A STORAGE MODEL WITH A TWO-STATE RANDOM ENVIRONMENT

by

Offer Kella

Department of Operations Research
Yale University
84 Trumbull Street
New Haven, CT 06520

and

Ward Whitt

AT&T Bell Laboratories
Room 2C-178
Murray Hill, NJ 07974-2070

July 13, 1990

Revised: June 3, 1991

To appear in *Operations Research*.

Abstract

Motivated by queues with service interruptions, we consider an infinite-capacity storage model with a two-state random environment. The environment alternates between ‘‘up’’ and ‘‘down’’ states. In the down state, the content increases according to one stochastic process; in the up state, the content decreases according to another stochastic process. We describe the steady-state behavior of this system under assumptions on the component stochastic elements. For the special case of deterministic linear flow during the up and down states, we show that the steady-state content is directly related to the steady-state workload or virtual waiting time in an associated G/G/1 queue, thus supplementing results of Gaver and Miller (1962), Miller (1963) and Chen and Yao (1990).

Subject classifications: Probability, regenerative processes: a storage model in a two-state random environment. Queues, approximations: a fluid model with random disruptions. Inventory/production, operating characteristics: a fluid model with alternating up and down times.

Gaver and Miller (1962), Miller (1963) and Chen and Yao (1990) studied a stochastic storage model that we believe is especially promising for operations research applications. The model is an infinite-capacity linear fluid model in a random environment, which can serve as an approximation for a G/G/1 queue when the major source of variability is the availability of the server rather than the interarrival times and service times. In particular, the environment is viewed as a regenerative stochastic process, with down times D_k having mean d and up times U_k having mean u, where the sequence of random vectors $\{(D_k, U_k) : k \geq 1\}$ is i.i.d. (but the joint distribution of (D_1, U_1) is general). During the down times, there is deterministic linear net flow into the buffer at rate λ; during up times, there is deterministic linear flow out of the buffer at rate $\mu - \lambda$ whenever the buffer is not empty ($\mu > \lambda$). (Think of an arrival rate λ and a service rate μ.) We require that $\lambda d < (\mu - \lambda) u$ to have long-run stability. Chen and Yao (1990) show that this fluid model is a good approximation for a D/D/1 queue in a random environment when the interarrival times and service times are relatively short compared to the up and down times.

Chen and Yao (1990) obtained a relatively complicated description of the steady-state buffer content in terms of an associated G/G/1 queue having service times λD_k and interarrival times $(\mu - \lambda) U_k$. Our first purpose is to supplement this analysis by showing that the steady-state buffer content, say Z, is in fact directly related to the steady-state workload (or virtual waiting time) in this G/G/1 queue, say V. In particular, *we show that* $(Z | Z > 0)$ *is distributed exactly as* $(V | V > 0)$, *and that the distribution of V has a relatively simple form despite the allowed dependence between service times and interarrival times.* We also determine $P(Z > 0)$, which typically does *not* coincide with $P(V > 0)$. Explicit formulas for the distribution of Z when U_1 and D_1 are independent and one of them has an exponential distribution are then immediate (corresponding to the classical M/G/1 and GI/M/1 models). Consequently, this stochastic storage model is remarkably tractable; hence its appeal.

Our second purpose is to consider more general models in the same spirit, in part, to better understand why the relationship above holds. Our first idea was to replace the deterministic linear flows by deterministic nonlinear flows, and indeed much of the analysis can be carried through for this case, provided that the flow is nondecreasing during the down periods and nonincreasing during the up periods. This covers the interesting case in which the up and down intervals are composed of subintervals with different linear rates on each subinterval.

Our next idea was to consider stochastic flows during the up and down periods, and this is the model we investigate first here. We define the general model in Section 1. We discuss the connection to the associated single-server queue in Section 2; this connection depends on the monotonicity of the sample paths of the two flow processes. The monotonicity is very natural during the down periods, but not so natural during the up periods, except in the case of the fluid model. However, the monotonicity during the up periods may sometimes be a reasonable approximation, which is worth considering because the analysis simplifies considerably.

