Computation Structures Group Memorandum Number 26 (REVISED)

MAC-M-403

EFFECTS OF SCHEDULING ON FILE MEMORY OPERATIONS*

by

PETER J. DENNING

Massachusetts Institute of Technology

Project MAC

---

* Presented at the 1967 Spring Joint Computer Conference, Atlantic City, New Jersey

EFFECTS OF SCHEDULING ON FILE MEMORY OPERATIONS*

by

Peter J. Denning

Massachusetts Institute of Technology

Cambridge, Massachusetts

---

*Work reported herein was supported in part by Project MAC, an M.I.T. research project sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract Number Nonr-4102(01).

INTRODUCTION

File system* activity is a prime factor affecting the throughput of any computing system, for the file memory is, in a very real sense, the heart of the system. No program can ever be executed without such secondary storage. It is here that input files are stored, that files resulting as the output of processes are written, that "scratchwork" required by operating processes may be placed. In a multiprogrammed computing system, there is a considerable burden on secondary storage, resulting both from constant traffic in and out of main memory (core memory), and from the large number of permanent files that may be in residence. It is clear that the demand placed on the file memory system is extraordinarily heavy; and it is essential that every part of the computing system interacting with file memory do so smoothly and efficiently.

---

*By "file memory" in this paper is meant secondary storage devices of fixed and movable head types. Drums are examples of fixed head devices, while both fixed and movable head disks exist. Tapes may be regarded as movable head devices.

In this paper we discuss some important aspects of file system organization relevant to maximizing the throughput, the average number of requests serviced per unit time. We investigate the proposition that the utilization of secondary memory systems can be significantly increased by scheduling requests so as to minimize mechanical access time. This proposition is investigated mathematically for secondary memory systems with fixed heads (such as drums) and for systems with moving heads (such as disks). In the case of fixed head drums the proposition is shown to be spectacularly true, with utilization

increases by factors of 20 easily attainable. For moving head devices such as disks this proposition is still true, but not dramatically so; optimistic predictions look for at best 40 per cent improvement, an improvement attained only with serious risk that some requests receive no service at all. The solution lies at an unsuspected place: a method of "scanning" the arm back and forth across the disk, which behaves statistically like the access-time-minimizer but provides far better response than first come first served policies.

Other analyses of drum and disk systems, but from different viewpoints, have appeared elsewhere[1,2,3]. The importance of the results obtained here is that they are relatively independent of the behavior of the processes generating requests.

THE MODEL OF THE FILE SYSTEM

For our purposes the computing system can be visualized as two basic entities, main memory and file memory. Only information residing in main memory can be processed. Since the cost of high-speed core memory is so high, there is at best limited availability of such memory space. A core allocation algorithm will be at work deciding which information is permitted to occupy main memory, and which must be returned to secondary memory; algorithms for doing so do not concern us here and are discussed elsewhere[4]. The essential point is that there may be considerable traffic between slow, secondary storage, and fast, core storage. System performance clearly depends on the efficiency with which these file memory operations take place. It is our concern here to discuss optimum scheduling policies for use by central file control.

We assume there is a basic unit of storage and transmission, called the page. Core memory is divided logically into blocks (pages); information is stored page by page on secondary storage devices; it follows that the unit of information transfer is the page. We will assume that requests for file system use are for single pages, and that a process which desires to transfer several pages will have to generate several page requests, one for each page. Single-page requests are desirable for two reasons. First, read requests will be reading pages from the disk or drum, pages which are not guaranteed to be stored consecutively; it is highly undesirable to force a data channel to be idle during the inter-page delays that might occur, when other requests might have been serviced in the meantime. And, secondly, the file system hardware is designed to handle pages.

By far the biggest obstacle to efficient file system operation is that these devices, mechanical in nature, are not random-access. This means that each request must experience a mechanical positioning delay before the actual information transfer can begin. It is instructive to follow a single request through the file system. The various quantities of interest are displayed in Figure 1 and defined below.

t_o - the transfer time; the time required to read or write one page from or to file memory.

a - the access time; the mechanical positioning delay, measured from the moment the request is chosen from the queue to the moment the page transfer begins.

t_s - the service time; $t_s = a + t_o$

W_q - the wait in queue; measured from the moment the request enters the queue to the moment it is chosen for service.

W_F - the total time a request spends in the file system.

If the device is a drum we may imagine that the switch arm is revolving at the same speed as the drum rotates, picking off the lead request in each queue as it sweeps by. This structure is equivalent to the following policy: choose as next for service that request which requires the shortest access time; this policy will be called the shortest access time first (SATF) policy. It should be clear that the FCFS policy is less efficient because it involves longer access times per request. It should also be clear that SATF does not discriminate unjustly against any request.

Now suppose that secondary memory is a disk. Due to physical constraints, it is not possible to move the switch arm of figure 2 between q_1 and q_m without passing $q_2, \cdots, q_{m-1}$. The arm does not rotate, it may only sweep back and forth. However it can move arbitrarily from one queue to any other. The operation of moving the arm is known as a seek; but the policy shortest seek time first (SSTF), which corresponds to moving the switch arm the shortest possible distance, is unsatisfactory. For, as we will see, SSTF is likely to discriminate unreasonably against certain queues. In order to enforce good service we will find that the best policy is to scan the switch arm back and forth across the entire range from q_1 to q_m. We shall refer to this policy as SCAN. It will result in nearly the same disk efficiency as SSTF, but provide fair service. Again it should be clear that SCAN should be better than FCFS, which results in wasted arm motion.

THE DRUM SYSTEM

Our first goal is to analyze a fixed head storage device. The analysis deals with a drum system, but is sufficiently general for extension to other devices of similar structure. The analysis we use is approximate, intended only to obtain the expected waiting times, but it should illustrate clearly the advantages of scheduling.

The drum we consider is assumed to be organized as shown in figure 3. We suppose that it is divided lengthwise into regions called fields; each field is a group of tracks; each track has its own head for reading and writing. The angular coordinate is divided into N sectors; the surface area enclosed in the intersection of a field and sector is a page. If the heads are currently being used for reading, they are said to be in read status (R-status); or if for writing, they are said to be in write status (W-status). Since there are two sets of amplifiers, one for reading, the other for writing, there is a selection delay involved in switching the status of the heads. Thus, suppose sector k is presently involved in a read (R) operation. If there is a request in the queue for a page on sector (k+1), and its operation is R, the request is serviced; but if the request is W, it cannot be satisfied until sector (k+2), since the write amplifiers cannot be switched in soon enough.

We suppose further that there is a drum allocation policy in operation which guarantees that there is at least one free page per sector. This might be done in the following way. A desired drum occupancy level (fraction of available pages used) is set as a parameter to the allocator. Whenever the current occupancy level exceeds the desired level, the allocator selects pages which have been unreferenced for the longest period of time and generates

requests to move them to a lower level of storage (for instance a disk). Since these deletion requests are scheduled along with other requests, there will be some delay before the space is available, so the desired occupancy level must be set less than 100 per cent. A level of 90 per cent appears sufficient to insure that the overflow probability (the probability that a given sector has every field filled) will be quite small. If this is done, two assumptions follow: first, that a W-request may commence at once (if the heads are in W-status), or at most after a delay of one sector (if the heads are in R-status); and second, it is highly probable (though not necessarily true) that a sequence of W-requests generated by a single process will be stored consecutively. The assumption that a W-request can begin at once is not unreasonable, and can be achieved at low cost.

We denote the probability that a request is an R-request by p and that it is a W-request by (1-p). That is,

$$\Pr[\text{a given request is R}] = p$$

$$\Pr[\text{a given request is W}] = 1 - p$$

We suppose that the access time per request is a random variable which can assume one of the N equally likely values

$$0, \frac{1}{N}T, \frac{2}{N}T, \ldots, \frac{N-1}{N}T$$

where T is the drum revolution time, N the number of sectors. When no attempt is made to minimize the access time, the expected access time is

$$E[a] = \sum_{k=0}^{N-1} \left(\frac{kT}{N}\right) \Pr[a = \frac{k}{N}] = \frac{N-1}{2N} T \tag{1}$$

Note that E[a] is not exactly $\frac{T}{2}$. The distribution function $F_a(u) = \Pr[a \leq u]$ is shown in Figures 4 and 5. Figure 4 is the actual distribution function,

while figure 5 is our approximation. This is used again in Appendix 2 for the formal derivation of the access time under the SATF policy.

There is some question whether this assumption of uniformity is valid; in fact it is not strictly so, for the following reason. In general a process will generate a sequence of single-page requests. If these are W-requests, and the drum allocation policy is functioning properly, there is a high probability these pages will be stored on consecutive sectors. It follows that a sizable subset of waiting R-requests, all originating from the same process, will want information which is stored on consecutive sectors, so that the access times will have higher probability of being small than a uniform distribution will assign. Although the use of a uniform distribution simplifies our calculations, it leads to conservative answers because the expectations obtained will tend to be too high.

If the service policy is FCFS the expected access time is just the access time for one request:

$$E[a] = \frac{N-1}{2N}T \tag{2}$$

Hence the utilization factor is

$$U_1 = \frac{\text{(page transfer time)}}{\text{(page transfer time)} + \text{(access time)}}$$

$$= \frac{\frac{T}{N}}{\frac{T}{N} + \frac{N-1}{N}\frac{T}{2}}$$

$$U_1 = \frac{2}{N+1} \tag{3}$$

where N is the number of sectors, and T the drum revolution time.

Rather than bore the reader with the detailed analysis of the SATF policy, we defer it to Appendix 2 and present immediately the result. The expected minimum access time when there are n requests in the queue is

$$(4) \qquad E[a] = \left[\frac{Tp}{n+1}\left(1 - \frac{1}{2N}\right)^{n+1} + \frac{T(1-p)}{N} + \frac{T(1-p)}{n+1}\left(1 - \frac{3}{2N}\right)^{n+1}\right] p^n$$

$$+ \frac{T}{N} p^{n+1} (1 - p^n) \left[\left(\frac{1}{p} - \frac{1}{N}\right)^n - \left(1 - \frac{1}{N}\right)^n\right]$$

E[a] is to be interpreted as follows: if a request is selected at a time when it is one of n waiting requests, it can expect to experience a delay of E[a] seconds before the page transfer begins. Hence the utilization factor under the SATF policy is

$$(5) \qquad U_2 = \frac{T/N}{E[a] + (T/N)} = \frac{T}{N\,E[a] + T}$$

where E[a] has been defined by equation (4). Note that the utilization factors U_1 and U_2 are not directly dependent on the input distribution (that is, the interarrival time distribution of requests) and depend only on these parameters:

n -- the number in the queue

N -- the number of sectors

p -- the probability a request is R.

