1973-03

Empirically Derived Micromodels for Sequences of Page Exceptions

Lewis, P.A.W.; Shedler, G.S.

P.A.W. Lewis and G.S. Shedler, IBM J. Res. Develop., March 1973
https://hdl.handle.net/10945/41303

P. A. W. Lewis
G. S. Shedler

Empirically Derived Micromodels for Sequences of Page Exceptions

Abstract: Based on a statistical analysis of actual computer program address traces, some results are presented of a study aimed at deriving empirically valid stochastic models for program reference patterns in a computer system operating under demand paging. For the address traces examined, a semi-Markov model for the (univariate) point process of page exceptions is formulated and found to be an adequate characterization of the data.*

1. Introduction

We present in this paper the initial results of an extensive study aimed at deriving by the statistical analysis of actual program traces empirically valid stochastic models for program reference patterns in a demand paged computer system having a two-level memory.

Several related stochastic processes can be studied to characterize page reference patterns:

1. Reference strings $\{R_i\}$, i.e., sequences of page references, where R_i is the name of the page referenced at time i. These can be thought of as a multivariate point process [1] in discrete time, the multivariate aspect being the fact that the events (references to a page) are of several types (different pages).
2. Distance strings $\{D_i\}$, e.g., sequences of stack distances for LRU (least recently used) replacement, as defined in [2], where D_i is the total number of distinct pages referenced since the last reference to R_i.
3. The point processes corresponding to page exceptions for various memory capacities c, i.e., (discrete) times i at which D_i exceeds the memory capacity. Denote this by $\{T_j(c)\}$, where $T_j(c)$ is the time of the jth page exception in memory of capacity c. It is also possible to consider page exception processes marked by the page name.

It is not necessarily simple to go (probabilistically) from one of these representations to another, nor is it clear yet which one is simplest to characterize probabilistically. It may, of course, be that one of the representations would be more convenient than any of the others in a particular application. The distance string representation suppresses page names, which may be advantageous in that the process should be more nearly stationary.

Modeling and analysis of page exceptions as a univariate, unmarked point process is a simpler task than that for the distance strings, since, especially for large memory size, available theory on rare events and thinning of point processes [3, Section 5.3] suggests that these relatively rare events (i.e., page exceptions) should be approximately a Poisson process.

This paper describes an attempt to model page exception processes. The initial basis for the modeling was the rare-event theory and its prediction of a Poisson process for the page exception process; an extensive analysis of data was then undertaken to reject or confirm and extend the model. The analysis showed quickly that the Poisson model for page exceptions was grossly inadequate, but, on the basis of a previous analysis of distance strings and certain sample characteristics from the present data, a semi-Markov model for the (univariate) point process of page exceptions was formulated and found to characterize the data adequately. We start, however, with a brief discussion of the modeling of the distance strings, since this provided the key to the derivation of the semi-Markov model.

Note that the paper does not attempt to derive multivariate models for the page exception processes at dif-

*Partial support by the Office of Naval Research through Contract NR042-288 and the Foundation Research Program at the Naval Postgraduate School is gratefully acknowledged.

ferent levels c. The analysis is limited to comparison of the estimated parameters of the univariate point process models for different capacities c. Nor have we examined the marked univariate point processes consisting of times of page exceptions and page name. These present far greater difficulties of statistical analysis. The main rationale for deriving models of the univariate page exception processes alone is their utility in analytic studies of multiprogrammed computers operating under demand paging (e.g., [4,5]).

2. Preliminary analysis of distance strings

In the sequel we shall use the term "distance string" to mean LRU distance string. Models for distance strings $\{D_i\}$ that assume that the D_i's are independent and identically distributed random variables on $1,2,\cdots,\mathcal{E}$, where $\mathcal{E}$ is the total number of pages, are intuitively suspect and for particular address traces have been shown formally (by Lewis and Yue in [6]) to be inadequate. Tests of independence of a *stationary* sequence of random variables will be used later in the paper and are therefore outlined here. The most useful are the following.

1. Nonparametric tests such as runs-up-and-down [7]. The nonparametric property is bought at the price of smaller power, i.e., probability of rejecting the hypothesis of independence when it is false.
2. Tests based on estimated serial correlation coefficients. The *estimated* serial correlations of lag k are, for a sequence of n observed values d_i of a process $\{D_i\}$,

$$\hat{\rho}_k = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(d_i-\bar{d})(d_{i+k}-\bar{d})}{\frac{1}{n}\sum_{i=1}^{n}(d_i-\bar{d})^2} = \frac{n-k}{n}\tilde{\rho}_k, \qquad (1)$$

where $\bar{d} = \sum d_i/n$. The alternative estimator $\tilde{\rho}_k$ is also sometimes used.

For first-order Markovian normal sequences [8,9] the test based on $\hat{\rho}_1$ is the asymptotically most powerful test. This is also true asymptotically for testing that a point process is a Poisson process (independent exponentially distributed intervals between events) against the alternative that the intervals between events have (marginal) exponential distributions, but first-order serial dependence, the independence being rejected if $\hat{\rho}_1$ is too large or too small [10].

Asymptotically $\hat{\rho}_1$ is normally distributed with $E(\hat{\rho}_1) \sim 0$, var $(\hat{\rho}_1) \sim (n-1)^{-1}$ under very general conditions when $\rho_1 = 0$.
3. Lastly, there are tests based on the periodogram, essentially testing for a flat spectrum [11,12]. The spectral density function of the sequence $\{D_i\}$ is, when it exists,

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\rho_k e^{-ik\omega} = \frac{1}{2\pi}\left\{1 + 2\sum_{k=1}^{\infty}\rho_k \cos(k\omega)\right\}$$
$$(-\pi \le \omega \le +\pi), \qquad (2)$$

with the inverse relationship

$$\rho_k = \int_{-\pi}^{\pi} \cos(k\omega) f(\omega)\, d\omega$$
$$= E\left\{\left[\frac{D_i - E(D_i)}{\sigma(D_i)}\right]\left[\frac{D_{i+k} - E(D_{i+k})}{\sigma(D_{i+k})}\right]\right\}$$
$$(k = 0, \pm1, \pm2, \cdots) \qquad (3)$$

Thus the ρ_k's are the Fourier coefficients of $f(\omega)$. Since $f(\omega)$ is an even function of ω, i.e., $f(\omega) = f(-\omega)$, it is usual to use only positive ω's and define $f_+(\omega) = 2f(\omega)$ for $0 \le \omega \le \pi$.

The periodogram (see [11,13]), which is the basis for estimation and testing that $f(\omega) = 1/2\pi$ (i.e., $\rho_k = 0$, $k \ne 0$), is

$$I(\omega) = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\hat{\rho}_k \cos(k\omega) \qquad (-\pi \le \omega \le \pi) \qquad (4)$$

It is thus the Fourier transform of $\hat{\rho}_k$ (see [13] for other interpretations). At the values $\omega_\ell = 2\pi\ell/n$, $\ell = 1,\cdots v$, where v is the largest integer less than or equal to $n/2$ minus one, the $I(\omega_\ell)$ would be independently exponentially distributed random variables if the $\{D_i\}$ were independent normally distributed random variables. Consequently [11, p. 48; 9, p. 76]

$$S_\ell' = \frac{\sum_{k=1}^{\ell} I(\omega_k)}{\sum_{k=1}^{v} I(\omega_k)} \qquad (\ell = 1,\cdots, v-1), \qquad (5)$$

are the order statistics from a random sample of size $(v-1)$ of uniform (0,1) random variables. The distribution theory is approximately true for independent but nonnormal $\{D_i\}$, and a test for independence can therefore be obtained by testing the uniformity of the S_ℓ''s with a Kolmogorov-Smirnov statistic [11, Ch. 6]. The distribution when the D_i's are independent exponentially distributed random variables is given by Lewis in [4].

Spectral tests were used by Lewis and Yue [6] to show that the distance strings they studied are correlated. Such dependence can be shown much more simply, but the estimated spectra have direct interpretations in terms of program behavior.

A more general alternative model for the distance strings is a representation of the D_i's as a first-order (Markov) chain, which is useful since it permits a formulation of "locality of reference" [5]. Asymptotic-theory maximum-likelihood tests of the hypothesis that the D_i's are a first-order Markov chain against a hypothesis of higher-order dependence are available [15], but are not

Table 1 Counts of one step transitions for stack distances for tape A with LRU replacement algorithm. Page size is 4K. The table entries are n_{jk}, the number of times $d_i = j$ and $d_{i+1} = k$ for $i = 1, \cdots, t_0 = 8{,}802{,}464$.

j/k	Counts of one step transitions for LRU distances 1	2	3	4	5	6	7
1	2,943,817	840,912	210,914	78,002	41,500	24,281	16,792
2	1,048,310	2,151,371	146,163	35,192	26,012	13,591	7,931
3	130,957	271,850	176,386	16,570	5,693	2,804	2,318
4	22,878	70,630	35,013	10,338	3,721	1,516	1,643
5	18,512	36,744	16,366	5,258	4,017	393	235
6	10,685	20,180	7,959	1,650	812	1,914	223
7	11,549	11,324	3,596	846	280	233	271
8	7,957	9,934	4,405	1,261	182	128	61
9	9,451	8,248	3,457	834	184	145	56
10	7,274	8,657	2,641	769	639	268	76

applicable because the transition matrix of the $\{D_i\}$ process is highly skewed and has many nonzero entries. The skewness is roughly the idea of locality and can be seen in the upper left hand corner of the transition matrix of the d_i's (Table 1) for the tape used in this study. The entry n_{jk} gives the number of times an observed distance k $(d_{i+1} = k)$ immediately followed an observed distance j $(d_i = j)$. The maximum likelihood estimator of the transition probability in a first-order Markov chain is $\hat{p}_{jk} = n_{jk}/n$.

