SIAM J. APPL. MATH.
Vol. 55, No. 3, pp. 784–799, June 1995

TIME-DEPENDENT BEHAVIOR OF FLUID BUFFER MODELS WITH MARKOV INPUT AND CONSTANT OUTPUT RATES*

THOMAS M. CHEN† AND VIJAY K. SAMALAM†

Abstract. The authors study a method for the analysis of fluid buffer models with Markov input rate and constant output rate. In telecommunications applications, the constant output rate typically represents a fixed-rate transmission channel. Given the initial system state, the conditional (time-dependent) survivor function and moments of the buffer content are characterized as the solution of an integral equation. The method depends on the solution of a first passage time problem for the cumulative input process (an integrated Markov process) which is described by a Fredholm integral equation of the second kind. Under appropriate conditions, the integral equation has a Neumann series solution that can be found by successive approximations. Numerical results are presented for a specific example of a buffer shared by a number of input sources varying as independent Ornstein–Uhlenbeck diffusion processes, which may be used, for instance, as a model for statistically multiplexed variable bit-rate video.

Key words. fluid buffer models, first passage time, integral equation of the second kind, Neumann series, Ornstein–Uhlenbeck diffusion

AMS subject classifications. 60K30, 60J60

1. Introduction. Fluid buffer models are natural for problems involving continuous flow, e.g., the control of dams (see Zuckerman [27]), the virtual waiting time in the G/G/1 queue (Prabhu [20]), or some stochastic problems arising in economics (Harrison [8]). In addition, fluid models are often useful as approximate models for certain queueing and inventory systems where the flow consists of discrete entities, but the behavior of individuals is not important to identify for performance analysis. For instance, rush-hour traffic congestion has been approximated by a deterministic fluid model (Newell [18]). A diffusion approximation for the GI/GI/1 queue can be made under heavy traffic conditions based on a Central Limit Theorem argument (see Gelenbe [5] and Kobayashi [13]). Another example is a queue with time-varying arrival rate that may temporarily exceed the service rate; fluid approximations are capable of capturing the queue fluctuations occurring during those overload intervals (see Norros et al. [19]).

In telecommunications applications, fluid buffer models are usually assumed to have a constant output rate which typically represents a fixed-rate transmission channel. An example is the stochastic fluid model with birth–death input rate process and constant output rate proposed by Anick et al. [1] for statistically multiplexed on–off data sources. The model has also been successfully applied to the steady-state analysis of statistically multiplexed packet voice (see Daigle and Langford [3]) and packet video (Maglaris et al. [16]). A related fluid model, where the input rate is an Ornstein–Uhlenbeck (O–U) diffusion process, has been considered for packet video (Simonian [24]).

Although fluid buffer models are often simple to describe, their analysis may be unexpectedly difficult. Most studies of fluid buffer models have been limited to the steady-state distribution of the buffer content, e.g., [1]. In this paper, we consider general fluid buffer models with Markov input and constant output rates which includes the models of [1], [3], [16], and [24]. Markov processes are a broad class of

* Received by the editors November 9, 1992; accepted for publication (in revised form) May 6, 1994.

† GTE Laboratories Inc., 40 Sylvan Road, Waltham, Massachusetts 02254.

stochastic processes useful for modeling many physical phenomena, and they often allow a tractable analysis. Instead of the usual approach involving partial differential equations, our approach depends on a sample path mapping and the solution of a first-passage time problem described by a Fredholm integral equation of the second kind. Under appropriate conditions, the integral equation has a series solution that can be calculated by successive approximations. The method is capable of yielding the conditional (time-dependent) survivor function and moments of the buffer content, as well as the steady-state distribution. The time-dependent behavior is important for some control applications and studies of the rate of convergence to a steady state. In comparison with the approach of partial differential equations, the method is not necessarily more tractable for explicit results, but integral equations are often more amenable for numerical solutions.

Numerical results are given for a specific example of a statistical multiplexer with input sources varying as independent O–U diffusion processes. This particular case arises, for instance, in the characterization of low-motion variable bit-rate video [16], [24]; the heavy-traffic limit of the stochastic fluid model when the number of sources is large (see Knessl and Morrison [12]); and the approximation of large video conference networks (Ustunel and Choukri [26]).

2. Model description. We consider a fluid buffer model with stochastic input rate X_t and constant output rate R. As mentioned, this situation is often encountered in telecommunications where the constant output rate represents a fixed-rate transmission channel. It is assumed that X_t is a Markov process such that the integrated process $\int_0^t X_\tau\, d\tau$, the cumulative input, has continuous sample paths (for instance, X_t may be piecewise-continuous). When the buffer is not empty, the rate of change in the buffer content or queue length, Q_t, depends on the difference $X_t - R$; and when the buffer is empty, it remains empty until X_t exceeds R. More specifically, Q_t evolves according to

$$\frac{d}{dt}Q_t = \begin{cases} X_t - R, & \text{if } Q_t > 0 \text{ or } X_t \geq R \text{ or both,} \\ 0, & \text{if } Q_t = 0 \text{ and } X_t \leq R. \end{cases} \tag{1}$$

It should be recognized that Q_t is a deterministic function of the sample path of X_t, but it is useful to consider (X_t, Q_t) as a bivariate Markov process representing the state of the system. Given the initial state $(X_0, Q_0) = (a, b)$, $b \geq 0$, our objective is the joint conditional survivor function

$$h(t, x, q|a, b)\, dx = Pr\{x < X_t < x + dx, Q_t > q|a, b\}. \tag{2}$$

From $h(t, x, q|a, b)$ we will obtain the conditional survivor function of Q_t,

$$H(t, q|a, b) = \int_{-\infty}^{\infty} h(t, x, q|a, b)\, dx = Pr\{Q_t > q|a, b\}, \tag{3}$$

and the conditional moments $E(Q_t^k|a, b)$.

