

## THE TAIL BEHAVIOR OF THE STATIONARY DISTRIBUTION OF A FLUID QUEUE WITH A GAUSSIAN-TYPE INPUT RATE PROCESS

Kazutomo Kobayashi  
NEC Corporation

Yukio Takahashi  
Tokyo Institute of Technology

(Received August 4, 1995; Revised October 18, 1995)

*Abstract* This paper deals with a fluid queue with a *Gaussian-type input rate process*. The Gaussian-type processes are ones defined as  $R_t = m + \int_{-\infty}^t h(t-s)dw_s$ , where  $m$  is a positive constant,  $w_t$  is a standard Wiener process and  $h(t)$  is an integrable function such that  $h(t)^2$  and  $H(t) = \int_t^{\infty} h(s)ds$  are also integrable. The class of Gaussian-type processes is wide enough to contain most of continuous time stochastic processes proposed so far for coded video traffic.

For the model, in this paper, the exponential decay property of the tail of the buffer content distribution is studied, and an upper bound and a lower one are given for the tail probability  $P(Q_{\infty} > x)$  of the buffer content distribution in the steady state. These bounds show that the tail probability decays exponentially with rate  $-\frac{C-m}{H(0)^2/2}$ , where  $H(0) = \int_0^{\infty} h(t)dt$  and  $C$  is the output rate of the fluid queue. This result guarantees, in a sense, the plausibility of the approximation formula  $P(Q_{\infty} > x) \approx B \exp\{-\frac{C-m}{H(0)^2/2}x\}$  proposed in the previous paper [*Performance Evaluation*, 1995].

### 1. Introduction

Asynchronous Transfer Mode (ATM) is a main technology for multiplexing and switching a variety of information such as voice, data, image and video, and the ATM technology has been adopted to High-speed Local Area Networks and to Broad-band ISDNs. Related to this movement, there have been a number of studies on performance analyses for ATM multiplexers.

In an ATM network, information is divided into blocks of small fixed size, referred to as cells, and transmitted. For the reason, the cell arrival process at a switching node or a multiplexer has strong correlation in inter-arrival times. This strong correlation makes the stochastic structure complicated and the performance analysis difficult.

The most popular approximation which makes the stochastic structure simpler is a *fluid queue model* in which the detailed behavior of cells is ignored and cell streams are regarded as fluid streams (Fig. 1). In the fluid queue model, the stochastic behavior of the *buffer content*  $Q_t$  is represented by the differential equation

$$\frac{dQ_t}{dt} = \begin{cases} R_t - C, & \text{if } Q_t > 0 \text{ or } R_t > C, \\ 0, & \text{otherwise,} \end{cases}$$

where  $C$  is a constant *output rate* and  $R_t$  is the *total input rate* at time  $t$ . Many fluid queue models studied so far assume On-Off types of traffic, in which sources have two states, On-state and Off-state, to alternate each other and each source pours fluid into the buffer at a constant rate during On-state [6, 1, 9]. However, coded video traffic, which is considered to be the main traffic in ATM networks, differs largely from On-Off types of traffic.

In some simulation studies, coded video traffic is modeled as a discrete time Gaussian process [10, 13, 7, 4]. Maglaris, et al. discuss a model in which coded video traffic is represented by a first-order autoregressive process [10]. With Maglaris' study as a motivation, Simonian analyzes a fluid queue model with an Ornstein-Uhlenbeck process, and obtains exact and approximate solutions [13]. Heyman, et al. discuss a model with a higher-order autoregressive process as a coded video input rate process [7], and Grünenfelder, et al. propose a model with an autoregressive moving average process [4]. These processes have an advantage that the parameters of them can be easily estimated from the empirical mean and the empirical autocovariance function. However, they have been used only for simulations, and according to the authors' knowledge, they have never been used for any theoretical analyses of fluid queue models.

There are studies on analyses of fluid queue models with Markovian-type input rate processes. Stern and Elwalid extend the On-Off source model to the one with a Markov modulated rate process (MMRP) as input rate [15]. Tanaka, et al. study the same model and obtain the transient state probabilities in an explicit form using eigenvalues and eigenvectors of a key matrix [16]. However, it is not easy to calculate the steady-state distribution from the results of these studies, except for simple fluid queue models. It is also difficult to estimate the parameters of the MMRP from actual measurement data.

In this paper, we discuss a fluid queue model with a *Gaussian-type input rate process* defined as  $R_t = m + \int_{-\infty}^t h(t-s)dw_s$ . The detailed definition is given in Section 3. The class of Gaussian-type processes is wide enough to contain most of continuous time stochastic processes proposed so far for coded video traffic such as continuous versions of autoregressive processes and autoregressive moving average processes.

For the model with a Gaussian-type input rate process, in the previous paper [8], the authors proposed an approximation formula  $P[Q_\infty > x] \approx B \exp[-\frac{C-m}{H(0)^2/2}x]$  for the tail probability of the buffer content distribution in the steady-state, where  $B$  is a positive constant and  $H(0) = \int_0^\infty h(t)dt$ . However, there was not given any theoretical proof for the exponential decay. The purpose of this paper is to justify the exponential decay by giving an upper bound and a lower bound which have the same exponential decay rate<sup>1</sup>. Specifically we prove that for sufficiently large  $x$ ,

$$\frac{1}{B_0 + B_1\sqrt{x}} \exp\left\{-\frac{C-m}{H(0)^2/2}x\right\} < P(Q_\infty > x) < (B_2 + B_3x) \exp\left\{-\frac{C-m}{H(0)^2/2}x\right\},$$

where  $B_0$ ,  $B_1$  and  $B_2$  are positive constants and  $B_3$  is a non-negative constant.

The paper is organized as follows. In Section 2, we discuss a fluid queue model for an ATM statistical multiplexer with coded video input, and we introduce *Gaussian-type input rate processes* in Section 3. Section 4 is devoted to our main theorem which gives an upper bound and a lower bound for the tail probability of the buffer content distribution.

