

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Performance modeling and analysis of ATM multiplexers
著者(和文)	小林和朝
Author(English)	Kobayashi Kazutomo
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第3131号, 授与年月日:1996年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第3131号, Conferred date:1996/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Performance Modeling and Analysis of ATM Multiplexers

by

Kazutomo KOBAYASHI

Submitted in partial fulfillment of
the requirement for the degree of
DOCTOR OF SCIENCE

Supervised by Prof. Yukio Takahashi
Department of Information Sciences
Tokyo Institute of Technology
Oh-okayama, Meguro-ku,
Tokyo 152, Japan

March 1996

Acknowledgments

I would like to express my sincere gratitude to Prof. Yukio Takahashi, who introduced me to the field of this study and led me toward the completion of this thesis. He taught me many things, not only stochastic processes and queueing theory but also an attitude toward theoretical researches. I am very thankful to have had an opportunity studying under his guidance.

I would like to thank Prof. Naoki Makimoto for his many valuable comments and suggestions. I also thank colleagues who have been studying with me at Takahashi's laboratory for their helpful comments and assistances. Especially, Mr. Toshiyasu Kurasugi have been helpful in preparing Chapter 4 of this thesis. I am also grateful to Mr. Mitsuharu Yano and Mr. Hidenobu Harasaki of NEC Corporation for providing me with a coded video traffic data. I wish to thank Prof. Mituaki Fujii, Prof. Masao Mori and Prof. Masakazu Kojima for the careful review of the entire manuscript and valuable comments.

Special thanks are due to Dr. Kojiro Watanabe of NEC USA and Dr. Issei Kino of NEC Corporation for introducing me to the field of performance evaluation of computers and communications. I greatly appreciate Dr. Kazumoto Iinuma, Mr. Masahiro Yamamoto, Mr. Katsuhiro Nakamura and Mr. Tomoyuki Fujita of NEC Corporation for the approval of my enrolling in the Ph. D. program and their encouragements. I also thank the members of System Basic Research Laboratory, NEC C&C Systems Research Laboratories for their supports.

I dedicate this thesis to my wife, Hitomi, in token of affection and gratitude.

Contents

1	Introduction and summary	1
1.1	Motivation and background	1
1.2	Input traffic models for ATM multiplexers	5
1.2.1	Markov-modulated cell stream models	6
1.2.2	Fluid stream models	6
1.3	Fluid queue models and their input rate processes	8
1.3.1	On-Off-type input rate processes	9
1.3.2	Ornstein-Uhlenbeck processes	11
1.3.3	Markov-modulated rate processes	12
1.3.4	Autoregressive processes	13
1.3.5	Gaussian-type input rate processes	15
1.3.6	Data analysis of VBR video traffic with scene changes	16
1.3.7	Long-range dependence of VBR video traffic	18
2	Diffusion approximation	20
2.1	Introduction	20
2.2	Gaussian-type input rate processes	22
2.3	Examples of input rate processes	26
2.4	A diffusion approximation for the fluid queue model	33
2.5	Numerical results and discussions	42
3	Tail behavior of stationary distribution	47
3.1	Introduction	47

3.2	A fluid queue model with a Gaussian-type input rate process	48
3.2.1	Coded video traffic	48
3.2.2	Model description	50
3.3	Bounds for the tail probability $P(Q_\infty > x)$	53
3.3.1	Proof of the upper bound	53
3.3.2	Proof of the lower bound	61
4	Data analysis and modeling of VBR video traffic with scene changes	63
4.1	Introduction	63
4.2	Data analysis of VBR video traffic with scenes	66
4.2.1	Transition of scene types	71
4.2.2	Length of scene	72
4.2.3	Average bit rate	74
4.2.4	Spike of scene	76
4.2.5	Body sequence of scene	77
4.2.6	Dependence between parameters	87
4.3	Modeling of VBR traffic	89
4.4	Simulated sequence of coding bit rates	91
4.4.1	Simulation of a single source	91
4.4.2	Simulation of multiple sources	102
4.5	Tail distribution of an ATM multiplexer	106
4.6	A discussion on long-range dependence	109
	Bibliography	113

Chapter 1

Introduction and summary

1.1 Motivation and background

Asynchronous Transfer Mode (ATM) has been recognized as a main technology for multiplexing and switching a variety of information, such as voice, data, image and video, at high speed, and it has been adopted by both high-speed local and wide area private networks and broadband public information highways. Especially, ATM-LANs, which are ATM networks for local area private networks, are already on the market these days.

The main characteristics of the ATM technology are the followings:

- A variety of information is carried with short fixed-length blocks called *cells*.
- Error recovery and flow control are absent in an ATM network¹.
- The cell route is determined when a call connection is established, and it is maintained for a series of information.

These characteristics make switching and buffer management of intermediate nodes

¹Recently, a flow control mechanism is considered for a new service class called *Available Bit Rate* (ABR)

simple and end-to-end transmission delays small enough to deal with real-time voice and video communications.