An informal test of the first-order hypothesis, which relates to the length of dependence in the sequence $\{D_i\}$, is to estimate directly [11, p. 71] a variance-time sequence

$$V_k = \text{var}\ (D_i + D_{i+1} + \cdots + D_{i+k-1})$$
$$= k\ \text{var}\ (D_i)\left\{1 + 2\sum_{\ell=1}^{k-1}\left(1 - \frac{\ell}{k}\right)\rho_\ell\right\}, \quad (6)$$

and also to estimate V_k using the estimated transition probabilities $\hat{p}_{jk}$. The length of dependence of the process for the distance strings studied by Lewis and Yue [6], estimated by observing the time k for $\hat{V}_k$ to reach a steady state, was much longer than those predicted by the first-order Markov model.

An alternative analysis that can be used to test for dependence of order greater than one, and which has potential for building an alternative model to the first-order Markov chain, is to examine the successive intervals O_k during which the D_i's equal 1. The interpretation of these runs of ones is that they are times during which the program references the same page. Under a first-order Markov chain model (including the model of independent $\{D_i\}$) the O_k's are a sequence of independent, geometrically distributed random variables:

$$\text{prob}\ \{O_k = x\} = (1 - p_{11})p_{11}^{x-1} \qquad (x = 1, 2, \cdots), \quad (7)$$

where p_{11} is the 1-1 transition probability in the chain; the distribution given by (7) is actually the "geometric plus one" distribution, the geometric distribution usually being defined on $0, 1, 2, \cdots$.

The geometric assumption is very strong, implying that no matter how long the program has been referencing the page, the time to exit is still geometric (see [16] for a discussion of this "lack of memory" property);

$$\text{prob}\ \{O_k = x + y | O_k > y\} = (1 - p_{11})p_{11}^{x-1} \qquad (x, y = 1, 2, \cdots).$$

This analysis, based on the sequence of runs of ones in the distance strings, is attractive because long dependence could be built into a model simply by making the O_k's nongeometric. Moreover the one-one transitions are the bulk of the transitions in the distance string. In Table 1 for a particular address trace, called tape A, there are a total of 2,943,817 1-1 transitions out of a total of 8,802,464 giving $\hat{p}_{11} = 0.3344$ and an estimated expected value for O_k under the first-order Markov chain model of $1/(1 - \hat{p}_{11}) = 1.502$.

Statistical analysis of the O_k's in tape A showed the marginal distribution of the O_k's to be highly positively skewed and nongeometric, and the sequence to have strong dependencies. In fact spectral analysis showed a strong quasicycle in the sequence reminiscent of those seen in autoregressive sequences [8]. This prompted a closer look at the transition matrix in Table 1 to see what this would reveal.

Two features stand out. Although 1-1 transitions dominate, there are almost as many 2-2 transitions. Moreover there are generally more transitions from j states, $j > 2$, to $k = 2$ than from j states, $j > 2$, to $k = 1$. Recalling that distance strings suppress page names and that we are considering LRU distance strings [2], an interpretation of the above is that there are at least two types of paging behavior. In one, the program stays mainly in one page, referring occasionally to a very small group of other pages (strong locality of reference); in the other the program refers to its pages almost at random (i.e., uniformly over the set of pages and with no time dependence). Thus locality becomes a dynamic, possibly two-state phenomenon. The analysis of page exceptions given below, however, suggests the possibility of having to include a third state in a model for the distance strings $\{D_i\}$.

This insight into the $\{D_i\}$ sequence will be pursued elsewhere. The importance at present is in explaining and modeling the sequences of page exceptions. Distance strings of several programs have been examined and appear to exhibit similar behavior.

3. Data and preliminaries

We display in this paper results for the particular address trace referred to as tape A. From the sequence of addresses traced, the distance string was derived by stack processing techniques [2] for a page size of 4096 bytes. Conventionally this page size is referred to as 4K. The data consisted of $t_0 = 8{,}802{,}464$ references to a total of 517 distinct pages. Results are given for three values of the memory capacity c; the values chosen for c (respectively, 76, 197, and 512) correspond to execution of the program in a relatively constrained, moderately constrained, and essentially unconstrained memory.

Much of the statistical analysis was done using the SASE IV program [17] developed for analyzing series of events. SASE IV can also be used for analyzing positive valued time series such as $\{D_i\}$.

4. Point processes—discrete and continuous time

Given a memory capacity c and a sequence of page references, a page exception can occur on any of the successive page references. We consider the page references to occur at equidistant time points (this is very nearly true), and the intervals between references are taken to be the unit of time in this study. The page exceptions therefore constitute a point process (series of events) in discrete time. Since the average times between page exceptions is so large ($\hat{\mu}$ in Table 2), it would be possible to treat the process as if it occurred in continuous time. Much of the continuous time analysis [11] has in fact been used; the main problem with using it throughout was the discovery that about 10% of the intervals between page exceptions had a value of 1024, and thus could only be modeled as a discrete component in the distribution of times between page exceptions.

In considering discrete point processes there are two main stochastic processes that can be and are considered. One is the binary sequence $\{\delta_i(c)\}$, where

$$\delta_i(c) = \begin{cases} 1 & \text{if event at time } i\,, \\ 0 & \text{if no event at time } i\,, \end{cases}$$

and we explicitly represent dependence on c, the memory size, if necessary. The use of 0 and 1 to represent the two states is convenient since then the cumulative process

$$N_i(c) = \sum_{k=1}^{i} \delta_i(c) \qquad (i = 1, \cdots) \tag{8}$$

gives the number of events (page exceptions) up to and including time i. The process $\{N_i(c)\}$ is called the *counting process* and is the basic representation of the point process. In particular, stationarity of the point process means that $\{N_i(c)\}$ has stationary increments. The spectrum of $\delta_i(c)$, following the definition Eq. (2), is denoted

Table 2 Sample characteristics of times-between-events. Page exception process for tape A with LRU replacement algorithm. Page size is 4K. Number of references $t_0 = 8{,}802{,}464$.

	Memory capacity c (pages) 76	197	512
n – number of page exceptions	1,807	820	517
$\hat{\mu}$ – estimated mean time between page exceptions	4,871	10,735	17,026
$\widehat{C(X)}$ – estimated coefficient of variation of times between page exceptions	3.34 (0.25)	3.27	3.70
$\hat{\gamma}_1$ – estimated coefficient of skewness of times between page exceptions	10.34 (1.95)	7.14	6.87
X_{max} – maximum time between page exceptions	333,374	420,786	704,921
X_{min} – minimum time between page exceptions	1	2	2
X_{med} – median time between page exceptions	1,024	1,024	1,024

by $g(\omega)$ and called the *count spectrum*, or *Bartlett spectrum*, of the process:

$$g(\omega) = \frac{1}{2\pi}\left\{1 + \sum_{k=1}^{\infty} \frac{E[\{\delta_i - E(\delta_i)\}\{\delta_{i+k} - E(\delta_{i+k})\}]}{\operatorname{var}\{\delta_i\}} \times \cos(k\omega)\right\} \qquad (-\pi \le \omega \le \pi)\,. \tag{9}$$

Note that frequency here is the reciprocal of periods T in page exception time: $\omega = 2\pi f = 2\pi/T$. Thus the highest possible frequency is $\omega = \pi$, or $T = 2$ page references, and, although we have assumed the density to exist, a delta function component in $g(\omega)$ at $\omega = \pi$ would represent a strict alternation of period 2, i.e., page exceptions every second reference.

The other process considered in the analysis of point processes is the sequence of times between page exceptions $\{X_j(c)\}$, or the (cumulative) times to page exceptions $T_j(c) = \sum X_\ell(c)$, where the sum on ℓ is from $\ell = 1$ to $\ell = j$. The two processes $\{N_i(c)\}$ and $\{T_j(c)\}$ are related by the identity

$$T_j(c) > i \quad \text{iff} \quad N_i(c) \le j - 1 \qquad (j = 1, 2, \cdots)\,, \tag{10}$$

so that

$$\text{prob}\,\{T_j(c) > i\} = \text{prob}\,\{N_i(c) \le j - 1\} \qquad (j = 1, 2, \cdots)\,. \tag{11}$$

In actual fact the sequence $\{X_j(c)\}$ is not stationary if $\{N_i(c)\}$ is stationary [11, Sect. 4.2]. The stationary interval process usually analyzed is the sequence of intervals following an *arbitrarily selected* page exception in the stationary process. We do not press this distinction here (see [11, Ch. 4] for details), and refer to both interval sequences as $X_j(c)$. The serial correlation coefficients for this sequence will be denoted, as in Eq. (2), by ρ_k and

Table 3 Tests for Poisson process and trend. Page exception process for tape A with LRU replacement algorithm. Page size is 4K. Number of references $t_0 = 8{,}802{,}464$.