## 2. Fluid Queue Model for ATM Statistical Multiplexer

We consider a fluid queue model for an ATM statistical multiplexer which accommodates multiple sources and serving with a single output line (see Fig. 1). If there are  $L$  sources and source  $i$  pours fluid into a buffer with time dependent rate  $R_t^{(i)}$ , then the *total input rate*  $R_t$  is given by  $R_t = \sum_{i=1}^L R_t^{(i)}$ . In our fluid queue model for the multiplexer, the buffer

<sup>1</sup>We note that K. Debicki and T. Rolski [3] obtain a similar upper bound for the same fluid queue model, independently

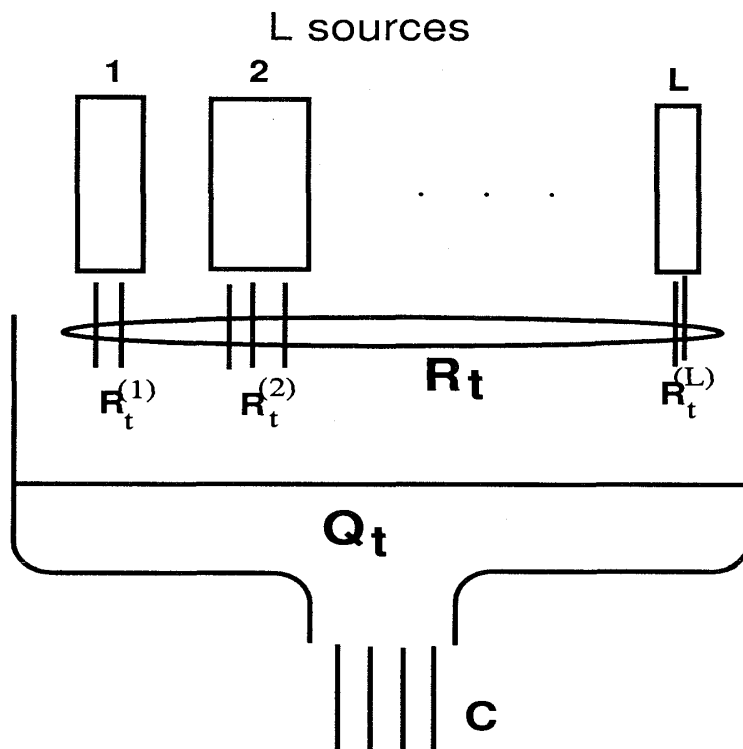


Figure 1: Fluid queue model for ATM statistical multiplexer

content  $Q_t$  varies according to the differential equation

$$\frac{dQ_t}{dt} = \begin{cases} R_t - C, & \text{if } Q_t > 0 \text{ or } R_t > C, \\ 0, & \text{otherwise,} \end{cases} \quad (2.1)$$

where  $C$  is constant *output rate*.

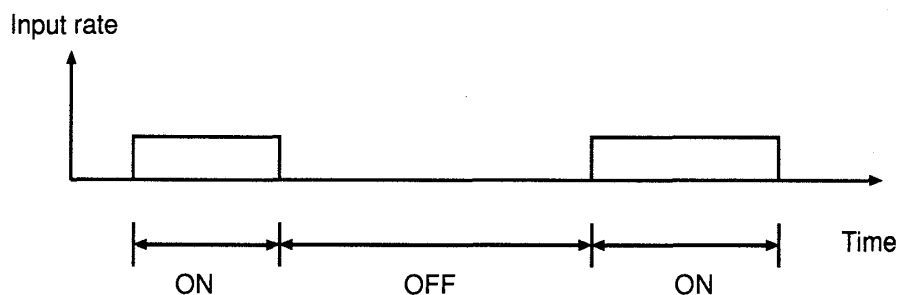


Figure 2: A On-Off fluid model

Many fluid queue models [6, 1, 9] assume On-Off type of traffic, in which each source has two states, On-state and Off-state, to alternate each other and pours fluid into the buffer at a constant rate during On-state (see Fig. 2). Traffic from voice sources or file-transfer sources might be modeled by On-Off type processes. However, coded video traffic, which is considered to be the main traffic in ATM networks, seems difficult to be modeled as an On-Off type process. Figure 3 shows an example of actual coded video traffic pattern. The vertical axis represents the input rate and the horizontal axis the time. It is much different from On-Off type of traffic.

