Free store distribution under random fit allocation: Part 2

C. M. Reeves

Department of Computer Science, University of Keele, Keele, Staffs. ST5 5BG

The paper extends the earlier analysis in Part I of external fragmentation in dynamic storage allocation under the random fit strategy. A duality is exposed between free store fragmentation and the clustering of reservations. A revised prediction, in the case of single word reservations, of a lower threshold of 48% store utilisation to yield a stable free store profile, is confirmed by simulation experiments. A new theoretical model is developed for low store utilisations.

(Received February 1979)

1. Introduction and summary

An earlier paper (Reeves, 1979) developed a mathematical analysis of external fragmentation in dynamic storage allocation under the random fit strategy. The size distribution of free fragments was represented by a generating function

$$\tilde{\phi}(a) = \sum_r \phi_r a^r \tag{1.1}$$

where ϕ_r denotes the proportion of free blocks having length r. Equations were derived relating $\tilde{\phi}$ to the analogous function $\tilde{b}$ describing the size distribution of incoming requests for storage space.

Key parameters in the theory were denoted by k and p_2. The block utilisation ratio k is the ratio of the number B of reservations to the number F of free fragments, and p_2 is the proportion of lone reservations, that is the ratio of the number of reserved blocks bounded on each side by a free block to the total number B of reserved blocks.

An argument was advanced to express p_2 as a function of k, namely

$$p_2 = (ke^{k-1})^{-1} \tag{1.2}$$

The derivation was felt to be unsatisfactory in that it was probabilistic in nature. The hope was expressed that it could be replaced by a direct analysis of the store allocation and deallocation mechanisms.

The case in which all requests are for single locations, namely

$$\tilde{b}(a) = a \tag{1.3}$$

was treated in detail and it was shown that the existence of a stable solution $\tilde{\phi}$ requires a store utilisation of at least 37%. The present paper continues the analysis. Section 2 contains the hoped for re-examination of the dependence of p_2 upon k. This emerges from a comprehensive analysis of the clustering of reservations into contiguous sequences. The analysis has a close similarity to the earlier analysis of the free store, with the number of reservations in a contiguous sequence corresponding with the number of adjacent locations comprising a free block. There are in a sense two dual systems co-existing in the store, loosely coupled by two functional relationships—one being Knuth's 50% rule, and the other the $p_2 - k$ dependence. The overall model has an attractive cohesion and symmetry which, alas, is obvious only in retrospect.

The revised form for p_2 is

$$p_2 = \frac{2}{k(1 + 2k)} \tag{1.4}$$

and in Section 3 its implications for the analysis of unit size requests are examined. The section is in essence a reworking of the corresponding part of the earlier paper. Surprisingly, the changes are qualitatively insignificant. All the main features persist. In particular there remains a lower limit on store utilisation for the existence of solutions. The position of this threshold rises however to 48%.

Section 4 explores the wider implications of the new model of reserved block clustering. A major feature of the model is its independence of $\tilde{b}$. The numbers of reservations in different contiguous sequences are distributed quite independently of any variations in the size of individual reservations. The proportion of sequences of given length is found to decrease with increasing length in approximately geometrical progression.

In order to confirm the many theoretical predictions of store behaviour, a number of simulation experiments were performed. Section 5 describes the simulation model which was coded in ALGOL 60 and run over Keele's link to the ICL 1906S computer at Liverpool University. The results obtained are summarised in Section 6. These broadly confirm the predictions at the higher utilisation levels.

At low utilisations the free store was observed to break up into one large segment and a number of small fragments. This suggested that an appropriate model for a lightly loaded store might be obtained by considering an infinite store finitely loaded. In Section 7 such a model is developed for the special case of unit length requests. The $p_2 - k$ dependence is unaffected but the 50% rule suffers minor modification. The predictions of the model are in good agreement with the results of simulation at low store loading. It is reasoned that the applicability of the model may extend up to about 25% utilisation. The middle range from 25% to 50% is a transitional region exhibiting considerable variability in its behaviour.

Two areas emerge for future study. Shore (1977) has noted significant departures from the 50% rule in his analysis of the results of earlier (Shore, 1975) simulations. In the present work the rule and its low utilisation modification are in reasonable accord with the experimental results over their whole range but they carry nothing like the conviction of, for example, the relationship of equation (1.4). The simulations disclose a consistently higher proportion of isolated free locations than is predicted.

For the future, it is perhaps most important to extend the models into the region of variable request sizes. The unit size special case has provided valuable insights but represents very much the nursery slopes of the problem. It may reasonably be expected that work on more general functions will shed light on such particular problems as the 50% rule and the transition region of store utilisation.

2. Equilibrium under random fit

Equations are derived in this section relating the equilibrium distribution of reserved blocks into contiguous sequences to the block utilisation ratio, k. The terminology and notation of the previous paper are carried over here, together with the necessary distinctions between average values and the actual values of random variables.

It will be remembered that in order to avoid consideration of the effects at the two ends of a linear store, the store is assumed to be circular. Previously the emphasis was upon the size of reservations and the lengths of the segments into which free store is fragmented. Here the interest centres upon the numbers of blocks irrespective of their size. Of particular concern is the clustering of reservations into contiguous sequences separated by isolated free blocks. (Two or more contiguous free blocks are automatically compacted into one).

The number of sequences of r reserved blocks, bounded at each end by a free block, is denoted by c_r. The total number of reserved blocks is B and of free blocks is F so that

$$\sum c_r = F \tag{2.1}$$

and

$$\sum_r rc_r = B \tag{2.2}$$

The block utilisation ratio referred to above is defined as

$$k = B/F \tag{2.3}$$

The probability that a randomly selected reserved sequence contains r blocks is denoted by σ_r where

$$\sigma_r = c_r/F \tag{2.4}$$

so that

$$\sum_r \sigma_r = 1 \tag{2.5}$$

and

$$\sum_r r\sigma_r = k \tag{2.6}$$

The generating function technique which proved valuable before is equally helpful here. The function corresponding to the σ distribution is defined as

$$\tilde{\sigma}(a) = \sum_r \sigma_r a^r \tag{2.7}$$

so that equations (2.5, 2.6) can be rephrased as

$$\tilde{\sigma}(1) = 1 \tag{2.8}$$

and

$$\tilde{\sigma}'(1) = k \tag{2.9}$$

To avoid possible confusion, it may be as well to remark again that here the terms in the expansion of $\tilde{\sigma}$ correspond to block counts, not block sizes as previously. As a corollary, all the argument of this section applies generally, whatever may be the distribution of reservations by size.

2.1 *Block reservation*

A request is satisfied, under the assumed random fit strategy, by choosing at random a sufficiently large free block. It is assumed that free blocks of sufficient size always exist. Following Knuth (1968), let p denote the probability that a request is satisfied by an oversize free block. The probability of an exact fit is then $1 - p$.