We may define a <u>throughput factor</u> Q to be the number of requests serviced per unit time:

$$(6) \qquad Q = \frac{1}{E[t_s]}$$

where $E[t_s]$ is the expected service time for the policy in force:

$$(7) \qquad E[t_s] = E[a] + \frac{T}{N}$$

For the sake of obtaining estimates of W_F, the wait one request expects to experience in the file system, we will assume that the number n in the queues stays close to its expectation:

$$n \approx E[n]$$

It is well known[5] that in a FCFS queue, the expected wait is

$$W_F \approx E[n]\, E[t_s]$$

where $E[t_s]$ according to equation (7) is the expected service time. In our case

$$W_F \approx n\, E[t_s] \approx n\left(\frac{T}{2} + \frac{T}{N}\right) = nT\left(\frac{N+2}{2N}\right) \tag{8}$$

To obtain an estimate of W_F under the SATF policy, we reason as follows. If there are n requests waiting, then n/N of them are waiting for a given sector (see Figure 2). So when a request enters the file system, it expects to wait T/2 for its sector to come into position, plus n/N additional drum revolutions before receiving service, and finally T/N for its own page transfer. Hence

$$W_F = \frac{T}{2} + \frac{n}{N}T + \frac{T}{N} = T\left(\frac{1}{2} + \frac{n+1}{N}\right) \tag{9}$$

Example

To dramatize the effects of scheduling, we give a numerical example. A typical high-speed drum has a revolution time of 16.7 ms (3600 rpm). Typically there are 4096 words written around the drum's circumference. We use a page size is 64 words, yielding 64 sectors around the drum. Typically the probability of a read will be greater than that of a write, so the choice of 0.7 for p is not unreasonable. We have the following for the parameters:

$T = 16.7$ ms $\qquad t_o = T/N = 0.26$ ms

$N = 64$ sectors $\qquad p = 0.7$

Substituting these parameters into equations (2) to (9) yields the following table. We have chosen $E[n] \approx n = 10$.

POLICY	milliseconds			Utilization U per cent	Throughput Q per second
	E[a]	$E[t_s]$	W_F		
FCFS	8.33	8.59	85.9	3	11.6
SATF	0.23	0.49	11.2	53	204.0

The relative improvement is

$$\frac{W_F\ [FCFS]}{W_F\ [SATF]} \approx \frac{nT\left(\frac{1}{2}+\frac{1}{N}\right)}{T\left(\frac{1}{2}+\frac{n+1}{N}\right)} = \frac{n(N+2)}{N+2(n+1)}$$

For the given parameters this is 7.66, which means that incoming requests see a drum that is 766 per cent faster! Note that when $n \gg N$, an overload condition, the improvement approaches a maximum of $(N+2)/2$.

Another calculation shows that U = 56 per cent for n = 20 under the SATF policy. It is apparent that with this rather simple policy the utilization of the drum is increased by a factor of almost 20! Another way to say this is that without scheduling, the drum is idle 97 per cent of the time, and with scheduling it is idle only 47 per cent of the time. It is apparent that scheduling for fixed head devices is well worth the cost, resulting in vastly increased throughput and utilization while providing fair service to requests.

As a final note we mention that some high-speed drums can switch head status so rapidly that the selection delay is negligible. In particular, a W-request may commence at once. The average behavior of the system under this assumption can be obtained as follows. First obtain the behavior of the system for read requests alone by setting p = 1 in equation (4) so that

(10) $$E[a \mid \text{R-request}] = \frac{T}{n+1}\left(1-\frac{1}{2N}\right)^{n+1}$$

Setting p = 1 in equation (4) cancels the effects of selection delay for W-requests used to derive E[a]. Since now

(11) $$E[a \mid \text{W-request}] = 0$$

if there is no selection delay, we obtain finally

(12) $$E[a] = \frac{Tp}{n+1}\left(1 - \frac{1}{2N}\right)^{n+1}$$

THE DISK SYSTEM

We turn our attention to movable head storage devices. Typical of equipment in this class is the movable arm disk, so we present an analysis for such a unit. There is sufficient generality that the results can be extended for use with other movable head devices of similar structure. First we will examine the behavior of a disk unit under access-time-minimizing scheduling policies and find that such policies do not provide comfortable increases in utilization. In fact they may give rise to very undesirable discrimination effects. Finally we will examine an alternative which leads to satisfactory disk behavior.

The system we consider is displayed in Figure 6. There are a series of disks of like radii equally spaced along a common axis. Both surfaces of each disk are coated with a magnetic substance, so that information may be stored on both sides. The entire assembly rotates in time T. Each disk has a set of tracks situated near the outer edge, so that each track is of maximum length. One can imagine that the k^{th} track of each disk is situated on the surface of a cylinder; thus, if the disk is W tracks wide, there are W concentric cylinders on which to store information. There is an assembly of movable arms, equipped with read/write heads. There is one set of heads for each side of each disk, sufficient to read or write one track, so each time a new track is requested, it is necessary to reposition the arms. The operation of positioning the arms is known as a seek. The machanism for doing this is hydraulic, and is therefore inherently slow. The seek time is illustrated in Figure 7, which indicates that a seek of one track requires a time S_{min}, while a seek the full width of the disk, (W-1) tracks, requires a time S_{max}. Typically S_{min} = 150 ms and S_{max} = 300 ms; thus, there is little difference between a seek of k tracks and a seek of (k+1) tracks. The bottleneck is getting the arms moving in the first place.

Once the arms have been positioned, any disk surface on the cylinder is addressable. We may regard the W concentric cylinders as a set of W drum systems by regarding each cylinder as a "drum". Unlike our drum model, we do not assume that there is always a free track on each cylinder on which to write a new page. This is because files stored on disks generally extend over several tracks, so it is futile to expect an availability of several tracks on the same cylinder. Therefore all requests are treated on an equal basis, without distinction between reads and writes.

Scheduling may be considered to operate on two levels. Waiting requests are sorted into W groups, one for each cylinder, in a manner indicated by Figure 2. The upper level of scheduling is to decide at which cylinder to position the arm. Once the arm is positioned, the lower level of scheduling comes into play. At this level, SATF scheduling could be used, but hardly seems worthwhile since the probability that more than one request is waiting for the same cylinder should, under normal load conditions, be quite small. So nothing is to be gained by scheduling policies other than FCFS within the cylinder queues. Thus the expected rotational delay is T/2. Our interest in disk scheduling will be confined to the level of deciding where to move the arms. In terms of pictures, we are interested in deciding how to move the rotary switch in Figure 2.

It is a common but questionable practice to chain tracks on the disk, the chaining information being placed at the end of each track; the next track in the chain is therefore unknown until the current track has been read. We assume that if the End of File (end of a chain) has not been reached, that a new request is generated and scheduled on an equal basis with other requests; thus a file which unfortunately has become chained far and wide across the disk need not tie up the arm for more than one track at a time.

In particular, a request requiring a shorter seek may be serviced in the interim between tracks of a given chain.

The disk waiting time parameters are shown in Figure 8. After a request enters the queue, it waits a time W_q until it is chosen for service. Once it has been chosen, a seek time s elapses. At the completion of the seek, an additional rotational delay r passes until the desired sector has come into position. Only then does transmission begin, and we assume it lasts for t seconds. The access time is

$$a = s + r \tag{13}$$

and the service time is

$$t_s = a + t = s + r + t \tag{14}$$

The waiting time in the system is

$$W_F = W_q + t_s \tag{15}$$

Our interest lies in deciding whether disk performance can be improved by planning the motion of the arm. Naturally we assume that, once the arms are positioned at a given cylinder, every request for that cylinder is served before another seek is initiated. The reason for this is apparent when one considers the magnitude of the minimum seek time, S_{min}, of Figure 7. Once every request for a given cylinder is satisfied, the arms are moved to a new cylinder; it would seem reasonable that a Shortest Seek Time First (SSTF) policy should be used. However this policy is unsatisfactory since it will

tend to discriminate against requests for the inner and outer cylinders. To see that this is true, suppose the arm is at the k^{th} track, where $k < W/2$, as indicated in Figure 9. Because requests are assumed to fall uniformly for any of the W tracks, it is apparent that there is more likelihood the arm will move toward the center of the track region; hence requests for the outer edges of the track region will tend not to be serviced. That is, since there are more requests in Region II than in Region I, the likelihood is greater that the arm moves into Region II.

Nevertheless it is entirely possible that the length of time a request for the extremities is delayed is still less than its normal wait under the FCFS policy. So before discarding the SSTF policy, we must analyze it to see whether the improvement in the utilization balances the undesirable effects. We will see that the improvement is marginal, that under heavy loading conditions undesirable discrimination effects occur. Our interest at this point is to analyze the SSTF policy; later we consider means to avoid anomalies.

Analysis of SSTF Policy

As before, we do not present the detailed analysis of the SSTF policy here; instead we defer it to Appendix 3. We outline here the assumptions made. The analysis has as its primary goal to obtain approximate figures and expressions for average seeks and waits, so as to obtain utilization factors and characterize the effects of scheduling.

We treat the problem as follows. There are (at a decision point) n cylinders requested, and the arm isat cylinder k, where k is equally likely to be any cylinder. We ask: what is the expected seek time? The answer we will obtain is dependent on n, the number of requests in the queue, and on the disk parameters. To answer the question we find the distribution function for the arm movement caused by a single request; then we assume there are n identically distributed arm-movement random variables and use the result of Appendix 1 to find the expected minimum arm movement.

In Figure 9 we show the arm positioned at track k. Suppose there is exactly one request waiting. How far will it cause the arm to move? To answer this we suppose that the arm does in fact move and later multiply by the probability that it moves. If there are n cylinders requested, the probability that the present cylinder is <u>not</u> requested is

$$\left(1 - \frac{1}{W}\right)^n \tag{16}$$

which is therefore just the probability a seek takes place. Now suppose the arm is positioned at cylinder k, where $k < W/2$. Given that the arm moves, we have the probability distribution of Figure 10. The arms move 1, 2, ... ,(k-1) positions, each with probability 2p, or k, k+1, ... , (W-k) positions, each with probability p. Here p is the probability a request is for a particular one of the remaining (W-1) cylinders:

$$p = \frac{1}{W-1} \tag{17}$$

To obtain a better understanding of Figure 10, it helps to imagine that Figure 9 has been folded on itself at position k. Since $k < W/2$, the

first (k-1) positions of the folded distribution are twice as probable as the remaining positions. It is clear that the situation $k > W/2$ is identical, so we need consider only $k < W/2$, obtaining $k > W/2$ by symmetry.

Figure 11 is the cumulative distribution $F_s(u)$ that the seek time s is less than or equal to u units:

$$F_s(u) = \Pr[s \leq u] \tag{18}$$

Figure 12 is a continuous approximation to $F_s(u)$, obtained by passing straight lines through the centers of each plateau in Figure 11.

In Appendix 3 these assumptions are used to obtain the expected minimum seek time under the SSTF policy. It is

$$E[s] = \left(\frac{W-1}{W}\right)^n \left[\frac{1}{2} + S_{min} + \frac{S_{max} - S_{min}}{2(n+1)}\left(1 + \frac{1}{n+2}\left(\frac{W}{W-1}\right)^{n+1}\right)\right] \tag{19}$$

The unscheduled (FCFS) seek time is obtained by setting $n = 1$ in equation (19). It is

$$E[s] = \frac{W-1}{W}\left[S_{min} + \frac{1}{2} + \frac{S_{max} - S_{min}}{4}\left(1 + \frac{1}{3}\left(\frac{W}{W-1}\right)^2\right)\right] \tag{20}$$

Finally the access time is

$$E[a] = E[s] + \frac{T}{2} \tag{21}$$

Saturation and Loading Effects

We have said nothing about what SSTF is to do in case of a "tie"; that is, when the minimum arm movement is the same in either direction. Suppose we were to adopt the rule that, in case of a tie, the arm moves away from the center of the disk region; in this way (we would reason) we can counteract the effects of outer-edge discrimination. Consider what would happen if, on the average, there were one request per cylinder. There would almost always be a tie for a motion of one track, so that the arm could drift to one edge of the track region and stay there. Any requests for the opposite edge might be overlooked indefinitely.