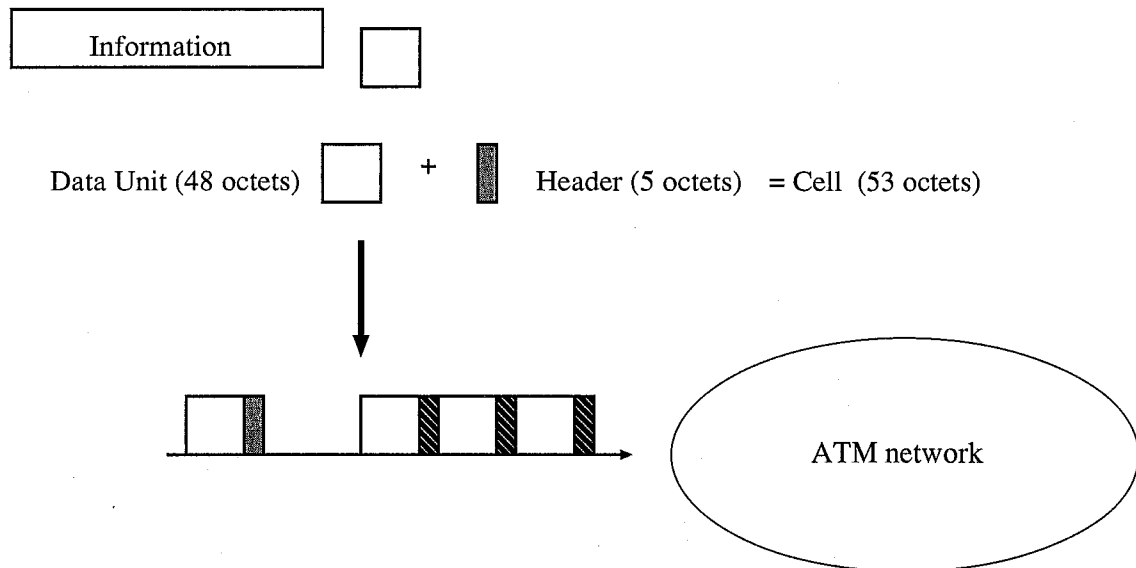


Figure 1.1: Concept of an ATM source

In an ATM source, information is divided into 48-octet data units, and a 5-octet header is attached to the top of each data unit. The resulting units, called cells, are transmitted into an ATM network (see Figure 1.1). The ATM network consists of ATM nodes and transmission lines that connect nodes and sources. The transmission rate of those lines is expected to be 156 Mbps or more. Each ATM node consists of a switch and several multiplexers² (see Figure 1.2), and each multiplexer has a finite buffer and a transmission facility which sends cells to the attached output line. All input lines for an ATM node are connected to the switch in the node. In the switch, cells from input lines are switched to corresponding multiplexers connected to their destination output lines. In a multiplexer, a transmission facility sends cells one by one to the output line with a constant service time. Excess cells are stored in the

²An actual ATM node might not be physically separated into a switch and multiplexers. But from the standpoint of performance modeling and analysis, an ATM node is usually modeled as a combination of a switch and multiplexers.

buffer in the order of arrival and served under the FCFS discipline. Since the buffer capacity is finite, arriving cells are lost when the buffer is full. Thus, a multiplexer functions as a queueing system with a finite buffer and a constant service time.

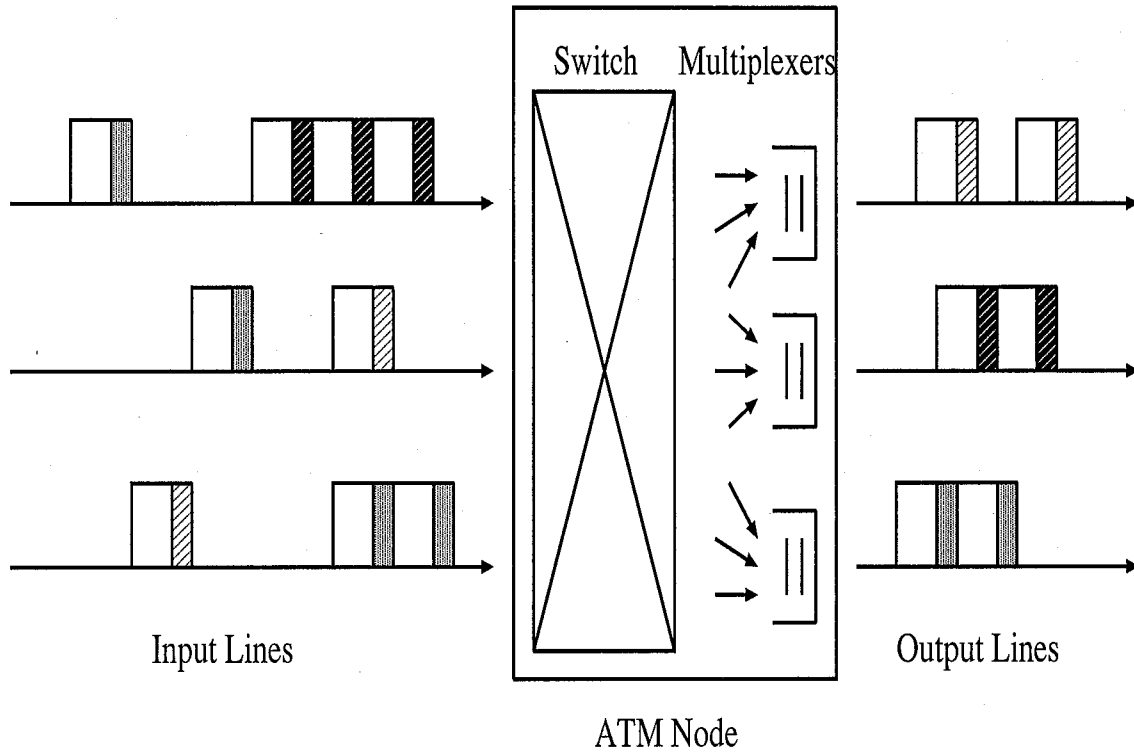


Figure 1.2: Concept of an ATM node

In this way, in ATM nodes, cell loss may occur at multiplexers with some positive probability. This probability is called *cell loss probability* and is the most important performance index for traffic design and control of ATM networks. In ATM networks, because of the lack of flow control and error recovery, each node is required to guarantee a cell loss probability of less than, say, 10^{-9} .

The cell loss probability is usually estimated by analysis of a single multiplexer model with an infinite buffer. This is justified by the following reasons:

- The route of cells is fixed for a series of information. Hence a cell loss probability of one multiplexer can be estimated independently of other multiplexers.

- Since the cell loss probability under consideration is very small, it can be approximated by the tail probability of the buffer content distribution for a multiplexer model with an infinite buffer.

In this thesis, we discuss a single multiplexer model that accommodates multiple sources, serves with constant service times, and has an infinite buffer. This model is called an *ATM multiplexer* hereafter.