In Section 3 we see what happens when the flow out is linear and deterministic. In this case, we show that the content process restricted to up periods coincides with a time-scaled version of the workload process in a single-server queue. (This is a partial explanation for the characterization of the full steady-state content Z for the fluid model.)

We first make stochastic assumptions in Section 4. In particular, we assume that the sequence of down and up times and associated flow processes is i.i.d. (The flow processes themselves are still general at first.) In Section 4 we characterize the steady-state behavior under successively stronger assumptions. In Section 5 we treat the linear fluid model of Gaver and Miller (1962), Miller (1963) and Chen and Yao (1990). Finally, in Section 6 we consider the extension arising when the down interval is composed of random subintervals with associated linear deterministic flow rates.

Before proceeding, we mention that we apply our results for the linear fluid model in Kella and Whitt (1992) to analyze the distributions of successive buffer contents in a tandem fluid network with stochastic input (in particular, nondecreasing Lévy process input). Perhaps the main contribution of Kella and Whitt (1992) is to determine the non-product-form steady-state distribution for the case of two tandem buffers. It turns out that, without any disruptions, the content of each buffer can be viewed as the linear fluid model in a random environment considered here. That application was our initial motivation for this work.

For additional background on fluid models, see Keilson and Rao (1970), Meyer, Rothkopf and Smith (1979, 1983), Newell (1982), Mitra (1988), Chen and Mandelbaum (1991), and references in these sources. For additional background on queues with service interruptions (or server vacations), see Doshi (1986, 1990).

1. The General Model

There are four elements in the general model: two sequences of strictly positive random variables, $\{D_k : k \geq 1\}$ and $\{U_k : k \geq 1\}$, and two sequences of continuous-time real-valued stochastic processes, $\{\{R_k(t) : t \geq 0\} : k \geq 1\}$ and $\{\{S_k(t) : t \geq 0\} : k \geq 1\}$, having right-continuous sample paths with left limits. We think of D_k and U_k as successive down and up times. The storage content grows according to $\{R_k(t) : t \geq 0\}$ during the k^{th} down time and decreases according to $\{S_k(t) : t \geq 0\}$ during the k^{th} up time.

In particular, let

$$T_k = D_1 + U_1 + \ldots + D_k + U_k \, , \; k \geq 1 \, , \tag{1}$$

with $T_0 = 0$; i.e., T_k is the k^{th} environment cycle time. We assume that the first down time begins at time $t = 0$. Then the net input process $\{Y(t) : t \geq 0\}$ can be defined by

$$Y(t) = \begin{cases} Y(T_k -) + R_{k+1}(t - T_k) , & T_k \le t < T_k + D_{k+1} \\ Y(T_k -) + R_{k+1}(D_{k+1} -) - S_{k+1}(t - T_k - D_{k+1}) , & T_k + D_{k+1} \le t < T_{k+1} , \end{cases} \quad (2)$$

for $k \ge 0$ with $Y(0-) = 0$ (where $X(t-) = \lim_{s \downarrow 0} X(t - s)$). Then the content process $\{Z(t) : t \ge 0\}$ is obtained by applying the reflection map to the net input process, i.e.,

$$Z(t) = Y(t) - \min\{0, \mathit{inf}\{Y(s) : 0 \le s \le t\}\} , \ t \ge 0 ; \quad (3)$$

see p. 19 of Harrison (1985). (Hence, $Z(0) = [R(0)]^+$, where $[x]^+ = \max\{x, 0\}$.)