This possibility that there may be times when there are many cylinder requests is the downfall of the SSTF policy. For under heavy loading conditions, the arm will tend to remain in one place on the disk, and pay little or no attention to other regions. Although SSTF provides an increase in utilization, it is not at all clear that it will provide reasonable service to the requests. For these reasons we feel the following policy is superior.

The SCAN Policy

As we mentioned in the discussion following Figure 2, a policy which insures reasonable service to requests is the SCAN policy, in which the switch arm is swept back and forth between q_1 and q_m, servicing any requests in the intervening queues as it passes by. We can think of operating the

disk arm as a "shuttle bus" between the inner and outer edges of the track region, stopping the arm at any cylinder for which there are waiting requests.

Using average-value arguments we can obtain estimates for the utilization factor U and waiting time W_F of a single request. First assume that there are sufficiently many cylinders and sufficiently few requests that the probability of finding more than one request for a given cylinder is much less than the probability of finding just one request. This is equivalent to assuming that the total number of waiting requests is the same as the number of cylinders requested. Further, assume that a given request falls randomly within the track region, and that when it arrives the arm is at some random position within the track region. Then the expected distance between the arriving request and the arm is W/3. Now with probability 1/2 the arm is moving toward the request, in which case the distance the arm must travel before reaching the request is W/3. With probability 1/2 the arm is moving away from the request, in which case the distance the arm must travel before reaching the request is $3(W/3) = W$. Hence the expected distance the arm moves before reaching the request is

$$\frac{1}{2}\left(\frac{W}{3}\right) + \frac{1}{2}W = \frac{2}{3}W \tag{22}$$

Next we obtain an estimate of the seek time per request. If there are n requests, these divide the track region in (n+1) regions of expected width W/(n+1). The seek time for one such distance is

$$E[s] = S_{min} + \frac{S_{max} - S_{min}}{n+1} \tag{23}$$

Assuming that, once the arm has stopped at a cylinder, there is an additional expected delay of $T/2$ for the starting address to rotate into position, we have for the access time

$$E[a] = E[s] + \frac{T}{2} \tag{24}$$

Now we can determine how long it takes the arm to cross the track region. It must make n stops and do a transfer of one track at each stop. So the time t between stops is

$$t = E[a] + T = E[s] + \frac{3}{2} T$$

or,

$$t = \frac{S_{max} - S_{min}}{n+1} + S_{min} + \frac{3}{2} T \tag{25}$$

Hence the time to cross W tracks is nt. The time W_F a single request must wait is, from equation (22),

$$W_F = \frac{2}{3} nt = \frac{2}{3} n \left[\frac{S_{max} - S_{min}}{n+1} + S_{min} \right] + nT \tag{26}$$

Example

To demonstrate the effects of scheduling, we present an example using the following parameters:

S_{min} = minimum seek time
= 150 ms

S_{max} = maximum seek time
= 300 ms

W = number of tracks
= 30

T = disk revolution time
= 60 ms

n = number of requests in queue
= 10

The utilization factor, or the fraction of time spent by the disk doing useful transmission, is

(27) $$U = \frac{t}{E[a] + t}$$

where t is the transfer time for a single request. We use $t = T$, that is, each transfer is a full track. Due to uncertainties resulting from possible saturation effects we do not venture to give estimates of W_F under the SSTF policy. Using $W_F = n\, E[t_s]$ under the FCFS policy, $Q = 1/E[t_s]$ for the throughput factor Q under all policies, and the disk equations just derived, we obtain the following table for the above parameters.

POLICY	milliseconds				Utilization U per cent	Throughput Q Per second
	E[s]	E[a]	$E[t_s]$	W_F		
FCFS	195	225	285	2850	21.0	3.5
SSTF	113	143	203	?	29.5	4.9
SCAN	164	194	254	1700	23.6	3.9

It is apparent that with 10 cylinders requested the expected utilization increase with SSTF over FCFS is

$$\frac{29.5}{21.0} = 1.40$$

so that the gain under SSTF is a not-too-impressive 40 per cent. It is the large minimum access time, S_{min}, that impairs efficiency. We conclude that, since the SSTF gain is marginal, the discrimination certain requests might suffer is not worth the risk. This substantiates our claim that SCAN is preferable. SCAN exhibits an improvement of

$$\frac{23.6}{21.0} = 1.12$$

or 12 per cent in utilization over FCFS; but the improvement in W_F,

$$\frac{2850}{1700} = 1.68$$

or 68 per cent is well worth the cost. It means that each request sees a disk that is 68 per cent faster. Finally note that the relative improvement

$$\frac{W_F\ [FCFS]}{W_F\ [SCAN]} \approx \frac{S_{min} + \frac{S_{max} - S_{min}}{3} + \frac{3}{2}T}{\frac{2}{3}\left(S_{min} + \frac{S_{max} - S_{min}}{n+1} + \frac{3}{2}T\right)}$$

approaches a maximum for

$$n \gg \frac{S_{max} - S_{min}}{S_{min} + \frac{3}{2}T}$$

which is obtained under normal load conditions.

CONCLUDING REMARKS

In the case of fixed head devices such as drums, we have seen that spectacular increases in utilization can be achieved by the rather simple expedient of Shortest Access Time First (SATF) scheduling. Requests serviced under the SATF policy are treated with the same degree of fairness as under the FCFS policy. Yet service is vastly improved.

We have seen, in the case of movable head devices such as disks, the impulse to use Shortest Seek Time First (SSTF) scheduling must be resisted. Not only may SSTF give rise to unreasonable discrimination against certain requests, but also during periods of heavy disk use it would result in poor service to requests far from the current arm position. Enforcing good response to requests by shuttling the arm back and forth across the disk under the SCAN policy surmounts these difficulties, at the same time providing little more arm movement than SSTF. Hence SCAN is the best policy for use in disk scheduling.

W_F, the time one request expects to spend in the file system, provides a simple, dramatic measure of file system performance. Figures 13 and 14 indicate comparative trends in W_F for two policies. As usual, n is the number in the queue, T the device revolution time, N the number of sectors on the drum, S_{max} and S_{min} the maximum and minimum disk seek times respectively. Figure 13 shows that even for small n, W_F under the SATF policy is markedly improved. The curves diverge rapidly. We call attention to the asymptotic slopes, which apply under heavy load conditions. Figure 14 leads to similar conclusions for the SCAN policy. There can be no doubt that these simple policies -- SATF for drum, SCAN for disk -- can significantly improve file memory operations and thus the performance of the computing system as a whole.

ACKNOWLEDGEMENT

The author wishes to thank Jack B. Dennis and Anatol W. Holt for helpful suggestions received while composing this paper.

APPENDIX 1 -- Expectation of the Minimum of Random Variables.

Consider a set of independent, identically distributed random variables $t_1, t_2, \ldots, t_n$ each having density function p(t). Define a new random variable

$$x = \min \{ t_1, t_2, \ldots, t_n \} \tag{1.1}$$

We wish to determine the expectation of x, E[x]. Now,

$$\begin{aligned} Pr[x > \alpha] &= Pr[t_1 > \alpha, t_2 > \alpha, \ldots, t_n > \alpha] \\ &= \left(Pr[t > \alpha] \right)^n \\ &= \left[F_t^c(\alpha) \right]^n \end{aligned}$$

where

$$F_t^c(\alpha) \equiv Pr[t > \alpha] = \int_\alpha^\infty p(t)\, dt \tag{1.2}$$

is the complementary cumulative distribution for t. Using the fact that

$$E[x] = \int_0^\infty Pr[x > u]\, du \tag{1.3}$$

we have

$$E[x] = \int_0^\infty \left(F_t^c(u) \right)^n du \tag{1.4}$$

APPENDIX 2. Analysis of SATF Policy.

Consider the SATF policy. As indicated by Figure 5, the distribution function for access time is

$$(2.1) \qquad F_a(u) = \begin{cases} 0 & u \le 0 \\ \frac{1}{2N} + \frac{u}{T} & 0 < u \le \alpha \\ 1 & \alpha < u \end{cases} \qquad \alpha = \left(N - \frac{1}{2}\right)\frac{T}{N}$$

where T is the drum revolution time, and N the number of sectors. Using Appendix 1,

$$(2.2) \qquad E_1[a] = \int_0^{\alpha} \left(1 - \frac{1}{2N} - \frac{u}{T}\right)^n du = \frac{T}{n+1}\left(1 - \frac{1}{2N}\right)^{n+1}$$

here $E_1[a]$ is the expected access time if every request is R, the heads are in R-status, and n requests are in the queue. If the heads are in W-status they must be switched, and no transfer can begin for T/N seconds, until one sector has passed. That is, the distribution function when a change in head status must be made is

$$(2.3) \qquad F_a(u) = \begin{cases} 0 & u \le \beta \\ \frac{1}{2N} + \frac{u}{T} & \beta < u \le \alpha \\ 1 & \alpha < u \end{cases} \qquad \beta = \frac{T}{N}, \quad \alpha = \left(N - \frac{1}{2}\right)\frac{T}{N}$$

From Appendix 1, the expected minimum access time is

$$(2.4) \qquad E_2[a] = \int_0^{\beta} du + \int_{\beta}^{\alpha} \left(1 - \frac{1}{2N} - \frac{u}{T}\right)^n du = \frac{T}{N} + \frac{T}{n+1}\left(1 - \frac{3}{2N}\right)^{n+1}$$

If there is one or more W-request and the heads are already in W-status, the access time will be zero. If the heads are in R-status, and if none of the waiting R-requests is for the present sector, the access time will be one transfer time, T/N. Otherwise there is an R-request for the present sector, and so the access time is zero. Thus we must consider several cases. In the following, n is the number of requests in the queue, p is the probability that a given request is R, N is the number of sectors, and T is the drum revolution time.

CASE 1. There are no W-requests in the queue, so that there are n R-requests. We must consider whether the heads are in R-status (with the same probability p that the last request was R) or in W-status (with probability 1-p).

CASE 1a. The heads are in R-status. The distribution function of access times is given by equation (2.1), with expectation $E_1[a]$ given in equation (2.2). $E_1[a]$ occurs with probability p.

CASE 1b. The heads are in W-status, so that at least one sector must pass before switching is completed. In this case no access of less than T/N can take place, so the distribution function is given by equation (2.3), with expectation $E_2[a]$ given by equation (2.4). $E_2[a]$ occurs with probability (1-p).

CASE 1 SUMMARY. The probability that every request is R is p^n. Thus if every request is R we have

$$E_R[a] = E_1[a]\, p + E_2[a]\, (1-p) \tag{2.5}$$

and $E_R[a]$ occurs with probability p^n.

CASE 2. There is at least one W-request, so that there are k = 0,1,...,(n-1) R-requests. We must consider two cases: the heads in W-status, and the heads in R-status.

CASE 2a. The heads are in W-status, so that there need be no switching of amplifiers, and the transfer may begin at once. In this case E[a] = 0, with the probability that the last request was W. This is just (1-p).