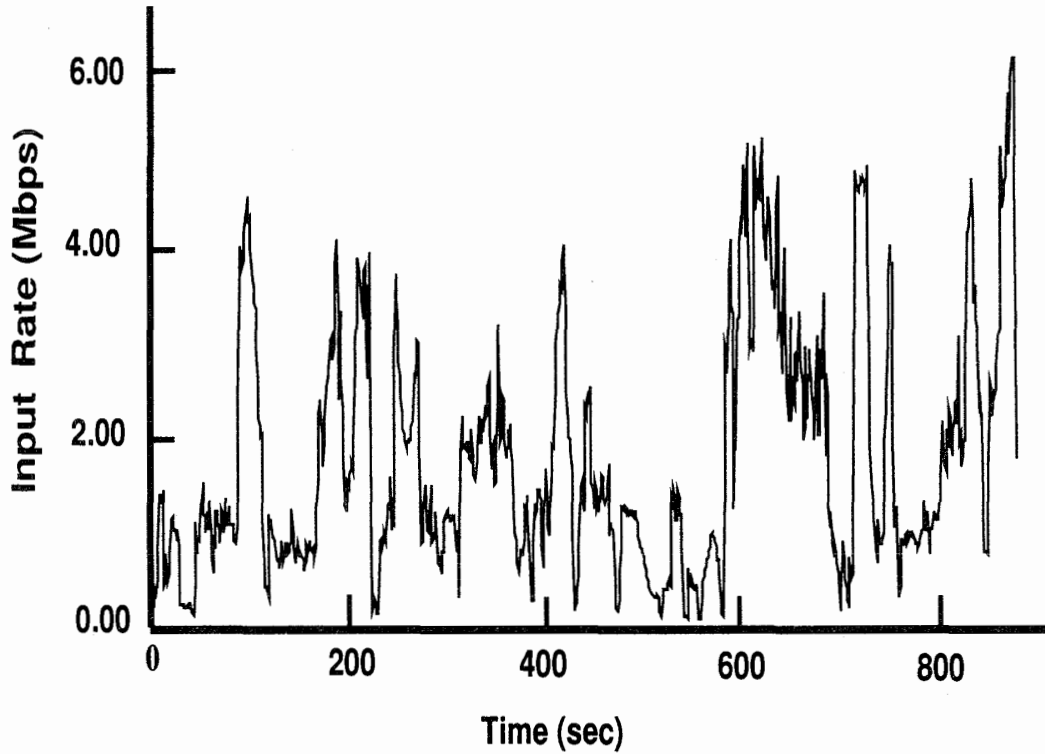


Figure 3: Coded Video Traffic

In [10, 7], video traffic is modeled as an autoregressive process

$$X_n = a_0 + \sum_{k=1}^K a_k X_{n-k} + \varepsilon_n,$$

where  $X_n$  is the input rate at discrete time  $n$ ,  $a_k$ 's are real constants and  $\varepsilon_n$ 's are independently and identically distributed random variables with zero mean. The random variables  $\varepsilon_n$ 's are usually assumed to be normally distributed. Then, the sequence  $\{\varepsilon_n\}$  is considered as a discrete version of the *white noise*. The parameters  $a_k, k = 1, \dots, K$ , are determined from the sample autocovariance function  $\Lambda(k)$  of actual measurement traffic data by using the *Yule-Walker* equation:

$$\begin{bmatrix} \Lambda(0) & \Lambda(1) & \dots & \Lambda(K-1) \\ \Lambda(1) & \Lambda(0) & \dots & \Lambda(K-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda(K-1) & \Lambda(K-2) & \dots & \Lambda(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} = \begin{bmatrix} \Lambda(1) \\ \Lambda(2) \\ \vdots \\ \Lambda(K) \end{bmatrix}.$$

The parameter  $a_0$  is given by  $m(1 - \sum_{k=1}^K a_k)$  where  $m$  is the sample mean.

In [4], coded video traffic is modeled as an autoregressive moving average process,

$$X_n = a_0 + \sum_{k=1}^K a_k X_{n-k} + \sum_{i=0}^M b_i \varepsilon_{n-i}.$$

The parameters  $a_k$ 's and  $b_i$ 's are also estimated from the sample mean and the sample autocovariance function.

In these processes,  $X_n$  can take negative values. So it is suggested in [4] that the input rate is given by  $\max(0, X_n)$ . However, if many video sources are multiplexed, then the standard deviation of  $X_n$  becomes relatively small compared with the mean of  $X_n$ , and the probability that  $X_n$  is negative becomes negligibly small. In such cases,  $X_n$  itself can be used as an input rate process.

These processes look complicated, but they can be rewritten in the same form

$$X_n = m + \sum_{i=-\infty}^n h_{n-i} \varepsilon_i. \quad (2.2)$$

In [11], the sequence  $\{h_n\}$  for the autoregressive moving average process is given by the inverse z-transform of

$$\mathcal{H}(z) = \frac{\sum_{i=0}^M b_i z^{-i}}{1 - \sum_{k=1}^K a_k z^{-k}}.$$

Gaussian-type Input Rate Processes which will be discussed in the next section are continuous versions of processes of type (2.2).

### 3. Gaussian-type Input Rate Processes

We discuss an input rate process  $R_t$  defined by

$$R_t = m + \int_{-\infty}^t h(t-s) dw_s, \quad (3.1)$$

where  $m$  is a positive constant,  $w_t$  is a standard Wiener process, and  $h(t)$  is a real valued measurable function such that  $h(t) = 0$  for  $t < 0$ . We assume that

$$\int_0^{\infty} |h(t)| dt < \infty, \quad (3.2)$$

$$\int_0^{\infty} h(t)^2 dt < \infty, \quad (3.3)$$

$$\int_0^{\infty} |H(t)| dt < \infty, \quad (3.4)$$

where

$$H(t) = \int_t^{\infty} h(s) ds. \quad (3.5)$$

Since  $|H(t)| \leq \int_0^{\infty} |h(t)| dt < \infty$ , it follows that

$$\int_0^{\infty} H(t)^2 dt \leq \int_0^{\infty} |h(t)| dt \int_0^{\infty} |H(t)| dt < \infty. \quad (3.6)$$

In order to avoid trivial cases, we also assume that  $m$  is less than  $C$  and  $h(t)$  is not a function taking zero almost everywhere, so that the integrals in the inequalities (3.2) to (3.4) are strictly positive.

The process  $R_t$  is the same as the one defined in our previous paper [8]. Here, let us call it a *Gaussian-type input rate process*. Clearly,  $R_t$  is a stationary Gaussian process with

- the mean  $m$ ,
- the autocovariance function  $\Lambda_R(\tau) = \int_0^{\infty} h(s+\tau)h(s)ds$ , and
- the variance  $\gamma \equiv \Lambda_R(0) = \int_0^{\infty} h(t)^2 dt$ .

The second term in the right-hand side of (3.1) is considered as the output of a linear system with a white noise input  $dw_t/dt$  and an impulse response  $h(t)$  [12]. Assumption (3.2) is the stability condition for the system, and assumption (3.3) is another essential condition in order that the integration in (3.1) exists. These assumptions are unavoidable conditions to discuss or to construct a linear system with a white noise input. Here, we assume (3.4), too. This assumption is made to ensure for the buffer content distribution in the fluid queue model to have an exponential decay property as will be seen later. However, the assumption is not strict for practical applications since most of important stable linear systems satisfy it (see Remark 1 below).

**Remark 1.** In [8], the authors gave some examples of  $h(t)$  satisfying (3.2), (3.3) and (3.4). The following is a summary of those examples.

- **Ornstein-Uhlenbeck process**

$$h(t) = \sigma e^{-\kappa t},$$

where  $\sigma$  is a real constant and  $\kappa$  is a positive constant.

- **Superposition of Ornstein-Uhlenbeck processes**

$$h(t) = \sum_{i=1}^N \sigma_i e^{-\kappa_i t},$$

where  $N$  is a positive integer,  $\sigma_i$ 's are real constants and  $\kappa_i$ 's are positive constants.

- **Continuous version of moving average process**

$$|h(t)| < M, \quad 0 \leq t \leq T, \quad \text{and} \quad h(t) = 0, \quad t > T, \quad (3.7)$$

where  $M$  and  $T$  are positive constants.