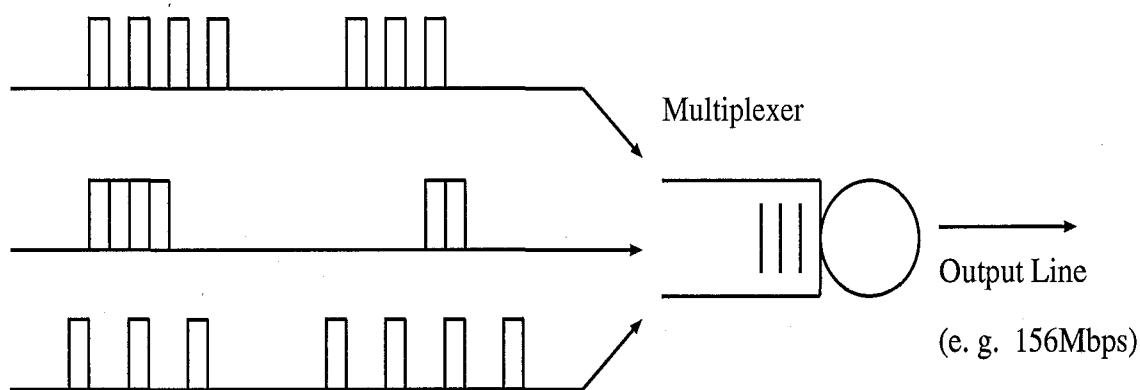


Figure 1.3: Queueing model of the ATM multiplexer

In an ATM multiplexer, the queueing mechanism is simple. However, cell arrival processes have strong correlations in interarrival times since information is divided into blocks of a small fixed size. This means that we can not apply the conventional queueing theory based on renewal processes. Moreover, since extremely rare events with a probability of 10^{-9} must be evaluated, “brute-force” computer simulations and numerical analyses of Markov chains with large state-spaces are also difficult to apply. Therefore, a new methodology is required to evaluate performance of an ATM multiplexer.

In subsequent sections of this chapter we will survey previous works which deal with this type of ATM multiplexer models and summarize results obtained in subsequent chapters.

1.2 Input traffic models for ATM multiplexers

The most important characteristic of ATM multiplexers is their input traffic. Let us consider a cell stream from a voice source as an example. The source sends cells consecutively at regular intervals during talk-spurts and stop sending during silence periods. In the cell stream, then, there exist two kinds of periods, *On-periods* and *Off-periods*. In an On-period cells arrive at regular intervals, and in an Off-period no cell arrives (see Figure 1.4). This type of traffic is called *On-Off-type traffic*.

Now we suppose that there are multiple cell streams from multiple sources and they are aggregated into a single input cell stream. At first sight, the aggregated cell stream might seem to be formulated as a Markov chain in a simple manner if we assume that On/Off-periods at each input line form an alternating phase-type renewal process and the streams from multiple sources are mutually independent. However, carefully observing the aggregated cell stream, one finds that even the derivation of a Markov chain is not easy unless the cell arrivals are synchronized. The assumption of synchronization is too restrictive for practical applications since cell intervals vary largely according to the sources.

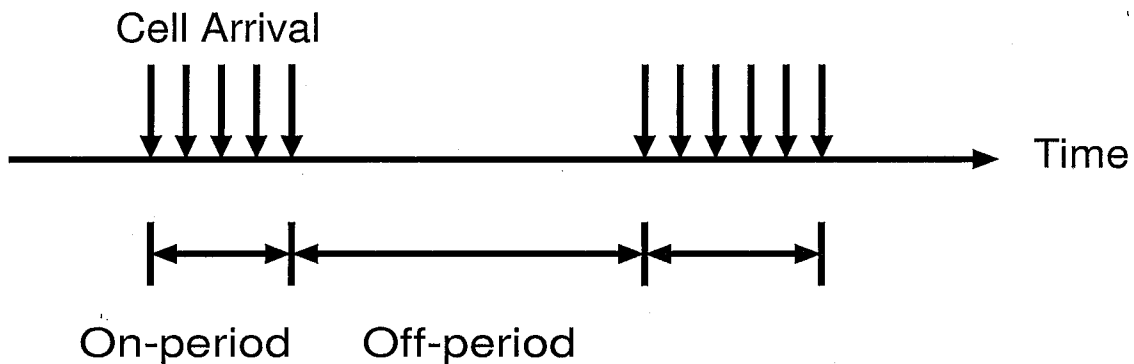


Figure 1.4: On-Off-type traffic

To cope with the difficulty, two types of approximate models are frequently used: *Markov-modulated cell stream models* and *fluid stream models*.

1.2.1 Markov-modulated cell stream models

“Markov-modulated” means that some parameters governing the behavior of a basic process are modulated subjecting to an environment Markov chain.

A simple model for On-Off-type traffic is the one in which a cell arrival process during On-periods is represented by a Poisson process and On/Off-periods are subjecting to exponential distributions. This model is called an *interrupted Poisson process* (IPP) [20] and is a renewal process with a 2nd-order hyperexponential distribution. The aggregated arrival process of multiple IPPs is no longer a renewal process. Instead it can be interpreted as a Markov-modulated Poisson process (MMPP) [12, 7] in which the arrival rate R_t varies according to a finite irreducible Markov process J_t . During any sojourn in state i , arrivals occur according to a Poisson process of rate r_i . The MMPP is parameterized by m arrival rates r_i ’s and an $m \times m$ transition rate matrix $\mathbf{M}$ governing the Markov process J_t , where m is the number of states of the Markov process.

Markovian arrival processes (MAP) [23] and phase-type Markov renewal processes (PH-MRP) [24] are proposed to extend the Poisson arrivals during sojourns in individual states into phase-type renewal arrivals.

In Markov-modulated cell stream models, as a general rule, the number of states becomes larger as the number of sources increases. In [12, 29], it is proposed that the aggregated cell stream is again approximated by a two-state MMPP whose parameters are estimated by matching several of the statistical characteristics.

1.2.2 Fluid stream models

In fluid stream models, detailed behavior of cells is ignored and cell streams are regarded as fluid streams with their input rates varying as some stochastic processes.

Let us consider On-Off-type traffic as an example. Each source pours fluid at a

constant rate during On-periods and suspends flow during Off-periods (see Fig. 1.5). The constant rate during On-periods is given by the reciprocal of the cell interarrival time during On-periods.

