

# MECHANIZED SEARCHING OF PHOSPHORUS COMPOUNDS. III. PUNCHED CARDS vs. RANDOM ACCESS COMPUTERS\*

By J. FROME, F. M. SIKORA, and M. GANNON

U. S. Department of Commerce, Patent Office, Washington, D. C.

During the past six months patent applications for organic phosphorus ester compounds have been mechanically searched by the two separate systems previously described. To date about 200 searches on the RAMAC and 102 searches on the multicolumn have been made on a file of 693 patents. While it is realized that this may possibly be too small a sample on which to base ultimate conclusions, certain trends are unmistakable and appear worthy of mention at this time.

The Office of Research and Development in the U. S. Patent Office is charged not only with investigating new methods for mechanizing the patent search for patent examiners, but also with making these new methods and their search files available to the public when practicable.

Therefore two search systems were developed and investigated. The two systems are, however, now compared in this report for the purpose of determining which system is better on an over-all basis. It is felt that this ultimate decision depends entirely upon the needs of the user. The two systems are compared merely to disclose the advantages and disadvantages of each, and also to reveal some idea of how pronounced these advantages and disadvantages actually are.

In order to judge a mechanized search system several important criteria are necessary: (1) cost, as to (a) file preparation, (b) machine rental, (c) machine operation; and (2) effectiveness, as to, (a) number of descriptors, (b) number of documents retrieved, (c) noise. (Noise: refers to those documents retrieved in a mechanized search which do not answer the question asked, on account of some peculiar feature of the system.)

The cost of mechanizing any specific body of chemical subject matter as well as the cost of the machine are the two most important factors to be considered in any mechanized search system.

A comparison between the cost of the two systems may best be illustrated by a table.

TABLE I

System	RAMP (Computer)	CAMP (Punched Card)
Machine used	RAMAC	Multicolumn sorter
Machine rental cost ratio	9	1
File cost ratio (cost not common to each)	2½	1
Operating personnel cost ratio	1 1/3	1
No. of cards used	100	1

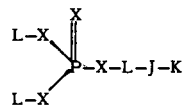
It is observed from this table that the computer rental is about nine times greater than that of the punch card machine. This is without doubt the most significant single cost factor at the present time.

File cost preparation in the computer system is also 2.5 times greater. This does not take into consideration the additional cost which is common to both, i.e., writing structural formulas.

Because of the fact that the ordinary punch card machine is much less complicated to operate, personnel with a minimum of machine training are used, whereas a school trained technician is required to operate the larger more complex electronic computers.

The other very important factor to be considered is the effectiveness of the system. The searches described are examples which illustrate the varying features of each system in terms of its effectiveness. The search questions have been purposely disguised so as not to reveal the subject matter of the patent applications involved.

(A) Actual Search:

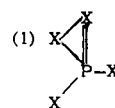


Questions Asked (RAMP)

(1) J-CH-1  
J-R-1

Questions Asked (CAMP)

Specifically



(1) Phosphorus nucleus

X<sub>3</sub>-L I node

X<sub>1</sub>-J II node

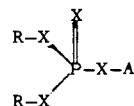
X<sub>1</sub>-K Term. node

RAMP

CAMP

1. Time to ask question	4 min.	3 min.
2. Time of search	3 min.	2 min.
3. No. of documents retrieved	19	12

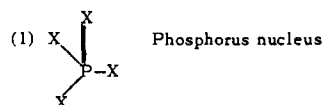
(B) Actual Search:



\*Presented before Division of Chemical Literature, ACS National Meeting, New York, September 13, 1960.

Questions Asked (RAMP)

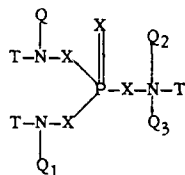
(1) P-X-A-1

Questions Asked (CAMP)X<sub>2</sub>-R 1 node

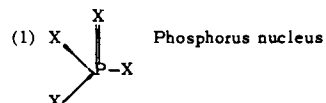
X-A 1 node

(Specifically)

	RAMP	CAMP
1. Time to ask question	4 min.	3 min.
2. Time of search	3 min.	2 min.
3. No. of documents retrieved	12	48

(C) Actual Search:Questions Asked (RAMP)

(1) T-CH-3  
P-X-N-3  
13 Frag

Questions Asked (CAMP)X<sub>3</sub>-N 1 nodeX<sub>3</sub>-QX<sub>3</sub>-T 11 node

	RAMP	CAMP
1. Time to ask question	4 min.	3 min.
2. Time of search	3 min.	2 min.
3. No. of documents retrieved	13	13

It will be observed that in the more complex ester type compound (A) searched, the CAMP system was more effective for the relatively same period of searching time in that it retrieved only 12 documents in comparison to the 19 which were retrieved in the RAMP system.

However, in the more simple type of ester compound (B) the situation was reversed with the RAMP system retrieving 12 documents in comparison to the 48 documents found by the CAMP system for a relatively similar time of search.

The intermediate type compound (C) gave exactly equivalent results in terms of number of documents retrieved over a similar period of time.

Table II discloses some over-all average results based on the 302 searches made in the 52 patent applications.

It is observed that for approximately equivalent times of search (2 minutes on the CAMP system and 3-4 minutes on the RAMP system)

TABLE II

	RAMP	CAMP
No. of documents retrieved/search	10	17
Noise	Little	Appreciable
No. of descriptors possible	50,000	960
Degree of specificity and genericity	Great	Limited

the RAMP system retrieved 10 patents per question while the CAMP system retrieved a surprisingly comparable 17.