- **Continuous version of autoregressive process**

$$h(t) = \sum_{i=1}^N \sigma_i t^{n_i} e^{-\kappa_i t} \cos(\omega_i t + \varphi_i), \quad t > 0, \quad (3.8)$$

where  $N$  is a positive integer,  $n_i$ 's are non-negative integers,  $\kappa_i$ 's are positive constants, and  $\sigma_i$ 's,  $\omega_i$ 's and  $\varphi_i$ 's are real constants.

- **Continuous version of autoregressive moving average process**

$$h(t) = \int_0^t g(t-s)f(s)ds, \quad (3.9)$$

where  $g(t)$  is a function of type (3.7) and  $f(t)$  is a function of type (3.8).

**Remark 2.** In actual ATM networks, many cell streams are multiplexed, namely, many input rate processes are superposed. Individual input rate processes may not be Gaussian-type, but the total input rate process can be expected to be close to the Gaussian-type one from the central-limit theorem. Therefore, Gaussian-type processes are thought to be widely applicable as input rate processes.

**Remark 3.** The mean  $m$  and the autocovariance function  $\Lambda_R(\tau)$  of an input rate process can be estimated by the sample mean and the sample autocovariance function of actual measurement data  $R_{i\Delta t}$ ,  $i = 1, 2, \dots, N$ , as follows:

$$m = \frac{1}{N} \sum_{i=1}^N R_{i\Delta t},$$

$$\Lambda_R(k\Delta t) = \frac{1}{N} \sum_{i=1}^{N-k} (R_{i\Delta t} - m)(R_{(i+k)\Delta t} - m).$$

Later we will see that the quantity  $H(0)^2/2$  is an important factor of decay rate of the tail probability  $P(Q_\infty > x)$ . Since

$$\begin{aligned} H(0)^2/2 &= \frac{1}{2} \int_0^\infty \int_0^\infty h(t)h(s)dt ds \\ &= \int_0^\infty \int_0^\infty h(t+s)h(s)ds dt \\ &= \int_0^\infty \Lambda_R(t)dt, \end{aligned} \quad (3.10)$$

it is easily estimated by

$$H(0)^2/2 = \sum_{k=1}^N \Lambda_R(k\Delta t)\Delta t.$$

In a fluid queue model with  $L$  independent input rate processes, the mean or the autocovariance function of the total input rate process are given by the sum of those of individual input rate processes. So the decay rate can be directly estimated from the empirical means and the empirical autocovariance functions of individual sources.

#### 4. Bounds for the Tail Probability $P(Q_\infty > x)$

In the previous paper [8], the authors proposed an approximation formula for the stationary tail probability as follows:

$$P(Q_\infty > x) \approx \frac{m - \eta}{C - \eta} \exp\left[-\frac{C - m}{H(0)^2/2}x\right], \quad (4.1)$$

where  $Q_\infty$  is a random variable subjecting to the steady-state distribution of the buffer content and  $\eta$  is the conditional expectation of  $R_t$  in the steady-state conditioned that  $R_t < C$ . In this section, we prove the following theorem which justifies, in a sense, the rate of the exponential decay in the approximation formula (4.1).

Let  $A_t$  be the *input process* defined as

$$A_t = \int_0^t R_s ds. \quad (4.2)$$

**Theorem 1.** *There are positive constants  $B_0$ ,  $B_1$  and  $B_2$  and a non-negative constant  $B_3$  such that for sufficiently large  $x$ ,*

$$\frac{1}{B_0 + B_1\sqrt{x}} \exp\left[-\frac{C - m}{H(0)^2/2}x\right] < P(Q_\infty > x) < (B_2 + B_3x) \exp\left[-\frac{C - m}{H(0)^2/2}x\right], \quad (4.3)$$

where  $H(0) = \int_0^\infty h(t)dt$ . If the covariance function  $\beta_t$  between  $A_t - Ct$  and  $R_0$  is non-negative, then the term  $B_3x$  can be omitted.

We shall prove the upper bound and the lower bound separately.

##### 4.1. Proof of the upper bound

To get the upper bound in (4.3), we employ the following result used in Simonian and Virtamo's paper [14].

*The stationary distribution of the buffer content satisfies*

$$P(Q_\infty > x) \leq \int_0^{+\infty} dt \int_{-\infty}^C (C - y)p_t(x, y)dy, \quad (4.4)$$

where  $p_t(x, y)$  is the joint density function of  $A_t - Ct$  and  $R_0$ .

In [14] (pp.1735 (2.10)), instead of  $p_t(x, y)dy$ , the expression  $p_t(x|y)d\Phi(y)$  is used in (4.4), where  $p_t(x|y) = \frac{d}{dx}P(A_t - Ct \leq x | R_0 = y)$  and  $\Phi(y) = P(R_0 \leq y)$ . Simonian and Virtamo apply (4.4) to the case of input rate varying as an Ornstein-Uhlenbeck process, giving  $h(t) = \sigma e^{-\kappa t}$  in our formulation, and obtain an upper bound in an explicit form.

Now let's consider  $A_t - Ct$ . From (4.2) and (3.1), we have

$$\begin{aligned} A_t - Ct &= (m - C)t + \int_0^t \int_{-\infty}^{\tau} h(\tau - s) dw_s d\tau \\ &= (m - C)t + \int_{-\infty}^0 \int_0^t h(\tau - s) d\tau dw_s + \int_0^t \int_s^t h(\tau - s) d\tau dw_s \\ &= (m - C)t + \int_{-\infty}^0 \{H(-s) - H(t - s)\} dw_s + \int_0^t \{H(0) - H(t - s)\} dw_s \\ &= -\mu t + H(0)\tilde{w}_t - \int_{-\infty}^t H(t - s) dw_s + \int_{-\infty}^0 H(-s) dw_s, \end{aligned}$$

where  $\mu = C - m$  and  $\tilde{w}_t = \int_0^t dw_s = w_t - w_0$ .