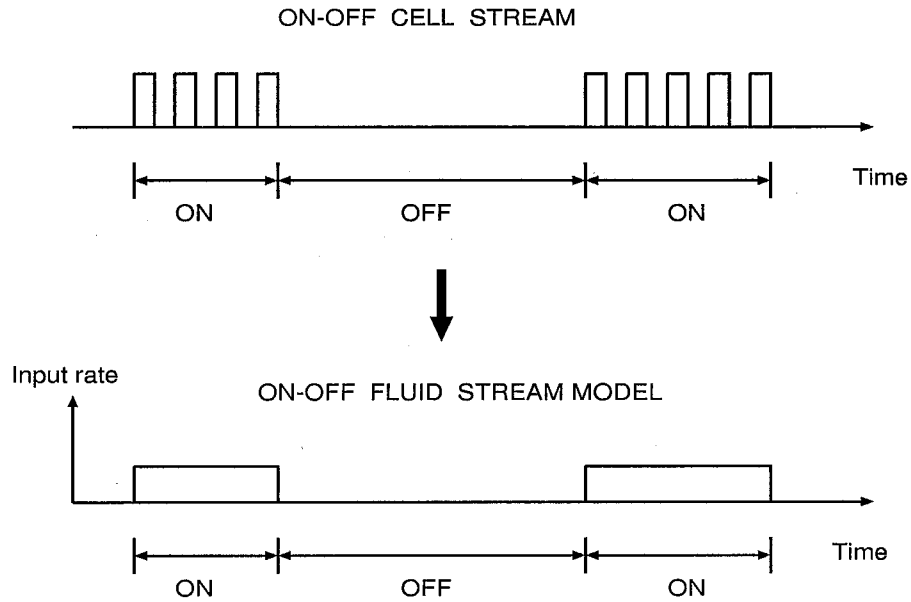


Figure 1.5: An On-Off fluid stream model

In ATM multiplexers, interarrival times of arrival cells are very short compared with variation of the input rate. This suggests that cell streams are well approximated by fluid streams. Fortunately the buffer capacity of ATM multiplexers is usually so large as to absorb the detailed behavior of cells. Thus fluid stream models are frequently used for analysis of ATM multiplexers. Most of the stream models treated in this thesis are of this type.

1.3 Fluid queue models and their input rate processes

Now we consider ATM multiplexer models in which cell streams are modeled as fluid stream models. In the multiplexer models, the buffer in an ATM multiplexer is represented as a bucket to store fluid in and the service facility is represented as a hole at the bottom of the bucket (see Figure 1.6). The buffer content of the ATM multiplexer corresponds to the volume of fluid in the bucket.

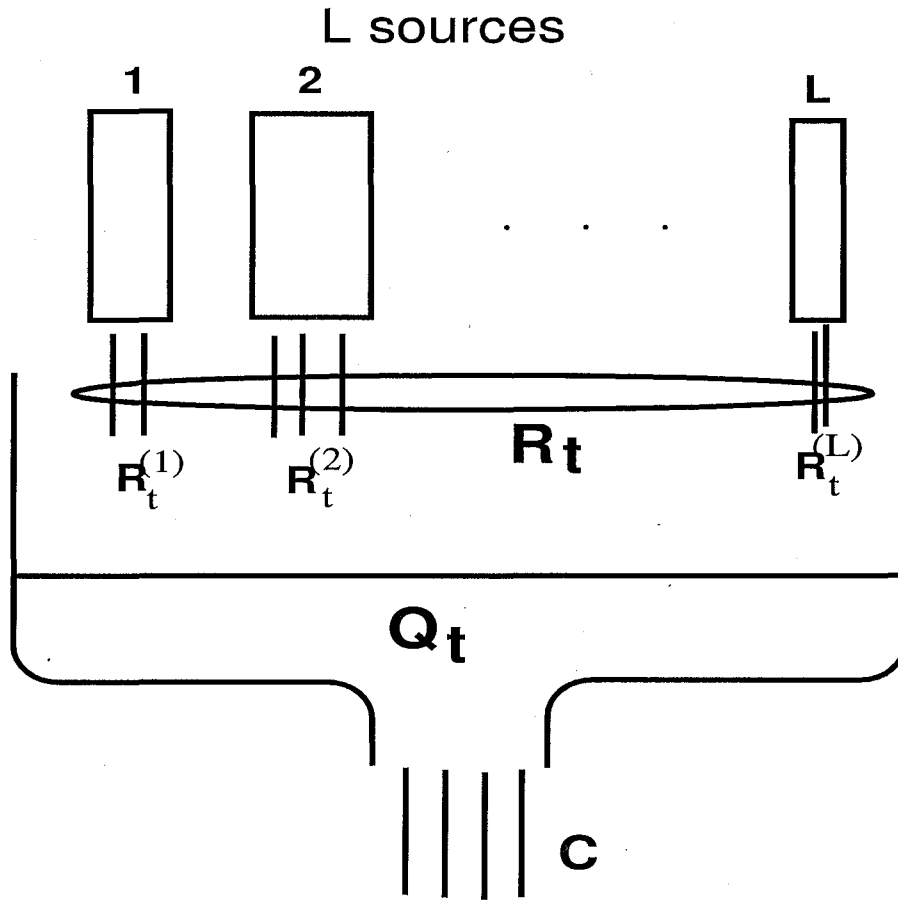


Figure 1.6: A fluid queue model

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 $R_t = \sum_{i=1}^L R_t^{(i)}$. The behavior of

the buffer content Q_t is then described by a 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*. This ATM multiplexer model is called a *fluid queue model*.

1.3.1 On-Off-type input rate processes

Now we consider a fluid queue model in which On-Off-type traffic is modeled by a fluid stream. In this model, each source has two states, On-state and Off-state, to alternate each other and pours fluid at a constant rate during On-state and suspends flow during Off-state (see Fig. 1.7).

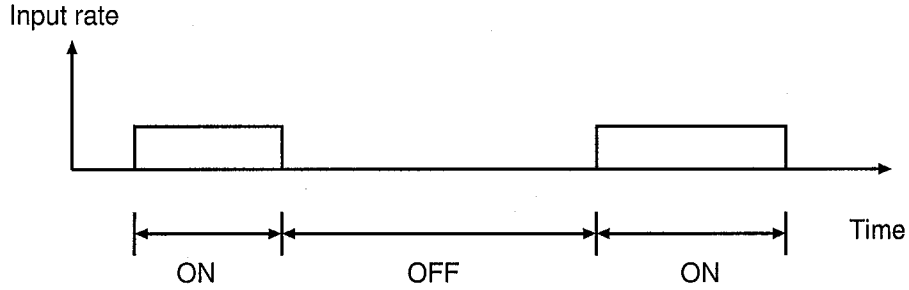


Figure 1.7: On-Off-type input rate processes

Let r_i be the On-state input rate of source i . Then the input rate process $R_t^{(i)}$ from source i is given by

$$R_t^{(i)} = \begin{cases} r_i, & \text{if source } i \text{ is in On-state at time } t, \\ 0, & \text{if source } i \text{ is in Off-state at time } t. \end{cases}$$

We will refer input rate processes of this type as *On-Off-type input rate processes* and sources generating On-Off-type input rate processes as *On-Off sources*.