3. Relation between Q_t and Y_t. In many cases, the usual approach of partial differential equations for the bivariate process (X_t, Q_t) is tractable only for the steady-state distribution; cf. [1], [16], and [24]. Instead, we shall introduce the integrated process

$$Y_t \equiv b + \int_0^t (X_\tau - R)\, d\tau \tag{4}$$

which has the same behavior as Q_t described in (1) except ignoring the zero boundary, and then relate Y_t to the actual Q_t. Note that Y_t has continuous sample paths by previous assumption. The transition density function $p(t,x,y|a,b)$ of the bivariate Markov process (X_t, Y_t) satisfies the same partial differential equations as the transition density of (X_t, Q_t) except freed from the complicated boundary condition at $Q_t = 0$ and thus may be simpler to solve. For instance, if X_t is a diffusion process with infinitesimal mean $\mu(x)$ and infinitesimal variance $\sigma^2(x)$, then $p(t,x,y|a,b)$ requires the solution of the backward equation

$$\text{(5)} \qquad \frac{\partial}{\partial t}p = \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}p + \mu(x)\frac{\partial}{\partial x}p + (x-R)\frac{\partial}{\partial y}p, \qquad -\infty < y < \infty$$

or the corresponding forward equation. This backward equation for (X_t, Y_t) is very similar to the backward equation for X_t alone except for the presence of the last term, and the usual solution methods can be attempted. (For examples of diffusion processes with solvable backward equations, see [10, Ch. 15].) The following approach can be applicable if $p(t,x,y|a,b)$ is known explicitly or computable.

Let the conditional probability density function of Y_t be denoted as $r(t,y|a,b) \equiv \int_{-\infty}^{\infty} p(t,x,y|a,b)\,dx$. Our approach is to obtain the survivor function of Q_t in terms of $p(t,x,y|a,b)$ and $r(t,y|a,b)$, which is possible by means of a direct mapping between the sample paths of Y_t and Q_t. Define the minimum of Y_t within the time interval $[0,t]$ as

$$\text{(6)} \qquad L_t \equiv \inf_{0\le\tau\le t} Y_\tau.$$

Then one can verify the sample path relation:

$$\text{(7)} \qquad Q_t = Y_t - \min(0, L_t)$$

(see Harrison [8]). On the right-hand side, Y_t represents the cumulative input minus the cumulative potential output, and the second term may be interpreted as the amount of potential output lost due to the buffer being empty.

It can be inferred from (4) and (7) that Q_t will remain bounded in the limit $t \to \infty$ if and only if

$$\text{(8)} \qquad \lim_{t\to\infty} \int_0^t (X_\tau - R)\,d\tau < \infty.$$

Assuming that X_t is ergodic in the sense that its time-average converges to its steady-state mean, one can substitute $(E(X_t) - R)t$ in the limit in (8); then it follows that

$$\text{(9)} \qquad \rho \equiv \frac{E(X_\infty)}{R} < 1$$

is a sufficient condition for the system to be stable, where ρ is the utilization factor. Not surprisingly, (9) means that the system is stable if the long-term average input rate is less than the output rate R.

From the relation (7), we can rewrite (2) as

$$
\begin{aligned}
&h(t,x,q|a,b)\,dx \\
&\quad= \int_{y=-\infty}^{\infty} Pr\{x < X_t < x+dx, y < Y_t < y+dy, \min(0,L_t) < y-q|a,b\} \\
(10)\qquad &\quad= dx\int_q^{\infty} p(t,x,y|a,b)\,dy \\
&\qquad+ \int_{y=-\infty}^{q} Pr\{x < X_t < x+dx, y < Y_t < y+dy, L_t < y-q|a,b\}.
\end{aligned}
$$

The unknown quantity is the joint probability distribution of (X,Y,L) in the last integral. Since the minimum of a random process is directly related to its first passage times (or hitting times), we have now essentially a first passage time problem for the process Y_t.

Define the first passage time $T_y \equiv \inf\{t > 0 : Y_t = y < b\}$ and the joint probability density function,

$$
(11)\qquad \phi(t,x,y|a,b)\,dt\,dx \equiv Pr\{t < T_y < t+dt, x < X_t < x+dx|a,b\}.
$$

That is, (11) is the probability that Y_t hits the level y from above for the first time during $(t,t+dt)$ with the slope $(x-R)$. From the Markov property of the bivariate process (X_t,Y_t), we can establish for $z \le b$

$$
\begin{aligned}
&Pr\{x < X_t < x+dx, y < Y_t < y+dy, L_t < z|a,b\} \\
(12)\qquad &\quad= dx\,dy\int_{-\infty}^{R}\int_0^t p(t-\tau,x,y|\xi,z)\phi(\tau,\xi,z|a,b)\,d\tau\,d\xi.
\end{aligned}
$$

The event $\{L_t < z\}$ in the left-hand side of (12) implies that Y_t must hit the level z with negative slope at some time $\tau < t$ because Y_t is continuous. The right-hand side of (12) is the total probability of all such sample paths of (X_t,Y_t). (For related results, see Goldman [6] and McKean [17].)

After substituting (12) into (10) and then observing that $p(t,x,y|a,b) = p(t,x,y+c|a,b+c)$, we obtain an expression for $h(t,x,q|a,b)$ in terms of the first passage time probability density function $\phi(t,x,y|a,b)$.

LEMMA 3.1. *For* $t > 0$, $-\infty < x < \infty$, $q \ge 0$,

$$
\begin{aligned}
&h(t,x,q|a,b) = \int_q^{\infty} p(t,x,y|a,b)\,dy \\
(13)\qquad &\qquad+ \int_{-\infty}^{R}\int_0^t p(t-\tau,x,q|\xi,0)f(\tau,\xi|a,b)\,d\tau\,d\xi
\end{aligned}
$$

where we have defined

$$
(14)\qquad f(t,x|a,b) \equiv \int_{-\infty}^{0} \phi(t,x,y|a,b)\,dy.
$$

Ignoring x on both sides of (13), the lemma expresses the probability of the event $\{Q_t > q\}$ as the sum of probabilities of the events $\{Y_t > q\}$ and $\{Y_t < q$ and $L_t < Y_t - q\}$. The last term in (13) recognizes that $Y_t = y < q$ and $L_t < y - q$ implies Y_τ

$(0 \le \tau \le t)$ must have crossed the level $(y-q)$ before time t; it sums the probability of all such sample paths identified by their first level crossing times.