2.1.1 *Oversize fit*

One of the neighbouring reserved sequences is augmented by the new reservation. The spatial distribution of reserved sequences is assumed random so that the probability is σ_r that it is initially of length r. As a result of the reservation, the counts of reserved blocks are amended to

$$c_r \leftarrow c_r - 1 \tag{2.10}$$

and

$$c_{r+1} \leftarrow c_{r+1} + 1 \tag{2.11}$$

2.1.2 *Exact fit*

The two neighbouring reserved sequences are merged with the new reservation to form a single sequence, and F is reduced by one. The lengths r and s of the initial sequences arise with probabilities σ_r and σ_s, assuming independence, and the counts are amended to

$$c_r \leftarrow c_r - 1 \tag{2.12}$$

$$c_s \leftarrow c_s - 1 \tag{2.13}$$

and

$$c_{r+s+1} \leftarrow c_{r+s+1} + 1 \tag{2.14}$$

Accordingly, the expected increment to $\tilde{c}$ following a single reservation is

$$\Delta^+\tilde{c} = p\sum_r \sigma_r(a^{r+1} - a^r) + (1-p)\sum_r\sum_s \sigma_r\sigma_s(a^{r+s+1} - a^r - a^s) \tag{2.15}$$

$$= p(a-1)\tilde{\sigma} + (1-p)(a\tilde{\sigma}^2 - 2\tilde{\sigma}) \tag{2.16}$$

2.2 *Block release*

All reserved blocks are equally likely to be released next. The three categories discussed previously are relevant here.

2.2.1 *Category* 0

These are the interior members of a reserved sequence. The effect of releasing the u^{th} such block in a sequence of r blocks, $1 \leqslant u \leqslant r - 2$ with the blocks numbered 0 to $r - 1$, is to split the sequence into two, one of length u and the other of length $r - u - 1$. F is increased by one and the counts are amended to

$$c_r \leftarrow c_r - 1 \tag{2.17}$$

$$c_u \leftarrow c_u + 1 \tag{2.18}$$

and

$$c_{r-u-1} \leftarrow c_{r-u-1} + 1 \tag{2.19}$$

2.2.2 *Category* 1

These are the end members of reserved sequences of length two or more. The release of such a block reduces the sequence length, r say, by one. The counts are amended to

$$c_r \leftarrow c_r - 1 \tag{2.20}$$

and

$$c_{r-1} \leftarrow c_{r-1} + 1 \tag{2.21}$$

2.2.3 *Category* 2

These are the lone reserved blocks. A release decreases c_1 by one and F by one.

$$c_1 \leftarrow c_1 - 1 \tag{2.22}$$

The expected increment to $\tilde{c}$ following a single release is thus seen to be

$$\Delta^-\tilde{c} = \frac{1}{B}\left\{\sum_{r>2} c_r \sum_{u=1}^{r-2}\left(a^u + a^{r-u-1} - a^r\right) + 2\sum_{r>1} c_r(a^{r-1} - a^r) - c_1 a\right\} \tag{2.23}$$

which, after some algebra, reduces to

$$\Delta^-\tilde{c} = \frac{1}{k(1-a)}\left\{2(a - \tilde{\sigma}) - a(1-a)\tilde{\sigma}'\right\} \tag{2.24}$$

2.3 *Equilibrium conditions*

In equilibrium, reservations and releases take place at the same mean rate. The conditions for equilibrium are thus seen to be

$$\Delta^+\tilde{c} + \Delta^-\tilde{c} = 0 \tag{2.25}$$

2.3.1 *The fifty percent rule revisited*

As a check on these derivations, setting $a = 1$ in equation (2.25) and using equation (2.1) yields

$$\Delta^+F + \Delta^-F = 0 \qquad (2.26)$$

which expresses the constancy of the count of free blocks. Substituting into equations (2.16, 2.24) gives

$$\Delta^+F = -(1 - p) \qquad (2.27)$$

and

$$\Delta^-F = \lim_{a \to 1} U/V \qquad (2.28)$$

where

$$U = 2(a - \tilde{\sigma}) - a(1 - a)\tilde{\sigma}' \qquad (2.29)$$

and

$$V = k(1 - a) \qquad (2.30)$$

Here both U and V tend to zero and so the limiting value of their ratio is obtained by differentiating each before proceeding to the limit.

$$\Delta^-F = \lim_{a \to 1} U'/V' = 1 - 2/k \qquad (2.31)$$

It then follows that

$$p = 2/k \qquad (2.32)$$

which is the equivalent of Knuth's 50% rule.

2.3.2 *The equilibrium equations*

The principal result, equation (2.32), of the last section, together with the substitution

$$x = 2/k \qquad (2.33)$$

introduced in the previous paper, allows the equilibrium conditions, equation (2.25), to be presented in the form

$$ax(1 - a)\tilde{\sigma}' - 2a(1 - a)(1 - x)\tilde{\sigma}^2 + 2(2 - 2a + a^2x)\tilde{\sigma} - 2ax = 0 \qquad (2.34)$$

A general discussion of this equation and its solution is deferred to Section 4.

2.4 *Lone reserved blocks*

The particular motivation for the work described in the present Section 2 was to determine the proportion of reserved blocks in category 2—i.e. lone reserved blocks. The previous paper denoted by n_2 the number of lone reserved blocks whereas in the present analysis the same count is denoted by c_1. The parameter p_2, denoting the probability that a random reserved block is lone, played a key role in the analysis of the lengths of free blocks and, using equation (2.33) of the previous paper and (2.4) of the present one, it follows that

$$p_2 = \sigma_1/k \qquad (2.35)$$

By extracting the coefficient of a in equation (2.34) it is seen that

$$\sigma_1 = \frac{2x}{x + 4} \qquad (2.36)$$

so that, in terms of k,

$$p_2 = \frac{2}{k(1 + 2k)} \qquad (2.37)$$

This is the relation that must replace equation (2.34) of the previous paper.

3. Unit size requests: a special case

This section indicates the changes required in the corresponding Section 3 of Part 1 in the light of the revised relationship between p_2 and k. For convenience of cross reference, in this section only, equations are numbered to correspond with those in the previous paper.

The condition on store utilisation in order that solutions shall exist is amended to