CASE 2b. The heads are in R-status. The access time then will be zero if there is an R-request for the present sector and T/N otherwise. This situation occurs with probability p. Now, for each k = 0,1,...,(n-1) the probability that there are k R-requests and (n-k) W-requests is

$$\binom{n}{k} p^k (1-p)^{n-k} \qquad \text{where } \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

The probability that none of these is for the present sector is

$$\left(1 - \frac{1}{N}\right)^k$$

That is, the access time is non-zero only if all k R-requests are not for the current sector. Hence the probability that there are k R-requests and none of them is for the present sector is

$$\left(1 - \frac{1}{N}\right)^k \binom{n}{k} p^k (1-p)^k$$

Since this is true for each k = 0,1,...,(n-1), we must sum over k.

$$\sum_{k=0}^{n-1} \binom{n}{k} \left(1 - \frac{1}{N}\right)^k p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \left[p\left(1 - \frac{1}{N}\right) \right]^k \left(1 - p\right)^{n-k} - p^n \left(1 - \frac{1}{N}\right)^n$$

$$= \left[1 - p + p\left(1 - \frac{1}{N}\right)\right]^n - p^n\left(1 - \frac{1}{N}\right)^n$$

$$= \left(1 - \frac{p}{N}\right)^n - p^n\left(1 - \frac{1}{N}\right)^n$$

Hence the probability

$$Q = \Pr\left\{\begin{array}{l}\text{no R-request for the present sector } \underline{\text{and}} \text{ heads} \\ \text{in R-status } \underline{\text{and}} \text{ at least one W-request}\end{array}\right\}$$

$$(2.6) \qquad Q = p^{n+1}\left[\left(\frac{1}{p} - \frac{1}{N}\right)^n - \left(1 - \frac{1}{N}\right)^n\right]\left(1 - p^n\right)$$

But Q is just the probability that $E[a] = \frac{T}{N}$. Collecting together the results of cases 1 and 2 we see that

$$(2.7) \qquad E[a] = E_R[a]\, p^n + \frac{T}{N} Q$$

so that

$$(2.8) \qquad E[a] = \left[\frac{Tp}{n+1}\left(1 - \frac{1}{2N}\right)^{n+1} + \frac{T(1-p)}{N} + \frac{T(1-p)}{n+1}\left(1 - \frac{3}{2N}\right)^{n+1}\right] p^n$$

$$+ \frac{T}{N} p^{n+1}\left(1 - p^n\right)\left[\left(\frac{1}{p} - \frac{1}{N}\right)^n - \left(1 - \frac{1}{N}\right)^n\right]$$

which concludes the derivation.

APPENDIX 3. Analysis of SSTF Policy.

From figure 11 we have for the distribution function for the arm movement for one request, conditioned on the arm moving:

$$(3.1)\qquad F_s(u) = \begin{cases} 0 & u \le \frac{1}{2} \\ 2pu - p & \frac{1}{2} < u \le (k - \frac{1}{2}) \\ pu + \frac{p}{2}(2k-3) & (k - \frac{1}{2}) < u \le (W - k + \frac{1}{2}) \\ 1 & (W - k + \frac{1}{2}) < u \end{cases}$$

This is true for $k < W/2$. By symmetry, the same is true for $k > W/2$, with k replaced by (W-k). From Appendix 1, the expected minimum seek time is

$$(3.2)\qquad E[s] = \int_0^\infty \left(1 - F_s(u)\right)^n du$$

$$= \int_0^{\frac{1}{2}} du + \int_{\frac{1}{2}}^{k-\frac{1}{2}} \left(1 + p - 2pu\right)^n du + \int_{k-\frac{1}{2}}^{W-k+\frac{1}{2}} \left(1 - \frac{p}{2}(2k-3) - pu\right)^n du$$

This integration is straightforward, so we do not reproduce it here. The result is

$$(3.3)\qquad E[s] = \frac{1}{2} + \frac{1}{2p}\left[1 - \left(1 - 2p(k-1)\right)^{n+1}\right]^{\frac{1}{n+1}} + \frac{1}{p(n+1)}\left(1 - 2p(k-1)\right)^{n+1}$$

Equation (3.3) applies for $k < W/2$. By symmetry (Figure 9) it applies for $k > W/2$ if we replace k by (W-k). The case $k < W/2$ occurs with probability 1/2, and so does the case $k > W/2$. Hence, collecting terms in equation (3.3),

$$(3.4) \quad E[s \mid k < \frac{W}{2}] = \frac{1}{2} + \frac{1}{2p(n+1)}\left[1 + \left(1 - 2p(k-1)\right)^{n+1}\right]$$

By symmetry, $E[s \mid k > \frac{W}{2}]$ is the same. Then

$$E[s] = \frac{1}{2} E[s \mid k < \frac{W}{2}] + \frac{1}{2} E[s \mid k > \frac{W}{2}]$$

$$(3.5) \quad E[s] = E[s \mid k < \frac{W}{2}]$$

Now, recalling that the head position k is assumed to be uniform for values of $k = 1,2,\ldots,W$, we can approximate its distribution function by a ramp of slope 1/W from $\frac{1}{2}$ to $W + \frac{1}{2}$. Conditioning on $k < \frac{W}{2}$, we can approximate the conditional distribution by a ramp of slope 2/W from $\frac{1}{2}$ to $\frac{1}{2}\left(W+1\right)$. Therefore we integrate $\left(\frac{2}{W} E[s \mid k < \frac{W}{2}]\right)$ on the interval $\left[\frac{1}{2}, \frac{1}{2}\left(W+1\right)\right]$. The result is, using $p = 1/(W-1)$:

$$(3.6) \quad E[s] = \frac{1}{2} + \frac{W-1}{2(n+1)}\left[1 + \frac{1}{n+2}\left(\frac{W}{W-1}\right)^{n+1}\right]$$

recalling that this is conditioned on the arm moving, we remove this condition by multiplying by the probability the arm moves:

$$\left(1 - \frac{1}{W}\right)^n = \left(\frac{W-1}{W}\right)^n$$

and obtain the following:

$$(3.7) \qquad E[s] = \left(\frac{W-1}{W}\right)^n \left[\frac{1}{2} + \frac{W-1}{2(n+1)} \left(1 + \frac{1}{n+2} \left(\frac{W}{W-1}\right)^{n+1} \right) \right]$$

Noting that the time required for (W-1) tracks is S_{max}, and the time for one track is S_{min} (if the arm moves), we have for the expected minimum seek time when n are in the queue:

$$(3.8) \qquad E[s] = \left(\frac{W-1}{W}\right)^n \left[\frac{1}{2} + S_{min} + \frac{S_{max} - S_{min}}{2(n+1)} \left(1 + \frac{1}{n+2} \left(\frac{W}{W-1}\right)^{n+1} \right) \right]$$

which concludes the derivation.

REFERENCES

[1]Weingarten, A. "The Eschenback Drum Scheme." Comm. ACM. Vol. 5, No. 7. (July, 1966).

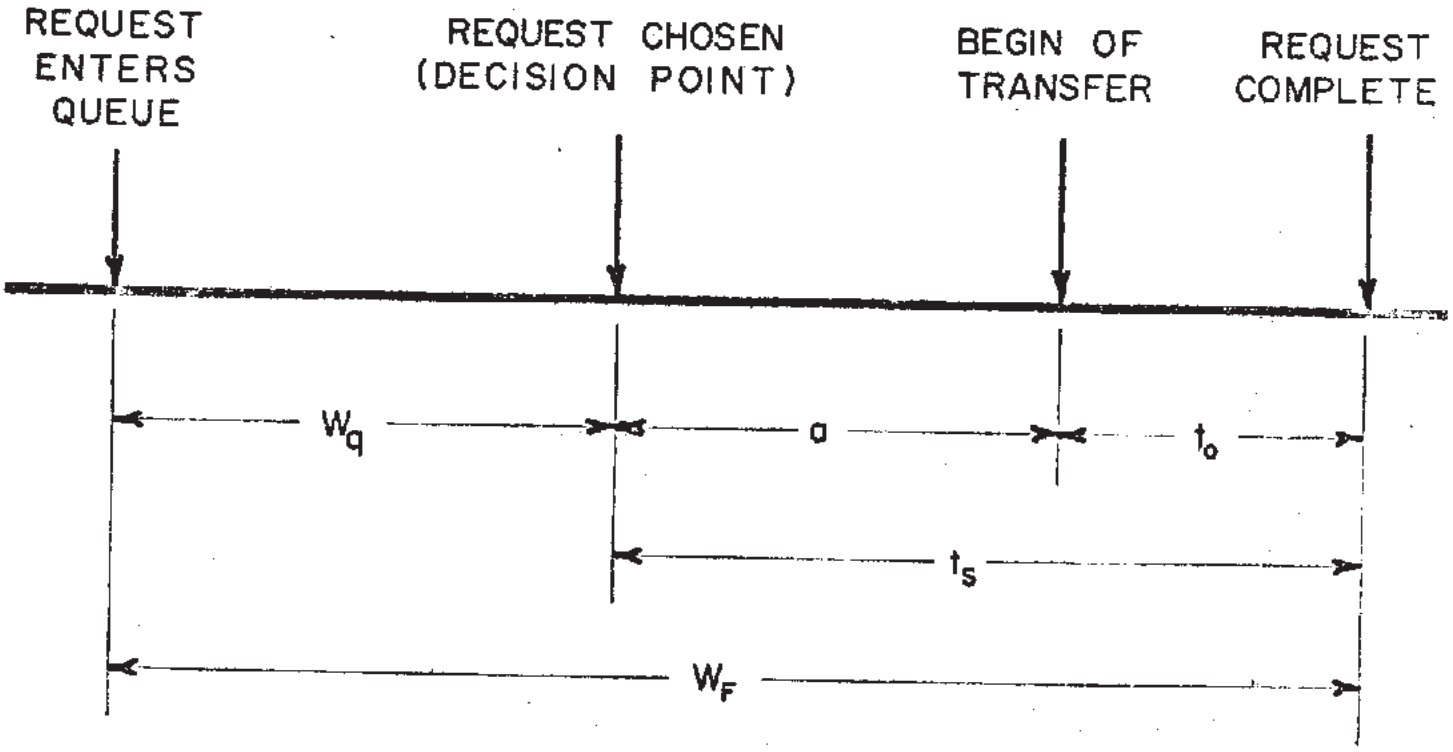
[2]Fife, D. W., and Smith, J. L. "Transmission Capacity of Disk Storage Systems with Concurrent Arm Positioning." IEEE Transactions on Electronic Computers. Vol. EC-14, No. 4. (August, 1965).