*Transformed data— Kolmogorov statistic**	*Memory capacity c (pages)* 76	197	512
	18.57	15.56	14.43
$U = \{S - t_0/2\}/\{t_0/(12n)^{\frac{1}{2}}\}$	−2.83	−8.67	−18.11

*Upper 1% point is 1.518.

the spectrum of intervals (2) by $f(\omega)$. Note here that frequency is related to *serial number j* of page exceptions, not page reference time i, and $\omega = \pi$ corresponds to an *alternation on exception number.*

The two spectra, beside their different interpretations, are not equivalent, although related, and analyses based on both will be used.

The null process for a discrete time point process corresponding to the Poisson process in continuous time is a Bernoulli process in which $\delta_i(c) = 0$ with probability p, independently of previous values of $\{\delta_i(c)\}$. For the Bernoulli process the ρ_k's equal 0 for $k \neq 0$, the spectrum is flat, $f(\omega) = 1/2\pi$, and the intervals between page exceptions are independent with identical geometric (plus one) distributions (i.e., a particular renewal process)

$$\text{prob}\ \{X_j(c) = x\} = (1-p)p^{x-1}$$
$$(x = 1, 2, \cdots; 0 < p < 1)\,. \qquad (12)$$

We have then for this "geometric plus one" distribution:

$$\mu = E\{X_j(c)\} = \frac{1}{(1-p)}\,,$$

$$\sigma = [\text{var}\ \{X_j(c)\}]^{\frac{1}{2}} = p^{\frac{1}{2}}/(1-p),$$

$$C(X) = \text{coefficient of variation} = \frac{[\text{var}\ \{X_j(c)\}]^{\frac{1}{2}}}{E\{X_j(c)\}}$$
$$= p^{\frac{1}{2}} < 1\,.$$

Note that as p becomes large, i.e., close to 1, the mean time between exceptions becomes large and $C(X)$ approaches 1, its value for the exponential distribution of times between events in a Poisson process. The coefficient of skewness γ_1 and coefficient of kurtosis γ_2 for this geometric distribution are complicated functions of p. For large p they are approximately equal to the values for an exponential distribution. For an exponential distribution $\gamma_1 = 2$, $\gamma_2 = 6$.

Two points should be made here. The serial correlation coefficient for a Bernoulli process must be interpreted with care. In particular, $\rho_k = 0$, $k \neq 0$ does *not* necessarily mean the process is Bernoulli. In a methodological sense the Bernoulli variable is as far away from the normal distribution as one can get, and statements and interpretations based on normal distribution theory can be very misleading.

5. Tests for trend and Poisson process

In estimating serial correlation coefficients, moments of marginal distribution of intervals, etc., in point processes, the assumption is made that the process is stationary. Lack of stationarity can make a hash of estimated parameters and therefore an analysis of page exceptions must begin with tests for trend in the data. In the case of gross inhomogeneity it might be required to characterize the trend; see [11,14,18].

We follow here tests for trend given in [11, Ch. 3], and in particular the test for a monotone trend in the rate of events based on the centroid of the times-to-events,

$$S = \sum_{j=1}^{n} T_j(c)/n\,,$$

where n is the number of page exceptions in the period of observation.

For a Poisson process (Bernoulli process with p close to 1) this statistic has, conditional on observing n events in the fixed period of observation, t_0, mean $t_0/2$ and variance $t_0^{\,2}/(12n)$. The normalized statistic

$$U = \frac{S - t_0/2}{t_0/(12n)^{\frac{1}{2}}}$$

is asymptotically normally distributed with mean 0 and variance 1.

It is particularly important to test the three series considered ($c = 76$, 197, 512) with a test for trend, since in the test runs the memory was started empty and pages brought in as requested by the program. Thus the page exception process up until the time the memory was filled could be expected to be nonstationary. For the three capacities, time to filling the memory was 199,363, 1,081,730, and 7,940,780 page references, respectively. Thus while for $c = 76$ and 197 there is probably only a transient effect, the effect will be marked for $c = 512$ and a stationary analysis of this data is to be viewed with caution.

The values of U for the three memory capacities are given in Table 3. The negative sign is consistent with a decreasing trend, but the statistic is probably not significantly large for $c = 76$, $c = 197$. This is because the variance of S under a Poisson hypothesis has been used; it can be shown to be inflated if the process is overdispersed relative to a Poisson process, although exact results are not known. The overdispersion in the data can be seen by examining the sampling characteristics of the three series given in Table 2. We return to this later in this section.

Tests for trends that are not smooth, e.g., changes between discrete levels, are performed in the SASE IV program [17], and confirm stationarity for levels 76 and 197. However, it should be borne in mind that for long programs discrete changes in the structure of the page exception processes will almost certainly occur. This is the case for a longer address trace (tape B) we have examined. An investigation of whether these level changes are stochastic but stationary would be in order, but constitutes a separate investigation. *The models derived in this paper should thus be considered to be micromodels.*

In summary the trend analysis suggests that for capacities of 76 and 197 the observed page exception processes can be treated as coming from stationary point processes. The data for $c = 512$ has to be treated with care as almost all sample characteristics except the spectra will be confounded by lack of stationarity. In the spectra these long term trends will primarily affect the low frequencies.

Returning to the tests for a Poisson process (Bernoulli process with p close to 1) for the page exception processes, little time has to be spent on a formal analysis. The skewness of the times between page exceptions is so marked in Table 2 that it is evident that the distribution is not exponential. If the geometric assumption is used, the maximum likelihood estimate of p under the assumption of independent intervals is $\hat{p} = (\hat{\mu} - 1)/\hat{\mu}$. For $c = 76$, $\hat{p} = 0.99979$ and $C(X)$ should be approximately one. However, the estimate of $C(X)$ is $\widehat{C(X)} = 3.34$. This estimate is obtained by averaging estimates in four time sections of the data. The resulting estimated variance of $\widehat{C(X)}$ (three degrees of freedom) is 0.25. Clearly any reasonable test would show $C(X) \neq 1$.

We return to analysis of the intervals in a later section after independence of the intervals has been examined.

Results of formal tests for a Poisson process are given in Table 3 (see [11, p. 152] for details). The upper 1% point of the Kolmogorov statistic is 1.628, well below the value 18.57 observed for capacity 76.

6. Tests for renewal process

The next step in the analysis of the data and the search for a model for the page exception processes is to test for a renewal process [19,11,9], i.e., that the intervals $X_j(c)$ are independent and identically distributed. These tests have been described in Section 2; the results for the data are given in Table 4. For a capacity of 76 the estimated serial correlation of lag 1, $\hat{\rho}_1$, using either the asymptotic variance $(n-1)^{-\frac{1}{2}}$ or the estimate of the variance from the sections (0.08, three degrees of freedom), is too large to be consistent with a true value $\rho_1 = 0$. Similarly the tests based on the cumulated periodogram Eq. (5) and described in Cox and Lewis [11, Ch. 6, p. 164] indicate rejection of the renewal hypothesis. The upper 1% point of $D_{n/2}$, for large n, is 1.518.

Table 4 Tests for dependence on serial number and dependence between intervals. Page exception process for tape A with LRU replacement algorithm. Page size is 4K. Number of references $t_0 = 8{,}802{,}464$.

	Memory Capacity c (pages) 76	197	512
n	1807	820	517
$\hat{\rho}_1$ – estimated serial correlation coefficient of order 1 for times between page exceptions	+0.188 (0.08)	+0.177	+0.130
$(n-1)^{\frac{1}{2}}\hat{\rho}_1$	+8.01 (1.7)	+5.11	+3.00
$\hat{\phi}_1$ – estimated partial serial correlation of order 1	0.035 (0.002)	0.065	0.002
Test for serial independence based on cumulated periodogram			
Kolmogorov-Smirnov statistic* – $D_{n/2}$	3.82	2.61	1.65
Anderson-Darling statistic** – $W^2_{n/2}$	36.43	14.91	6.94

*Upper 1% point is 1.518.
**Upper 1% point is 3.857.

Using the results informally there is an indication of decreasing dependence as the memory capacity increases, the estimated serial correlation coefficients, for instance, decreasing as c increases. (Again $c = 512$ must be used carefully; a smooth trend will generally inflate the value of the estimate $\hat{\rho}_1$.) This trend in the value of $\hat{\rho}_1$ may indicate that the rare event theory is becoming valid as c increases. In addition Table 2 indicates that the intervals are becoming less skewed, but not inordinately so. Thus a Poisson hypothesis would be out of the question.

The estimated partial serial correlation given in Table 4 will be defined and discussed later.

The next step is to find, if possible, a model for the page exception processes in which the dependence between intervals is accounted for. To do this we examine first in more detail the marginal interval process $\{X_j(c)\}$.

7. Interval properties

A further search for a model, the Poisson and renewal hypothesis having been rejected, could start with either examination of the interval sequence $\{X_j(c)\}$ or the counting sequence $\{N_i(c)\}$. Actually the semi-Markov model we describe below was derived from the analysis of the distance strings $\{D_i\}$ and the marginal and second-order joint properties of the intervals $\{X_j(c)\}$, the second-order properties of counts being used to verify the model.

We attempt now to describe this process of modeling, although the description is of necessity truncated.