$$0{\cdot}4849 \leqslant \theta \leqslant 1 \qquad (3.21)$$

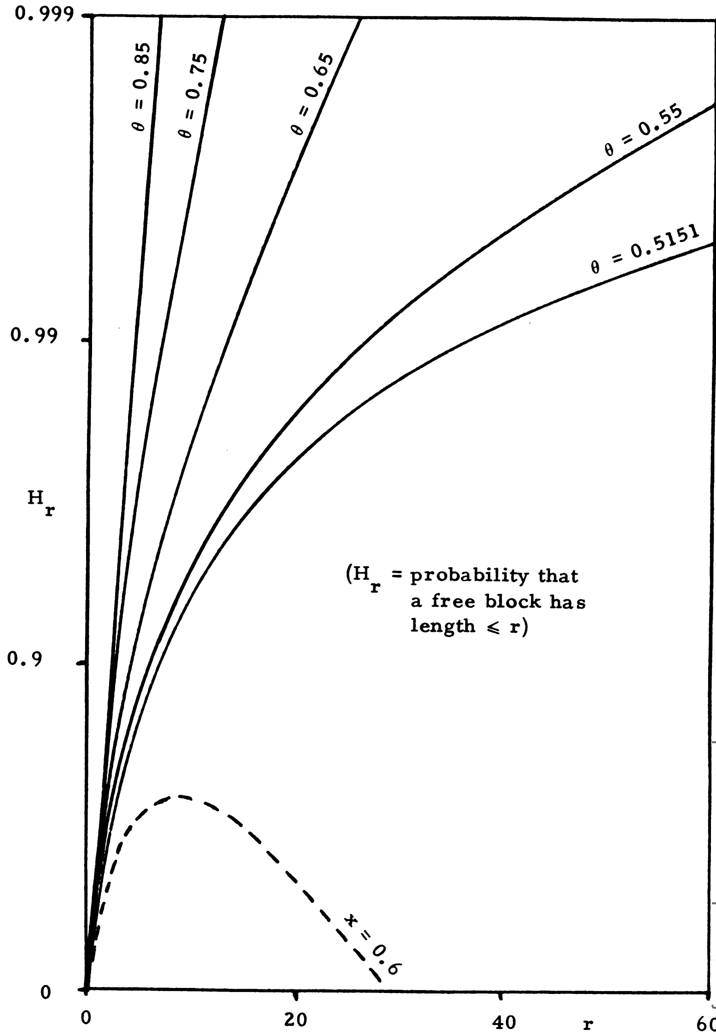

Fig. 1

Table 1 Parameter sets for plotting H_r curves

θ	x	k	p_2	$\bar{r}$
0·95	0·0950	21·0515	0·0022	1·1080
0·85	0·2553	7·8326	0·0153	1·3822
0·75	0·3770	5·3046	0·0325	1·7682
0·65	0·4608	4·3400	0·0476	2·3369
0·55	0·5065	3·9484	0·0569	3·2306
0·4849	0·5151	3·8825	0·0588	4·1249
—	0·6	3·3333	0·0783	—

with corresponding ranges for the other major parameters as follows

$$3{\cdot}8825 \leqslant k < \infty \qquad (3.22)$$

$$0{\cdot}5151 \geqslant x \geqslant 0 \qquad (3.23)$$

$$0{\cdot}0588 \geqslant p_2 \geqslant 0 \qquad (3.24)$$

$$4{\cdot}1249 \geqslant \bar{r} \geqslant 1 \qquad (3.25)$$

It is predicted therefore that no stable solution exists for less than 48% store utilisation. The high utilisation approximation is unchanged. The analysis in the earlier paper of the complete solution is also unchanged, though the revised form for p_2 is now to be understood. This has the effect of replacing **Fig. 1** and **Table 1** by those in the present paper.

The remarkable feature of these various changes is their qualitative insignificance. The numbers have changed somewhat: in particular, the lower limit of store utilisation has moved from 37% to 48%. The important features are the existence of such a threshold and the general characteristics of the distributions. These have survived.

4. The general sequence—length distribution

This section returns to treat equation (2.34) which determines the equilibrium distribution of the lengths of contiguous sequences of reserved blocks. The treatment is general in that it is not restricted to any particular distribution of the size of individual reservations. Sequence length in this context means the number of blocks, not the amount of store occupied. In particular therefore, there is no *a priori* reason for anticipating a more restrictive range for the parameter x (such as equation (3.23) provides) than that implied by the 50% rule (equations 2.32, 2.33) namely

$$0 \leqslant x \leqslant 1 \tag{4.1}$$

By forming a series expansion in powers of a of the left hand side of equation (2.34) and setting each coefficient to zero, a set of recurrence equations is produced as follows, for $n \geqslant 1$.

$$(nx + 4)\sigma_n - (nx - x + 4)\sigma_{n-1} + 2x\sigma_{n-2} - 2(1 - x)(S_{n-1} - S_{n-2}) - 2x\delta_{n1} = 0 \tag{4.2}$$

where S_n is the coefficient of a^n in $\tilde{\sigma}^2$, namely

$$S_n = \sum_{u=1}^{n-1} \sigma_u \sigma_{n-u} \tag{4.3}$$

and δ is the Kronecker delta.

The case $n = 1$ has already been used in Section 2.4 to relate p_2 and k. More generally these equations may be used with successive n values to generate the sequence of σ values for any x. By equation (2.6) the mean sequence length is k.

4.1 *The extreme case* $x = 1$

The equilibrium equation, equation (2.34), takes a simple form in the case $x = 1$ corresponding to minimal block utilisation ratio. Here the nonlinear term in $\tilde{\sigma}^2$ drops out, leaving

$$a(1 - a)\tilde{\sigma}' + 2(2 - 2a + a^2)\tilde{\sigma} - 2a = 0 \tag{4.4}$$

This may be solved analytically to give

$$\tilde{\sigma} = \frac{1}{2a^4}\left\{3(1 - a)^2 e^{2a} + 2a^3 + 3a^2 - 3\right\} \tag{4.5}$$

and then the solution may be expanded to powers of a to give

$$\sigma_n = 3 \,.\, 2^{n+1}\frac{n(n + 3)}{(n + 4)!} \tag{4.6}$$

4.2 *Numerical solutions*

As with the free block distribution, the chosen representation for the sequence length distribution is in terms of its cumulative distribution function L defined by

$$L_n = \sum_{u=1}^{n} \sigma_u \tag{4.7}$$

Fig. 2 shows the variation of L_n with n for a range of x values computed from the recurrence equations (4.2) for successive values of n. For the sake of consistency with Fig. 1, the same logarithmic scale was used in plotting L_n. The degree of linearity of each curve was quite unexpected. It is most pronounced when x is small, corresponding to high store utilisation. This is the region of greatest operational economy and so it is worth investigating the form of an analytical approximation for use in preference to numerical solutions.

4.3 *Two geometrical approximations*

The name 'geometrical approximation' is appropriate both to acknowledge that it was suggested by examination of a graph and because the linearity observed in the log scale indicates that the points plotted are in geometrical progression. The situation is closely analogous to that in the high utilisation approximation in Section 3.3. The appropriate relation is

$$L_n^* = 1 - \eta^n \tag{4.8}$$

where asterisks, both here and subsequently, indicate an approximation. It follows that

$$\sigma_n^* = \eta^{n-1} - \eta^n \tag{4.9}$$

so that

$$\tilde{\sigma}^* = \frac{(1 - \eta)a}{1 - a\eta} \tag{4.10}$$

At this stage a criterion of fit is required in order that η may be related to x. Two possibilities suggest themselves.

Firstly, with the objective of obtaining a fit over the whole range of n, η may be chosen to give the theoretical mean sequence length. By equation (2.9) this requires

$$\tilde{\sigma}^{*\prime}(1) = 2/x \tag{4.11}$$

leading to

$$\eta = 1 - x/2 \tag{4.12}$$

The case $n = 1$ is of particular significance since it is the value of σ_1 which gives the parameter p_2 for use in the analysis of free block lengths as set out in Section 2.4. There is therefore a case for approximating L_n in such a way as to preserve equation (2.36). This leads to an alternative choice of η, namely

$$\eta = 1 - \frac{2x}{x + 4} \tag{4.13}$$

which gives a mean sequence length of $k + \frac{1}{2}$ in place of k.

A comparison of values of L_n for various n and x is made in **Table 2.** Here L_n refers to the numerical solution, L_n^* to the approximation determined by equation (4.12) and L_n^{**} to that determined by equation (4.13).