Since  $A_t - Ct$  and  $R_0$  are jointly Gaussian random variables,  $p_t(x, y)$  is written as

$$p_t(x, y) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma_t|}} \exp\left[-\frac{1}{2}(x + \mu t, y - m)\Sigma_t^{-1}(x + \mu t, y - m)^T\right], \quad (4.5)$$

where  $\Sigma_t$  is a covariance matrix of the vector random variable  $(A_t - Ct, R_0)$ . The matrix  $\Sigma_t$  is written as

$$\Sigma_t = \begin{pmatrix} \alpha_t & \beta_t \\ \beta_t & \gamma \end{pmatrix},$$

where  $\alpha_t$  is the variance of  $A_t - Ct$ ,  $\beta_t$  the covariance between  $A_t - Ct$  and  $R_0$ , and  $\gamma$  the variance of  $R_0$ . Then  $\alpha_t$ ,  $\beta_t$  and  $\gamma$  are written in terms of  $h(t)$  and  $H(t)$  as follows:

$$\alpha_t = H(0)^2 t - 2H(0) \int_0^t H(s) ds + 2 \int_0^\infty H(s)^2 ds - 2 \int_0^\infty H(t + s)H(s) ds, \quad (4.6)$$

$$\beta_t = - \int_0^\infty H(t + s)h(s) ds + \int_0^\infty H(s)h(s) ds, \quad (4.7)$$

$$\gamma = \int_0^\infty h(s)^2 ds. \quad (4.8)$$

**Lemma 1.**  $\alpha_t$ ,  $\beta_t$  and  $\gamma$  have the following properties:

1.  $\alpha_t$  and  $\beta_t$  are absolutely continuous functions of  $t$  such that  $\alpha_0 = 0$  and  $\beta_0 = 0$ .
2.  $\gamma$  and  $\alpha_t$  for  $t > 0$  are positive. Also  $\beta_t^2 < \alpha_t \gamma$  for  $t > 0$ .
3.  $\lim_{t \rightarrow \infty} \beta_t = H(0)^2/2 > 0$ .
4. Let  $\xi_t$  be a function such that  $\alpha_t = H(0)^2(t + \xi_t)$ , then  $|\xi_t|$  is bounded, and hence  $\lim_{t \rightarrow \infty} \alpha_t/t = H(0)^2$ .
5.  $\beta_{\Delta t} = \gamma \Delta t + o(\Delta t)$  for small  $\Delta t$ .
6.  $\alpha_{\Delta t} = \gamma \{\Delta t\}^2 + o(\{\Delta t\}^2)$  for small  $\Delta t$ .

*Proof.* 1. Since  $H(t)$  is absolutely continuous,  $\alpha_t$  and  $\beta_t$  are also absolutely continuous. It is obvious from (4.6) and (4.7) that  $\alpha_0 = 0$  and  $\beta_0 = 0$ .



2. Since  $\gamma$  and  $\alpha_t$  are the variances of  $R_0$  and  $A_t - Ct$ , it is obvious that  $\gamma \geq 0$  and  $\alpha_t \geq 0$ . From the assumption that  $h(t)$  is not a function taking zero almost everywhere, it is easily checked that  $\gamma > 0$  and  $\alpha_t > 0$  for  $t > 0$ . Since  $\beta_t^2/\alpha_t\gamma$  is the correlation coefficient between  $R_0$  and  $A_t - Ct$ , the inequality  $\beta_t^2 \leq \alpha_t\gamma$  holds for  $t \geq 0$ . The equality does not hold for  $t > 0$  since  $h(t)$  is not equal to zero almost everywhere.

3. Since  $\lim_{t \rightarrow \infty} H(t+s) = 0$ , the limit of (4.7) is given by

$$\begin{aligned} \lim_{t \rightarrow \infty} \beta_t &= \int_0^\infty H(s)h(s)ds \\ &= \int_0^\infty \Lambda_R(s)ds \\ &= H(0)^2/2 > 0. \end{aligned}$$

4. We evaluate  $H(0)^2\xi_t = \alpha_t - H(0)^2t$ . From (4.6), we have

$$\begin{aligned} |\alpha_t - H(0)^2t| &\leq 2|H(0)| \int_0^t |H(s)|ds + 2 \int_0^\infty H(s)^2ds + 2 \left| \int_0^\infty H(t+s)H(s)ds \right| \\ &\leq 2 \int_0^\infty |h(s)|ds \int_0^\infty |H(s)|ds + 2 \int_0^\infty H(s)^2ds \\ &\quad + 2\sqrt{\int_0^\infty H(s)^2ds \int_0^\infty h(s)^2ds}. \end{aligned}$$

From (3.2), (3.3), (3.4) and (3.6), the right-hand side of the above inequality is finite.

5. By differentiating (4.7), we have

$$\beta'_t = \int_0^\infty h(t+s)h(s)ds = \Lambda_R(t).$$

It is easily checked from (3.3) that  $\Lambda_R(t)$  is continuous at  $t = 0$ . Note that  $\beta_0 = 0$  and  $\beta'_0 = \Lambda_R(0) = \gamma$ . Using Taylor's formula, we have

$$\begin{aligned} \beta_{\Delta t} &= \beta_0 + \Delta t\beta'_0 + o(\Delta t) \\ &= \gamma\Delta t + o(\Delta t). \end{aligned}$$

6. By differentiating (4.6), we have

$$\begin{aligned} \alpha'_t &= H(0)^2 - 2H(0)H(t) + 2 \int_0^\infty h(t+s)H(s)ds \\ &= H(0)^2 - 2 \int_0^\infty H(t+s)h(s)ds \\ &= 2\beta_t. \end{aligned}$$

From  $\alpha_0 = 0$ , it follows that

$$\begin{aligned} \alpha_{\Delta t} &= 2 \int_0^{\Delta t} \beta_s ds \\ &= \gamma\{\Delta t\}^2 + o(\{\Delta t\}^2). \quad \square \end{aligned}$$

Now we return to (4.5). Rearranging the term in the brackets, we have

$$\begin{aligned} p_t(x, y) &= \frac{1}{\sqrt{(2\pi)^2(\gamma\alpha_t - \beta_t^2)}} \exp\left[-\frac{\gamma(x + \mu t)^2 - 2\beta_t(x + \mu t)(y - m) + \alpha_t(y - m)^2}{2(\gamma\alpha_t - \beta_t^2)}\right] \\ &= \frac{1}{\sqrt{(2\pi)^2(\gamma\alpha_t - \beta_t^2)}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t} - \frac{\{y - m - \frac{\beta_t}{\alpha_t}(x + \mu t)\}^2}{2(\gamma - \frac{\beta_t^2}{\alpha_t})}\right]. \end{aligned} \quad (4.9)$$