We offer an interpretation of $f(t,x|a,b)$. For $x \le R$, (14) is by definition

$$f(t,x|a,b)\,dx\,dt = \int_{y=-\infty}^{0} Pr\{t < T_y < t+dt, x < X_t < x+dx|a,b\}\,dy. \tag{15}$$

It is relevant to apply a geometric argument made by Rice [22] and observed later by Siegert [23] and Goldman [6]. For all $t < T_y (y \le b)$, note that $Y_{t+dt} = Y_t + (X_t - R)dt > y$, and the event of Y_t crossing the level y from above with slope $(x-R)$ for the first time during $(t, t+dt)$ is equivalent to the event $\{Y_t = L_t \text{ and } y < Y_t < y + (R-x)dt\}$. This implies the substitution $dy = (R-x)dt$ in (15), which leads to

$$\begin{aligned} &f(t,x|a,b)\,dx\,dt \\ &\quad= \int_{y=-\infty}^{0} (R-x) Pr\{x < X_t < x+dx, y < Y_t = L_t < y+dy|a,b\}\,dt. \end{aligned} \tag{16}$$

Finally, after integration over y and recalling (7), one can make the interpretation

$$\begin{aligned} f(t,x|a,b)\,dx &= (R-x)Pr\{x < X_t < x+dx, Y_t = L_t \le 0|a,b\} \\ &= (R-x)Pr\{x < X_t < x+dx, Q_t = 0|a,b\}. \end{aligned} \tag{17}$$

Through Lemma 3.1 we have translated the original problem into a first passage time problem by expressing $h(t,x,q|a,b)$ in terms of the unknown function $f(t,x|a,b)$. Now it remains to solve for $f(t,x|a,b)$.

4. Fredholm integral equation for $f(t,x|a,b)$. The usual approach for finding first passage time probabilities via partial differential equations can be difficult, even numerically, due to complicated boundary conditions (e.g., see Hagan [7]). We follow a different approach here based on the fact that the probability density function $\phi(t,x,y|a,b)$ also satisfies an integral equation.

LEMMA 4.1. *For $x \le R$, $y \le b$, $t > 0$, the first passage time density $\phi(t,x,y|a,b)$ satisfies the following integral equation:*

$$\begin{aligned} \phi(t,x,y|a,b) &= (R-x)p(t,x,y|a,b) \\ &\quad - \int_{-\infty}^{R}\int_0^t (R-x)p(t-\tau,x,0|\xi,0)\phi(\tau,\xi,y|a,b)\,d\tau\,d\xi. \end{aligned} \tag{18}$$

Proof. Define

$$\Psi(t,x,y|a,b)\,dx\,dy \equiv Pr\{x < X_t < x+dx, y < Y_t = L_t < y+dy|a,b\}. \tag{19}$$

Recall the previous discussion concerning the interpretation of $f(t,x|a,b)$ which established that

$$(R-x)\Psi(t,x,y|a,b)\,dt\,dx = \phi(t,x,y|a,b)\,dt\,dx \tag{20}$$

for $y \le b$, $x \le R$. Equation (12) may be rewritten as

$$\begin{aligned} Pr\{x < X_t < x+dx, y < Y_t < y+dy, L_t \ge z|a,b\} &= p(t,x,y|a,b)\,dx\,dy \\ &\quad - dx\,dy\int_{-\infty}^{R}\int_0^t (R-\xi)p(t-\tau,x,y|\xi,z)\Psi(\tau,\xi,z|a,b)\,d\tau\,d\xi. \end{aligned} \tag{21}$$

Multiplying by $(R - x)$ and taking the limit as $z \to y^-$ yields the result, noting that $y \leq L_t \leq Y_t < y + dy$ implies $L_t = Y_t$. Related results are in [6], [23]. □

The right-hand side of (18) may be interpreted as the probability that Y_t crosses the level y from above with slope $(x - R)$ during $(t, t + dt)$ minus the probability that Y_t has also crossed y at least once before time t. The resulting difference is therefore the probability that Y_t crosses y during $(t, t + dt)$ for the first time.

Equation (18) may be recognized as a Fredholm integral equation of the second kind (with two independent variables, t and x). The second kind of integral equation is characterized by the presence of the unknown function $\phi(t, x, y|a, b)$ both inside and outside of the integral operator. Fredholm integral equations have fixed limits in the integral term (strictly speaking, equation (18) is Volterra in the t variable, but Volterra equations are generally regarded as special cases of Fredholm equations). In the integral equation, we may identify the "kernel"

$$K(t, x, \xi) \equiv (R - x)p(t, x, 0|\xi, 0) \tag{22}$$

and the "forcing function," $(R - x)p(t, x, y|a, b)$.

Because (18) is valid for every $y \leq b$, we can integrate over $-\infty < y < 0$ to derive an integral equation for $f(t, x|a, b)$.

PROPOSITION 4.2. *For $x \leq R$, $t > 0$, the function $f(t, x|a, b)$ satisfies the integral equation*

$$f(t, x|a, b) = F(t, x|a, b) - \int_{-\infty}^{R} \int_0^t K(t - \tau, x, \xi) f(\tau, \xi|a, b)\, d\tau\, d\xi, \tag{23}$$

where the forcing function is

$$F(t, x|a, b) \equiv \int_{-\infty}^{0} (R - x)p(t, x, y|a, b)\, dy. \tag{24}$$

This is also a Fredholm integral equation of the second kind (with two independent variables).

Interpretation. Ignoring x on both sides, the left-hand side of (23) can be interpreted as the probability of the event $Q_t = 0$, or equivalently, the event $\{Y_t < 0$ and $Y_t = L_t\}$. The right-hand side involves the probability of the event $\{Y_t < 0\}$ minus the event $\{Y_t < 0$ and $Y_t \neq L_t\}$. The integral term on the right hand side arises from the fact that $Y_t \neq L_t$ implies that $Y_t = L_\tau = Y_\tau$ for some $\tau < t$.

5. Neumann series solution and successive approximations. A series solution usually exists for integral equations of the second kind (see Korn [14, p. 499]). If (23) is repeatedly substituted into itself, it will yield an infinite series

$$\begin{aligned} f(t, x|a, b) = F(t, x|a, b) &- \int_{-\infty}^{R} \int_0^t K_1(t - \tau, x, \xi) F(\tau, \xi|a, b)\, d\tau\, d\xi \\ &+ \int_{-\infty}^{R} \int_0^t K_2(t - \tau, x, \xi) F(\tau, \xi|a, b)\, d\tau\, d\xi - \cdots \end{aligned} \tag{25}$$

known as the Neumann series solution, where the so-called iterated kernels are

$$\begin{aligned} K_1(t, x, \xi) &= K(t, x, \xi), \\ K_n(t, x, \xi) &= \int_{-\infty}^{R} \int_0^t K_1(t - \tau, x, \eta) K_{n-1}(\tau, \eta, \xi)\, d\tau\, d\eta, \qquad n \geq 2. \end{aligned} \tag{26}$$