We will also consider content processes $\{Z_u(t) : t \ge 0\}$ and $\{Z_d(t) : t \ge 0\}$ restricted to up and down intervals, respectively; e.g., $Z_u(t)$ represents $Z(s)$ where s is the time when the cumulative up time first reaches t. For this purpose, define an environment state indicator process by

$$I(t) = \begin{cases} 1 , & T_k + D_{k+1} \le t \le T_{k+1} \\ 0 & T_k \le t < T_k + D_{k+1} \end{cases} \quad (4)$$

for $k \ge 0$. Processes depicting the cumulative up and down time during the interval $[0, t]$ can then be defined by

$$C_u(t) = \int_0^t I(s)\, ds , \ t \ge 0 , \quad (5)$$

and

$$C_d(t) = t - C_u(t) , \ t \ge 0 . \quad (6)$$

Now define inverse processes by

$$C_u^{-1}(t) = \inf\{s \ge 0 : C_u(s) > t\} , \ t \ge 0 , \quad (7)$$

assuming that $C_u(t) \to \infty$ as $t \to \infty$, and similarly for $C_d^{-1}(t)$. Then let

$$Z_u(t) = Z(C_u^{-1}(t)) , \ t \geq 0 , \tag{8}$$

and

$$Z_d(t) = Z(C_d^{-1}(t)) , \ t \geq 0 . \tag{9}$$

Note that (7)–(8) makes $Z_u(U_1) = Z(D_1 + U_1 + D_2)$ as opposed to $Z(D_1 + U_1)$.

2. An Associated Single-Server Queue

Under an extra assumption on the basic stochastic processes $\{R_k(t) : t \geq 0\}$ and $\{S_k(t) : t \geq 0\}$, we can relate our model to an associated single-server queue with unlimited waiting space and the first-in first-out discipline. In the queueing model, let a 0^{th} customer arrive at an empty system at time $t = 0$. Moreover, let the service time of customer $k - 1$ be $R_k(D_k -)$ and let the interarrival time between customers $k - 1$ and k be $S_k(U_k -)$. Let $\{W_k : k \geq 0\}$ be the associated sequence of waiting times, defined as usual by

$$W_k = [W_{k-1} + R_k(D_k -) - S_k(U_k -)]^+ , \ k \geq 1 , \tag{10}$$

where $W_0 = 0$.

Theorem 2.1. If the sample paths of $\{R_k(t) : k \geq 0\}$ and $\{S_k(t) : t \geq 0\}$ are nondecreasing for all k w.p.1, then

$$Z(T_k -) = W_k , \ k \geq 0 , \quad \text{w.p.1.}$$

Proof. Under the monotonicity assumptions, we can represent the content process by

$$Z(t) = \begin{cases} Z(T_k -) + R_{k+1}(t - T_k) , & T_k \leq t < T_k + D_{k+1} \\ [Z(T_k -) + R_{k+1}(D_{k+1} -) - S_{k+1}(t - T_k - D_{k+1})]^+ , & T_k + D_{k+1} \leq t < T_{k+1} \end{cases} \tag{11}$$

for $k \geq 0$ with $Z(0-) = 0$. (This can be established from (2) and (3) by induction on k. ■

Our main results later are connections between the content processes $\{Z(t) : t \geq 0\}$ and

$\{ Z_u(t) : t \geq 0 \}$ and the workload (or virtual waiting time) process of this single server queue. The workload $V(t)$ depicts the remaining service time of all customers in the system at time t, i.e.,

$$V(t) = [W_{N(t)} + R_{N(t)+1}(D_{N(t)+1}-) - t + \sum_{k=1}^{N(t)} S_k(U_k-)]^+ \ , \ t \geq 0 \ , \tag{12}$$

where

$$N(t) = \max \{ k \geq 0 : \sum_{j=1}^{k} S_j(U_j-) \leq t \} \ , \ t \geq 0 \ . \tag{13}$$

3. Linear Decline

A second nice connection to the single-server queue occurs when we assume the process $\{ S_k(t) : t \geq 0 \}$ is deterministic and linear, i.e.,

$$S_k(t) = st \ , \ t \geq 0 \ . \tag{14}$$

Theorem 3.1. If, in addition to the monotonicity assumptions of Theorem 2.1, (14) holds, then

$$Z_u(t) = V(st) \ , \ t \geq 0 \ ,$$

where $\{ V(t) : t \geq 0 \}$ is the workload process of the single-server queue in (12).