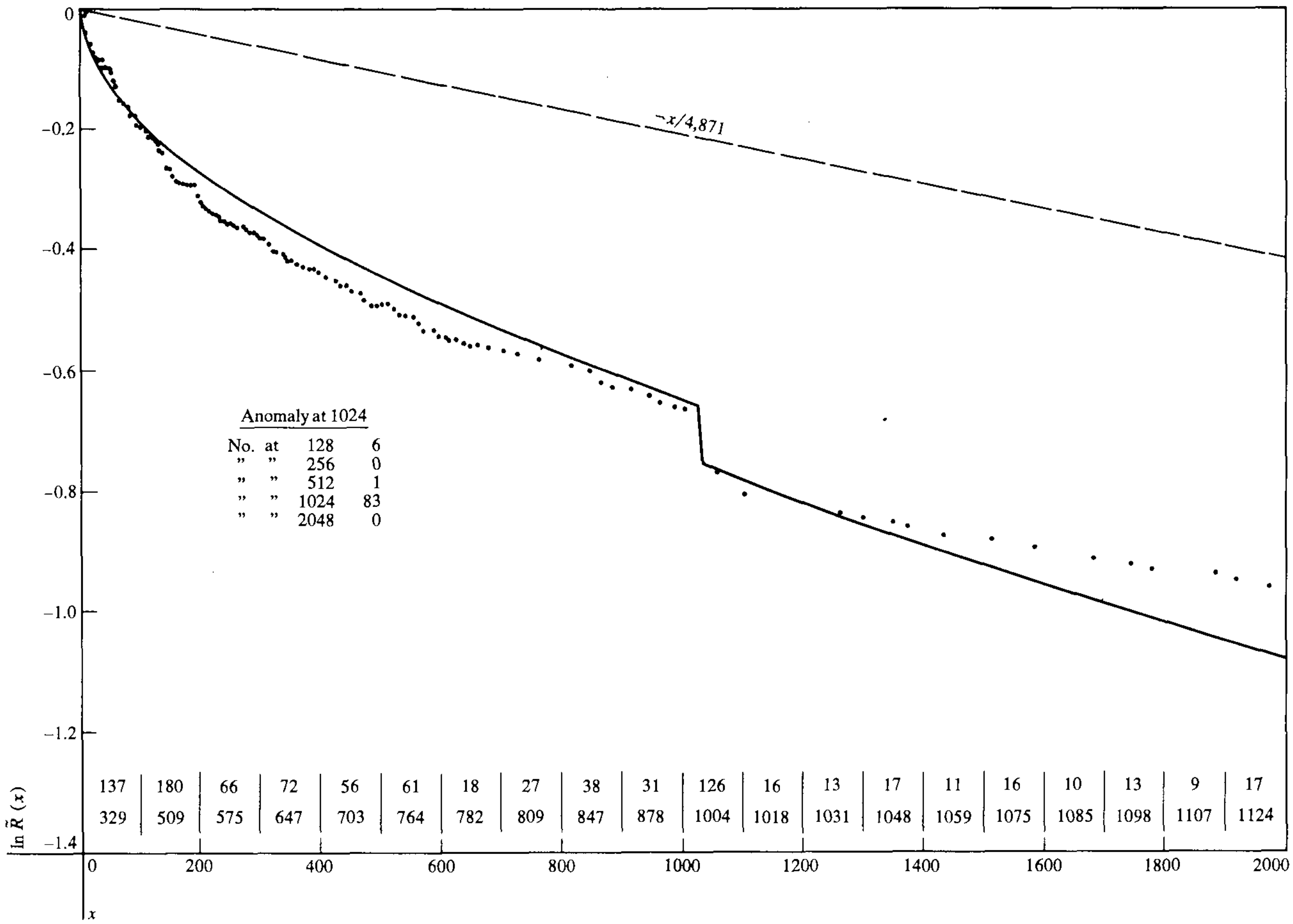

Figure 1 Empirical log-survivor function ln R(x) for page exception process for tape A with LRU replacement algorithm. Page size 4K, capacity = 76, n = 1807, $\hat{\mu} = 4{,}871$ references, $\widehat{C(X)} = 3.34$, $\hat{\gamma}_1 = 10.34$, $\hat{\gamma}_2 = 151.22$. Solid line is the fitted semi-Markov log survivor function with estimated parameters $\hat{\pi}_1 = 0.066$, $\hat{\gamma} = 0.950$, $\hat{p}_1 = 0.99997813$, $\hat{p}_2 = 0.99977748$, $\hat{k} = 0.45$.

• *Marginal distribution of intervals*

We discuss first in further detail the estimated marginal distribution of intervals, the sampling properties being given in Table 2. The striking feature is the extreme skewness of the distribution of the intervals between page exceptions. For $c = 76$, we have estimates $\widehat{C(X)} = 3.34$, $\hat{\gamma}_1 = 10.34$ and $\hat{\gamma}_2 = 151.22$, compared to the values for an exponential distribution of $C(X) = 1.0$, $\gamma_1 = 2$, $\gamma_2 = 6$. Moreover the maximum time between exceptions for $c = 76$ is 333,374, the minimum value is 1, compared to a mean of 4,871 and a median of 1024.

Graphical presentation for such a skewed empirical distribution is difficult. On a scale such that the largest observation is included, the logarithm of the empirical survivor function, $\tilde{R}(x)$, where

$$\tilde{R}(x) = \frac{\text{number of intervals greater than } x}{\text{total number of intervals, } n} \quad (x = 1, 2, \cdots), \tag{13}$$

drops rapidly and then becomes linear for large x. If in fact the sample came from a geometric distribution, this plot would be linear, within the limits imposed by sampling fluctuations, since

$$R(x) = p^x = e^{-x} \ln (1/p) \qquad (x = 1, 2, \cdots), \tag{14}$$

$$\ln R(x) = x \ln p \approx -x/\mu, \tag{15}$$

the approximation holding with p close to one.

The nonlinearity of the plot is consistent with the rejection of the Poisson (geometric) hypothesis. The initial part of the log survivor function plot for $c = 76$ is shown in Fig. 1. At the bottom of the figure we give a numerical, rather than a graphical, histogram for an interval of length 100. Both the individual cell counts and the cumulative cell counts are given. The plot encompasses only 2,000/333,374 = 0.006 of the observed *range* of the intervals between page exceptions, but a fraction 1124/1807 = 0.62 of the observed number of intervals. Note

Table 5 Estimated parameters for semi-Markov model. Page exception process for tape A with LRU replacement algorithm. Page size is 4K. Number of references $t_0 = 8{,}802{,}464$.

	Memory Capacity c (pages) 76	197	512
$\hat{\alpha}_1$ – estimated transition prob. of type 1|type 1	0.458	0.453	0.346
$\hat{\alpha}_2$ – estimated transition prob. of type 2|type 2	0.962	0.941	0.937
$\hat{\pi}_2$ – estimated proportion of type 2 intervals	0.934	0.903	0.913
$\hat{\mu}_1$ – mean of type 1 interval, assumed geometric	45,715*	82,422**	155,230***
$\hat{\mu}_2$ – mean of type 2 interval	1,973.1	3,033.0	3,780.5
$\hat{k}$ – parameter of negative binomial distribution for type 2 interval	0.45	0.45	0.35
$\hat{\gamma}$ – proportion of type 2 intervals of length other than 1024	0.950	0.880	0.898
$\hat{\mu}_2'$ – mean negative binomial	2022.8	3307.4	4,092.7

*Based on conditional mean greater than 30,000
**Based on conditional mean greater than 50,000
***Based on conditional mean greater than 75,000

the large number of very short intervals, e.g., there are 192 intervals less than $x = 50$.

There are several other outstanding features in the plot.

1. The plot is roughly convex, being always below the dashed line $-x/\hat{\mu} = -x/4{,}871$, suggesting that the model for the marginal distribution of intervals should have a decreasing hazard and an exponential tail (for details, see [11, p. 140]). For a continuous variable, models with these properties are the mixed exponential and the gamma distribution. Discrete analogs are, respectively, the mixed geometric and the negative binomial plus one:

$$p(x) = \text{prob}\ \{X_j = x\} = \pi p^{x-1}(1-p) + (1-\pi) r^{x-1}(1-r) \quad (0 < \pi < 1, x = 1, 2, \cdots)$$

$$(0 < p < 1;\ 0 < r < 1) \quad \text{(mixed geometric)} \quad (16)$$

$$p(x) = \text{prob}\ \{X_j = x\} = \binom{k + x - 2}{x - 1} p^{x-1}(1-p)^k \quad (k > 0, x = 1, 2, \cdots)$$

$$(0 < p < 1) \text{ (negative binomial plus one)}\ . \quad (17)$$

Note that for the negative binomial plus one

$$E(X) = 1 + \frac{kp}{(1-p)}, \quad (18)$$

$$\text{var}\ (X) = \frac{kp}{(1-p)^2}, C^2(X) = \frac{kp}{[1 - p(1-k)]^2}. \quad (19)$$

For the moment we prefer a mixed model such as the mixed geometric (exponential) for two reasons. First, there is a suggestion of a two-state phenomenon from the distance string analysis, and second, a mixture model is generated by the semi-Markov model for the page exception times, the semi-Markov model being strongly suggested by the second-order joint properties of intervals given in the next subsection.