The answers to the RAMP system also revealed very little noise. By "noise" we mean those documents retrieved in a mechanized search which do not answer the question asked, on account of some peculiar feature of the system.

The answers in the CAMP system, however, did reveal a considerable amount of noise, undoubtedly as a result of the compositing feature of putting all information describing a total of 25 compounds (average) on one punch card.

Table II also points out the difference in number of descriptors which can be used to describe the organic phosphorus compounds in each system.

For example the 50,000 descriptors in the RAMP system enable us to define each alkyl radical in a compound specifically in terms of whether it is an ethyl or a butyl radical, also whether it is straight or branched chain, and finally whether it is alkyl. In the CAMP system the 960 descriptors limit us to define the above radicals in terms of a generic descriptor "alkyl" only.

Similarly the radical "bicycloheptane" in the RAMP system is defined both as bicycloheptane and cycloalkyl whereas in the CAMP system it is defined only as cycloalkyl.

It is thus seen that the RAMP system, because of its degree of specificity, is particularly outstanding for finding relatively simple type specific organic phosphorus compounds such as triethyl phosphate, tricresyl phosphate, etc., and separating them from their homologs. Of course it also can retrieve complex compounds.

The results obtained from testing both systems at various stages of file size also forecast some future trends as illustrated in Table III.

TABLE III

	RAMP	CAMP
Future increase in file expected	Great	Great
Increased time to search	Little	Great
Memory	Limited	Unlimited
Increased noise	Little	Greater

A distinct advantage of the RAMP system is that as the organic phosphorus patent file

increases, the time of search in the RAMP system will increase only slightly, as a general rule. This is subject to the exception that when a descriptor common to a large percentage of all organic phosphorus compounds (P=O-l for example) is asked, the time of search will in such instances be considerably increased by the size of the file. However, by skillful manipulation in asking the questions the searcher can avoid this situation in 90% of his searches.

With regard to the CAMP system, as presently set up for phosphates and thiophosphates, the time of search is exactly proportional to the size of the file.

The present collection of organic phosphorus compound patents indicates that the total number of patents in this field has more than doubled in the past ten years. Thus we can expect that if the total number of patents again doubles in the next ten or fifteen years, the time of search will increase only slightly in the RAMP system while it is expected to double in the CAMP system.

Table II also discloses an advantage of the CAMP system over the RAMP system with respect to the memory feature.

In the RAMP system where we now use the RAMAC 305 the system is machine limited. We at present have the ability to store only 5 million alphanumeric characters. While 5 million characters sounds like a large number, we have discovered that this number is exceeded with relative ease when coding organic elements and radicals in the manner of the present system. However, if the memory size of the machines increases as expected, this limitation will be removed. Additional use of another random access machine with a larger memory will avoid this problem.

The CAMP system on the other hand is unlimited as to the number of documents it will hold. Since the CAMP system memory is in effect the number of punched cards used its memory is literally unlimited. Again this statement is subject to the proviso that one must understand that his time of search is exactly proportional to the size of his file.

A further advantage of the RAMP system is that as the size of its file grows, little increase in noise can be expected, while in the CAMP system the amount of noise shows a considerable increase as the file grows larger and more closely related compounds are patented.

Table IV summarizes the advantages of each system.

TABLE IV ADVANTAGES

<u>RAMP</u>	<u>CAMP</u>
Large number of descriptors	Large memory
Inverted file	Low cost
Fast	Fast
No real increase in time as file grows	
Varying scope unlimited	
Little noise	

The inverted file feature of the RAMAC is advantageous in that its time of search does not increase in direct proportion to the size of the file.

The 3-4 minute average search time for the RAMAC indicates that it is definitely fast enough to far exceed the speed of manual searching and, furthermore, that it will eventually exceed the speed of the CAMP system when a larger file is developed.

The fact that questions on the RAMP can be varied in scope from generic to specific and also the factor that a large number of questions can be asked of it quickly and efficiently, makes the RAMP system particularly advantageous for anyone interested in asking a great number of widely divergent questions. Patent Office use, for example, best illustrates the type of use which takes full advantage of the varying scope features of the RAMP system.

The outstanding feature of the CAMP system, however, without doubt is its low cost. First of all the low rental cost of the machine makes the system particularly attractive. Second it is expected that the Patent Office file of cards for this system eventually may be made available to the public.

A third outstanding feature of the CAMP system is its speed. This, of course, is best illustrated by the fact that at present it takes only two minutes to search a file of 693 patent documents. The speed of the system depends essentially on the speed of the particular punch card sorter used.

Table V summarizes the disadvantages of each system.

TABLE V DISADVANTAGES

<u>RAMP</u>	<u>CAMP</u>
High Cost	Fair amount of noise
Limited memory	Increase in time as file increases
	Limited dictionary

As previously stated the limited memory aspect poses a real problem for RAMP when it uses the RAMAC 305. Also its high cost may in many cases make the system prohibitive for certain users.

CAMP's disadvantages are not believed to be particularly serious at the present time but it is recognized nevertheless that they are factors to consider.

In conclusion it is believed that our analysis indicates that when cost is of extreme importance and the field of search is limited, the CAMP punch card system appears desirable.

In the final analysis, the user must consider the various factors involved, weigh them carefully, and then decide which system best suits his needs.