Let s_n denote the nth term of the Neumann series (25). Note that (25) is an alternating series because the integrals of probability density functions are always nonnegative. By Leibniz' test for alternating series, the Neumann series is convergent if

$$\text{(i)}\quad |s_n| \geq |s_{n+1}|, \qquad n \geq 1, \\ \text{(ii)}\quad \lim_{n\to\infty} s_n = 0. \tag{27}$$

Furthermore, by the ratio test, the Neumann series is absolutely convergent if

$$\lim_{n\to\infty} \left| \frac{s_{n+1}}{s_n} \right| < 1. \tag{28}$$

Unfortunately, the physical interpretation of the Neumann series (25) is not simple, and it is not clear how to verify the conditions for convergence. However, it is possible to make a general qualitative statement about convergence. Roughly speaking, if the process Y_t has a strong negative drift such that $Y_t = L_t$ at any time with high probability, then the kernel $K(t, x, \xi)$ in (23) will be small and the Neumann series may be expected to converge more quickly (as the integral terms in (25) will be small). In general, this will be the usual case because Y_t should have a negative drift if the buffer is stable and satisfies (9). On the other hand, if Y_t does not exhibit a significant negative drift, then the kernel in (23) may be large and the Neumann series may be slow to converge.

In practice, the solution to integral equations of the second kind can be calculated iteratively by the method of successive approximations. With the initial approximation $\hat{f}_0(t, x|a, b) = F(t, x|a, b)$, the nth successive approximation is calculated as

$$\begin{aligned} \hat{f}_n(t, x|a, b) &= F(t, x|a, b) \\ &\quad - \int_{-\infty}^{R} \int_0^t K(t-\tau, x, \xi) \hat{f}_{n-1}(\tau, \xi|a, b)\, d\tau\, d\xi, \qquad n \geq 1. \end{aligned} \tag{29}$$

One can verify that $\hat{f}_n(t, x|a, b)$ is the partial sum of the first $n+1$ terms of the Neumann series. Hence if the Neumann series is convergent, we may expect the successive approximations $\hat{f}_n(t, x|a, b)$ to approach the actual solution in the limit as n increases.

It is usually impractical to calculate the entire infinite series; therefore, we are primarily concerned with the approximation error in the nth successive approximation:

$$e_n(t, x|a, b) \equiv |f(t, x|a, b) - \hat{f}_n(t, x|a, b)|. \tag{30}$$

PROPOSITION 5.1. *An upper bound for the approximation error is*

$$e_n(t, x|a, b) \leq |s_{n+2}|. \tag{31}$$

Proof. The nth successive approximation is

$$\begin{aligned} \hat{f}_n(t, x|a, b) &= F(t, x|a, b) - \cdots \\ &\quad + (-1)^n \int_{-\infty}^{R} \int_0^t K_n(t-\tau, x, \xi) F(\tau, \xi|a, b)\, d\tau\, d\xi, \end{aligned} \tag{32}$$

whereas the exact expression, from repeated substitution of (23) into itself, differs only in the last term:

$$f(t,x|a,b) = F(t,x|a,b) - \cdots \\ + (-1)^n \int_{-\infty}^{R} \int_0^t K_n(t-\tau,x,\xi) f(\tau,\xi|a,b)\, d\tau\, d\xi. \tag{33}$$

The approximation error is the following difference:

$$\begin{aligned} e_n(t,x|a,b) &= \left| \int_{-\infty}^{R} \int_0^t K_n(t-\tau,x,\xi)[f(\tau,\xi|a,b) - F(\tau,\xi|a,b)]\, d\tau\, d\xi \right| \\ &= \left| \int_{-\infty}^{R} \int_0^t K_n(t-\tau,x,\xi) \right. \\ &\qquad \left. \cdot \int_{-\infty}^{R} \int_0^{\tau} K(\tau-\gamma,\xi,\eta) f(\gamma,\eta|a,b)\, d\gamma\, d\eta\, d\tau\, d\xi \right| \\ &= \left| \int_{-\infty}^{R} \int_0^t K_{n+1}(t-\tau,x,\xi) f(\tau,\xi|a,b)\, d\tau\, d\xi \right|. \end{aligned} \tag{34}$$

Note that $f(t,x|a,b) \leq F(t,x|a,b)$, and therefore

$$e_n(t,x|a,b) \leq \left| \int_{-\infty}^{R} \int_0^t K_{n+1}(t-\tau,x,\xi) F(\tau,\xi|a,b)\, d\tau\, d\xi \right|. \tag{35}$$

The right-hand side of (35) may be recognized as $|s_{n+2}|$. □

The error bound can be calculated conveniently, during the process of iterating successive approximations, as the difference between the nth and $(n+1)$th successive approximation

$$|s_{n+2}| = |\hat{f}_{n+1}(t,x|a,b) - \hat{f}_n(t,x|a,b)|. \tag{36}$$

As expected, it is apparent from (36) that the error bound (and hence the error) will vanish for large n if the Neumann series is convergent.

6. Survivor function and moments of Q_t. After obtaining $f(t,x|a,b)$, the survivor function and moments of Q_t can be found immediately by integration.

PROPOSITION 6.1. *For $t > 0$, $q \geq 0$, the conditional survivor function of Q_t is*

$$H(t,q|a,b) = \int_q^{\infty} r(t,y|a,b)\, dy + \int_{-\infty}^{R} \int_0^t r(t-\tau,q|\xi,0) f(\tau,\xi|a,b)\, d\tau\, d\xi \tag{37}$$

and the conditional kth moment of Q_t is

$$E(Q_t^k|a,b) = \int_0^{\infty} y^k r(t,y|a,b)\, dy \\ + \int_{-\infty}^{R} \int_0^{\infty} kq^{k-1} \int_0^t r(t-\tau,q|\xi,0) f(\tau,\xi|a,b)\, d\tau\, dq\, d\xi. \tag{38}$$

If the approximation $\hat{f}_n(t,x|a,b)$ is used instead to calculate an approximation $\hat{H}_n(t,q|a,b)$ for $H(t,q|a,b)$, then the approximation error can be shown to be

$$|H(t,q|a,b) - \hat{H}_n(t,q|a,b)| = \int_{-\infty}^{R} \int_0^t r(t-\tau,q|\xi,0) e_n(\tau,\xi|a,b)\, d\tau\, d\xi, \tag{39}$$

which is bounded by

$$|H(t,q|a,b) - \hat{H}_n(t,q|a,b)| \le |\hat{H}_{n+1}(t,q|a,b) - \hat{H}_n(t,q|a,b)|. \tag{40}$$