2. Note the jump in the plot at $x = 1024$. For capacity $c = 76$ there were in fact 83 intervals between page exceptions with this value. This is examined in Table 5 where the parameter $\hat{\gamma}$ is such that $(1 - \hat{\pi}_1)\hat{\gamma}$ is the proportion of intervals of length other than 1024. The feature appears at all capacities; it appears to be a situtation in which the program references sequentially each of the (4-byte) words in a (4K-byte) page.

We return to the problem of incorporating this feature into the model later; the problem is whether it is a feature of one state in a two-state model, or a separate state or mode of referencing pages. The short table in Fig. 1 shows that this feature does not recur at harmonics or subharmonics of 1024.

• *Second-order joint properties of intervals*

The sample (estimated) second-order joint moments of the intervals are shown in Table 4. We have already remarked that the estimated serial correlation coefficient of lag 1, $\hat{\rho}_1$, is significantly large for all capacities. Another outstanding feature is that the estimated partial serial correlations (Table 4), $\hat{\phi}_1$, are very small, suggesting first-order Markov dependence of intervals.

The partial serial correlation of order one is [20, p. 64–5] simply the correlation between X_j and X_{j+2} with X_{j+1} fixed:

$$\phi_1 \triangleq \text{corr}\ \{(X_j|X_{j+1}), (X_{j+2}|X_{j+1})\} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}. \quad (20)$$

Note that with first-order Markov dependence $\rho_k = \beta^k$,

$|\beta| < 1$, so that

$$\phi_1 = \frac{\beta^2 - (\beta)^2}{1 - \beta^2} = 0 . \tag{21}$$

The semi-Markov model for intervals is not quite Markovian (hence its name) as we will show below, but it does give a small partial correlation coefficient. This and the observed interval properties and the results of the analysis of the distance strings suggest adopting a two-state semi-Markov process as a tentative model for the data. We now give the details of the model and the results of fitting the model to the data.

8. Two-state semi-Markov model for univariate point processes

A two-state semi-Markov model for univariate point processes is discussed in detail by Cox and Lewis [11, p. 194–7]. We summarize these results here and give extensions. Note that the model is usually used as a model for bivariate point processes (e.g., [1]).

The model is defined on the intervals $\{X_j\}$ rather than on the counts $\{\delta_i\}$. Thus we suppose there are two types of intervals with probability density functions (continuous time) or probability distributions (discrete time), $p_1(x)$ and $p_2(x)$. The model postulates that there is a two-state Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} \alpha_1 & 1 - \alpha_1 \\ 1 - \alpha_2 & \alpha_2 \end{pmatrix}, \tag{22}$$

so that given that X_{j-1} has p.d.f. $p_1(x)$, the probability that X_j has p.d.f. $p_2(x)$ is $1 - \alpha_1$, the probability that X_j has p.d.f. $p_1(x)$ is α_1, etc., independently of the type or length of previous intervals.

Note that while this is the underlying structure, it is assumed that the type of interval is not observable, i.e., we have a univariate point process. Further, while the probabilistic properties of the bivariate process are simple to derive, those of the univariate process are more difficult. Moreover, like the renewal process, the two-state semi-Markov process has very special independence properties that are not likely to be true in real situations. It is thus simply a convenient *model*, not gospel truth; possible relaxations of the assumptions are discussed later.

Now the equilibrium distribution associated with $\mathbf{P}$, the probability vector solution of the matrix equation $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, is

$$\pi_1 = \frac{1 - \alpha_2}{2 - \alpha_1 - \alpha_2}, \qquad \pi_2 = 1 - \pi_1 = \frac{1 - \alpha_1}{2 - \alpha_1 - \alpha_2}, \tag{23}$$

so that in equilibrium the probability is π_1 that at the time when an event occurs we choose the next interval with p.d.f. $p_1(x)$, etc. Thus the marginal p.d.f. of intervals is

$$p(x) = \pi_1 p_1(x) + \pi_2 p_2(x) . \tag{24}$$

Also

$$\mu = E(X) = \pi_1\mu_1 + \pi_2\mu_2 , \tag{25}$$

$$\sigma^2 = \text{var } (X) = \pi_1\sigma_1^{\,2} + \pi_2\sigma_2^{\,2} + \pi_1\pi_2(\mu_1 - \mu_2)^2, \tag{26}$$

where μ_1, $\sigma_1^{\,2}$ and μ_2, $\sigma_2^{\,2}$ are the mean and variance of intervals of the first and second types, respectively.

In Cox and Lewis [11, p. 196–7] an elementary derivation of the serial correlation coefficients ρ_k is given:

$$\rho_k = \frac{(\mu_1 - \mu_2)^2\pi_1\pi_2}{\pi_1\sigma_1^{\,2} + \pi_2\sigma_2^{\,2} + \pi_1\pi_2(\mu_1 - \mu_2)^2}\, \beta^k \triangleq M\beta^k$$
$$(k = 1,2;\cdots) , \tag{27}$$

where $\beta = \alpha_1 + \alpha_2 - 1$. Note that $\beta = \alpha_1 - (1 - \alpha_2) \le \alpha_1 \le 1$, since α_1 is a probability; in fact $|\beta| \le 1$. We have $\beta = -1$ when $\alpha_1 = 0$ and $1 - \alpha_2 = 1$, the case of strict alternation between the two types of intervals. When $\alpha_1 = (1 - \alpha_2)$, $\beta = 0$ and therefore $\rho_k = 0$ for $k \ne 0$. In this case $\alpha_1 + \alpha_2 = 1$, and the process can be shown to be a renewal process with the mixture interval p.d.f. Eq. (24).

Note that while the ρ_k decrease geometrically, they are modulated by M, which is zero, for instance, if $\mu_1 = \mu_2$ or $p_1(x) = p_2(x)$, the latter case giving a renewal process. Otherwise M lies between zero and one. It is close to one, for example, if the difference in the location of the two types of intervals $(\mu_2 - \mu_1)$ is large relative to the dispersions σ_1 and σ_2 of the intervals.

The spectrum $f_+(\omega)$, for this univariate semi-Markov process is derived from (2) and (27), and is

$$f_+(\omega) = \frac{1}{\pi}\left[1 + 2M\beta\left\{\frac{\cos \omega - \beta}{1 + \beta^2 - 2\beta \cos \omega}\right\}\right]. \tag{28}$$

The outstanding feature of the serial correlations and the spectrum is that they depend on $p_1(x)$ and $p_2(x)$ only through *their first two moments. This characteristic of the two-state semi-Markov process makes it attractive for modeling the page-exception process.* It also says that a good deal of the detail of the process is not given by the second-order joint moments of intervals.

The serial partial correlation of order one, ϕ_1, defined by (20) is

$$\phi_1 = \frac{M\beta^2 - M^2\beta^2}{1 - M^2\beta^2} = \frac{M\beta^2(1 - M)}{1 - M^2\beta^2} \le \frac{1}{4}\frac{\beta^2}{1 - (M\beta)^2}$$
$$= \frac{1}{4(1/\beta^2 - M^2)} . \tag{29}$$

Thus no matter whether ρ_1 is positive or negative, ϕ_1 is positive and relatively small. *Note in particular that there is not first-order Markovian dependence in the intervals between events in the semi-Markov model* *for univariate point processes*, i.e., given X_i, the distribution of X_{i+1} is not completely determined.

Properties of the counting process $\{\delta_i\}$ can be derived by straight renewal-theoretic methods. They are, however, messy and depend on the functional form of $p_1(x)$ and $p_2(x)$. The spectrum of counts, in particular, follows from results in Cox and Lewis [11, p. 197] and will not be given here.

9. Estimation of parameters in the model

Estimation of parameters of the semi-Markov model is intimately connected with the detailed assumptions made in the model, and has been done in an ad hoc manner for the three capacities $c = 76$, 197 and 512. The results were shown in Table 5.

We now describe this estimation process, giving details of the further assumptions made, these being primarily the functional form of $p_1(x)$ and $p_2(x)$. Recalling the detailed discussion of the empirical log survivor function in Section 7, the linear tail of the empirical plot, plus the rare-event theory, suggested using a geometric-plus-one distribution for $p_1(x)$, this being the interval, referred to as type 1, with the largest mean value:

$$p_1(x) = p_1^{x-1}(1-p_1) \qquad (0 < p_1 < 1; x = 1,2,\cdots), \tag{30}$$

$$\mu_1 = \frac{1}{1-p_1}. \tag{31}$$

To obtain the convex shape of the log survivor function in Fig. 1, it would be sufficient to mix $p_1(x)$ with another geometric distribution, but the large number of short intervals and the extreme skewness of the interval between page exception distribution suggested using a distribution for the type 2 interval, $p_2(x)$, which is more skewed than the geometric distribution and has a decreasing hazard. A fairly arbitrary choice, made to keep the number of parameters down, is the "negative-binomial-plus-one distribution" (17):

$$p_2(x;k) = \binom{k+x-2}{x-1} p_2^{x-1}(1-p_2)^k$$
$$(0 < p_2 < 1; k > 0; x = 1,\cdots). \tag{32}$$

It remains to take care of the discrete component at $x = 1024$. As remarked above, this could be the result of a third state, but on the same rough grounds it was decided to include this in with the type 2 intervals as a mixture. Thus we have, in toto, for the marginal distribution of the intervals between page exceptions

$$\text{prob}\ \{X = x\} = p_X(x)$$
$$= \pi_1 p_1(x) + (1-\pi_1)[\gamma p_2(x;k) + (1-\gamma)\delta(x-1024)]$$
$$(x = 1, 2, \cdots), \tag{33}$$

where $1 - \gamma$ is the probability that a type 2 interval comes from the distinct degenerate distribution with all its mass at $x = 1024$ and

$$\delta(x-1024) = \begin{matrix} 1 & x = 1024 \\ 0 & \text{otherwise.} \end{matrix} \tag{34}$$

There are now six parameters to estimate:

P
- (i) α_1, the transition probability of type 1|type 1 intervals, (22);
- (ii) α_2, the transition probability of type 2|type 2 intervals, (22);

$p_1(x)$
- (iii) $\mu_1 = 1/(1-p_1)$, the mean of the type 1 interval;

$p_2(x)$
- (iv) $\mu_2 = 1 + kp_2/(1-p_2)$, the mean of the negative binomial distribution in the type 2 interval distribution mixture, (33);
- (v) k, the second parameter of the negative binomial distribution in the type 2 interval distribution mixture, (33);
- (vi) γ, where $1 - \gamma$ is the probability that a type 2 interval comes from the distribution with all its mass at 1024, (34).

