750 lb	4,500 lb
TOTAL FLOOR AREA	21.0 sq ft	6.5 sq ft	73.7 sq ft
COST WITHOUT DCU	$ 105 K		
VIDEO CAPABILITY	No	Yes	Yes
PHOTO CAPABILITY	No	Yes	Yes
INTERFACE	1 DCU per 8 Telex I	1 RCU	1 RCU (No change)

* 1 RACE MAGAZINE = 1.5 TELEX I

Without Prepositioning

ADVANTAGES OF RACE

- Extremely Large and Expandable Storage Capacity
- Equivalent or Better Access Time as Telex I
- Less Floor Area
- Cheaper Per Unit Storage
- Lighter in Weight
- Capable of Storing

  Digital
  Video
  Microfilm

- Combinations of Digital and Video or Microfilm
- Modular Expandability in Capacity
- Magazines Replaced Easily and Fast
- Modular Expandability in Read/Write Heads (Ability to read or write from 1 to 96 channels simultaneously)
- Simpler Interface Requirements than Telex
- Future Enhancements

  Increased Bit Density
  Increased Transfer Rate
  Ruggedized Portable Units