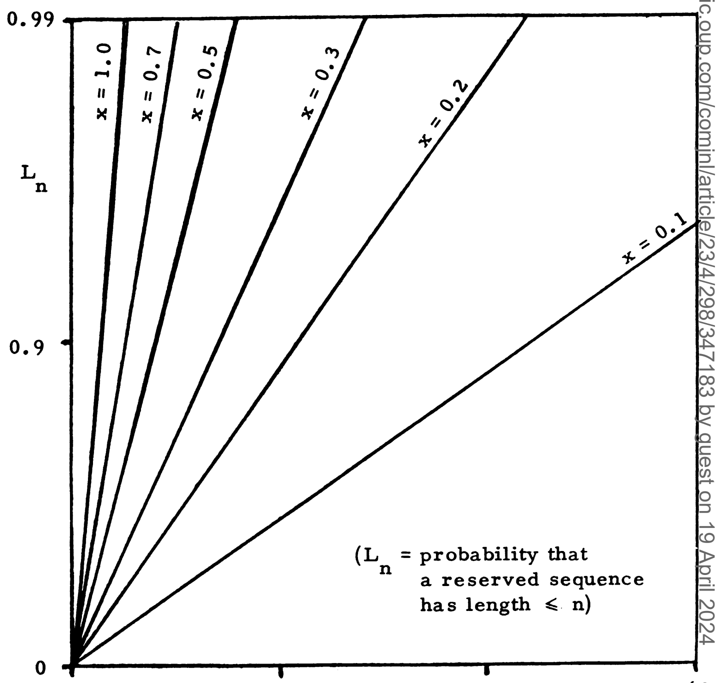

Fig. 2

Table 2 Some sample L-values for comparison

x	n	L_n	L_n^*	L_n^{**}
1·0	1	0·4000	0·5000	0·4000
	3	0·9048	0·8750	0·7840
	6	0·9983	0·9844	0·9533
0·4	1	0·1818	0·2000	0·1818
	15	0·9668	0·9648	0·9507
	30	0·9989	0·9988	0·9976
0·1	1	0·0488	0·0500	0·0488
	30	0·7855	0·7854	0·7769
	60	0·9541	0·9539	0·9502

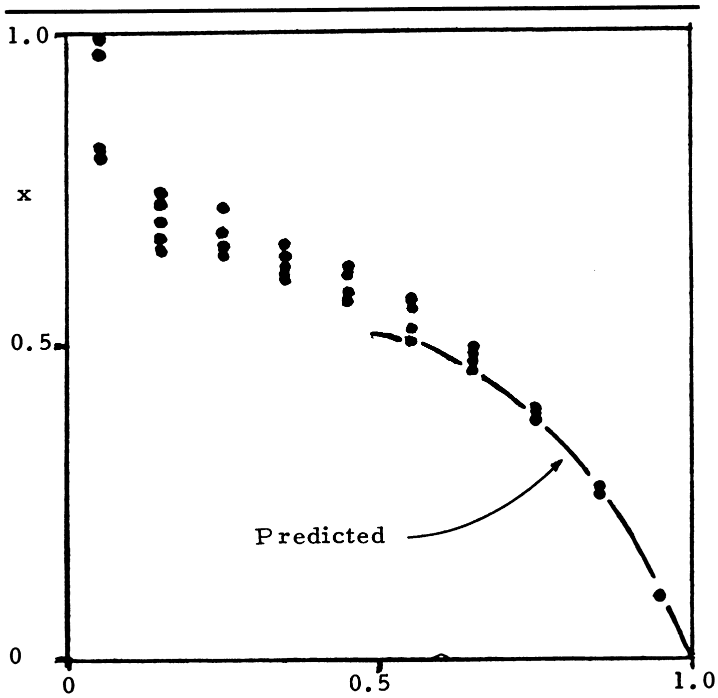

Fig. 3

5. A simulation model for unit size requests

A simulation model, restricted to unit size requests, is now developed for comparison with the analytical treatment. So that the results obtained by its use, and described in Section 6, may carry conviction, it seems desirable to give a brief description of the model in the present section.

Each cycle of the simulation adjusts the store configuration for the effects of one random request followed by one random release. In this way the contribution to the variance in the results that would arise from temporal fluctuation in store utilisation is eliminated. It is the spatial distributions in equilibrium that are of interest. An experiment consists of three phases

(*a*) initial loading to achieve the desired store utilisation

(*b*) a warm-up period of a specified number of cycles to achieve a suitable 'realistic' configuration

(*c*) the run proper, consisting of a specified number of metered cycles, where metering consists of accumulating the necessary sums after each cycle to enable means of the components of the f and c counts to be estimated, f_r being the number of free blocks of length r and c_r the number of sequences of r reservations.

The size of the store, N, and number of reservations, B, are fixed parameters of each experiment. The number of free blocks, F, will fluctuate. The store configuration at the start of each cycle is conveniently represented by a pair of integer arrays U, V [0: $F - 1$] with subscripts corresponding to spatial position in the circular store. Here U_r is the length of the rth free block, and V_r is the length of the rth sequence of contiguous reserved blocks (each of unit length). The assumed ordering of the store is

$$\ldots\ \ldots\ U_{F-1}\ V_{F-1}\ U_0\ V_0\ U_1\ V_1 \ldots\ \ldots$$

5.1 *Reservations*

The free block to which a request is allocated is selected at random. Thus

$$r \leftarrow \text{int}(\text{F} \times \text{random}) \qquad (5.1)$$

where *random* at each reference yields a fresh sample value from the uniform distribution in (0.1), and the function *int* yields the integer part of its argument.

Having thus determined that the rth free block is to be used, it is necessary to test whether it produces an exact fit ($U_r = 1$) or is oversize ($U_r > 1$). In the latter case a decision is required whether to allocate space for the request from the left or right end. Shore (1977) has noted the importance of this decision and so the opportunity is created here for exploring three strategies,

(*a*) systematic placement in which the left hand end is always used,

(*b*) alternating placement in which alternately left and right hand ends are used, starting with the left hand end, and

(*c*) random placement in which the end is selected at random with equal probabilities.

It may be noted that as alternatives to the random fit allocation strategy of equation (5.1), two other strategies may easily be explored. In the 'first fit' case, equation (5.1) is replaced by

$$r \leftarrow 1 \qquad (5.2)$$

and in 'next fit' by

$$r \leftarrow (r + 1) \bmod F \qquad (5.3)$$

where $r = 1$ initially. Strictly of course each of these two alternatives should be accompanied by a change in the distribution of releases but these variations alone will give some useful indications of performance.

5.2 *Release*

The block to be released is selected at random. Thus

$$r \leftarrow 0;\ s \leftarrow \text{int}((B + 1) \times \text{random}) + 1;$$
$$\textbf{while } V_r < s \textbf{ do begin } s \leftarrow s - V_r;\ r \leftarrow r + 1 \textbf{ end};\ s \leftarrow s - 1 \qquad (5.4)$$

This selects the s^{th} block in the r^{th} sequence, where $0 \leqslant r \leqslant F - 1$ and $0 \leqslant s \leqslant V_r - 1$. The action to be taken depends upon which of the three categories of Section 2.2 is applicable.

$$\begin{array}{l} \textbf{if } V_r = 1 \textbf{ then } \text{category 2 } \textbf{else} \\ \textbf{if } s = 0 \textbf{ then } \text{category 1 left } \textbf{else} \\ \textbf{if } s = V_r - 1 \textbf{ then } \text{category 1 right } \textbf{else} \\ \text{category 0.} \end{array} \qquad (5.5)$$