The estimated parameters are shown in Table 5. The parameter μ_1 was estimated as the slope of the linear tail of the log survivor function; this involved an eyeball judgement of where the linearity set in; for $c = 76$ this point was taken to be $x = 30{,}000$ and actually $\hat{\mu}_1$ was taken to be the estimated conditional mean of the observations greater than 30,000. Specifically, take all observed intervals X with value greater than 30,000, subtract 30,000 from each such observation and take the average. We are assuming here that $\pi_2 \sum p_2(x)$, where the sum is for x greater than 30,000, is so small that with very high probability all the observations greater than 30,000 came from the tail of the geometric distribution $p_1(x)$.

The parameter γ was estimated by setting $\hat{\pi}_2(1-\hat{\gamma})$ equal to the proportion of intervals between page exceptions with length 1024. This involves a very small bias of picking up intervals from $p_1(x)$ or $p_2(x;k)$.

The remaining four parameters can be estimated by the method of moments, i.e., solving the equations for ρ_1, $E(X)$, $E(X^2)$, $E(X^3)$, using estimates of these quantities from Tables 2 and 4. The equation for ρ_1 is given in (27); $E(X) = \mu$ in terms of the parameters, using (33) and (18) is, for example,

$$\mu = \frac{\pi_1}{1-p_1} + (1-\pi_1)\left[\gamma\left\{1 + \frac{kp_2}{1-p_2}\right\} + (1-\gamma)1024\right]. \tag{35}$$

This estimation method is not a simple procedure and there is no guarantee of a unique solution. Consequently a modified method was used in which estimates of α_1, α_2 and p_2 were obtained using $\hat{p}_1 = \hat{\mu}_1 - 1/\hat{\mu}_1$, $\hat{\gamma}$, and a fixed

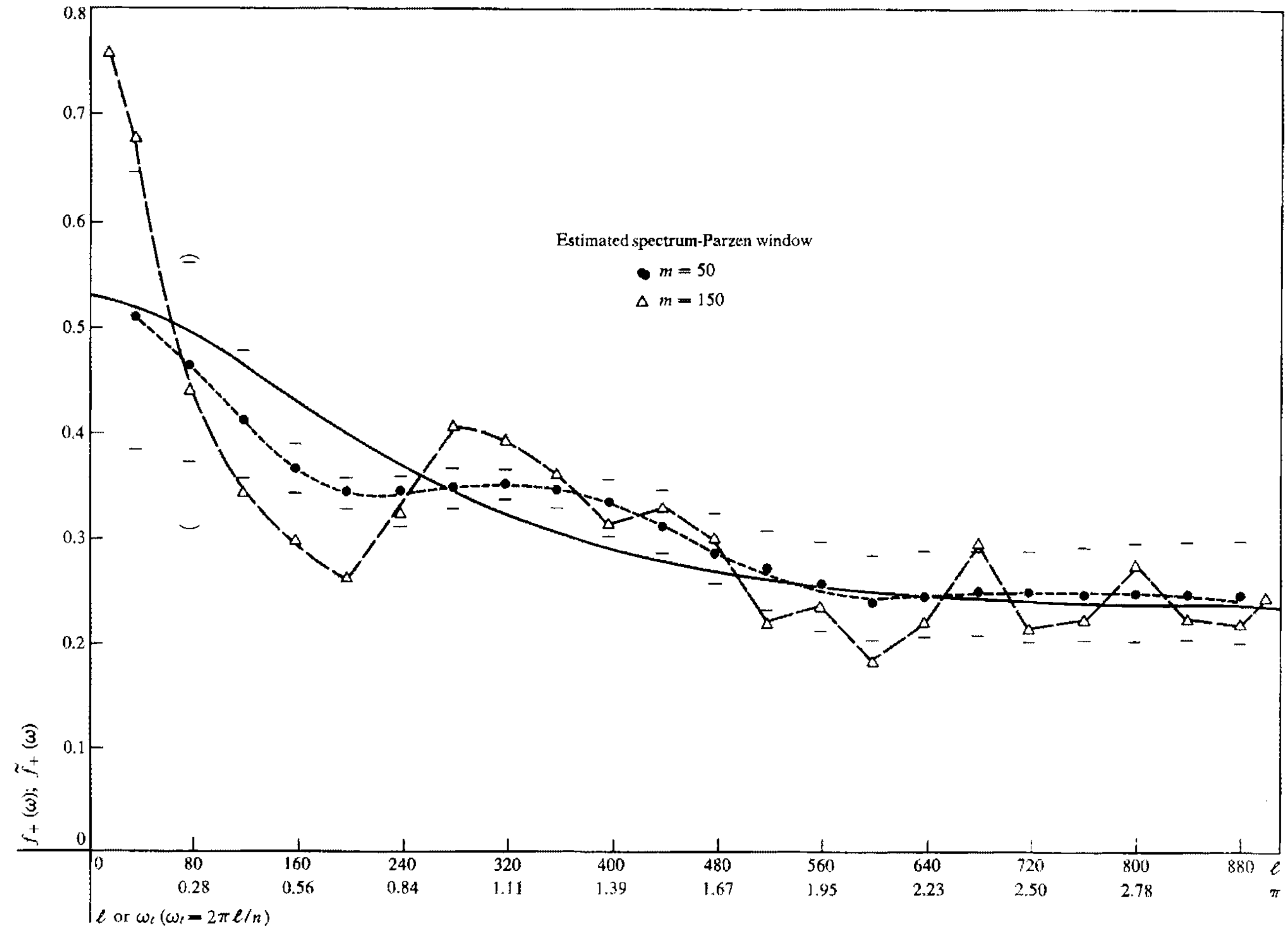

Figure 2 Fitted semi-Markov spectrum of intervals for page exception process for tape A with LRU replacement algorithm. Page size 4K, capacity = 76, n = 1807. Theoretical (solid line) spectrum for semi-Markov process with estimated parameters $\hat{\pi}_1 = 0.066$, $\hat{k} = 0.45$, $\hat{\gamma} = 0.950$, $\hat{\alpha}_2 = 0.962$, $\hat{\alpha}_1 = 0.458$, $\hat{\mu}_1 = 45{,}715$, $\hat{\mu}_2 = 1{,}973.1$.

k. Estimates for several values of k were then used in an expression for the log survivor function and that k that gave the best fit to the empirical log survivor function (Fig. 1) was chosen. The fitted log survivor function is shown as a solid line in Fig. 1.

The estimated parameters are shown in Table 5, and again the nonstationarity for capacity $c = 512$ should be recalled. Nevertheless the overall impression is that the estimated structural parameters $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{k}$ and $\hat{\gamma}$ are fairly consistent over the range of capacities examined, *the main difference at different capacities being the scale of the process, measured by* $\hat{\mu}_2'$ *and* $\hat{\mu}_1$.

We have not used serial correlations beyond the first in this parameter estimation scheme. The serial correlation of lag two enters into the partial serial correlation coefficient of order one, ϕ_1, given by (20). For capacity $c = 76$ this was estimated (Table 4) to be 0.035. Using the estimated parameter values in (29), we get $\hat{\beta} \sim 0.42$, $\hat{M} \sim 0.45$ and $\hat{\phi}_1 \sim 0.02$. The values are roughly comparable, indicating a reasonable goodness-of-fit of the model. This is explored further in the next two sections.

10. Test of fit based on the spectrum of intervals

We now consider the fit of the model, using the estimated parameters, by examining the computed and estimated spectrum of intervals for capacity $c = 76$.

The computed spectrum $\hat{f}_+(\omega)$, obtained by using the estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{k}$ and $\hat{\gamma}$ from Table 5 to calculate the constants in the expression (28), is shown in Fig. 2. It has the characteristic shape of the spectrum of a first-order process with positive dependence.