Substituting  $p_t(x, y)$  in (4.4) with (4.9), we have

$$\begin{aligned} P(Q_\infty > x) &\leq \frac{1}{2\pi} \int_0^\infty dt \cdot \frac{1}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] \\ &\quad \times \int_{-\infty}^C \frac{C - y}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}} \exp\left[-\frac{\{y - m - \frac{\beta_t}{\alpha_t}(x + \mu t)\}^2}{2(\gamma - \frac{\beta_t^2}{\alpha_t})}\right] dy. \end{aligned} \quad (4.10)$$

We let

$$\tilde{y} = \frac{y - m - \frac{\beta_t}{\alpha_t}(x + \mu t)}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}} \quad \text{and} \quad z_t = \frac{\mu - \frac{\beta_t}{\alpha_t}(x + \mu t)}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}}.$$

Then the right-hand side of (4.10) is represented as

$$\begin{aligned} &\frac{1}{2\pi} \int_0^\infty dt \cdot \frac{1}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] \sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}} \int_{-\infty}^{z_t} (z_t - \tilde{y}) \exp\left[-\frac{\tilde{y}^2}{2}\right] d\tilde{y} \\ &= \frac{1}{2\pi} \int_0^\infty dt \cdot \frac{1}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] \sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}} \left\{ z_t \int_{-\infty}^{z_t} \exp\left[-\frac{\tilde{y}^2}{2}\right] d\tilde{y} - \exp\left[-\frac{z_t^2}{2}\right] \right\} \end{aligned}$$

For  $0 < t < \infty$ , the following inequality holds:

$$\gamma - \frac{\beta_t^2}{\alpha_t} < \gamma, \quad \int_{-\infty}^{z_t} \exp\left[-\frac{\tilde{y}^2}{2}\right] d\tilde{y} < \sqrt{2\pi}, \quad \exp\left[-\frac{z_t^2}{2}\right] > 0.$$

Hence we have

$$P(Q_\infty > x) < \sqrt{\frac{\gamma}{2\pi}} \int_0^\infty \frac{z_t}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] dt. \quad (4.11)$$

**Lemma 2.** *Let  $x_0$  be an arbitrarily fixed positive number. Then, there exist a positive constant  $D_1$  and a non-negative constant  $D_2$  such that for  $x > x_0$  and  $t > 0$*

$$z_t = \frac{\mu - \frac{\beta_t}{\alpha_t}(x + \mu t)}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}} \leq D_1 + D_2 x. \quad (4.12)$$

If  $\beta_t \geq 0$  for all  $t > 0$ , then we may put  $D_2 = 0$ .

*Proof.* We write  $z_t$  as

$$z_t = f(t) + g(t)(x - x_0),$$

where

$$f(t) = \frac{\mu(1 - \frac{\beta_t}{\alpha_t}t) - \frac{\beta_t}{\alpha_t}x_0}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}}, \quad g(t) = \frac{-\frac{\beta_t}{\alpha_t}}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}}.$$

From Lemma 1,  $\gamma - \beta_t^2/\alpha_t$  is positive for  $t > 0$  and bounded from above by  $\gamma$ . We note that  $\lim_{t \rightarrow 0}(\gamma - \frac{\beta_t^2}{\alpha_t}) = 0$ ,  $\lim_{t \rightarrow 0}(1 - \frac{\beta_t}{\alpha_t}) = 0$ , and  $\lim_{t \rightarrow 0} \beta_t/\alpha_t = +\infty$ . Hence we have

$$\lim_{t \rightarrow 0} f(t) = -\infty, \quad \lim_{t \rightarrow 0} g(t) = -\infty.$$

On the other hand, it is easily checked that

$$\lim_{t \rightarrow \infty} f(t) = \frac{\mu}{2\sqrt{\gamma}}, \quad \lim_{t \rightarrow \infty} g(t) = 0,$$

since  $\lim_{t \rightarrow \infty} \beta_t/\alpha_t = 0$ ,  $\lim_{t \rightarrow \infty} \beta_t^2/\alpha_t = 0$ ,  $\lim_{t \rightarrow \infty} \beta_t = H(0)^2/2$  and  $\lim_{t \rightarrow \infty} \alpha_t/t = H(0)^2$ . Functions  $f(t)$  and  $g(t)$  are continuous and they are bounded from above. If we set  $D_2 = \sup_{t>0} g(t)$  and  $D_1 = \sup_{t>0} f(t) - D_2x_0$ , then  $D_1$  and  $D_2$  are finite and the inequality (4.12) holds.

If  $\beta_t \geq 0$  for all  $t > 0$ , then  $g(t) \leq 0$  and hence  $D_2 = 0$ .  $\square$

From the inequality (4.11) and Lemma 2, we have

$$P(Q > x) < (D_1 + D_2x) \sqrt{\frac{\gamma}{2\pi}} \int_0^\infty \frac{1}{\sqrt{\alpha_t}} \exp[-\frac{(x + \mu t)^2}{2\alpha_t}] dt. \quad (4.13)$$

To evaluate the integrand in the right-hand side of (4.13), we shall replace  $\alpha_t$  with  $H(0)^2(t + \xi_t)$ . From Lemma 1,  $\xi_t$  is a bounded function and we let  $\bar{\xi}$  is an upper bound of  $|\xi_t|$ .

**Lemma 3.** *Let  $\delta$  be an arbitrary positive number. Then, for  $x > \mu\bar{\xi} + \delta/2$  and  $t > 0$ ,*

$$\frac{1}{\sqrt{H(0)^2(t + \xi_t)}} \exp[-\frac{(x + \mu t)^2}{2H(0)^2(t + \xi_t)}] < \frac{D_3}{\sqrt{H(0)^2(t + \bar{\xi})}} \exp[-\frac{(x + \mu t - \delta)^2}{2H(0)^2(t + \bar{\xi})}], \quad (4.14)$$

where  $D_3 = H(0)/\sqrt{2\mu\delta e}$ .