Proof. Both processes have jumps up of $R_{k+1}(D_{k+1}-)$ at times $U_1 + \ldots + U_k$ and otherwise decrease linearly at rate s, with a reflecting barrier at the origin. ■

4. Steady State

We now investigate the steady-state behavior of our model under stochastic assumptions. In particular, we assume

Main Independence Assumption. $\{ (D_k, U_k, \{ R_k(t) : t \geq 0 \}, \{ S_k(t) : t \geq 0 \}) : k \geq 1 \}$ is an i.i.d. sequence with $E[D_1] = d < \infty$, $E[U_1] = u < \infty$, $E[R_1(D_1-)] < \infty$ and $0 < E[S_1(U_1-)] < \infty$.

Let $\Rightarrow$ denote convergence in distribution. Let G/G/1 denote the single-server queueing model in which the sequence of ordered pairs of interarrival times and service times is stationary (without any independence assumed). Let GI/G/1 denote the special case in which all the random variables are independent. We will typically be concerned with G/G/1 with extra independence (but not GI/G/1). As a consequence of Theorem 2.1, we have

Theorem 4.1. Under the Main Independence Assumption and the monotonicity assumptions of Theorem 2.1, there exists a proper random variable Z_e such that

$$Z(T_k -) \Rightarrow Z_e \quad \text{as } k \to \infty$$

if and only if $\rho \equiv E[R_1(D_1 -)]/E[S_1(U_1 -)] < 1$. Moreover, Z_e is distributed as the steady-state waiting time in a G/G/1 queue with (possibly dependent) service time $R_1(D_1 -)$ and interarrival time $S_1(U_1 -)$.

Proof. Apply results for the GI/G/1 queue, e.g., Propositions 1.1 and 1.2 on p. 181 of Asmussen (1987). The GI/G/1 theory applies because $\{R_k(D_k -) - S_k(U_k -) : k \geq 1\}$ is i.i.d. ■

Using the regenerative structure of the environment, we can also establish limits for the continuous-time process. For any nonnegative random variable X with finite mean, let X^* be a random variable with the stationary-excess (or equilibrium-residual-life) distribution associated with the distribution of X; i.e.,

$$P(X^* \leq x) = \frac{1}{EX} \int_0^x P(X > s)\, ds \ . \tag{15}$$

Let $\stackrel{d}{=}$ denote equality in distribution.

Remark (4.1) Below we want to consider $R_1(D_1^*)$. When $\{R_1(t) : t \geq 0\}$ and D_1 are independent, the intended meaning is clear, but in general they are dependent. Let $R_1(D_1^*)$ have the distribution

$$P(R_1(D_1^*) > x) = \frac{1}{ED_1} E \int_0^{D_1} 1_{\{R_1(t) > x\}} \, dt$$
$$= \frac{1}{ED_1} \int_0^\infty P(R_1(t) > x \, , \, D_1 > t) \, dt$$
$$= \frac{1}{ED_1} \int_0^\infty P(R_1(t) > x | D_1 > t) P(D_1 > t) \, dt$$
$$= \int_0^\infty P(R_1(t) > x | D_1 > t) \, dP(D_1^* \le t) \ .$$

Theorem 4.2. In addition to the assumptions of Theorem 4.1 with $\rho < 1$, suppose that D_1, U_1 and $D_1 + U_1$ have non-lattice distributions. Then