An On-Off-type input rate process is parameterized by an On-state input rate and the distributions of On-periods and Off-periods. In voice sources, the On-period

and the Off-period distributions are set by talk-spurt and silence distributions, respectively, and the input rate during On-periods is set by the reciprocal of the cell interarrival time during On-periods.

Pioneer works on fluid queue models with On-Off sources are done by O. Hashida and M. Fujiki [11], and D. Anick, D. Mitra and M. M. Sondhi [2]. They assume a single type of sources and exponential On/Off distributions. The “single type” means that all the sources have the same parameters. Let $1/\lambda$, $1/\mu$, r and L be the mean Off-period, the mean On-period, the On-state input rate and the number of sources, respectively. The input rate process R_t is formulated by the input rate matrix $\mathbf{R} = \text{diag}\{0, r, 2r, \dots, Lr\}$ and the transition rate matrix,

$$\mathbf{M} = \begin{bmatrix} -L\lambda & \mu & & & \\ L\lambda & -(L-1)\lambda - \mu & 2\mu & & \\ & (L-1)\lambda & \vdots & \vdots & \\ & & 2\lambda & -\lambda - (L-1)\mu & L\mu \\ & & & \lambda & -L\mu \end{bmatrix}.$$

The stationary buffer distribution is obtained in a closed form [2]. From this result, the tail of the buffer distribution is given as

$$P(Q_t > x) \sim Be^{-\gamma x},$$

where

$$\gamma = \frac{\lambda}{C/L} - \frac{\mu}{r - C/L}.$$

Here, $f(x) \sim g(x)$ indicates $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$. The multiplicative constant B is represented as

$$B = \rho^L \prod_{z_i \in Z} \frac{z_i}{z_i + \gamma},$$

where ρ is the traffic intensity given by $N\lambda/(\lambda + \mu)$ and Z is the set of negative eigenvalues of the matrix $(\mathbf{R} - C\mathbf{I})^{-1}\mathbf{M}$ except for $-\gamma$.

L. Kosten [19] and H. Kobayashi [15] discuss fluid queue models with On-Off sources of multiple types.

1.3.2 Ornstein-Uhlenbeck processes

A. Simonian [34] discusses a fluid queue model with input rate varying as an Ornstein-Uhlenbeck process and obtains the exact stationary distribution of the buffer content. This process is defined by the stochastic differential equation

$$dR_t = \kappa(m - R_t)dt + \sigma dw_t,$$

where κ , m and σ are constants and w_t is a standard Wiener process. This process is a continuous version of a first-order autoregressive process, and approximates well the total input rate process from On-Off sources of a single type with exponential On/Off distributions when the number of sources are large.

A fluid queue model with On-Off sources of multiple types is discussed by L. Kosten [19] and H. Kobayashi [15]. However from their results it is difficult to obtain not only the exact distribution of the buffer content but also the asymptotic behavior. To cope with this difficulty, it is proposed in H. Kobayashi and Q. Ren [16] that a total input rate process from On-Off sources of multiple types is approximated by a superposition of Ornstein-Uhlenbeck processes.

Let there be N_k On-Off sources of type- k , $k = 1, \dots, K$, and Off- and On-periods of type- k sources be exponentially distributed with parameters λ_k and μ_k , respectively;

$$\begin{aligned} \lambda_k^{-1} &= \text{the mean length of an Off-period of a type-}k \text{ source,} \\ \mu_k^{-1} &= \text{the mean length of an On-period of a type-}k \text{ source.} \end{aligned}$$

We also let r_k be the input rate of a type- k source at On-state. The aggregated input rate process of type- k is a birth-and-death process on $\{0, 1, \dots, N_k\}$. As N_k tends to infinity and r_k tends to 0 in such a way that $N_k r_k$ remains constant, the birth-and-death process converges to an Ornstein-Uhlenbeck process. In the sense, the aggregated input rate process of type- k can be approximated by an Ornstein-Uhlenbeck process $R_t^{(k)}$.

Using this approximation³, the total input rate process is represented by a superposition of independent K Ornstein-Uhlenbeck processes:

$$R_t = \sum_{k=1}^K R_t^{(k)},$$

where

$$dR_t^{(k)} = \kappa_k(m_k - R_t^{(k)})dt + \sigma_k dw_t^{(k)}, \quad k = 1, \dots, K,$$

and $w_t^{(k)}$, $k = 1, \dots, K$, are mutually independent Wiener processes.

An approximate formula for the tail distribution of the buffer content is obtained in H. Kobayashi and Q. Ren [16]:

$$P(Q_t > x) \approx B \exp\left[-2\frac{C - m}{b}\right],$$

where

$$m = \sum_{k=1}^K \frac{N_k r_k \lambda_k}{\lambda_k + \mu_k}, \quad b = \sum_{k=1}^K \frac{2N_k r_k^2 \lambda_k \mu_k}{(\lambda_k + \mu_k)^3}.$$

The multiplicative constant B is given by

$$B = \frac{1}{\theta \sqrt{2\pi}} e^{-\frac{\theta^2}{2}}, \quad \theta = \frac{C - m}{\sigma_R}, \quad \sigma_R^2 = \sum_{k=1}^K \frac{2N_k r_k^2 \lambda_k \mu_k}{(\lambda_k + \mu_k)^2}.$$

from a heuristic discussion that $P(Q_t > 0) \approx P(R_t > C)$.

1.3.3 Markov-modulated rate processes

T. E. Stern and A. I. Elwalid [36] introduce Markov-modulated rate processes (MMRP) as input rate processes for dealing with video traffic as well as voice traffic. In a MMRP, the input rate R_t varies according to an m -state irreducible Markov process J_t . Thus, the MMRP is parameterized by the transition rate matrix $\mathbf{M}$ governing the Markov process J_t and the input rate matrix $\mathbf{R} = \text{diag}\{r_1, \dots, r_m\}$. On-Off-type input rate processes are included in the class of MMRP. Note that

³In [16], the approximate process is obtained by applying the Taylor series expansions for an original partial differential equation and retaining the first- and second-order terms.

the behavior of the input rate R_t is the same as that of the arrival rate of the MMPP introduced in Markov cell stream models. The difference is that the MMPP generates a cell stream whereas the MMRP generates a fluid stream.