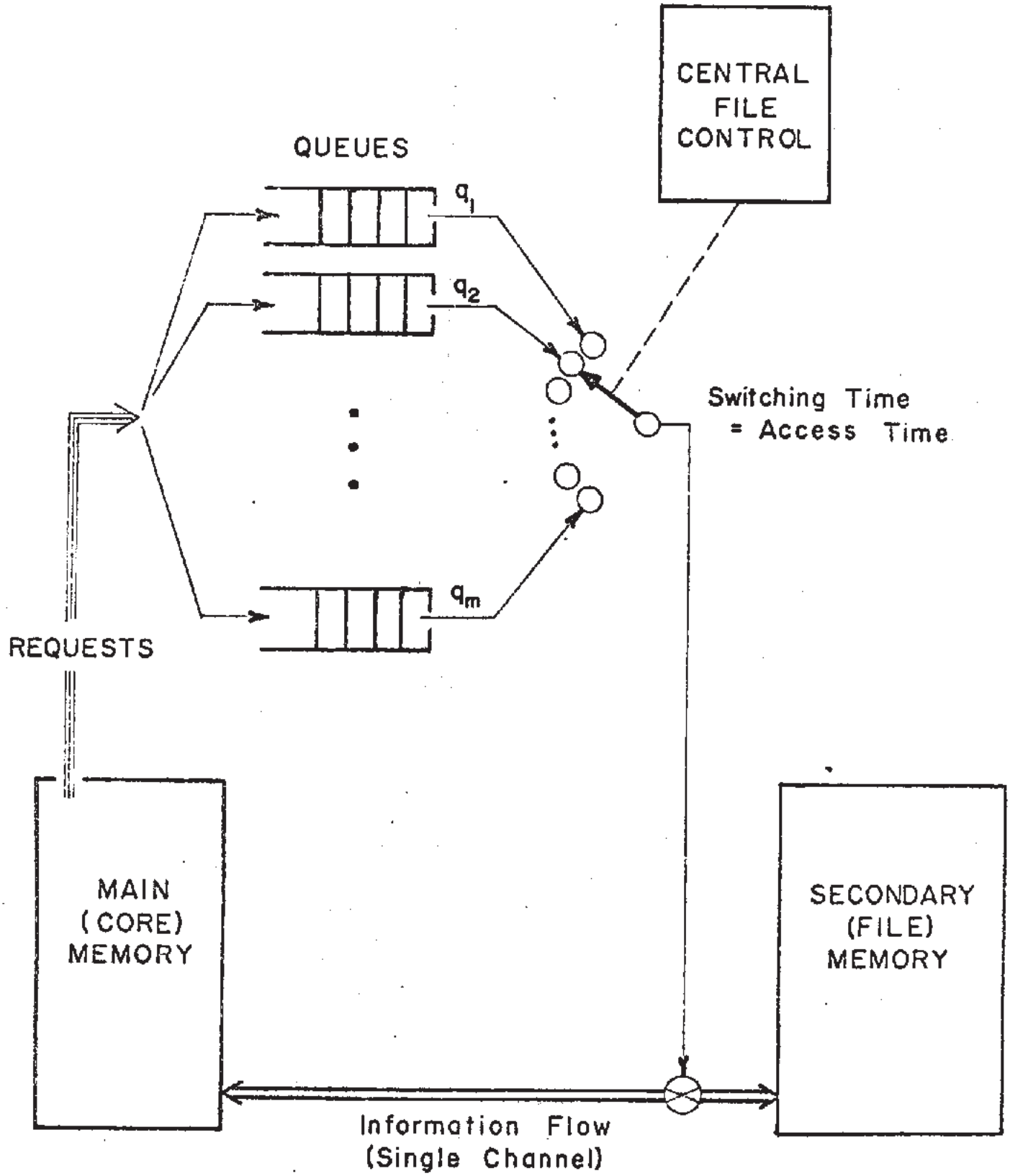
[3]Denning, P. J. "Queueing Models for File Memory Operations." M.I.T. Project MAC Technical Report MAC-TR-21. (October, 1965)

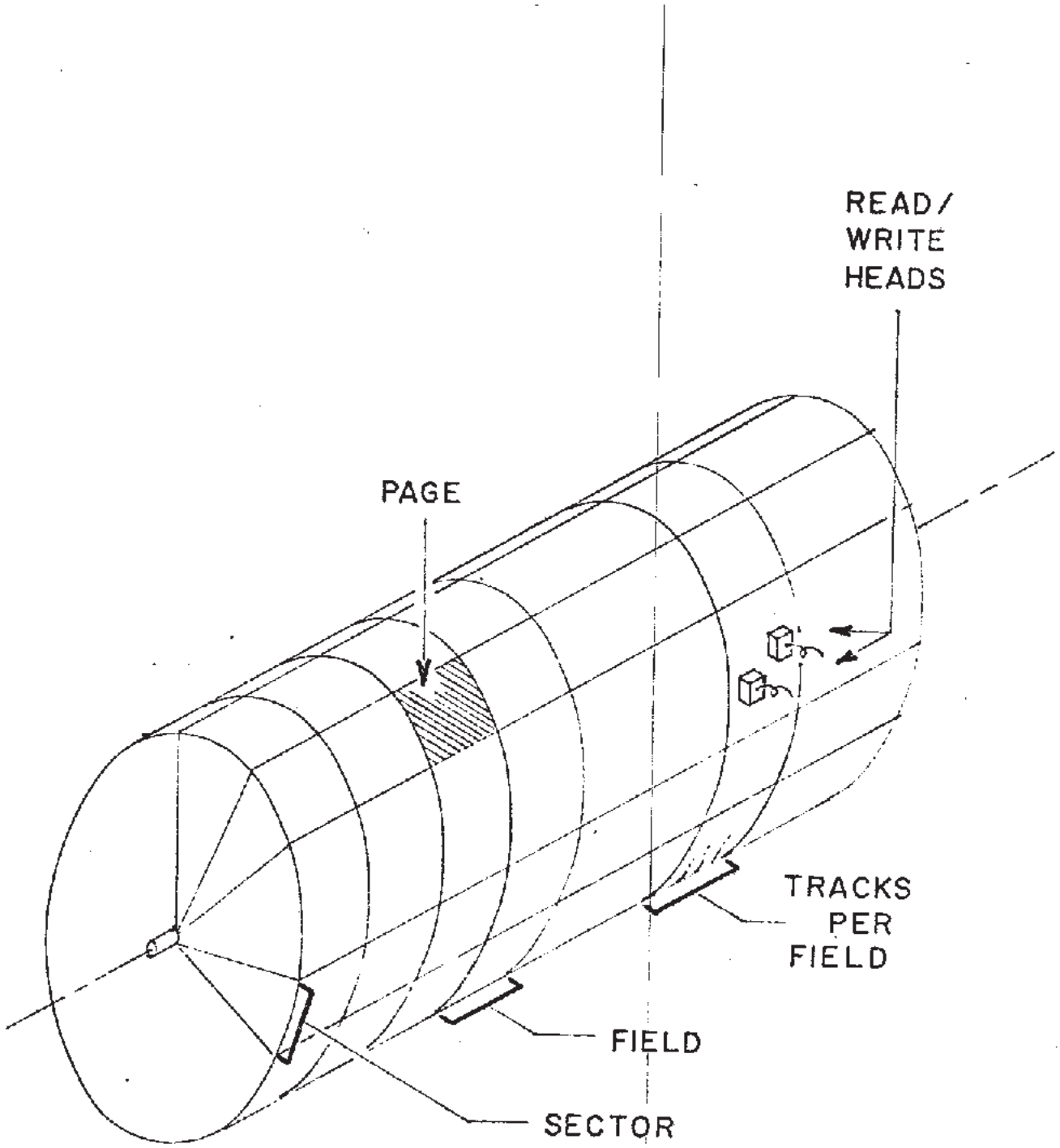
[4]Belady, L. A. "A Study of Replacement Algorithms for a Virtual-Storage Computer." IBM Systems Journal. Vol. 5, No. 2,(1966).

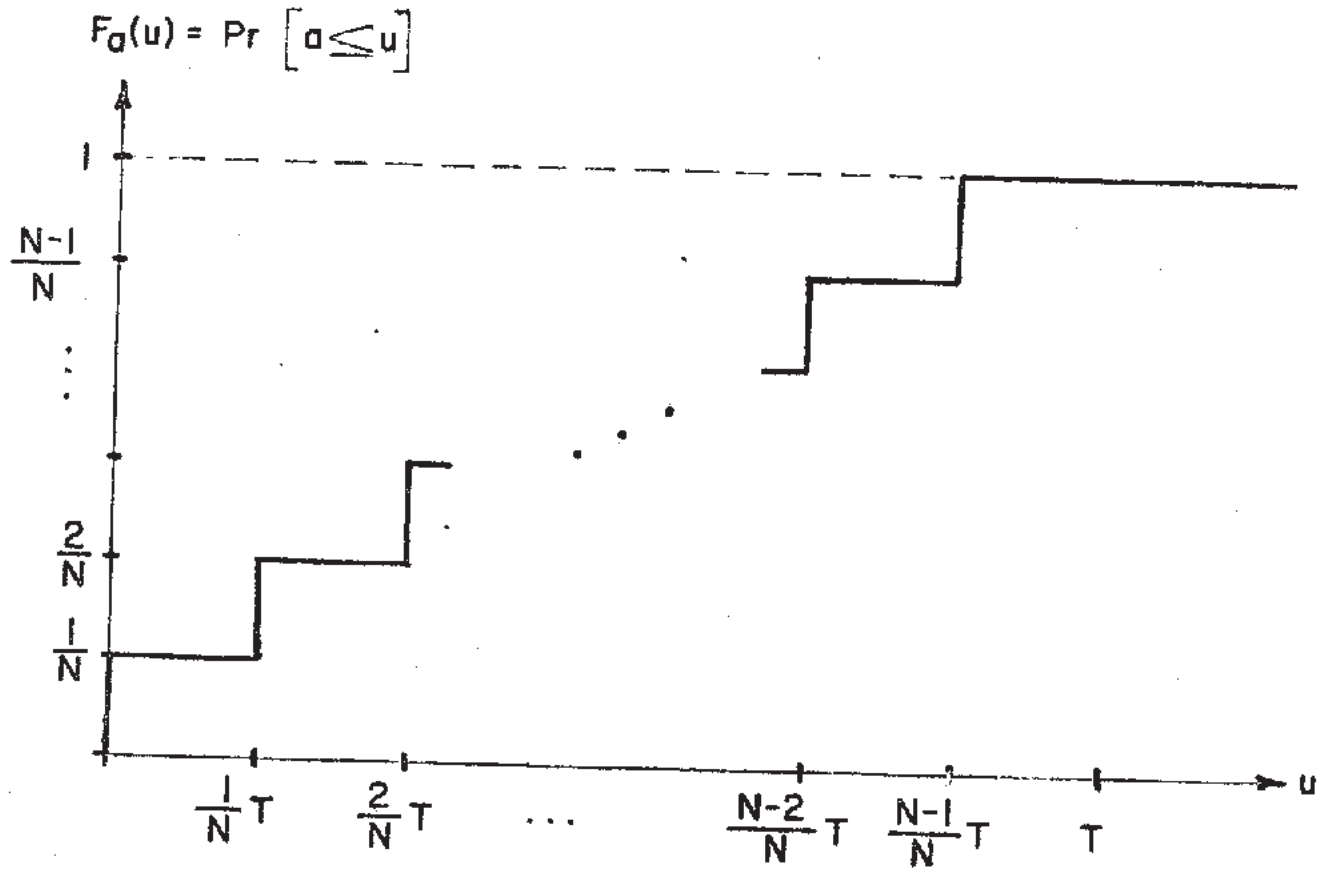
[5]Saaty, T. L. Elements of Queueing Theory. (McGraw-Hill Book Co., Inc., New York, 1961). pp 40 ff.