(a) $Z_d(t) \Rightarrow Z_d \overset{d}{=} Z_e + R_1(D_1^*)$,
(b) $Z_u(t) \Rightarrow Z_u$ and
(c) $[Z(t), I(t)] \Rightarrow [Z, I]$ as $t \to \infty$,

where Z_e is from Theorem 4.1, $R_1(D_1^*)$ is independent of Z_e, $Z_d \overset{d}{=} (Z | I = 0)$, $Z_u = (Z | I = 1)$,

$$P(Z > z) = \frac{u}{u+d} P(Z_u > z) + \frac{d}{u+d} P(Z_d > z) \tag{16}$$

and

$$P(I = 1) = 1 - P(I = 0) = \frac{u}{u+d} \ .$$

Proof. Apply basic regenerative process theory as on pp. 125-127 of Asmussen (1987). See Green (1982) for an explicit treatment of the subintervals. ■

Remark (4.2) The independence in the Main Independence Assumption can be relaxed; see Wolff (1988). Also $D_1 + U_1$ is automatically nonlattice if in addition D_1 and U_1 are independent.

Theorem 4.3. If, in addition to the assumptions of Theorem 4.2,

$$U_k, D_k, \{R_k(t) : t \ge 0\} \quad \text{and} \quad \{S_k(t) : t \ge 0\}$$

are mutually independent, then

$$Z_u \stackrel{d}{=} [Z_e + R_1(D_1-) - S_1(U_1-^*)]^+$$

with the component random variables being independent.

Proof. This is a standard result for the GI/G/1 queue; see p. 189 of Asmussen (1987). ■

We can obtain a useful further characterization of Z_u under the linearity condition (14).

Theorem 4.4. If (14) holds in addition to the assumptions of Theorem 4.2, then Z_u is distributed as the steady-state workload in the associated G/G/1 queue; i.e.,

$$P(Z_u > 0) = \rho \equiv E[R_1(D_1-)]/su \tag{17}$$

and

$$P(Z_u > x | Z_u > 0) = P(Z_e + R_1(D_1-)^* > z)$$

where $R_1(D_1-)^*$ is independent of Z_e.

Proof. Apply Theorem 3.5, p. 189, of Asmussen (1987) and Theorem 3.1 here, noting that the standard GI/G/1 argument holds if the service time and waiting time of customer k are independent. Alternatively, apply (4.5.1) and (4.5.7) of Franken et al. (1981). ■

From Theorems 4.2 and 4.4, we obtain a nice characterization of Z in terms of $R_1(D_1^*)$ and $R_1(D_1-)^*$.

Corollary 1. Under the assumptions of Theorem 4.4,

$$P(Z > z) = \frac{u}{u+d} \rho P(Z_e + R_1(D_1-)^* > z) + \frac{d}{u+d} P(Z_e + R_1(D_1^*) > z)$$

for ρ in (17), with $R_1(D_1-)^*$ and $R_1(D_1^*)$ independent of Z_e.

Corollary 2. Under the assumptions of Theorem 4.4,

$$E[Z_u] = \rho E[Z_e] + \frac{E[R_1(D_1-)^2]}{2E[R_1(D_1-)]} .$$

It is natural to develop approximations for $E[Z_u]$ and $E[Z]$ by exploiting approximations for $E[Z_e]$; e.g.,

$$E[Z_e] \approx \frac{\tau\rho(c_a^2 + c_s^2 - 2c_{as}^2)}{2(1-\rho)} \ , \tag{18}$$

where $\tau = ER_1(D_1-)$, while c_s^2 and c_a^2 are the squared coefficients of variation (variance divided by the square of the mean) of $R_1(D_1-)$ and sU_1, respectively, while

$$c_{as}^2 = \frac{\mathrm{Cov}(R_1(D_1-), sU_1)}{E[R_1(D_1-)]E[sU_1]} \ ; \tag{19}$$

see Fendick and Whitt (1989). (The squared coefficients of variation of sU_1 and U_1 coincide. Thus (18) is exact when U_1 is independent of D_1 and U_1 has an exponential distribution.) Alternatively, we can use refinements; e.g., see Whitt (1989), Fendick and Whitt (1989) and references cited there.