Under the assumption that the transition rate matrix $\mathbf{M}$ is reversible, T. E. Stern and A. I. Elwalid [36] obtain the tail distribution of the buffer content of the form

$$P(Q_t > x) = \sum_n d_n e^{z_n x}$$

where z_n 's are negative eigenvalues of $(\mathbf{R} - C\mathbf{I})^{-1}\mathbf{M}$. The coefficients d_n are given from these eigenvalues and the corresponding eigenvectors. Tanaka, et al. [37] study the same model and obtain non-stationary state probabilities in an explicit form without the assumption of the reversibility of $\mathbf{M}$. However, except for simple cases, it is not easy to calculate the exact and asymptotic steady-state distribution from the results of these studies.

1.3.4 Autoregressive processes

Voice and data traffic is well modeled as an On-Off-type input rate process. However, coded video traffic is much different from them. Figure 1.8 shows an example of an actual coded video traffic pattern. The vertical axis represents the input rate and the horizontal axis the time.

B. Maglaris et al. [25] and D. P. Heyman et al. [14] model coded video traffic as an autoregressive processes of the form

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

where X_n is the input rate at discrete time n , a_k 's are real constants and ε_n 's are mutually independent and identically distributed random variables with zero mean and finite variance. The random variables ε_n 's are usually assumed to be normally distributed. The sequence $\{\varepsilon_n\}$ is considered to be a discrete version of the *white noise*.

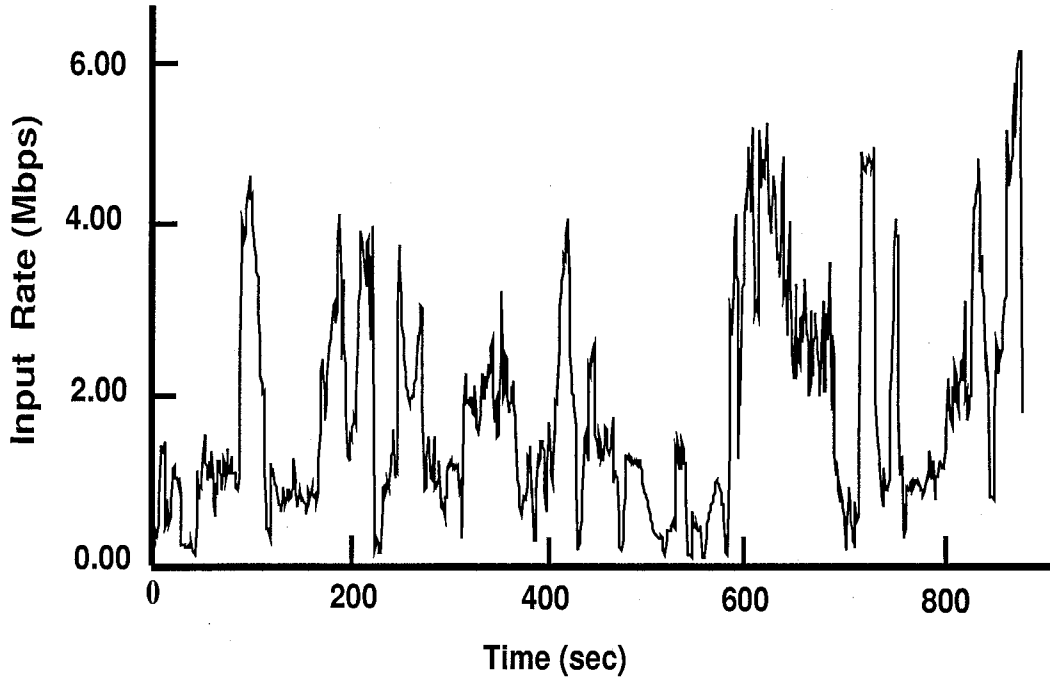


Figure 1.8: Coded video traffic

R. Grünfelder et al. [10] model coded video traffic as an autoregressive moving average process,

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

using the well-established digital filter theory.

The advantage of those processes is that their parameters can be easily estimated from an empirical mean and an empirical autocovariance function. However, they have been considered inconvenient for queueing analysis because they are multi-dimensional *continuous state* Markov processes. Consequently, they have been used only for simulations.

1.3.5 Gaussian-type input rate processes

In order to deal with not only voice and data traffic but also coded video traffic, we introduce a new type input rate processes called Gaussian-type input rate processes in Chapter 2 and discuss a fluid queue model with a Gaussian-type input rate process in Chapters 2 and 3. A stochastic process R_t is a *Gaussian-type input rate process* if it is of the form

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

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.

In Chapter 2, we show that the class of Gaussian-type input rate processes covers continuous versions of both autoregressive processes and autoregressive moving average processes as well as Ornstein-Uhlenbeck processes and their superpositions. Moreover, we demonstrate that a total input rate process from general On-Off sources with possibly different parameters and different non-exponential On- and Off-period distributions can be approximated by a Gaussian-type input rate process. Therefore, the class of Gaussian-type input rate processes is wide enough to contain most of continuous time stochastic processes proposed so far for voice and data traffic and coded video traffic.

For a fluid queue model with a Gaussian-type process, we derive in Chapter 2 an approximate formula

$$P(Q_t > x) \approx \frac{m - \eta}{C - \eta} \exp\left[-\frac{C - m}{H(0)^2/2}x\right], \quad \text{for } x > 0, \quad (1.1)$$

where $H(0) = \int_0^\infty h(s)ds$ and η is the conditional expectation of R_t in the steady-state conditioned that $R_t < C$. This formula is derived by approximating the behavior of Q_t by a diffusion process x_t on $\mathbf{R}(-\infty, \infty)$ that has no reflecting barrier.

The formula (1.1) contains four parameters. Three parameters other than C can be calculated from the mean and the autocovariance function of the input rate

process R_t . This means that we can estimate the cell loss probability directly from measurement data of the input traffic without parameter fitting on $h(t)$. Hence, the formula (1.1) gives us a powerful tool for estimating the cell loss probability.