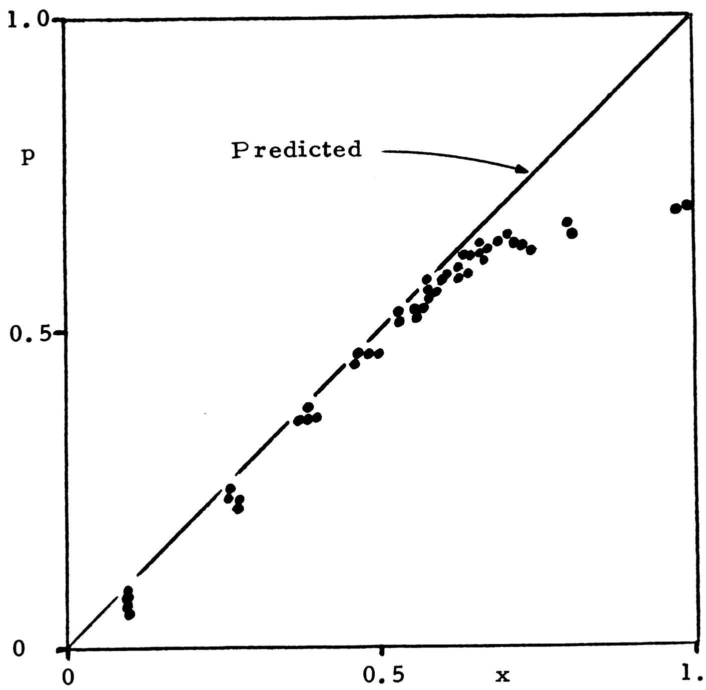

Fig. 4

5.3 *Metering*

The quantities of interest at the end of each cycle are the counts f_r of free blocks, and c_r of reserved sequences, of length r. These are initialised in two integer arrays at the end of the warm-up phase. During the different stages of each cycle, these counts are updated in the manner described in Sections 2.2 and 2.3 of the previous paper and Sections 2.1 and 2.2 of this paper.

At the end of the warm-up period, when the arrays f and c are initialised, two further arrays, Sf and Sc, are cleared. These accumulate the sums of the corresponding f and c elements over the individual cycles of the experiment. Contributions are added in after each cycle is complete. Finally, after an experiment of Ne cycles, say, means are estimated for f_r and c_r by

$$\bar{f}_r = Sf_r/Ne, \bar{c}_r = Sc_r/Ne \quad (5.6)$$

From these values, estimates of F, x, H_r, L_r, etc. are readily derived.

6. Results from the simulation model

The simulation program was run in conditions corresponding to the random fit alternating placement strategies. Two store sizes were considered: $N = 100, 200$. Ten store utilisation levels were selected: $\theta = 0.05\ (0.10)\ 0.95$. Two types of initial configuration were taken: one with $F = 1$ so that all reservations and all free store were compacted, and one in which each of the reserved store and free store was divided into ten equal fragments. Runs were replicated using different initial seeds for the random number generator. Each experiment was for 10,000 cycles.

Fig. 3 shows the observed relationships between θ and x with the curve indicating the theoretical relationship derived from equations (3.4, 3.17, 3.18, and 3.30) of Part 1 together with equation (2.37) for p_2 in the present paper. Agreement is good at high utilisations. At lower utilisations, the variability of the corresponding x value is seen to increase.

In **Fig. 4** is shown the relationship between p and x. The theoretical dependence is expressed by the 50% rule: $x = p$. Agreement is close for θ above the theoretical threshold but is seen to diminish consistently at lower utilisations. Even where agreement is good, the p value appears biased to fall somewhat short of the predicted value.

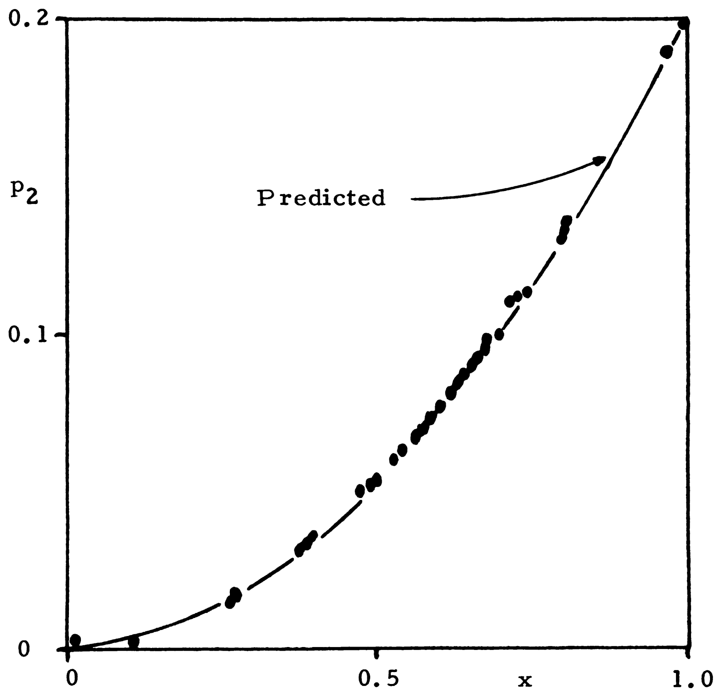

Fig. 5

The variation of p_2 with x is shown in **Fig. 5.** Agreement with the prediction

$$p_2 = \frac{x^2}{x + 4} \quad (6.1)$$

derived from equation (2.37) is gratifyingly close over the whole range of x.

Turning next to the observed distributions H and L of free block length and reserved sequence length, **Figs. 6** and **7** show a representative set of results for a store size $N = 200$. The two

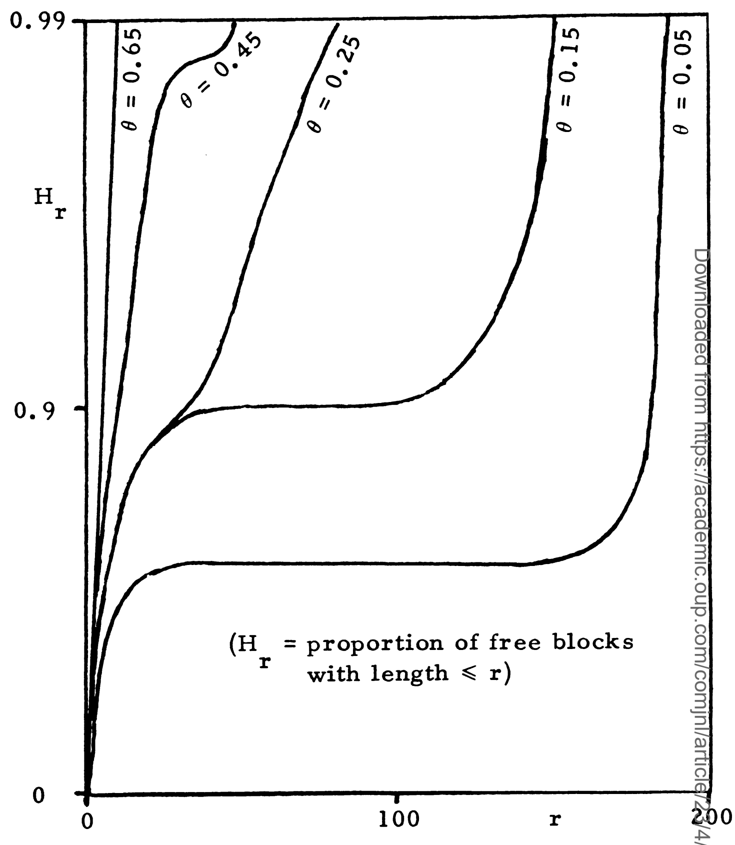

Fig. 6(*a*)