5. The Linear Fluid Model with Random Disruptions

If, in addition to (14), we assume that

$$R_k(t) = rt \ , \ t \geq 0 \ , \tag{20}$$

with $s > 0$ and $r > 0$, then we obtain the linear fluid model with random disruptions considered by Gaver and Miller (1962), Miller (1963) and Chen and Yao (1990). For this special case, they established Theorem 4.1 and obtained an expression for the Laplace transform for the steady-state content Z. However, from the analysis above, we obtain a much more elementary expression for the distribution of Z. We apply the following elementary lemma. (We omit the proof.)

Lemma 5.1. Under (20),

$$R_1(D_1^*) = rD_1^* \stackrel{d}{=} (rD_1)^* = R_1(D_1-)^* \ .$$

Theorem 5.1. Suppose that (14) and (20) hold in addition to the assumptions of Theorem 4.2.

Then $(Z_u | Z_u > 0)$ and $(Z | Z > 0)$ are both distributed the same as $(V | V > 0)$ where V is the steady-state workload in a G/G/1 queue with (possibly dependent) service times rD_k and interarrival times sU_k; i.e.,

$$P(Z_u > z | Z_u > 0) = P(Z > z | Z > 0) = P(Z_e + rD_1^* > z) \ .$$

where D_1^* is independent of Z_e. Moreover,

$$P(Z_u > 0) = P(V > 0) = \rho = rd/su \ .$$

Corollary. Under the assumptions of Theorem 5.1,

$$P(Z > z) = \left[\frac{u}{u+d} \right] \left[\frac{s+r}{r} \right] P(V > z)$$

$$= \left[\frac{u}{u+d} \right] \left[\frac{s+r}{r} \right] \rho P(Z_e + rD_1^* > z) \ , \ z \geq 0 \ ,$$

where V is the steady-state workload in the G/G/1 queue with (possibly dependent) service times rD_k and interarrival times sU_k.

Using (18) with the Corollary to Theorem 5.1, we obtain the approximation (which is exact when U and D are independent and U has an exponential distribution)

$$E[Z] \approx \left[\frac{u}{u+d} \right] \left[\frac{s+r}{r} \right] \left[\frac{\rho^2 (c_U^2 + c_D^2 - 2c_{UD}^2)}{2(1-\rho)} + \frac{\rho (c_D^2 + 1)}{2} \right] rd \ , \tag{21}$$

where c_D^2 and c_U^2 are the squared coefficients of variation of D_1 and U_1; and $c_{UD}^2 = \mathrm{cov}(U, D)/E[U]E[D]$; see (1)–(5) of Whitt 1989). From $\rho = rd/su$, we can rewrite (21) as

$$E[Z] \approx \left[\frac{a+\rho}{a+1} \right] \left[\frac{\rho^2 (c_U^2 + c_D^2 - 2c_{UD}^2)}{2(1-\rho)} + \frac{\rho (c_D^2 + 1)}{2} \right] su \tag{22}$$

where $a = d/u$. For d/u and r/s fixed, a and ρ are fixed, so that $E[Z]$ is approximately directly proportional to s and u.

6. Another Special Case: Down Subintervals

Another special case of Section 4 occurs when the down time D_k is composed of random n subintervals, i.e., $D_k = D_{k1} + \ldots + D_{kn}$ with $E[D_{ki}] = d_i$, and $\{R_k(t) : t \geq 0\}$ can be represented as $\{R_{ki}(t) : t \geq 0\}$ during down subinterval i. In addition, suppose that $R_{ki}(t) = r_i t$, $t \geq 0$, where $r_i > 0$ for each i and (14) holds. Moreover, suppose that $D_{11}, \ldots, D_{1n}$ are independent with non-lattice distributions. Then we can focus on the processes $\{Z_{di}(t) : t \geq 0\}$ restricted to the i^{th} down subinterval. Instead of Theorem 4.2(a), we have

$$Z_{di}(t) \Rightarrow Z_{di} \stackrel{d}{=} Z_e + \sum_{j=1}^{i-1} r_j D_{1j} + r_i D_{1i}^* \; , \; 1 \leq i \leq n \; , \tag{23}$$