Similarly, let $\hat{E}_n(Q_t^k|a,b)$ denote the approximate conditional moment of Q_t calculated by using $\hat{f}_n(t,x|a,b)$ instead of $f(t,x|a,b)$. The approximation error is bounded by

$$\begin{aligned}
&|E(Q_t^k|a,b) - \hat{E}_n(Q_t^k|a,b)| \\
&\quad = \int_{-\infty}^{R}\int_0^{\infty} kq^{k-1}\int_0^t r(t-\tau,q|\xi,0)e_n(\tau,\xi|a,b)\,d\tau\,dq\,d\xi \\
&\quad \le |\hat{E}_{n+1}(Q_t^k|a,b) - \hat{E}_n(Q_t^k|a,b)|.
\end{aligned} \tag{41}$$

The method is capable of yielding the steady-state distribution of Q_t if the steady-state exists in the limit $t \to \infty$. There is a simple way to calculate the steady-state survivor function and moments of Q_t by taking advantage of the Final Value Theorem of Laplace transforms [15], assuming that the appropriate Laplace transforms exist. Let the Laplace transform of a function with respect to the time variable t be denoted by a * superscript and complex variable s; for example, $f^*(s,x|a,b) \equiv \int_0^\infty e^{-st}f(t,x|a,b)\,dt$.

Taking Laplace transforms in (23) yields the next corollary.

COROLLARY 6.2. *For $x \le R$, the Laplace transform $f^*(s,x|a,b)$ satisfies the integral equation*

$$f^*(s,x|a,b) = F^*(s,x|a,b) - \int_{-\infty}^{R} K^*(s,x,\xi)f^*(s,\xi|a,b)\,d\xi, \tag{42}$$

where the forcing function is

$$F^*(s,x|a,b) \equiv \int_{-\infty}^{0}(R-x)p^*(s,x,y|a,b)\,dy \tag{43}$$

and the kernel is

$$K^*(s,x,\xi) = (R-x)p^*(s,x,0|\xi,0). \tag{44}$$

Again, this is a Fredholm integral equation of the second kind (with one independent variable now). Note that the Laplace transformation has reduced the integral term from second order to first order.

The Neumann series solution is

$$\begin{aligned}
f^*(s,x|a,b) = F^*(s,x|a,b) &- \int_{-\infty}^{R} K_1^*(s,x,\xi)F^*(s,\xi|a,b)\,d\xi \\
&+ \int_{-\infty}^{R} K_2^*(s,x,\xi)F^*(s,\xi|a,b)\,d\xi - \cdots,
\end{aligned} \tag{45}$$

where the iterated kernels are

$$\begin{aligned}
&K_1^*(s,x,\xi) = K^*(s,x,\xi), \\
&K_n^*(s,x,\xi) = \int_{-\infty}^{R} K_1^*(s,x,\eta)K_{n-1}^*(s,\eta,\xi)\,d\eta, \qquad n \ge 2.
\end{aligned} \tag{46}$$

As before, the partial sums of the Neumann series can be calculated by the method of successive approximations.

After $f^*(s,x|a,b)$ is obtained, the steady-state survivor function and moments of Q_t follow directly from Laplace transforms of (37) and (38).

COROLLARY 6.3. *The steady-state survivor function of Q_t is*

$$Pr\{Q_\infty > q\} = \lim_{s\to 0} s \int_{-\infty}^{R} r^*(s,q|\xi,0) f^*(s,\xi|a,b)\, d\xi \tag{47}$$

and the kth moment of Q_t is

$$E(Q_\infty^k) = \lim_{s\to 0} s \int_{-\infty}^{R} \int_{0}^{\infty} kq^{k-1} r^*(s,q|\xi,0) f^*(s,\xi|a,b)\; dq\, d\xi \tag{48}$$

if these limits exist.

Proof. The results follow from taking Laplace transforms of (37) and (38), and applying the Final Value Theorem. The first terms on the right-hand sides of (37) and (38) vanish because Y_t should drift to $-\infty$ as $t \to \infty$ with probability one if the buffer is stable. □

In practice, of course, it would be sufficient to evaluate (47) and (48) for a very small value of s. Although the steady-state distribution of Q_t may be found by this method, there are simpler approaches if only steady-state results are desired (cf. [25]).

7. Multiplexed O–U source rates. In this section, we consider a particular example of a fluid buffer shared by N independent and identically distributed input sources. This model may represent the output buffer of a packet switch [11]. The constant output rate represents the fixed-rate transmission channel, and the inputs may be on–off data sources or variable bit-rate video sources. The time-dependent behavior of the queue is of interest in the situation where the source rates are controlled by feedback from the switch based on the level of queue occupancy [9].

In particular, each source rate $X_t^{(n)}$ is assumed to vary as an O–U diffusion process with infinitesimal mean, $\mu_n(x) = \beta - \alpha x$, and infinitesimal variance, $\sigma_n^2(x) = \sigma^2$, where α, β, and σ are positive parameters. That is, $X_t^{(n)}$ is the solution of the stochastic differential equation

$$dX_t^{(n)} = (\beta - \alpha X_t^{(n)})dt + \sigma dW_t^{(n)}, \tag{49}$$

where the $W_t^{(n)}$ $(n = 1, ..., N)$ are independent standard Brownian motions. This model may represent statistically multiplexed variable bit-rate video, for example, where each O–U process is a continuous-time approximation for a first-order autoregressive video source [16], [24].

It can be shown that the total input rate $X_t = X_t^{(1)} + \cdots + X_t^{(N)}$ is then an O–U process as well with infinitesimal mean $\mu(x) = N\beta - \alpha x$ and infinitesimal variance $\sigma^2(x) = N\sigma^2$. From (49), we have

$$\begin{aligned} dX_t &= dX_t^{(1)} + \cdots + dX_t^{(N)} \\ &= \left[(\beta - \alpha X_t^{(1)}) + \cdots + (\beta - \alpha X_t^{(N)})\right]dt + \sigma\left[dW_t^{(1)} + \cdots + dW_t^{(N)}\right] \\ &= (N\beta - \alpha X_t)dt + \sqrt{N}\sigma dW_t, \end{aligned} \tag{50}$$

where W_t is a standard Brownian motion. Thus we consider the buffer with input rate X_t as equivalent to the fluid buffer with N independent O–U sources. From (9), the system is stable if $\rho = N\beta/\alpha R < 1$.