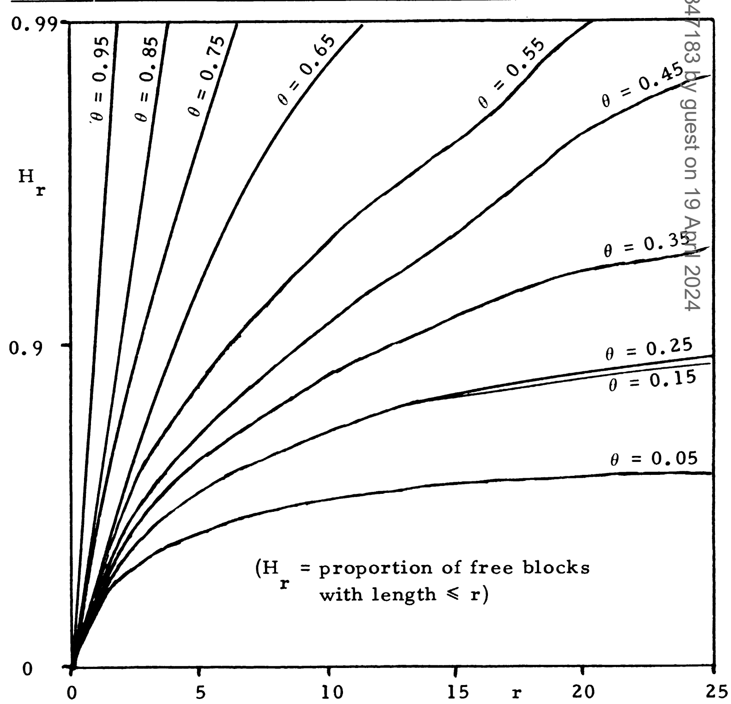

Fig. 6(*b*).

versions of Fig. 6 differ in range and scale of r. Agreement with the predicted distributions is seen to be satisfactory wherever the latter are defined, though it is best at the higher store utilisations. In particular it is observed that the theoretical L distribution appears valid over the full range of x.

At low utilisations, below the threshold $\theta = 0{\cdot}4849$, the H distribution exhibits quite clearly an unexpected, though in retrospect unsurprising, behaviour. The free store breaks up into one large block and a number of small fragments. The horizontal sections of the H curve are at levels $H = 1 - \frac{1}{F}$, indicating that just one of the F free blocks has a length approaching N. The variation of F with θ in the corresponding experiments at $N = 200$ is shown in **Fig. 8.**

The curves correspond to just one simulation run at each θ value. They are consistent with other runs except that in the middle utilisation region there is considerable variability in the H curves, particularly at the upper range of r.

7. A theoretical model for low store utilisation

The theories that have been developed thus far, both in this and in the previous paper, have assumed spatial homogeneity of the store in the statistical sense. The various blocks have been assumed to be well mixed up so that one region of store is much the same as any other. As a consequence the store size N has not appeared explicitly in predictions; the significant parameters have been the ratios θ, k and x.

The observation that at low utilisations one free block grows to a size dependent upon N, upsets this apple cart. It is now seen that, in the low utilisation region, θ is an insensitive parameter for describing the system. Of more significance is B, the number of reservations. It may be surmised that for given B, the distribution of reservations into sequences, and the distribution of the smaller fragments of free store, are largely unaffected by the total store size. An increase in N would, one might imagine, merely extend the length of the large free block by the corresponding amount.

Consider therefore the distribution of B reservations in an infinite store, where all reservations are of unit length. (The case of variable length reservations is a problem for a future occasion). Suppose the reservations lie in F sequences, $F \leqslant B$,

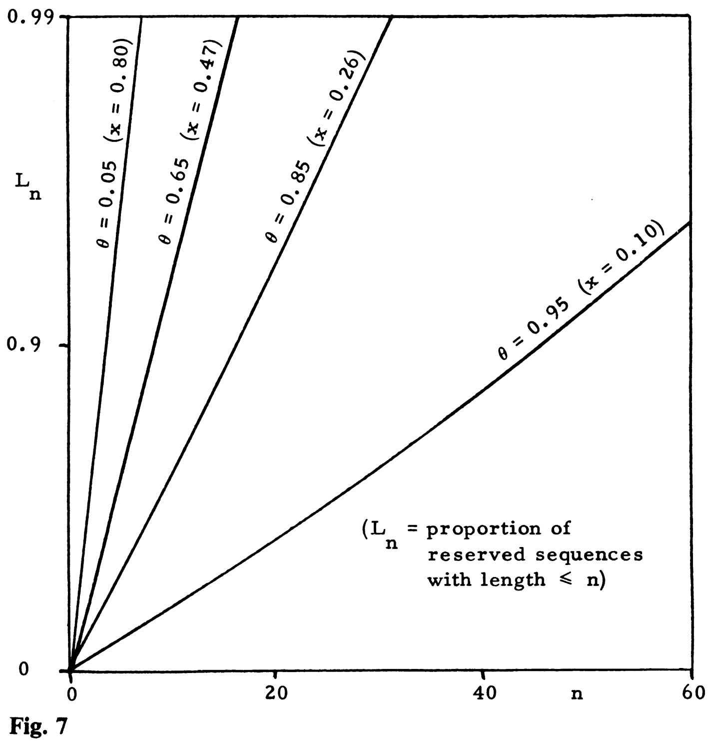

Fig. 7

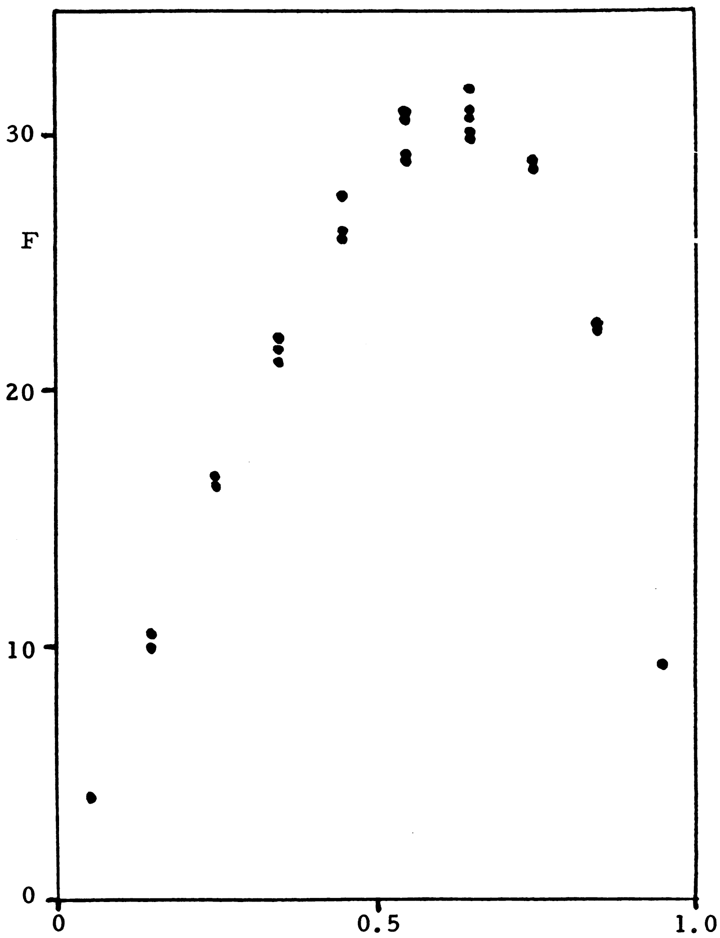

Fig. 8

thus fragmenting the free store into $F - 1$ finite fragments and one unbounded residue. Let there be f_r fragments of length r. Thus

$$1 + \sum_r f_r = F \tag{7.1}$$

The probability that a random fragment is of length r is $\phi_r = f_r/F$ and the probability that a random fragment is unbounded is $1/F$.