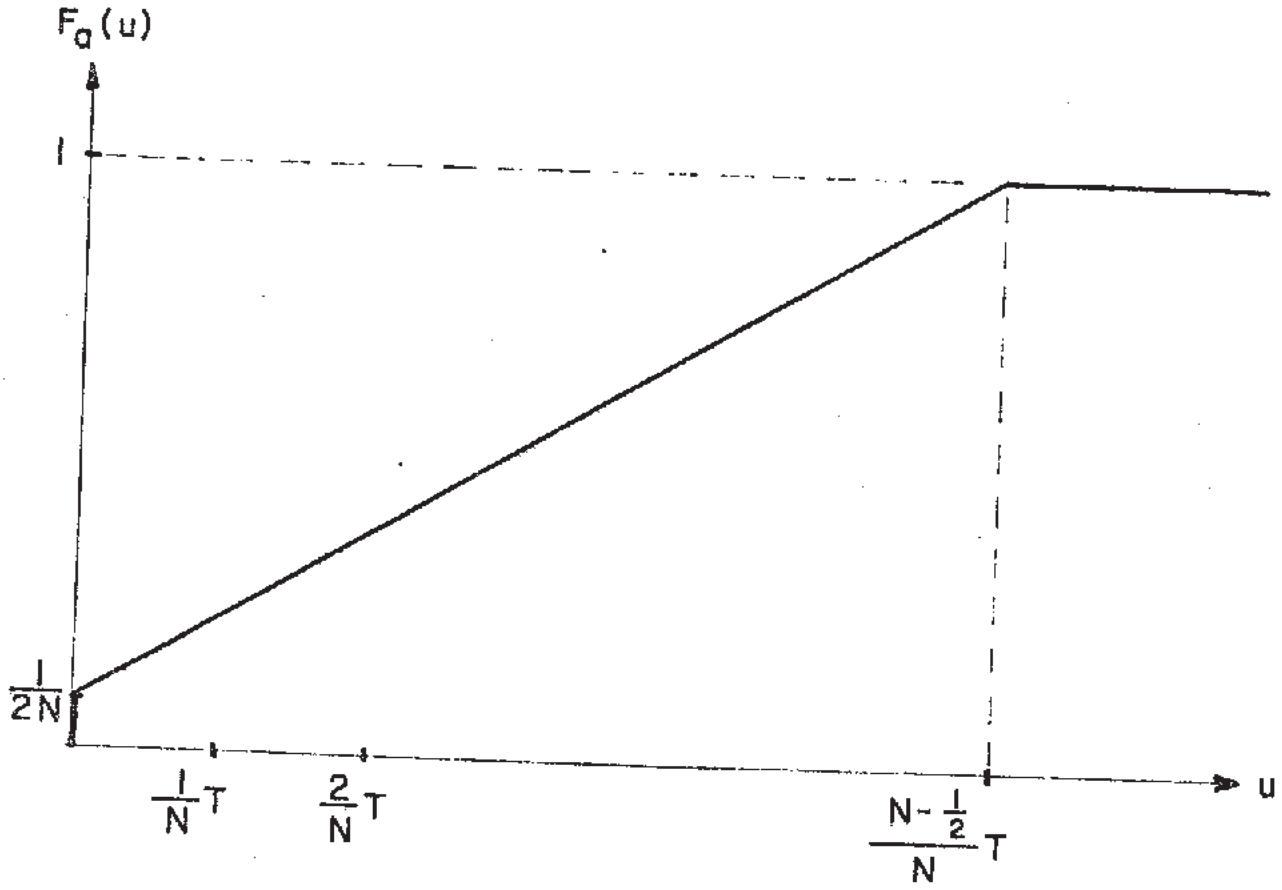
LIST OF FIGURE CAPTIONS

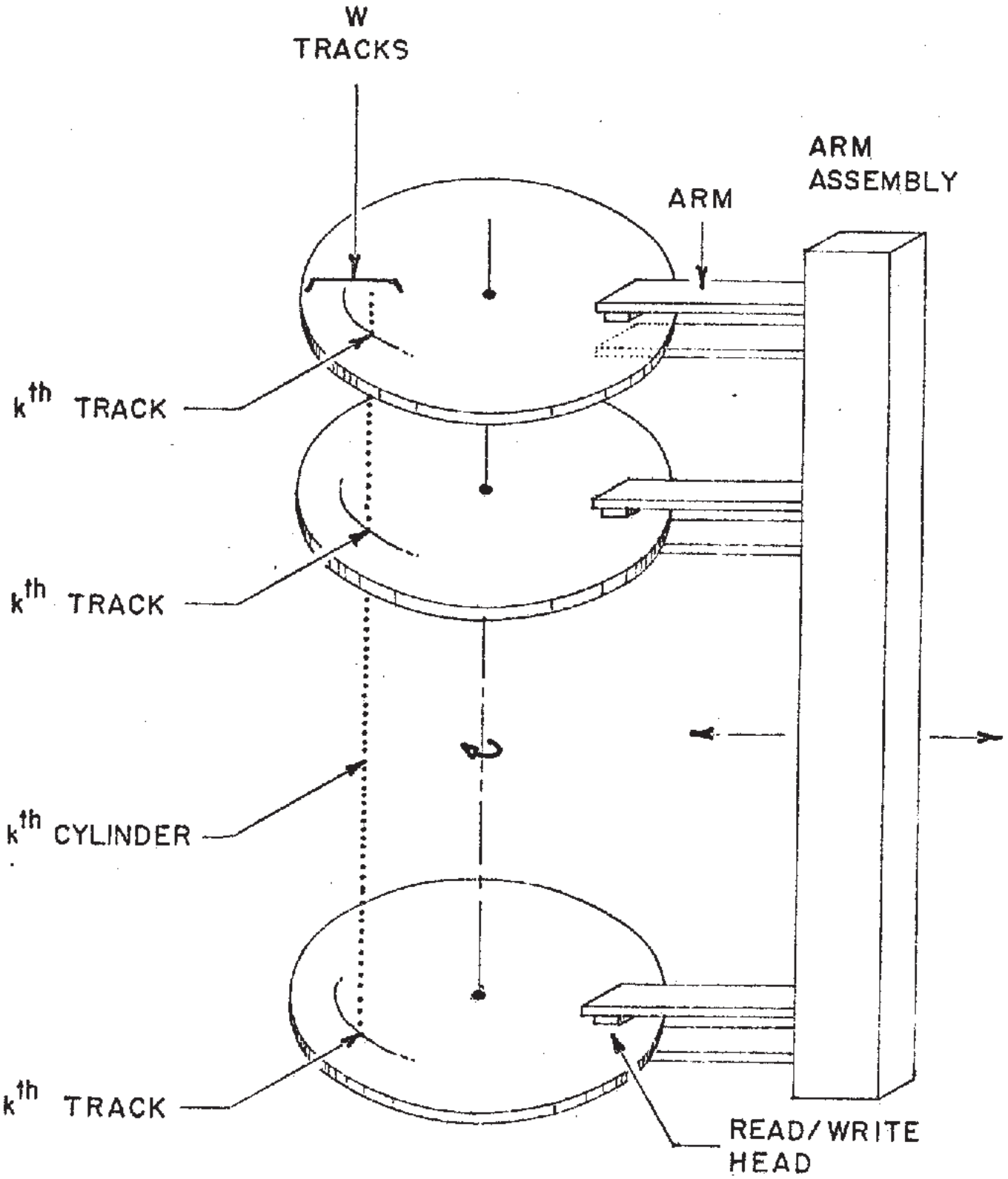

REQUEST ENTERS QUEUE
REQUEST CHOSEN (DECISION POINT)
BEGIN OF TRANSFER
REQUEST COMPLETE
W_q
a
t_o
t_s
W_F

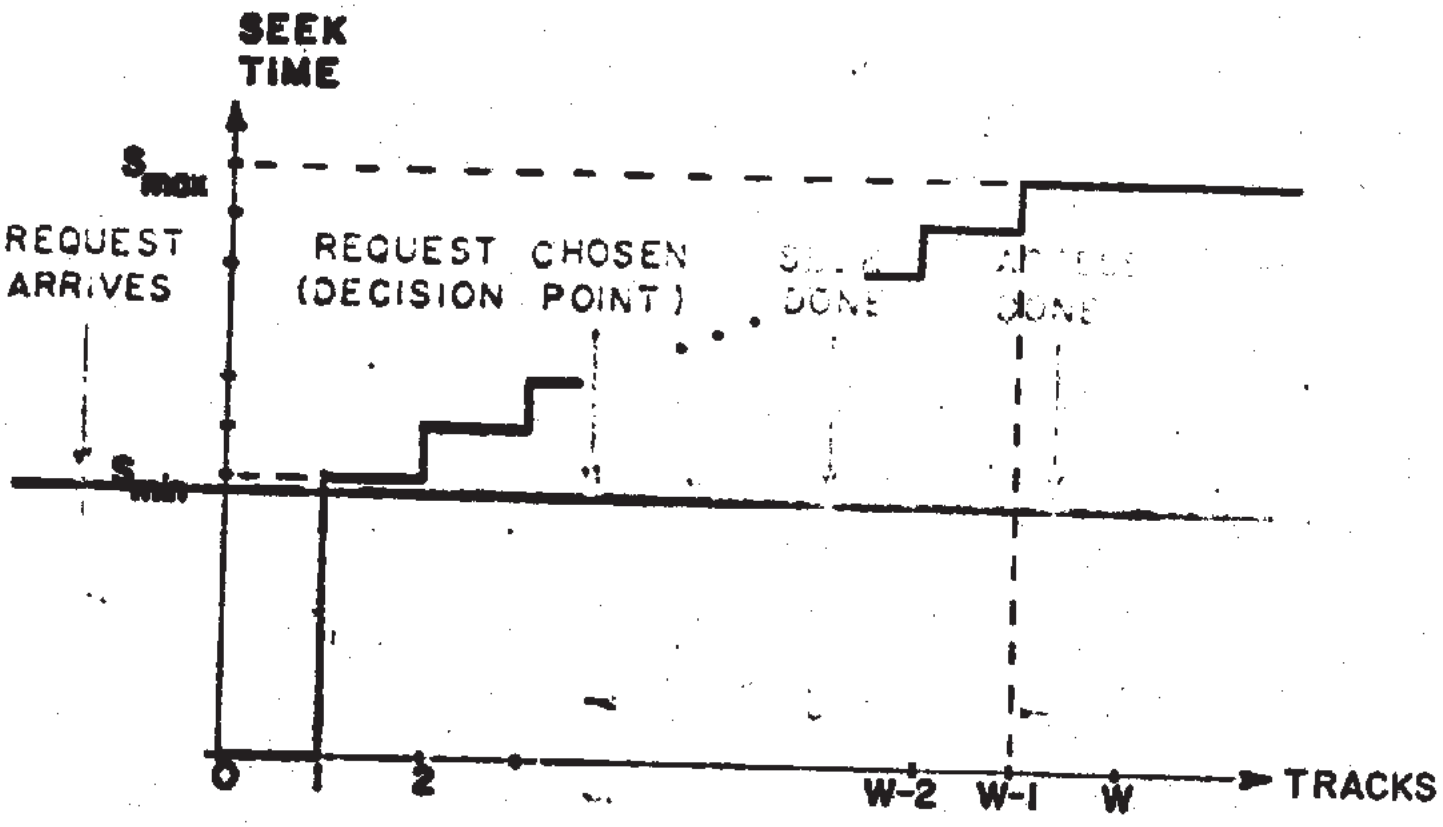

CENTRAL
FILE
CONTROL
QUEUES
q_1
q_2
q_m
Switching Time
= Access Time
REQUESTS
MAIN
(CORE)
MEMORY
SECONDARY
(FILE)
MEMORY
Information Flow
(Single Channel)