It is known that X_t is a Gaussian Markov process and (X_t, Y_t) is a jointly Gaussian Markov process (see Arnold [2]). The conditional mean and variance of X_t are

$$\begin{aligned} m_x &= E(X_t|a) = (a - \rho R)e^{-\alpha t} + \rho R, \\ v_x &= \mathrm{var}(X_t|a) = \frac{N\sigma^2}{2\alpha}(1 - e^{-2\alpha t}), \end{aligned} \tag{51}$$

and those of Y_t are

$$\begin{aligned} m_y &= E(Y_t|a,b) = b + \alpha^{-1}(a - \rho R)(1 - e^{-\alpha t}) + R(\rho - 1)t, \\ v_y &= var(Y_t|a,b) = \frac{N\sigma^2}{2\alpha^3}[(3 - e^{-\alpha t})(e^{-\alpha t} - 1) + 2\alpha t]. \end{aligned} \tag{52}$$

The conditional probability density function of Y_t is

$$r(t, y|a, b) = (2\pi v_y)^{-1/2} \exp\{-(y - m_y)^2/2v_y\}. \tag{53}$$

The conditional joint density function of (X_t, Y_t) is

$$p(t, x, y|a, b) = r(t, y|a, b)(\pi u)^{-1/2} \exp\left\{ - \left[x - m_x - \frac{w}{v_y}(y - m_y) \right]^2 /u \right\}, \tag{54}$$

where

$$w \equiv \frac{N\sigma^2}{2\alpha^2}(1 - e^{-\alpha t})^2; \quad u \equiv 2\left(v_x - \frac{w^2}{v_y} \right). \tag{55}$$

Not surprisingly, the time constant of m_y and v_y depends on the parameter α, which determines the autocorrelation of the Y_t process. It can be seen that the mean of Y_t (normalized as before by dividing by R) is unaffected by N but the standard deviation of Y_t (and hence the coefficient of variation of Y_t) decreases by $N^{-1/2}$. This is, of course, an example of the Central Limit Theorem for the sum of independent Gaussian processes. Thus, a larger value of N reduces the random fluctuations in Y_t, which in turn reduces Q_t (recall the relation between Y_t and Q_t); this is the reason for statistical multiplexing with large values of N.

We must solve an integral equation for the first passage time density of a bivariate Gaussian Markov process (cf. [4] for an integral equation for the one-dimensional case). We used the method of successive approximations (29) to solve for $f(t, x|a, b)$. In this case, the forcing function in (23) is

$$F(t, x|a, b) = \frac{R - x}{\sqrt{2\pi v_x}} \exp\{-(x - m_x)^2/2v_x\} \Phi\left(- \frac{v_x m_y + w(x - m_x)}{\sqrt{v_x(v_y - w^2)}} \right), \tag{56}$$

where $\Phi(x)$ is the standard normal distribution function, and the kernel in (23) is

$$K(t, x, \xi) = (R - x)p(t, x, 0|\xi, 0). \tag{57}$$

In the Neumann series (25), the integral terms were too complicated for explicit expressions, so the successive approximations were computed using numerical integration (see Press et al. [21, Ch. 4]). Fortunately, serious numerical difficulties were not encountered because all functions were smooth. Over the range of parameters studied, we found that the initial approximation $f(t,x|a,b) \approx \hat{f}_0(t,x|a,b) = F(t,x|a,b)$ was reasonably close, i.e., the first term of the Neumann series (25) was dominant. Typically, only about two successive approximations were necessary to obtain an accuracy of several digits. An example of the convergence of the successive approximations $\hat{f}_n(t,x|a,b)$ is shown in Fig. 1. The reason for the fast convergence of the Neumann series is because Y_t demonstrates a strong negative drift even for a utilization factor up to 0.8.

After calculating $f(t,x|a,b)$ by successive approximations, the survivor function and moments of Q_t were found by numerically integrating (37) and (38). In the following results, we chose $\alpha = 3.90$, $\beta = 2.03$, and $\sigma^2 = 0.42$ to match the statistics of the autoregressive video source model measured in [16]. The initial conditions were $X_0 = E(X_\infty)$ and $Q_0 = 0$; naturally, the rate of convergence to steady state depends on the proximity of the initial state (X_0, Q_0) to the expected steady state. Figure 2

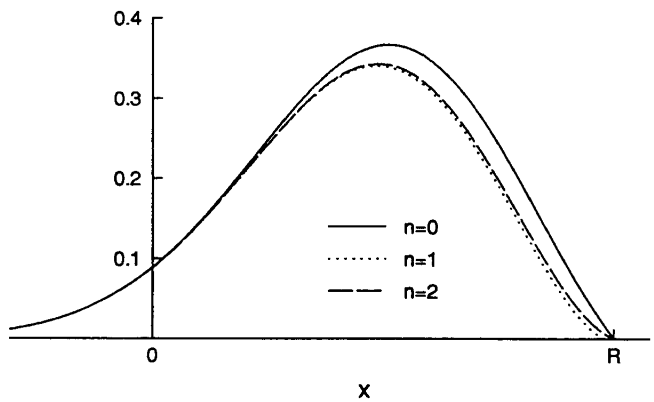

FIG. 1. *Successive approximations $\hat{f}_n(t,x|a,b)$ for $t = 1$ sec, $a = \rho R = E(X_\infty)$, $b = 0$, $N = 1$, and $\rho = 0.8$.*

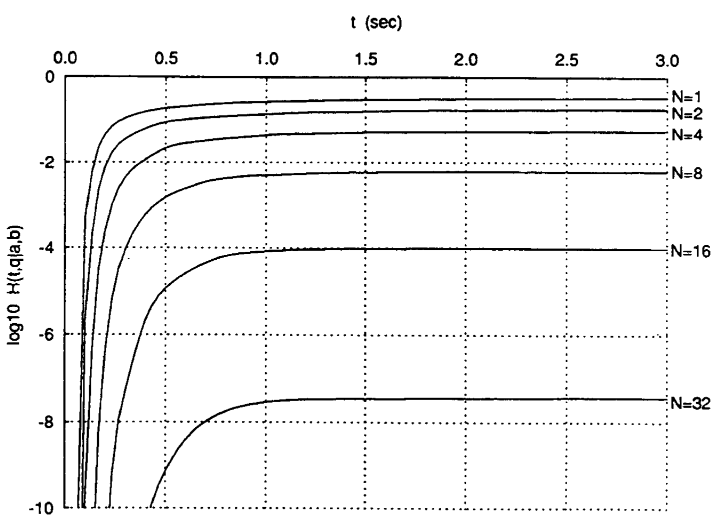

FIG. 2. *Conditional survivor function of Q_t as function of t, for $q = 32$ msec, $a = \rho R = E(X_\infty)$, $b = 0$, and $\rho = 0.8$.*

shows the survivor function $H(t, q|a, b)$ as a function of t for q =32 ms and $\rho = 0.8$. In Figs. 3–4, the conditional mean and variance of Q_t are plotted as functions of t for $\rho = 0.8$. Figure 5 shows the survivor function $H(t, q|a, b)$ as a function of q for $\rho = 0.8$, N =1, 2, 4. The accuracy of all results shown were verified by computer simulations of the fluid buffer.