The concept of generating functions is carried over as before with

$$\tilde{\phi}(a) = \sum_r \phi_r a^r \tag{7.2}$$

An important difference now arises however in that

$$\tilde{\phi}(1) = 1 - 1/F \tag{7.3}$$

7.1 *Reservations*

A free block is selected at random and the reservation is allocated to one end, thus extending the length of its neighbouring sequence. Thus

$$\begin{array}{ll} \text{With probability } \phi_1, & \{f_1 \leftarrow f_1 - 1\} \\ \text{With probability } \phi_r, r > 1, & \{f_r \leftarrow f_r - 1; f_{r-1} \leftarrow f_{r-1} + 1\} \\ \text{With probability } 1/F, & \{\text{null}\} \end{array} \tag{7.4}$$

The expected change in $\tilde{f}$ per reservation is

$$\begin{aligned} \Delta^+\tilde{f} &= -a\phi_1 + \sum_{r>1} \phi_r(a^{r-1} - a^r) \\ &= (a^{-1} - 1)\,\tilde{\phi} - \phi_1 \end{aligned} \tag{7.5}$$

which reproduces equation (3.5) of Part 1.

7.2 *Releases*

A reservation is selected at random for release. The category probabilities p_0, p_1, p_2 are defined as before.

With probability p_0, $\{f_1 \leftarrow f_1 + 1\}$
With probability $p_1\,\phi_r$, $\{f_r \leftarrow f_r - 1; f_{r+1} \leftarrow f_{r+1} + 1\}$
With probability p_1/F, {null}
With probability $p_2\Phi_{rs}$, $\{f_r \leftarrow f_r - 1; f_s \leftarrow f_s - 1; f_{r+s+1} \leftarrow f_{r+s+1} + 1\}$
With probability $2p_2\,\phi_r/F$, $\{f_r \leftarrow f_r - 1\}$ (7.6)

Here Φ_{rs} denotes the probability that the (ordered) pair of free blocks, neighbouring the lone reservation being released, have lengths r, s.

$$\Phi_{rs} = \frac{fr}{F}\cdot\frac{fs}{F-1} = \frac{F}{F-1}\,\phi_r\,\phi_s \text{ if } r \neq s$$
$$= \frac{fr}{F}\cdot\frac{f_r - 1}{F-1} = \frac{F}{F-1}\,\phi_r\left(\phi_r - \frac{1}{F}\right) \text{ if } r = s \quad (7.7)$$

By purely topological considerations, as before, (i.e. equations (2.23, 2.24) of Part 1).

$$p_1 = x - 2p_2 \quad (7.8)$$

and

$$p_0 = 1 - p_1 - p_2 = 1 - x + p_2 \quad (7.9)$$

where

$$x = 2F/B \quad (7.10)$$

The expected change in $\tilde{f}$ per release is

$$\Delta^-\tilde{f} = p_0 a + p_1 \sum_r \phi_r\,(a^{r+1} - a^r)$$
$$+ p_2 \sum_r \sum_s \Phi_{rs}\,(a^{r+s+1} - a^r - a^s) - \frac{2p_2}{F}\sum_r \phi_r a^r \quad (7.11)$$

which reduces to

$$\Delta^-\tilde{f} = (p_0 + p_1\tilde{\phi} + p_2\,\tilde{\phi}^2)a - x\tilde{\phi}$$
$$+ \frac{ap_2}{F-1}\left\{\tilde{\phi}^2 - \tilde{\phi}(a^2)\right\} + \frac{2p_2}{F(F-1)}\,\tilde{\phi} \quad (7.12)$$

which now differs significantly from equation (3.6) of the previous paper.

7.3 *Equilibrium conditions*

The equilibrium conditions are as before, namely

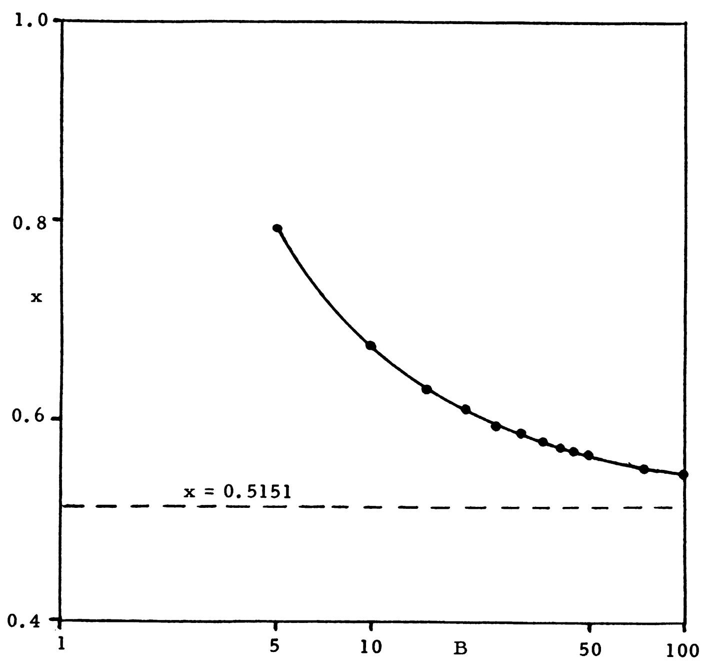

Fig. 9

$$\Delta^+\tilde{f} + \Delta^-\tilde{f} = 0 \quad (7.13)$$

The substitution $a = 1$ enables ϕ_1 to be derived as before, and here leads to

$$p \equiv 1 - \phi_1 = x - \frac{2p_2}{F^2} \quad (7.14)$$

which replaces the 50% rule.

Eliminating ϕ_1 now leads to equilibrium equations in the form

$$a^{-1}\tilde{\phi} = U_1 a + U_2 + (U_3 a + U_4)\tilde{\phi} + U_5 a\tilde{\phi}^2 + U_6 a\tilde{\phi}(a^2) \quad (7.15)$$

where

$$U_1 = -1 + x - p_2$$
$$U_2 = 1 - x + 2p_2/F^2$$
$$U_3 = -x + 2p_2$$
$$U_4 = 1 + x - \frac{2p_2}{F(F-1)}$$
$$U_5 = -\frac{F}{F-1}\,p_2$$
$$U_6 = \frac{p_2}{F-1} \quad (7.16)$$

On comparing coefficients of successive powers of a, it follows that

$$\phi_1 = U_2$$
$$\phi_2 = U_1 + U_4\phi_1$$
$$\phi_3 = U_3\phi_1 + U_4\phi_2$$
$$\phi_4 = U_3\phi_2 + U_4\phi_3 + U_5\phi_1^2 + U_6\phi_1$$
$$\phi_5 = U_3\phi_3 + U_4\phi_4 + U_5(\phi_1\phi_2 + \phi_2\phi_1)$$
$$\phi_6 = U_3\phi_4 + U_4\phi_5 + U_5(\phi_1\phi_3 + \phi_2\phi_2 + \phi_3\phi_1) + U_6\phi_2$$
etc. (7.17)

7.4 *Derivation of* p_2

Reconsideration of Section 2 in the case of low F values indicates that in equation (2.15) the product $\sigma_r\sigma_s$ should be replaced in a manner analogous to the introduction of Φ_{rs} in Section 7.2. It follows that equation (2.34) is replaced by

$$ax(1 - a)\tilde{\sigma}' - 2a(1 - a)\,(1 - x)\tilde{\sigma}^2 + 2(2 - 2a + a^2x)\tilde{\sigma}$$
$$- 2ax - \frac{2a(1 - a)\,(1 - x)}{F - 1}(\tilde{\sigma}^2 - \tilde{\sigma}(a^2)) = 0 \quad (7.18)$$

Extracting the coefficient of a yields however the same expression, equation (2.36), for σ_1 and hence the same dependence of p_2 upon x as quoted in equation (6.1). This result is of course supported by the simulation results in Fig. 5.