*Proof.* Let  $F(t)$  be the ratio of the right and the left sides of (4.14), namely,

$$F(t) = \frac{1}{D_3} \sqrt{\frac{t + \bar{\xi}}{t + \xi_t}} \exp[-\frac{(x + \mu t)^2}{2H(0)^2(t + \xi_t)} + \frac{(x + \mu t - \delta)^2}{2H(0)^2(t + \bar{\xi})}].$$

Since  $t + \xi_t < t + \bar{\xi}$ , it is less than

$$\frac{1}{D_3} \sqrt{\frac{t + \bar{\xi}}{t + \xi_t}} \exp[-\frac{\delta(x + \mu t - \delta/2)}{H(0)^2(t + \xi_t)}].$$

The numerator in the brackets is divided as  $\delta(x + \mu t - \delta/2) = \delta(x - \mu\bar{\xi} - \delta/2) + \mu\delta(t + \bar{\xi})$ , and we have

$$F(t) < \frac{1}{D_3} \sqrt{\theta} \exp[-\frac{\mu\delta}{H(0)^2}\theta] \exp[-\frac{\delta(x - \mu\bar{\xi} - \delta/2)}{H(0)^2(t + \xi_t)}],$$

where  $\theta = (t + \bar{\xi})/(t + \xi_t)$ .

The function  $\sqrt{\theta} \exp[-\frac{\mu\delta}{H(0)^2}\theta]$  is concave on  $[0, \infty)$  and has maximum value  $D_3 = H(0)/\sqrt{2\mu\delta e}$ . Hence  $F(t) < 1$  and this proves the inequality (4.14).  $\square$

**Lemma 4.**

$$\int_0^\infty \frac{1}{\sqrt{t+\bar{\xi}}} \exp\left[-\frac{(x+\mu t-\delta)^2}{2H(0)^2(t+\bar{\xi})}\right] dt < D_4 \exp\left[-\frac{\mu}{H(0)^2/2}x\right], \quad (4.15)$$

where  $D_4 = \frac{\sqrt{2\pi}H(0)}{\sqrt{\mu}} \exp[\zeta^2(\frac{\delta}{\mu} + \bar{\xi})]$ .

*Proof.* Let  $\tilde{t} = t + \bar{\xi}$ ,  $\zeta = \sqrt{\mu}/H(0)$  and  $\tilde{x} = (x - \delta - \mu\bar{\xi})/\mu$ . Then the left-hand side of (4.15) is written as

$$e^{-\zeta^2\tilde{x}} \int_{\bar{\xi}}^\infty \frac{1}{\sqrt{\tilde{t}}} \exp\left[-\frac{\zeta^2}{2}\left(\tilde{t} + \frac{\tilde{x}^2}{\tilde{t}}\right)\right] d\tilde{t}.$$

By extending the integral region to  $(0, \infty)$  and by applying the well-known formula

$$\int_0^\infty \frac{e^{-at-b^2/t}}{\sqrt{t}} dt = \sqrt{\frac{2\pi}{a}} e^{-2b\sqrt{a}},$$

we have an upper bound

$$\frac{\sqrt{2\pi}}{\zeta} \exp[-2\zeta^2\tilde{x}] = D_4 \exp\left[-\frac{\mu}{H(0)^2/2}x\right],$$

where

$$D_4 = \frac{\sqrt{2\pi}}{\zeta} \exp[\zeta^2(\frac{\delta}{\mu} + \bar{\xi})]. \quad \square$$

Applying Lemmas 3 and 4 to (4.13), we have for  $x > \mu\bar{\xi} + \delta/2$ ,

$$\begin{aligned} P(Q > x) &< D_4(D_1 + D_2x) \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\mu}{H(0)^2/2}x\right] \\ &= (B_2 + B_3x) \exp\left[-\frac{\mu}{H(0)^2/2}x\right], \end{aligned}$$

where  $B_2 = D_4D_1\sqrt{\gamma/2\pi}$  and  $B_3 = D_4D_2\sqrt{\gamma/2\pi}$ .

If  $\beta_t \geq 0$  for all  $t > 0$ , then we may set  $D_2 = 0$  and hence  $B_3 = 0$ .

**Remark 4.** Recall that  $\beta_0 = 0$  and  $\beta'_t = \Lambda_R(t)$ . Then,  $\beta_t$  is rewritten as

$$\beta_t = \int_0^t \Lambda_R(s) ds.$$

In case of an Ornstein-Uhlenbeck process,  $\Lambda_R(t)$  is always positive and hence  $\beta_t$  is always positive. Hence  $B_3$  in the upper bound may be set to zero. When actual measurement data  $R_{i\Delta t}$ ,  $i = 1, 2, \dots, N$ , are given, we can also judge whether  $\beta_t$  is always positive or not, by accumulating the sample autocovariance of  $R_{i\Delta t}$ ,  $i = 1, 2, \dots, N$ .

## 4.2. Proof of the lower bound

**Lemma 5.** *The tail distribution of the buffer content  $Q_\infty$  in the steady-state satisfies*

$$P(Q_\infty > x) \geq \max_{t \geq 0} P(A_t - Ct > x). \quad (4.16)$$

*Proof.* According to [2, 5], the solution of the differential equation (2.1) with  $Q_0 = 0$  is written as

$$\begin{aligned} Q_t &= A_t - Ct - \inf_{0 \leq s \leq t} \{A_s - Cs\} \\ &= \sup_{0 \leq s \leq t} \{A_t - A_s - C(t-s)\}. \end{aligned}$$

Let  $\tilde{t} = t - s$ , then

$$Q_t = \sup_{0 \leq \tilde{t} \leq t} \{A_{\tilde{t}} - C\tilde{t}\}.$$

Thus

$$P(Q_\infty > x) = \lim_{t \rightarrow \infty} P(\sup_{0 \leq \tilde{t} \leq t} \{A_{\tilde{t}} - C\tilde{t}\} > x) = P(\sup_{t \geq 0} \{A_t - Ct\} > x).$$

Since  $\{\sup_{t \geq 0} \{A_t - Ct\} > x\} \supset \{A_t - Ct > 0, t \in [0, \infty)\}$ , we have for arbitrary  $t > 0$

$$P(Q_\infty > x) \geq P(A_t - Ct > x).$$

Hence we have the inequality (4.16).  $\square$

Since  $A_t - Ct$  is a Gaussian process,  $P(A_t - Ct > x)$  is given by

$$P(A_t - Ct > x) = \bar{\Phi}(y_t), \quad (4.17)$$

where

$$y_t = \frac{x + \mu t}{\sqrt{\alpha_t}} \quad \text{and} \quad \bar{\Phi}(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty \exp[-\frac{u^2}{2}] du.$$

Note that  $\xi_t$  is a function such that  $\alpha_t = H(0)^2(t + \xi_t)$ . From Lemma 1, there exists  $\bar{\xi}$  such that  $|\xi_t| < \bar{\xi} < \infty$ . For  $t > \bar{\xi}$  let

$$\tilde{y}_t = \frac{x + \mu t}{\sqrt{H(0)^2(t - \bar{\xi})}},$$

then we have the following lemma.