In Chapter 3, in order to justify the decay rate of the approximate formula (1.1), it is proved that for sufficiently large x ,

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

where B_0 , B_1 and B_2 are positive constants and B_3 is a non-negative constant. The upper bound and the lower bound have the same exponential decay rate as the formula (1.1). In this sense, this result guarantees the plausibility of the approximate formula (1.1).

We note that K. Debicki and T. Rolski [6] obtain a similar upper bound for the same fluid queue model, independently.

1.3.6 Data analysis of VBR video traffic with scene changes

In ATM networks, coded video traffic, called *variable-bit-rate* (VBR) video traffic, is considered to become main traffic since video transmissions require larger bandwidth than voice and data ones.

In previous works, coded video traffic is frequently modeled as a discrete time Gaussian process [25, 14, 10, 27], which was discussed in Subsection 1.3.4. The model is based on analyses of data in video teleconference or video phone, and it takes no account of scene changes. However, in most VBR video traffics there are fluctuations in bit rates caused by scene changes, and they look too complex to be modeled as such processes.

In Chapter 4 we statistically analyze a certain real VBR video traffic data. In the data, we recognize that “spikes” appear at moments of scene changes and the behavior of the traffic differs from scene to scene. So we analyze the traffic data by

taking account of scene changes.

First we divide the sequence of the traffic data into subsequences corresponding to scenes, and classify all the subsequences into five types of scenes from a viewpoint of camera work and degree of the motion of subjects. For each scene type, we examine distributions and statistical behaviors of the scene length, the average rate of scene and the spike of scene. The body of scene is analyzed using autoregressive (AR) models.

Based on the analysis, we develop a simulation model consisting of three sub-models, a Markovian scene transition model, a spike model and an AR model. In our model, successive scene types are selected according to a Markov chain from the set of the five types of scenes, and the parameters of the spike model and the AR model for a scene are selected randomly according to scene-type dependent distributions estimated from the data.

A sequence of traffic generated by the simulation model is compared with the original sequence in four aspects; the marginal distribution, the autocorrelation function, the R/S function and the spectral density function. The R/S function is used to test for long-range dependence of the sequence, which will be discussed in the next subsection. In these aspects, the simulated traffic is recognized to resemble the real one very well.

Using the simulation model, we examine the tail behavior of the steady-state queue length distribution of an ATM multiplexer with multiple sources. The simulation results indicate that the approximation formula (1.1) derived in Chapter 2 provides good approximate values when the number of sources is large, say more than 150. When the number of sources is 50, the formula underestimates the simulated tail distribution but the decay rate is similar. In this sense, the approximation formula provides a useful tool to analyze and estimate the congestion in buffers of ATM multiplexers.

1.3.7 Long-range dependence of VBR video traffic

J. Beran et al. [4] analyze 20 large sets of actual VBR video traffic data, using R/S analysis and time-variance analysis. Based on the analyses, they say “*long-range dependence* is an inherent characteristic of VBR video traffic”. A long-range dependent process [5] is a wide-sense stationary stochastic process having autocovariance function $\Lambda(k), k \geq 0$, such that

$$\Lambda(k) \sim k^{-D} L(k), \text{ as } k \rightarrow \infty, \quad (1.2)$$

where $0 < D < 1$ and $L(k)$ is a slowly varying function. Hence for a long-range dependent process the autocovariance function is non-summable, namely,

$$\sum_{k=0}^{\infty} \Lambda(k) = \infty.$$

The input rate processes discussed so far, for example, discrete time Gaussian processes [25, 14, 10, 27], Markovian processes [11, 2, 19, 36] and Gaussian-type input rate processes introduced in Chapter 2, are not long-range dependent processes. *Self-similar* processes such as *fractional Brownian motions* [26] and *fractional ARIMA processes* [9] are typical long-range dependent processes. Beran et al. indicate that VBR video traffic will be modeled as a self-similar process.

However, to their indication there are some questions including

1. Is VBR video traffic long-range dependent in the sense of the definition (1.2)?
2. Does VBR video traffic have self-similar nature?

The final section of Chapter 4 is devoted to a discussion on these questions. We observe that a real VBR video traffic exhibits long-range dependence in R/S analysis. But, this results doesn't necessarily mean that it is long-range dependent in the sense (1.2). To make it clear, we need to find causes that make the data show long-range dependence in R/S analysis.

If the VBR video traffic had self-similar nature, it would be long-range dependent in the sense (1.2). However, the analysis and the modeling of the VBR video traffic data seem to suggest that the traffic doesn't have self-similar nature and that it is not long-range dependent in the sense (1.2). In particular, our simulation model gives us a clue of this point. Our simulation model does not use any processes having self-similar nature, but a simulated sequence exhibits long-range dependence in R/S analysis as well as the original sequence.

In this thesis, we don't find what makes VBR video traffic long-range dependent in R/S analysis. By examining our model further, we will be able to find causes and make clear whether the traffic is long-range dependent or not in the rigorous sense. This is a subject for future study.

Chapter 2

Diffusion approximation

2.1 Introduction

In this chapter, we deal with a fluid queue model described in Chapter 1. In the model, the stochastic behavior of the buffer content Q_t is represented by the differential equation

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

where C is a constant *output rate* and R_t is an *input rate process*.

Typical fluid queue models so far assume On-Off sources. This type of 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 and suspends flow during Off-state. However, coded video traffic, which is considered to be the main traffic in ATM networks, differs from On-Off-type traffic. Thus, we should consider more general input rate processes, whose class is wide enough to deal with coded video traffic as well as On-Off-type traffic.

H. Kobayashi and Q. Ren [16] propose that a total input rate process from On-Off sources of multiple types is approximated by a superposition of Ornstein-

Uhlenbeck processes, which is one of Gaussian processes, to obtain computationally feasible solutions for a fluid queue model with On-Off sources of multiple types.

On the other hand, coded video traffic is frequently modeled as a discrete time Gaussian process [25, 34, 14, 10]. B. Maglaris, et al. [25] model coded video traffic as a first-order autoregressive process, and D. P. Heyman, et al. [14] approximate it by a higher-order autoregressive process. R. J. Grünenfelder, et al. [10] model coded video traffic by an autoregressive moving average process.

In this chapter, we introduce a new type of input rate processes defined as

$$R_t = m + \int_{-\infty}^t h(t-s) dw_s, \quad (2.2)$$

where m is a constant, w_t is a standard Wiener process, and $h(t)$ is a real integrable function such that $h(t) = 0$ for $t < 0$. We will refer this type of input rate processes as *Gaussian-type input rate processes*.