From these results, we may observe that the system seems to approach steady state fairly quickly from the given initial conditions, on the order of seconds, with a rate strongly dependent on α. As mentioned earlier, the dependence on α is not surprising because α determines the autocorrelation of Y_t. The rate of convergence to steady-state appears weakly dependent, if at all, on N. As expected, the survivor function and moments of Q_t do depend greatly on N. The results in Figs. 2–4 suggest

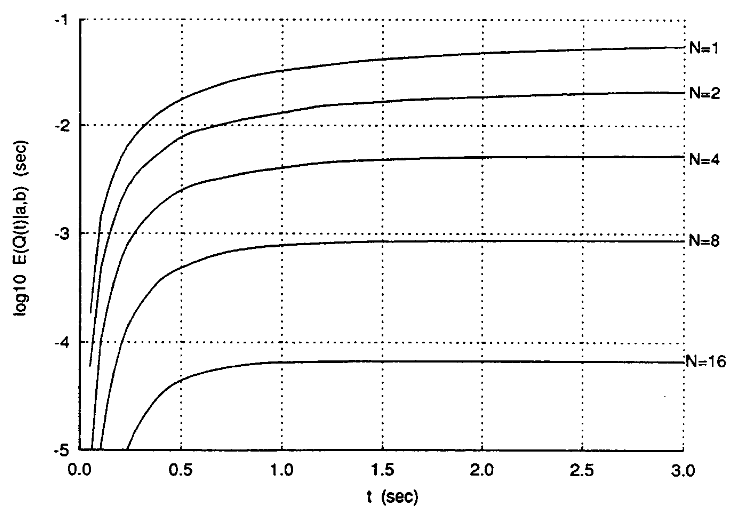

FIG. 3. *Conditional mean of Q_t as function of t, for q = 32 msec, $a = \rho R = E(X_\infty)$, $b = 0$, and $\rho = 0.8$.*

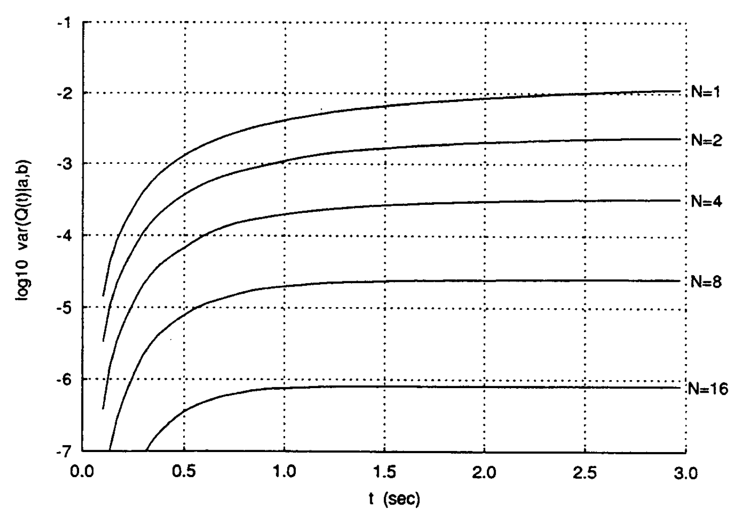

FIG. 4. *Conditional variance of Q_t as function of t, for q = 32 msec, $a = \rho R = E(X_\infty)$, $b = 0$, and $\rho = 0.8$.*

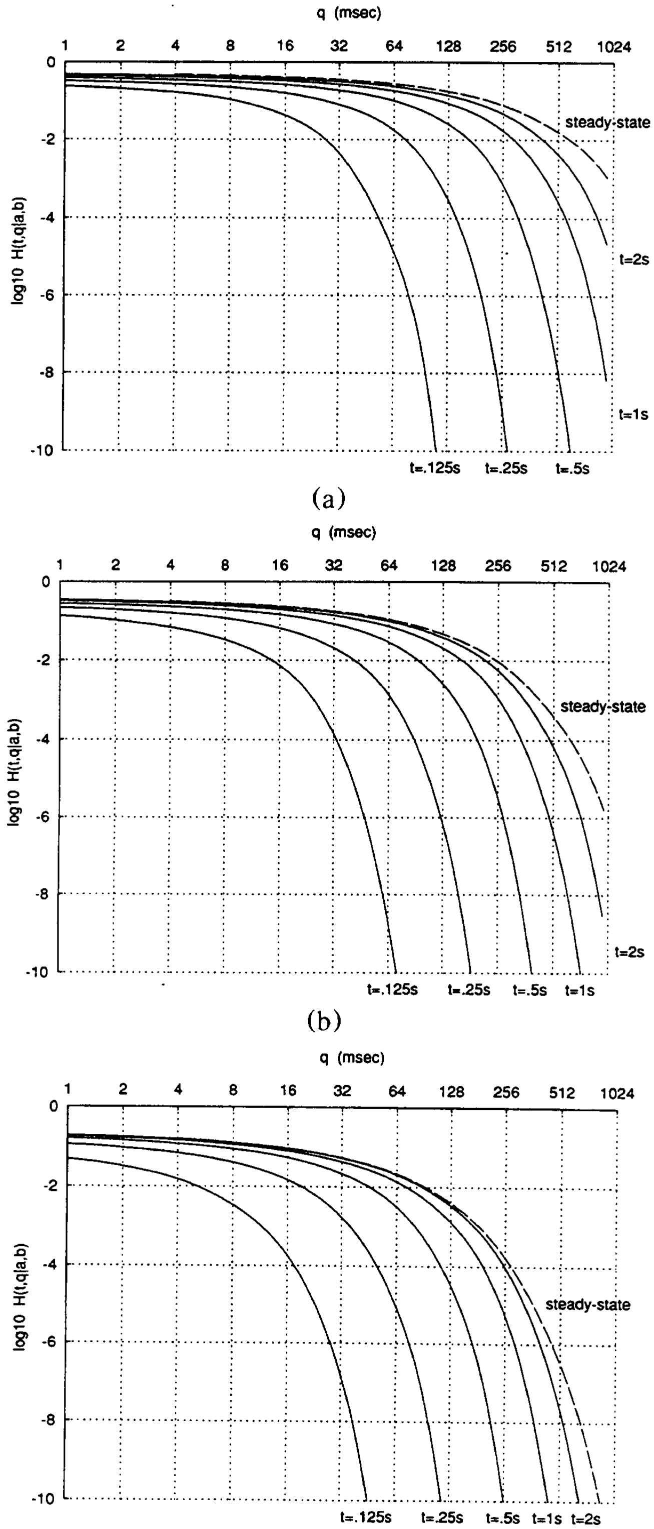

FIG. 5. *Conditional survivor function of* Q_t *as function of* q, *for* $a = \rho R = E(X_\infty)$, $b = 0$, $\rho = 0.8$, *and* (a) $N = 1$, (b) $N = 2$, *and* (c) $N = 4$.