Two smoothed estimates $\tilde{f}_+(\omega)$ of the spectrum of intervals are shown in Fig. 2. These were obtained using a Parzen lag window λ_j [11, p. 108] directly on the estimated serial correlations $\tilde{\rho}_j$ given by (1):

$$\tilde{f}_+(\omega) = \frac{1}{\pi}\left\{1 + \sum_{j=1}^{m} \lambda_j \tilde{\rho}_j \cos(j\omega)\right\}, \qquad (36)$$

where (e.g., [11, p. 108]):

$$\lambda_j = \begin{cases} 1 - \dfrac{6j^2}{m^2(1-j/m)} & (j \le m/2), \\ 2(1-j/m)^3 & (m/2 < j \le m), \\ 0 & (j > m). \end{cases} \qquad (37)$$

The reason for smoothing directly on the estimated correlation coefficients instead of on the periodogram is that this is the way it was written into the SASE IV program. The two methods are equivalent [13].

The coefficient of variation of individual estimates $\tilde{f}_+(\omega_\ell)$, $\omega_\ell = 2\pi\ell/n$, using the Parzen window with a given m is approximately (m/n). Thus as m decreases (bandwidth increases) the coefficient of variation decreases, the curve gets smoother, and individual estimates become more biased. The question of which m gives the best "resolution" is purely empirical; estimates for $m = 50$ and $m = 150$ are shown as black circles and triangles, respectively. The $m = 50$ window seems to give the best resolution.

The estimates were actually obtained by applying the Parzen window estimate (36) to four distinct sections of the intervals between events and averaging the four estimates obtained at each frequency $\omega_\ell = 2\pi\ell/(n/4)$ (see [13] for details). This allows one to obtain estimates (three degrees of freedom) of the standard deviation of the averaged estimates of $\tilde{f}_+(\omega_\ell)$. The estimated standard deviations are shown in Fig. 2 as horizontal lines ($m = 50$), one estimated standard deviation on either side of the smoothed estimated spectral value. These are not true confidence intervals but are shown only to give an idea of the precision of the smoothed spectral estimate.

One such estimate for $m = 150$ is shown at $\ell = 80$ as round brackets about the estimated value $\tilde{f}_+(\omega_{80})$, designated by Δ. It should be approximately $(150/50)^{\frac{1}{2}} = (3)^{\frac{1}{2}}$ as wide as the band for $m = 50$.

Within the limits of the sampling fluctuations the estimated and computed spectra tally well. A formal test of fit [11, p. 172] based on the adjusted cumulated periodogram (from Eq. (4))

$$S_\ell'' = \frac{\sum_{k=1}^{\ell} I(\omega_k)/\hat{f}(\omega_k)}{\sum_{k=1}^{v} I(\omega_k)/\hat{f}(\omega_k)}$$

was not made because of computational limitations. The estimated and computed spectra for capacities $c = 197$ and $c = 512$ have not been displayed in the paper.

Note that this is a relatively weak form of goodness-of-fit since the theoretical spectrum (28) does not depend on detailed assumptions about $p_1(x)$ and $p_2(x)$, and the computed spectrum $\hat{f}_+(\omega)$ has been adjusted to fit $\tilde{f}_+(\omega)$ to some extent by estimating the parameters in the semi-Markov model from the data.

A far more critical test of the model is obtained by comparing estimated and computed spectra of counts, and this we do next.

11. Tests of fit based on the spectrum of counts

We now consider the count spectrum $g(\omega)$ given by (9), the spectrum of the binary count sequence $\{\delta_i(c)\}$, for capacity $c = 76$. Again we will normalize and actually work with a spectrum for positive ω; $g_+(\omega) = 2\pi g(\omega)$. For a Bernoulli process (Poisson) $g_+(\omega) = 1$ for all ω.

Again it is important to note that this is a spectrum for page reference time, not serial number of the page exceptions. Thus $\omega = 2\pi/T$, where T is the period of the cycle in page reference time, $T = 2$ being the shortest possible period, corresponding to page exceptions every other reference.

With the highly skewed marginal distribution of intervals between page exceptions seen in this study, there is an extreme problem of resolution of spectral components. Thus with the many short intervals there might be significant dynamic effects with periods of 2, 3, etc; there might also be dynamic effects around periods corresponding to the estimated mean time between page exceptions, to the mean times between page exceptions of the two types of intervals and to the period 1024 of the discrete component in the intervals. The latter is of particular interest as the estimated spectrum should tell us whether to model the discrete (1024) component as a third state in a semi-Markov model, rather than lumping it with type 2 intervals.

Let t_j be the observed value of $T_j(c)$, the number of page references to the jth page exception. Then the periodogram for $\{\delta_i(c)\}$ is, using (9) and (4),

$$2I(\omega_\ell) = 1 + 2\sum_{k=1}^{t_0-1} \hat{\rho}_k \cos(\omega_\ell k) \sim \left|\sum_{j=1}^{n} e^{i\omega_\ell t_j}\right|^2, \qquad (38)$$

where

1. t_0(=8,802,464) is the number of page references observed, or the total time of observation;
2. n is the number of page exceptions observed in $(0,t_0]$;
3. here $\omega_\ell = 2\pi\ell/t_0$, and not $2\pi\ell/n$ as for the spectrum of intervals;
4. the period T_ℓ corresponding to ω_ℓ, since $\omega_\ell = 2\pi/T_\ell = 2\pi\ell/t_0$, is $T_\ell = t_0/\ell$;
5. the quantity within the absolute value sign is the finite Fourier transform of $\delta_i(c)$ or the finite Fourier-Stieljes transform of $N_i(c)$:

$$\int_0^{t_0} e^{i\omega_\ell u}\, dN_u(c) = \sum_{k=0}^{t_0} \delta_k(c) e^{i\omega_\ell k} = \sum_{j=1}^{n} e^{i\omega_\ell t_j},$$

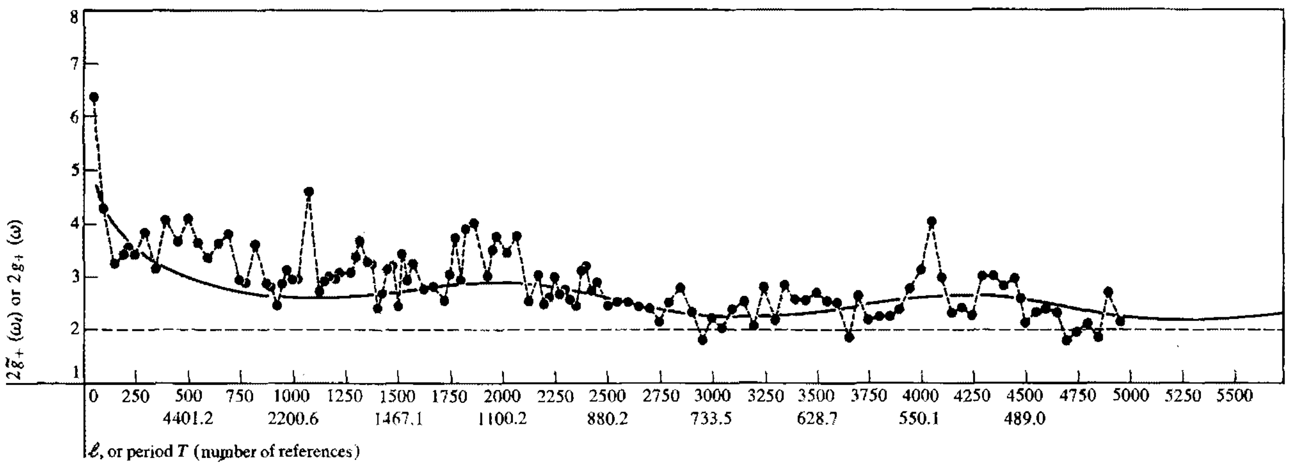

Figure 3 Fitted semi-Markov spectrum of counts for page exception process for tape A with LRU replacement algorithm. Page size 4K, capacity = 76, n = 1807. Smoothed spectrum, average of four sections, each smoothed with quadratic window. 50 points (dashed line). Theoretical spectrum (solid line) for semi-Markov process with estimated parameters $\hat{\gamma} = 0.950$, $\hat{k} = 0.45$, $\hat{p}_2 = 0.99977748$, $\hat{p}_1 = 0.99997813$, $\hat{\alpha}_1 = 0.458$, $\hat{\alpha}_2 = 0.962$.

since $\delta_k(c)$ only equals one where $k = t_j$, for some j.

No further details of the estimation of $g_+(\omega)$ are given here (see [11, Ch. 5; 21]). In Fig. 3 we give an estimate $2\tilde{g}_+(\omega_\ell)$ of $2g_+(\omega)$ obtained by smoothing the periodogram $2I(\omega_\ell)$. The smoothing involved a 50-point quadratic window [11, p. 131] applied directly to periodograms for four nonoverlapping sections of the data. The four estimates were then averaged, at each ω_ℓ, to give a final estimate.

Again the sectioning was done to save computation time and to obtain estimates of the variances of the individual estimates $2\tilde{g}_+(\omega_\ell)$, and the adequacy of the amount of smoothing was judged empirically.

The estimate $2\tilde{g}_+(\omega_\ell)$ in Fig. 3 is shown as a jagged dashed line, the straight line dashed segments connecting the individual estimates. It was only computationally feasible to go out to $\ell = 4900$, or $\omega_\ell = 2\pi \times 4900/8{,}802{,}464 \sim 0.0035\pi$. Up to that point the estimated spectrum lies above the value 2, which is the theoretical value for a Bernoulli process.