where $Z_e, D_{11}, \ldots, D_{1,i-1}$ and D_{1i}^* are mutually independent. Moreover,

$$R_1(D_1 -) = r_1 D_{11} + \ldots + r_n D_{1n} \; , \tag{24}$$

so that $R_1(D_1 -)^*$ is determined by (15) and (24). Finally, $P(Z > z)$ is given by (16) with

$$P(Z_d > z) = \sum_{i=1}^{n} \left[\frac{d_i}{d} \right] P(Z_e + \sum_{j=1}^{i-1} r_j D_{1j} + r_i D_{1i}^* > z) \; , \quad z \geq 0 \; , \tag{25}$$

and $P(Z_u > z)$ as in Theorem 4.4, using (24).

Remark (6.1) The analysis is much more difficult when the up interval is divided into subintervals.

Acknowledgment. We thank Michael Rothkopf for bringing his paper and the papers by Gaver and Miller (1962) and Miller (1963) to our attention.

References

Asmussen, S. 1987. *Applied Probability and Queues.* John Wiley and Sons, New York.

Chen, H. and A. Mandelbaum. 1991. Discrete flow networks: bottleneck analysis and fluid approximations. *Math. Opns. Res.,* to appear.

Chen, H. and D. D. Yao. 1990. A fluid model for systems with random disruptions. *Opns. Res.,* to appear.

Doshi, B. T. 1986. Queueing systems with vacations – a survey. *Queueing Systems, Theory and Applications* 1, 29-66.

Doshi, B. T. 1990. Single server queues with vacations. In *Stochastic Analysis of Computer and Communications Systems* (H. Takagi ed.) North-Holland, Amsterdam, 217-265.

Fendick, K. W. and W. Whitt. 1989. Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue. *Proceedings IEEE* 77, 171-194.

Franken, P., D. König, U. Arndt and V. Schmidt. 1981. *Queues and Point Processes,* Akademie-Verlag, Berlin.

Gaver, D. P., Jr. and R. G. Miller, Jr. 1962. Limiting distributions for some storage problems. In *Studies in Applied Probability and Management Science* (K. J. Arrow, S. Karlin and H. Scarf eds.), Stanford University Press, 110-126.

Green, L. 1982. A limit theorem on subintervals of interrenewal times. *Opns. Res.* 30, 210-216.

Harrison, J. M. 1985. *Brownian Motion and Stochastic Flow Systems,* John Wiley and Sons, New York.

Keilson, J. and S. S. Rao. 1970. A process with chain dependent growth rate. *J. Appl. Prob.* 7, 699-711.

Kella, O. and W. Whitt. 1992. A tandem fluid network with Lévy input. *Queues and Related Models,* (I. Basawa and U. Bhat eds.), North-Holland, to appear.

Meyer, R. R., M. H. Rothkopf and S. A. Smith. 1979. Reliability and inventory in a production-storage system. *Management Sci.* 25, 799-807.

Meyer, R. R., M. H. Rothkopf and S. A. Smith. 1983. Erratum to ‘‘Reliability and Inventory in a production-storage system.’’ *Management Sci.* 29, 1346.

Miller, R. G., Jr. 1963. Continuous time stochastic storage processes with random linear inputs and outputs. *J. Math. Mech.* 12, 275-291.

Mitra, D. 1988. Stochastic theory of a fluid model of multiple failure-susceptible producers and consumers coupled by a buffer. *Adv. Appl. Prob.* 20, 646-676.

Newell, G. F. 1982. *Applications of Queueing Theory,* Chapman and Hall, London.

Whitt, W. 1989. An interpolation approximation for the mean workload in a GI/G/1 queue. *Opns. Res.* 37, 936-952.

Wolff, R. W. 1988. Sample-path derivations of the excess, age and spread distributions. *J. Appl. Prob.* 25, 432-436.