The modification of the 50% rule can best be related to the familiar form by eliminating p_2 from equation (7.14) and rearranging as

$$\frac{F}{B} = \frac{1}{2}p + \frac{1}{B(B + \frac{1}{2}F)} \quad (7.19)$$

7.5 *A numerical solution*

For given B, the value of F is obtained by the method of binary chopping applied to the interval $\left(1, \frac{B}{2}\right)$ within which F is known to lie. For each trial value of F, the values of x, p_2 and the U coefficients are readily determined. Equations (7.17) then enable the sequence of ϕ elements to be computed. The sequence is terminated if, for some n, (i) $\phi_n < 0$ or (ii) $\phi_1 + \phi_2 + \cdots + \phi_n > 1 - \frac{1}{F}$. Some initial experimentation indicates that these conditions correspond respectively to F being (i) too high, or (ii) too low.

It is of interest to consider the limiting behaviour as $F \to \infty$. It is seen that, in the limit, equations (7.5, 7.12 for $\Delta^+\tilde{f}$ and scale.

The computed H distributions for a range of B values are shown in **Fig. 10**, using as before a logarithmic scale for the H axis. The observed characteristics of the simulation runs are successfully reproduced at low store utilisations.

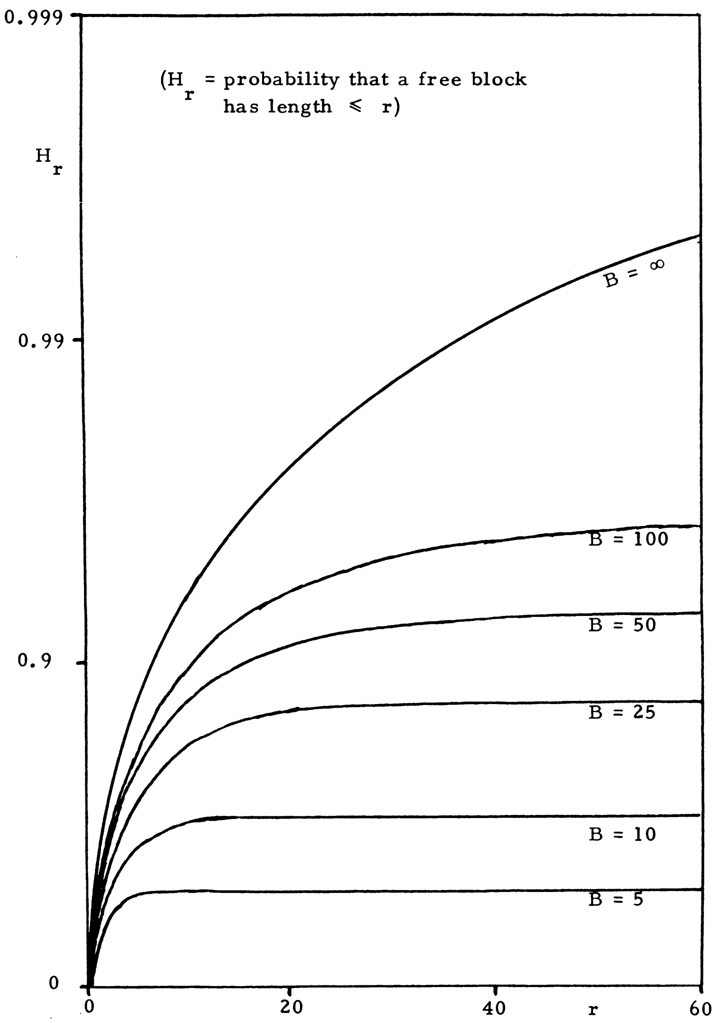

Fig. 10

$\Delta^{-}\bar{f}$ tend towards equations (3.5, 3.6) of the previous paper in the high utilisation situation. It is an obvious inference that the previous high utilisation model merges with the present low utilisation model at the threshold configuration corresponding to $x = 0{\cdot}5151$. This inference is supported by **Fig. 9**, showing the computed variation of x with B, where B is plotted on a log scale.

7.6 *Validity of the model*

An estimate of the range of θ values over which the low utilisation model is applicable for given N is provided by the hypothesis that the long free block should be at least as great as the total space, λ say, taken up by the B reservations and the $F - 1$ smaller fragments of free store. Now

$$\lambda = B + \sum_r r f_r \tag{7.20}$$

$$= B + F\tilde{\phi}'(1) \tag{7.21}$$

where, by differentiating equation (7.15) by a and then setting $a = 1$,

$$\tilde{\phi}'(1) = \frac{1 - x}{2p_2} \tag{7.22}$$

so that

$$\lambda = \left(\frac{4 + x - x^2}{2x^2}\right)F \tag{7.23}$$

In the limiting case of a large store where, for any given θ, each of N, B and F are large and $x \to 0{\cdot}5151$, it follows that

$$\frac{\lambda}{B} = \frac{4 + x - x^2}{4x} \to 2{\cdot}0624 \tag{7.24}$$

The validity hypothesis corresponds to $\lambda \leqslant \frac{N}{2}$ which requires $\theta \leqslant 0{\cdot}2424$. Thus the tentative expectation would be that the low utilisation model is applicable up to 25% store utilisation. Above 50% utilisation the high utilisation model has been demonstrated. The middle range from 25% to 50% is a transitional region with a considerable variability in behaviour.

As a check on this reasoning, the case $N = 200$ leads to predictions $\lambda = 13{\cdot}96, 51{\cdot}40, 90{\cdot}57$ for $\theta = 0{\cdot}05, 0{\cdot}15, 0{\cdot}25$ and hence to the size of the long free block being 186·04, 148·60, 109·43 respectively. These are verified by the position of the steepest parts of the H curve in Fig. 6(a) for $\theta = 0{\cdot}05$ and $\theta = 0{\cdot}15$ at the right hand end where H_r makes its final rise to include the long free block. The case $\theta = 0{\cdot}25$ is seen to lie already within the transitional region.

References

KNUTH, D. E. (1968). *The art of computer programming*, Vol. 1: Fundamental algorithms, (1st edn), § 2.5 Dynamic storage allocation, pp. 435-455.

REEVES, C. M. (1979). Free store distribution under random fit allocation, Part 1, *The Computer Journal*, Vol. 22, pp. 346-351.

SHORE, J. E. (1975). On the external storage fragmentation produced by first-fit and best-fit allocation strategies, *CACM*, Vol. 18, pp. 433-440.

SHORE, J. E. (1977). Anomalous behaviour of the fifty-percent rule in dynamic memory allocation, *CACM*, Vol. 20, pp. 812-820.

Books reviewed in this issue