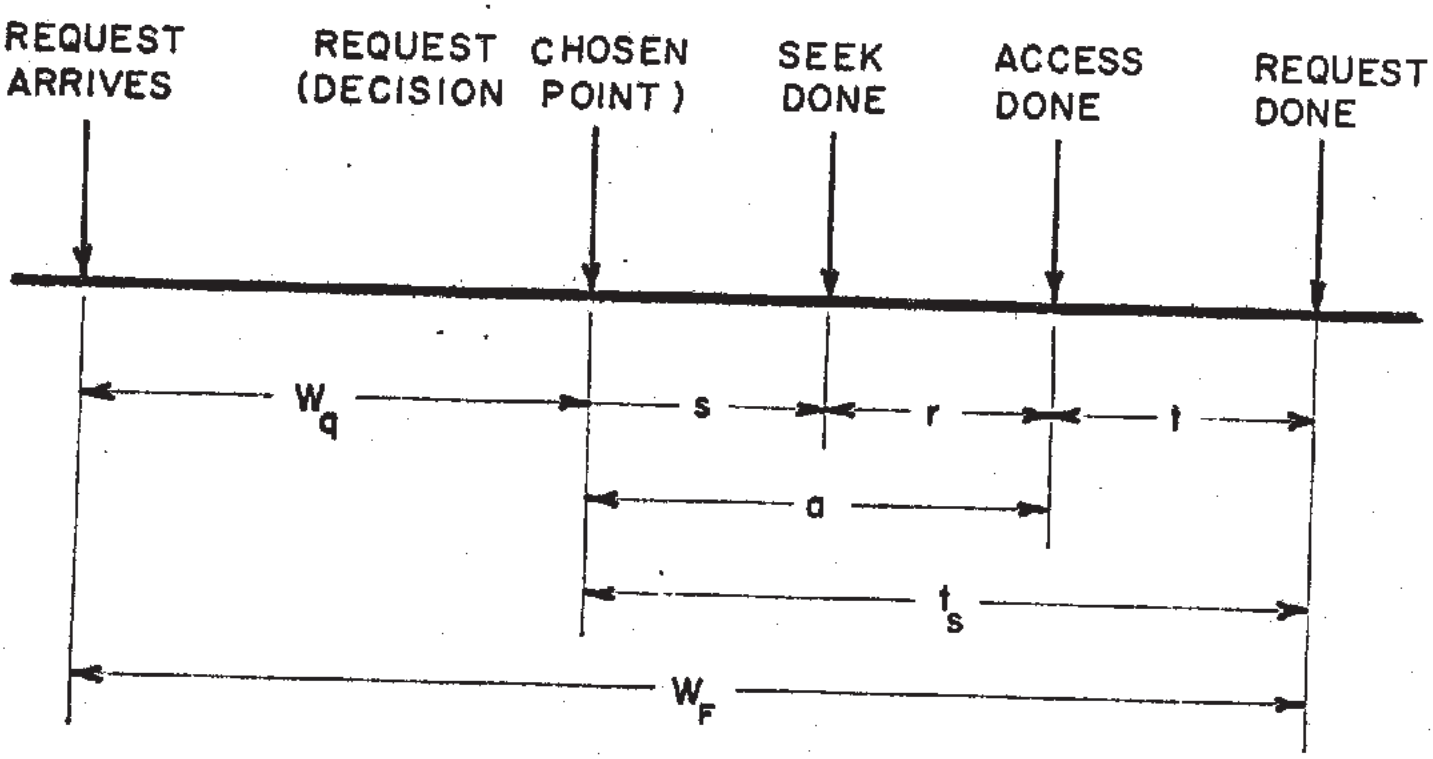

READ/
WRITE
HEADS
PAGE
TRACKS
PER
FIELD
FIELD
SECTOR

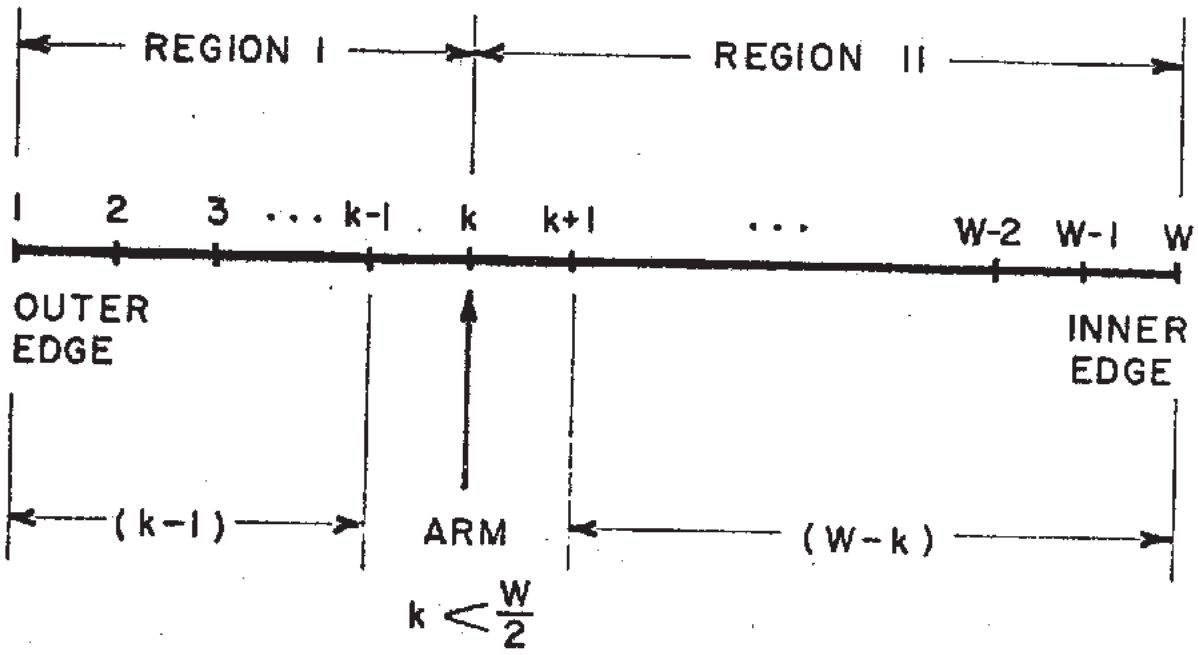

$F_a(u) = Pr\left[a \leq u\right]$
1
$\frac{N-1}{N}$
$\frac{2}{N}$
$\frac{1}{N}$
$\frac{1}{N}T$
$\frac{2}{N}T$
$\frac{N-2}{N}T$
$\frac{N-1}{N}T$
T
u

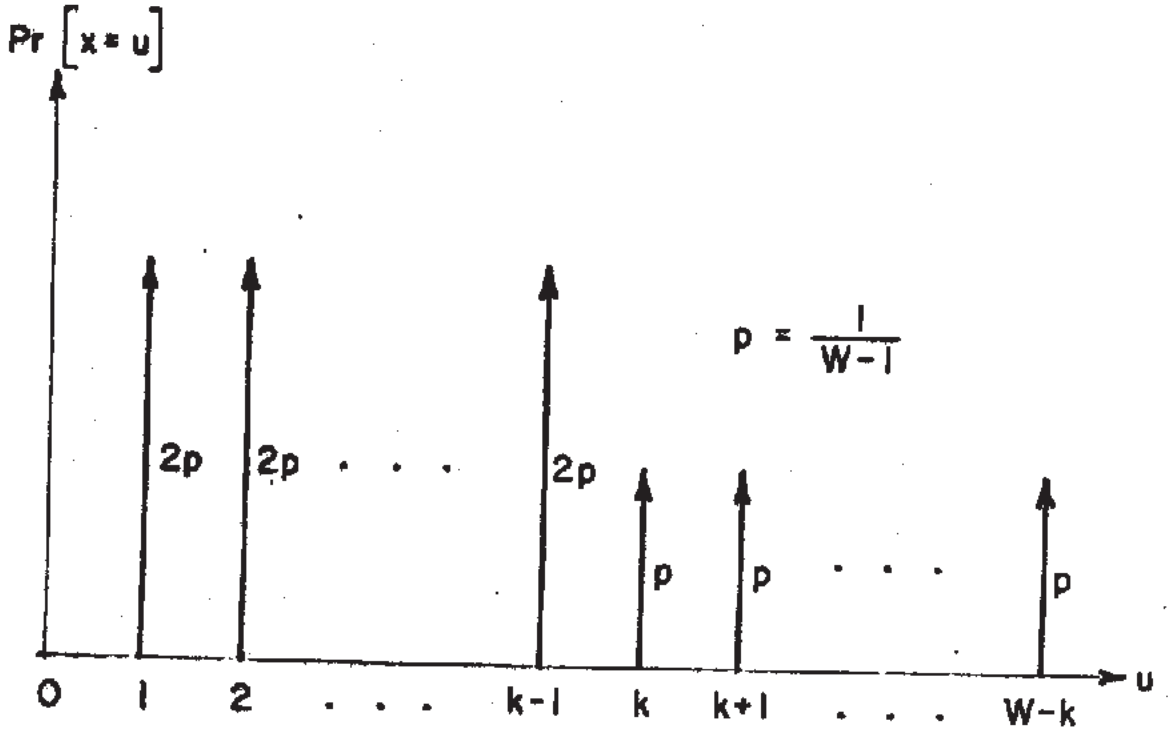

$F_a(u)$
1
$\frac{1}{2N}$
$\frac{1}{N}T$
$\frac{2}{N}T$
$\frac{N-\frac{1}{2}}{N}T$
u