that the logarithm of the survivor function and moments of Q_t depend roughly linearly with N.

We also computed the steady-state survivor function of Q_t from (47) with $s = 10^{-3}$. These steady-state results were entirely consistent with previous results from [16].

8. Conclusions. For the class of fluid buffer models with Markov input and constant output rates, we have studied an approach for obtaining the conditional distribution of buffer contents by solving a first passage time problem described by a Fredholm integral equation of the second kind. In comparison with the usual approach via partial differential equations, the method is not necessarily advantageous for analytical results, but integral equations are often easier for numerical solutions. The time-dependent behavior of the buffer is important for some control problems and studies of the rate of convergence to steady-state. The method is also capable of yielding the steady-state distribution but is not the simplest approach if only steady-state results are desired.

For the particular example of a statistical multiplexer shared by independent O–U input sources, we were able to numerically compute not only the time-dependent but also the steady-state distribution and moments of buffer contents.

We believe that this approach can be extended in future work to obtain explicit solutions for certain cases; to analyze fluid models with finite buffer spaces; and to obtain first time-to-buffer-overflow probabilities.

Acknowledgments. The authors are indebted to the anonymous referees for their thorough reviews and constructive comments. In particular one reviewer provided a careful exposition of the relation between the integral equations in this paper and the usual partial differential equations, and derived a proof of Proposition 1 through integration of parts of the partial differential equations.

REFERENCES

[1] D. Anick, D. Mitra, and M. Sondhi, *Stochastic theory of a data-handling system with multiple sources*, Bell Sys. Tech. J., 61 (1982), pp. 1871–1894.

[2] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, John Wiley, New York, 1974.

[3] J. Daigle and J. Langford, *Models for analysis of packet voice communications systems*, IEEE J. on Sel. Areas in Commun., SAC-4 (1986), pp. 847–855.

[4] J. Durbin, *The first-passage density of a continuous Gaussian process to a generalized boundary*, J. Appl. Probab., 22 (1985), pp. 99–122.

[5] E. Gelenbe, *On approximate computer system models*, J. Assoc. Comput. Mach., 22 (1975), pp. 261–269.

[6] M. Goldman, *On the first passage of the integrated Wiener process*, Ann. Math. Stat., 42 (1971), pp. 2150–2155.

[7] P. Hagan, C. Doering, and C. Levermore, *The distribution of exit times for weakly colored noise*, J. Statist. Physics, 54 (1989), pp. 1321–1352.

[8] J. Harrison, *Brownian Motion and Stochastic Flow Systems*, John Wiley, New York, 1985.

[9] H. Kanakia, P. Misra, and A. Reibman, *An adaptive congestion control scheme for real-time packet video transport*, Computer Commun. Rev., 23 (1993), pp. 20–31.

[10] S. Karlin and H. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.

[11] M. Karol, M. Hluchyj, and S. Morgan, *Input versus output queueing on a space-division packet switch*, IEEE Trans. Comm., COM-35, (1987), pp. 1347–1356.

[12] C. Knessl and J. Morrison, *Heavy-traffic analysis of a data-handling system with many sources*, SIAM J. Appl. Math., 51 (1991), pp. 187–213.

[13] H. Kobayashi, *Application of the diffusion approximation to queueing networks*, J. Assoc. Comput. Mach., 21 (1974), pp. 316–328.

[14] G. KORN AND T. KORN, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York, 1968.

[15] W. LEPAGE, *Complex Variables and the Laplace Transform for Engineers*, Dover, New York, 1961.

[16] B. MAGLARIS, S. ANASTASSIOU, P. SEN, G. KARLSSON, AND J. ROBBINS, *Performance models of statistical multiplexing in packet video communications*, IEEE Trans. Comm., 36 (1988), pp. 834–844.

[17] H. MCKEAN, JR., *A winding problem for a resonator driven by a white noise*, J. Math. Kyoto U., 2 (1963), pp. 227–235.

[18] G. NEWELL, *Applications of Queueing Theory*, Chapman and Hall, London, 1982.

[19] I. NORROS, J. ROBERTS, A. SIMONIAN, AND J. VIRTAMO, *The superposition of variable bit rate sources in an* ATM *multiplexer*, IEEE J. on Sel. Areas in Commun., 9 (1991), pp. 378–387.

[20] N. PRABHU, *Stochastic Storage Processes*, Springer-Verlag, New York, 1980.

[21] W. PRESS, B. FLANNERY, S. TEUKOLSKY, AND W. VETTERLING, *Numerical Recipes in C: The Art of Computing*, Cambridge Univ. Press, Cambridge, UK, 1988.

[22] S. RICE, *Mathematical analysis of random noise - part* 3, Bell Sys. Tech. J., 24 (1945), pp. 46–108.

[23] A. SIEGERT, *On the first passage time probability problem*, Phys. Rev., 81 (1951), pp. 617–623.

[24] A. SIMONIAN, *Stationary analysis of a fluid queue with input rate varying as an Ornstein-Uhlenbeck process*, SIAM J. Appl. Math., 51 (1991), pp. 828–842.

[25] A. SIMONIAN AND J. VIRTAMO, *Transient and stationary distributions for fluid queues and input processes with a density*, SIAM J. Appl. Math., 51 (1991), pp. 1732–1739.

[26] A. USTUNEL AND T. CHOUKRI, *Diffusion approximation of videoconference networks*, Queueing Systems, 6 (1990), pp. 405–424.

[27] D. ZUCKERMAN, *Two-stage output procedure of a finite dam*, J. Appl. Probab., 14 (1977), pp. 421–425.