**Lemma 6.**

$$P(Q_\infty > x) > \bar{\Phi}(\min_{t > \bar{\xi}} \tilde{y}_t). \quad (4.18)$$

*Proof.* From  $|\xi_t| < \bar{\xi} < \infty$ , we have

$$y_t = \frac{x + \mu t}{\sqrt{H(0)^2(t + \xi_t)}} < \frac{x + \mu t}{\sqrt{H(0)^2(t - \bar{\xi})}} = \tilde{y}_t, \quad \text{for } t > \bar{\xi}.$$

Since  $\bar{\Phi}(y)$  is a decreasing function of  $y$ ,  $\bar{\Phi}(y_t)$  is greater than  $\bar{\Phi}(\tilde{y}_t)$  for all  $t > \bar{\xi}$ . Then, from (4.16) and (4.17), we have

$$P(Q_\infty > x) \geq \max_{t > \bar{\xi}} \bar{\Phi}(y_t) > \max_{t > \bar{\xi}} \bar{\Phi}(\tilde{y}_t) = \bar{\Phi}(\min_{t > \bar{\xi}} \tilde{y}_t). \quad \square \quad (4.19)$$

By using Lemma 6, we can evaluate a lower bound. Since  $\tilde{y}_t$  is a convex function of  $t$ ,

$$\min_{t > \bar{\xi}} \tilde{y}_t = \frac{2\sqrt{\mu x + \mu^2 \bar{\xi}}}{H(0)}.$$

Using (4.19) and the inequality

$$\sqrt{2\pi}\Phi(y) = \int_y^\infty \exp\left[-\frac{u^2}{2}\right] du \geq \frac{1}{y+1} \exp\left[-\frac{y^2}{2}\right],$$

we have

$$P(Q_\infty > x) > \frac{1}{\sqrt{2\pi}\left(\frac{2\sqrt{\mu x + \mu^2 \bar{\xi}}}{H(0)} + 1\right)} \exp\left[-\frac{\mu x + \mu^2 \bar{\xi}}{H(0)^2/2}\right].$$

Since  $\sqrt{ax+b} < \sqrt{ax} + \sqrt{b}$  for  $a > 0$ ,  $b > 0$  and  $x > 0$ , we obtain the desired lower bound

$$P(Q_\infty > x) > \frac{1}{B_0 + B_1\sqrt{x}} \exp\left[-\frac{C-m}{H(0)^2/2}x\right], \quad (4.20)$$

where  $B_0 = \sqrt{2\pi}\left(\frac{2\mu\sqrt{\bar{\xi}}}{H(0)} + 1\right) \exp\left[-\frac{\mu^2\bar{\xi}}{H(0)^2/2}\right]$  and  $B_1 = \frac{2\sqrt{2\pi}\mu}{H(0)} \exp\left[-\frac{\mu^2\bar{\xi}}{H(0)^2/2}\right]$ .

## References

- [1] D. Anick, D. Mitra and M. M. Sondhi, Stochastic Theory of a Data-Handling System with Multiple Sources, *The Bell System Technical Journal* **61**(8) (1982) 1871-1894.
- [2] V. E. Benes, *General Stochastic Processes in the Theory of Queues* (Addison-Wesley, Reading, MA, 1963).
- [3] K. Debicki and T. Rolski, A Gaussian Fluid Model, *Queueing Systems* **20** (1995) 433-452.
- [4] R. Grünfelder, J. P. Cosmas, S. Manthorpe and A. Odinma-Pkafor, Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queueing System Performance, *IEEE J. Sel. Areas in Commun.* **9**(3) (1991) 284-293.
- [5] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems* (Wiley, New York, 1985).
- [6] O. Hashida and M. Fujiki, Queueing Models for Buffer Memory in Store-and-Forward System, *Proc. ITC-7*, (1973) 323/1-323/7.
- [7] D. P. Heyman, A. Tabatabai and T. V. Lakshman, Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks, *IEEE Trans. on Circuits and Systems for Video Tech.* **2**(1) (1992) 49-59.
- [8] K. Kobayashi and Y. Takahashi, Steady-State Analysis of ATM Multiplexer with Variable Input Rate through Diffusion Approximation, *Performance Evaluation* **23**(2) (1995) (to appear).
- [9] L. Kosten, Stochastic Theory of Data Handling Systems with Groups of Multiple Sources, *Performance of Computer-Communication Systems* (North-Holland, 1984) 321-331.
- [10] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. D. Robbins, Performance Models of Statistical Multiplexing in Packet Video Communications, *IEEE Trans. on Commun.* **36**(7) (1988) 834-843.
- [11] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, New Jersey, 1975).
- [12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).

- [13] A. Simonian, Stationary Analysis of a Fluid Queue with Input Rate Varying as an Ornstein-Uhlenbeck Process, *SIAM J. Appl. Math.* **51**(3) (1991) 828-842.
- [14] A. Simonian and J. Virtamo, Transient and Stationary Distributions for Fluid Queues and Input Processes with a Density, *SIAM J. Appl. Math.* **51**(6) (1991) 1732-1739.
- [15] T. E. Stern and A. I. Elwalid, Analysis of Separable Markov-Modulated Rate Models for Information-Handling Systems", *Adv. Appl. Prob.* **23** (1991) 105-139.
- [16] T. Tanaka, O. Hashida and Y. Takahashi, Transient Analysis of Fluid Model for ATM Statistical Multiplexer, *Performance Evaluation* **23**(2) (1995) 145-162.

Kazutomo KOBAYASHI

C&C Research Laboratories, NEC Corporation  
4-1-1 Miyazaki, Miyamae-ku, Kawasaki 216, Japan  
e-mail: koba@sbl.cl.nec.co.jp

Yukio TAKAHASHI

Mathematical and Computing Sciences  
Tokyo Institute of Technology  
2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan  
e-mail: yukio@is.titech.ac.jp