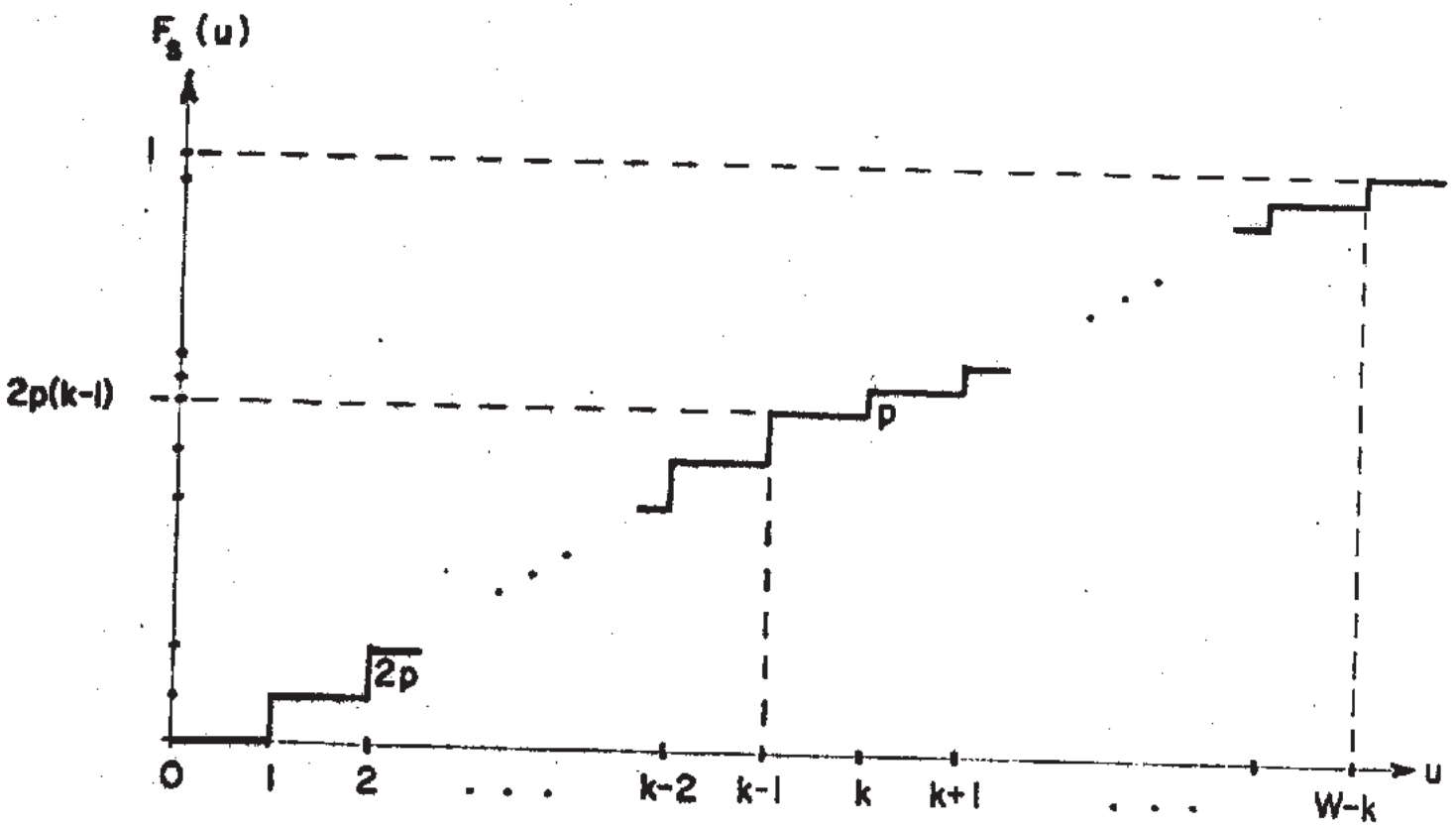

W
TRACKS
ARM
ASSEMBLY
ARM
k^th TRACK
k^th TRACK
k^th CYLINDER
k^th TRACK
READ/WRITE
HEAD

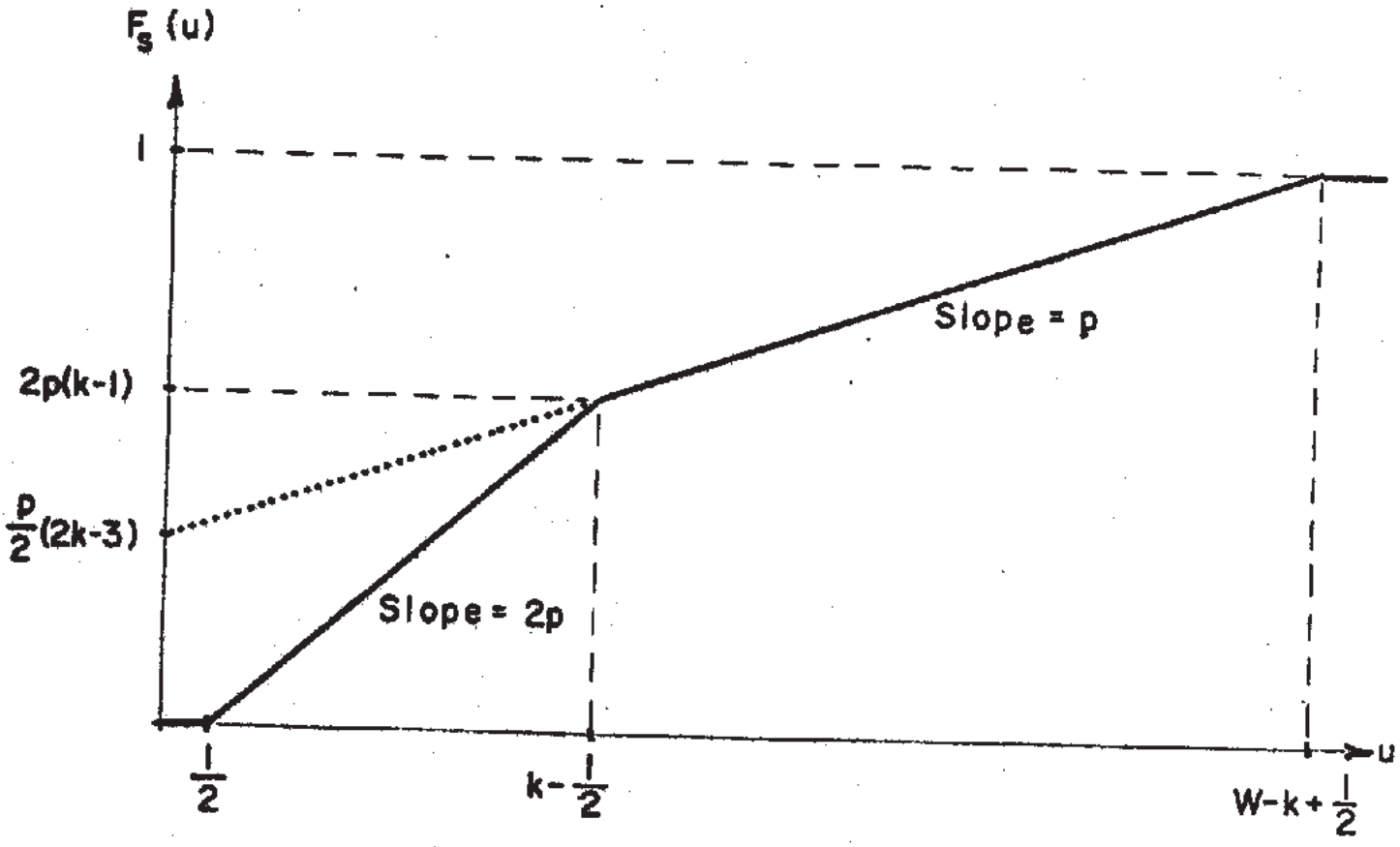

SEEK TIME
S_{max}
REQUEST ARRIVES
REQUEST CHOSEN (DECISION POINT)
S_{min}
0
2
W-2
W-1
W
TRACKS

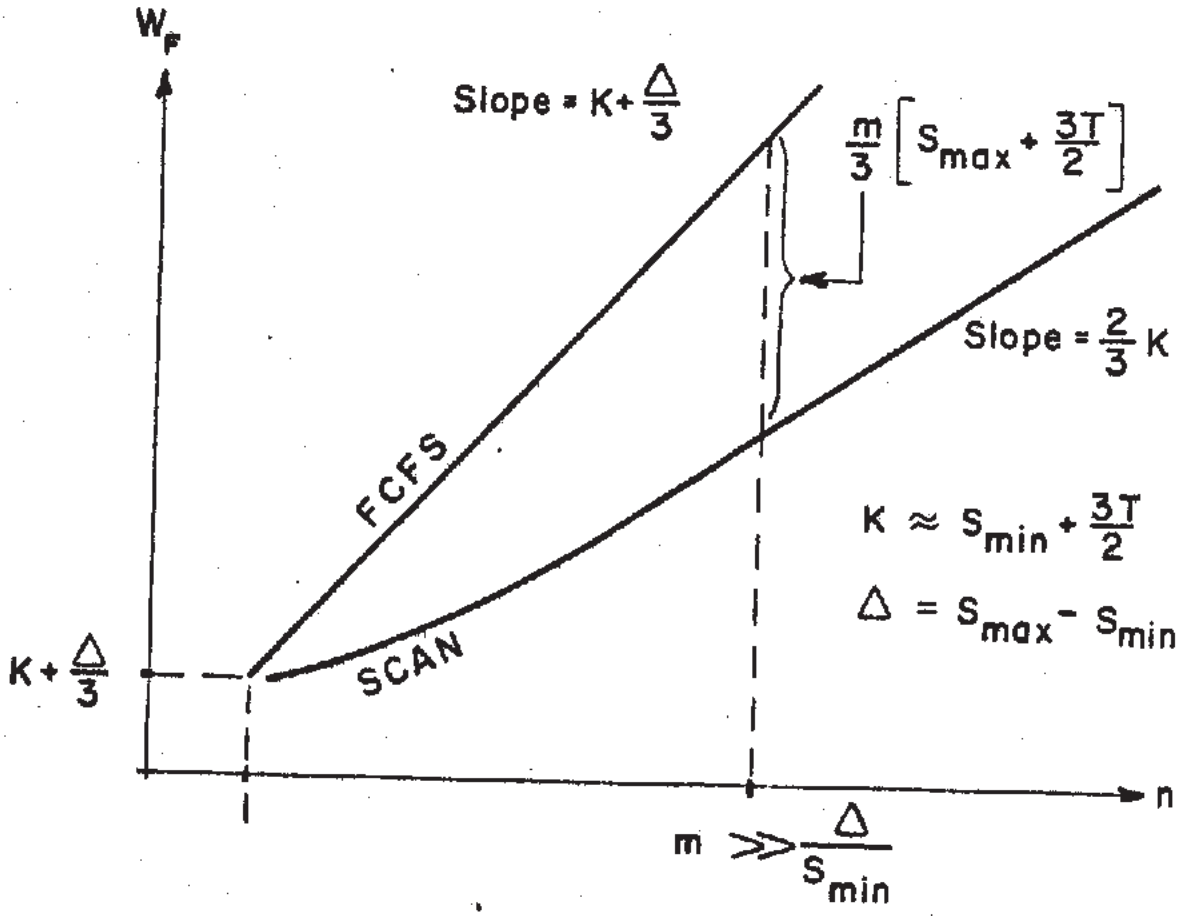

REQUEST ARRIVES
REQUEST (DECISION
CHOSEN POINT)
SEEK DONE
ACCESS DONE
REQUEST DONE
W_q
s
r
t
a
t_s
W_F


REGION I
REGION II
1
2
3
k-1
k
k+1
W-2
W-1
W
OUTER EDGE
INNER EDGE
(k-1)
ARM
(W-k)
$k < \frac{W}{2}$


Pr [x = u]
2p
2p
2p
p
p
p
p = 1/(W-1)
0
1
2
k-1
k
k+1
W-k
u


$F_s(u)$
1
2p(k-1)
p
2p
0
1
2
. . .
k-2
k-1
k
k+1
. . .
W-k
u


$F_s(u)$
1
$2p(k-1)$
$\frac{p}{2}(2k-3)$
Slope = p
Slope = 2p
$\frac{1}{2}$
$k-\frac{1}{2}$
$W-k+\frac{1}{2}$
u


W_F
Slope = $K + \frac{\Delta}{3}$
$\frac{m}{3}\left[S_{max} + \frac{3T}{2}\right]$
Slope = $\frac{2}{3}K$
FCFS
SCAN
$K \approx S_{min} + \frac{3T}{2}$
$\Delta = S_{max} - S_{min}$
$K + \frac{\Delta}{3}$
n
$m \gg \frac{\Delta}{S_{min}}$