The computed spectrum of counts, $2\hat{g}_+(\omega)$, for the two-state semi-Markov model, using the estimated parameters, is shown as a solid line in Fig. 3. In general it agrees very well with the estimated spectrum.

The main anomaly in the figure is the small peaks in the estimated spectrum at periods of approximately $T = 2 \times 1024$, $T = 1024$, and $T = 1024/2 = 512$. These are multiples or submultiples of the length 1024 of the distinct group of intervals between page exceptions seen in Fig. 1. *The conclusion from the spectral analysis would be that these intervals should be modeled as a separate state in the semi-Markov model, instead of being lumped in with state 2.*

Computation of the spectra of counts for the other capacities was not done. It should be noted that these computations are extremely expensive if the whole spectrum is required. It now seems possible, however, to use fast Fourier transform techniques [22] and this type of analysis should become more feasible.

12. Summary and conclusions

The result of this study has been the following.

1. On the basis of observation of data on page reference patterns and (LRU) page exceptions, we have postulated a two-state semi-Markov model for the univariate page exception process for a given memory capacity c.
2. Parameters of the model have been estimated in an ad hoc manner from the data.
3. The goodness-of-fit of the model to the data has been examined by estimating the spectra of counts and intervals and comparing these to theoretical spectra. The fit is generally good, the main exception being the conclusion that the distinct group of intervals of length 1024 should be modeled dynamically as a third state in the model.
4. The effect of the change in the capacity c on the model seems to be roughly to change the scale of the distributions of times between exceptions without changing structural parameters, e.g., transition probabilities, drastically.

Limitations of the study have been the following.

1. The ad hoc method of estimating the parameters in the model. The main need here is to formalize the estimation procedure.
2. Estimates of parameters from sections of the data are needed to examine the sensitivity of the estimation procedure.
3. The study needs to be done for more capacities c in order to get a better idea of the change of the parameters with c. Other (smaller) page sizes should also be studied. The effect of other replacement algorithms is also of interest.
4. Eventually a study of the joint properties of page exception processes $\{T_j(c)\}$ at different capacities c would have to be done. It might be simpler, rather than to deal directly with multivariate page exception processes, to go back to the distance strings $\{D_i\}$ and model them.
5. The page exception process of only one other tape was examined, showing roughly the same characteristics on a different scale.

The model for the page exception process itself has utility for driving models of uniprogrammed or multiprogrammed paging machines (see, e.g., [4]). Recalling that it is at best a model that fits the details observable from finite samples, there are also other limitations.

1. There is a need to relate parameter values not only to different capacities c, but to some measurable characteristic of the program. This is a difficult calssification problem.
2. There is a need to model the changes in parameters over time. As remarked before, the model is a *micro-model in time*, and dynamic changes with time will occur. Are these simply changes in parameter values, or is the model particular to certain sections of the data?
3. Details of the model may not be correct and it might perhaps have been better to use generalizations such as the pseudo-Markov model of Ekholm [23]. This, however, introduces additional parameters, and it does not seem possible to differentiate such fine points from the available data.

Finally we remark on theoretical underpinnings for the model. Semi-Markov models, like renewal models, have very special dependency structure that seldom occurs in practice. Allusion has been made to the theory of rare events to suggest a Poisson process. In the face of two-state phenomena in the basic distance string process, will the "rare" events of page exceptions $\{D_i > c\}$ lead to approximate semi-Markov models? The question is difficult and will be addressed elsewhere. We note, however, two points.

In available "rare-event" theory, events are deleted independently with probability p, and as $p \to 1$ the process of remaining events on a suitably scaled time axis, goes to a Bernoulli or Poisson process (see [3] for details). We know of no work in which the deletion process is more general.

However, a situation in which a type of thinning occurs on a self-generated basis is in level-crossing theory [24]. Again Poisson or Bernoulli limits occur. For discrete time, join the points $\{D_i, i\}$ and $\{D_{i+1}, i+1\}$ by a straight line. If the line goes up across a level c, we say we have an upcrossing event at $i + 1$, if it goes down across a level c we have a downcrossing at $i + 1$; see [25]. The page exception process we have considered in this paper is slightly different. A page exception occurs at each up-crossing time plus the times i between an up-crossing and the successive down-crossing. No theory exists for this process, although for high c this process is very close to the up-crossing process. No other limits than Poisson limits are known.

One special case is that in which the $\{D_i\}$ are independent. Then the page exception process is a Bernoulli process. If the D_i have two distributions that occur during different alternating periods of random duration, and if the D_i are independent, given the period, the level crossings are a doubly stochastic Bernoulli process [26]. It is known then that the spectrum of counts is just the spectrum of the alternating process, but is not known how close the exception process is to a semi-Markov process.

Acknowledgment

The authors gratefully acknowledge the programming assistance of Mrs. Sandra L. Keating and some helpful technical criticism from C. A. Galtieri.

References

1. D. R. Cox and P. A. W. Lewis, "Multivariate Point Processes," *Proc. 6th Berkeley Symposium on Math. Stat. and Prob.* **3**, L. LeCam, J. Neyman, E. L. Scott, (eds.), University of California Press, Berkeley, 401–448 (1972).
2. R. L. Mattson, J. Gecsei, D. R. Slutz and I. L. Traiger, "Evaluation techniques for storage hierarchies," *IBM Syst. J.* **9**, 2, 78–117 (1970).
3. D. Daley and D. Vere-Jones, "A Summary of the Theory of Point Processes." *Stochastic Point Processes*, P. A. W. Lewis, (ed.), John Wiley and Sons, Inc., New York, 1972, 299–383.
4. P. A. W. Lewis and G. S. Shedler, "A Cyclic-Queue Model of System Overhead in Multiprogrammed Computer Systems," *J. ACM* **18**, 2, 199–220 (1971).
5. G. S. Shedler and C. Tung, "Locality in Page-Reference Strings," *SIAM J. on Computing* **1**, 3, 218–241 (1972)).
6. P. A. W. Lewis and P. C. Yue, "Statistical Analysis of Program Reference Patterns in a Paging Environment," *Proc. of IEEE Conference on Computers*, Boston, 1971.
7. M. G. Kendall, and A. Stuart, *The Advanced Theory of Statistics*, Vol. 3, 2nd Edition, Griffih, London, 1968.

8. T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley and Sons, Inc., New York, 1971.
9. W. Feller, *An Introduction to Probability Theory and Its Applications*. Vol. II, 2nd Edition, John Wiley and Sons, Inc., New York, 1971.
10. P. A. P. Moran, "Testing for Serial Correlation with Exponentially Distributed Variates," *Biometrika* **54**, 395–401 (1967).
11. D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*. Methuen, London, and Barnes and Noble, New York, 1966.
12. J. Durbin, "Tests of Serial Independence Based on the Cumulated Periodogram," *Bull. I.S.I.* **42**, 1039–1048 (1969).
13. J. W. Cooley, P. A. W. Lewis and P. D. Welch, "The Application of the Fast Fourier Transformation to the Estimation of Spectra and Cross Spectra," *J. Sound Vib.* **12**, 3, 339–352 (1970).
14. P. A. W. Lewis, "Recent Results in the Statistical Analysis of Univariate Point Processes," *Stochastic Point Processes*, P. A. W. Lewis, (ed.). John Wiley and Sons, Inc., New York, 1972, 1–54.
15. P. Billingsley, *Statistical Interference for Markov Processes*, University of Chicago Press, Chicago, 1961.
16. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd Edition, John Wiley and Sons, Inc., New York, 1968.
17. P. A. W. Lewis, A. M. Katcher and A. H. Weis, *SASE IV – An Improved Program for the Statistical Analysis of Series of Events*. IBM Research Report RC-2365, 1969.
18. M. Brown, "Statistical Analysis of Non-Homogeneous Poisson Processes," in *Stochastic Point Processes*, P. A. W. Lewis, (ed.), John Wiley and Sons, Inc., New York, 1972, 67–89.
19. D. R. Cox, *Renewal Theory*, Methuen, London, 1962.
20. G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco, 1970.
21. P. A. W. Lewis, "Remarks on the Theory, Computation and Application of the Spectral Analysis of Series of Events," *J. Sound Vib.* **12**, 353–75 (1970).
22. A. S. French and A. V. Holden, "Alias-free Sampling of Neuronal Spike Trains," Kybernetik **8**, 165–71 (1971).
23. A. Ekholm, "A Generalization of the Two-state Two Interval Semi-Markov Model," *Stochastic Point Processes*, P. A. W. Lewis, (ed.), John Wiley and Sons, Inc., New York, 1972, 272–284.
24. M. R. Leadbetter, "Point Processes Generated by Level Crossings," *Stochastic Point Processes*, P. A. W. Lewis, (ed.), John Wiley and Sons, Inc., New York, 1972, 436–467.
25. H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. John Wiley and Sons, Inc., New York, 1967.
26. J. Grandell, "Statistical Inference for Doubly Stochastic Poisson Processes," *Stochastic Point Processes*, P. A. W. Lewis (ed.). John Wiley and Sons, Inc., New York, 1972, 90–121.

Received December 22, 1972

Dr. Lewis is located at the Naval Postgraduate School, Monterey, California; Mr. Shedler is at the IBM Research Division Laboratory, Monterey and Cottle Roads, San Jose, California 95193.