We show in Section 2.2 that under a mild condition the input process $A_t = \int_0^t R_s ds$ is approximated in a form

$$A_t \sim at + \sqrt{b} \tilde{w}_t, \quad \text{for large } t, \quad (2.3)$$

where $\tilde{w}_t$ is a standard Wiener process such that $\tilde{w}_0 = 0$. The parameter a is equal to m and b is determined from the autocovariance function of R_t .

We also show in Section 2.3 that the representation (2.2) of R_t is fairly general. Processes such as Ornstein-Uhlenbeck processes, superpositions of Ornstein-Uhlenbeck processes and time continuous versions of autoregressive and autoregressive moving average processes, can be represented in the form (2.2). Moreover, a total input rate process from general On-Off sources with possibly different parameters and different non-exponential On- and Off-period distributions can be approximated by a Gaussian-type input rate process.

Exploiting the asymptotic property (2.3), we derive an approximation formula in Section 2.4 for the stationary distribution of the buffer content Q_t through a new

diffusion approximation method:

$$P[Q_t > x] \approx \frac{a - \eta}{C - \eta} \exp[-2\frac{C - a}{b}x], \quad \text{for } x > 0, \quad (2.4)$$

where C is the constant output rate and η is the conditional expectation of R_t in the steady-state conditioned that $R_t < C$. This formula contains four parameters. Three parameters other than C can be calculated from the mean and the autocovariance function of the input rate process R_t .

The formula (2.4) is derived in the following steps. First, we approximate the input process and the output process separately by one-dimensional diffusion processes, and then by taking the difference of them we derive a diffusion process x_t approximating Q_t . This process x_t is represented by a stochastic differential equation without any reflecting barrier and hence takes a value on the whole line. When $x_t < 0$, we interpret that $Q_t = 0$. The right hand side of (2.4) is the corresponding stationary tail probability of the process x_t .

Numerical results and some discussions on related approximation formulas are presented in Section 2.5. The approximation formula (2.4) is similar to the one obtained by H. Kobayashi and Q. Ren [16], but the multiplicative constant $B = (a - \eta)/(C - \eta)$ is different from their constant B' . Numerical results show that these two constants are very close to each other in a wide range of parameters, but B' exceeds 1 in case of heavy traffic, while B remains less than 1. In this sense, we think that our formula provides a more preferable approximation.

2.2 Gaussian-type input rate processes

In this section, we introduce and discuss Gaussian-type input rate processes. A stochastic process R_t is a *Gaussian-type input rate process* if it is of the form

$$R_t = m + \int_{-\infty}^t h(t-s) dw_s. \quad (2.5)$$

Here, 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 $t < 0$ and that

$$\int_0^\infty |h(t)|dt < \infty, \quad \int_0^\infty h(t)^2 dt < \infty, \quad (2.6)$$

and

$$\int_0^\infty |H(t)|dt < \infty, \quad (2.7)$$

where

$$H(t) = \int_t^\infty h(s)ds. \quad (2.8)$$

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

$$\int_0^\infty H(t)^2 dt \leq \int_0^\infty |h(t)|dt \int_0^\infty |H(t)|dt < \infty. \quad (2.9)$$

Clearly, R_t is a Gaussian process since it is given as a linear transform of the Wiener process, and it is also stationary in the wide sense because the mean $E[R_t]$ is constant m and the autocovariance $E[(R_{t+\tau} - m)(R_t - m)]$ depends only on τ . If we denote the autocovariance by $\Lambda_R(\tau)$, then it is given by

$$\Lambda_R(\tau) = \int_0^\infty h(s+\tau)h(s)ds. \quad (2.10)$$

The stationary distribution of R_t is given by the normal distribution $N(m, \sigma_R^2)$, where $\sigma_R^2 = \Lambda_R(0)$.

Now let us consider the *input process*

$$A_t = \int_0^t R_s ds. \quad (2.11)$$

From (2.5), we have

$$\begin{aligned} A_t &= m t + \int_0^t \int_{-\infty}^\tau h(\tau - s) dw_s d\tau \\ &= m t + \int_{-\infty}^0 \int_0^t h(\tau - s) d\tau dw_s + \int_0^t \int_s^t h(\tau - s) d\tau dw_s \\ &= m t + H(0)\tilde{w}_t - Z_t \end{aligned} \quad (2.12)$$

where

$$\tilde{w}_t = \int_0^t dw_s = w_t - w_0 \quad \text{and} \quad Z_t = \int_{-\infty}^t H(t-s)dw_s - \int_{-\infty}^0 H(-s)dw_s.$$

We will show that Z_t in (2.12) is negligible for large t . Note that A_t is not a stationary process, but is a Gaussian process since integration is a linear operator. The mean of A_t is equal to mt and the autocovariance function $\Lambda_A(t+\tau, t)$ for $\tau \geq 0$ is given by

$$\begin{aligned} \Lambda_A(t+\tau, t) &= E[\{A_{t+\tau} - m(t+\tau)\}\{A_t - mt\}] \\ &= H(0)^2 E[\tilde{w}_{t+\tau}\tilde{w}_t] - H(0)E[\tilde{w}_{t+\tau}Z_t] \\ &\quad - H(0)[\tilde{w}_t Z_{t+\tau}] + E[Z_{t+\tau}Z_t], \end{aligned}$$

where

$$\begin{aligned} E[\tilde{w}_{t+\tau}\tilde{w}_t] &= t, \\ E[\tilde{w}_{t+\tau}Z_t] &= \int_0^t H(t-s)ds, \\ E[\tilde{w}_t Z_{t+\tau}] &= \int_0^t H(t-s)ds, \\ E[Z_{t+\tau}Z_t] &= \int_0^\infty H(s+\tau)H(s)ds + \int_0^\infty H(s)^2ds \\ &\quad - 2 \int_0^\infty H(t+s)H(s)ds. \end{aligned}$$

From (2.7) and (2.9), $E[\tilde{w}_{t+\tau}Z_t]$, $E[\tilde{w}_t Z_{t+\tau}]$ and $E[Z_{t+\tau}Z_t]$ converge to finite limits as $t \rightarrow \infty$. On the other hand, $E[\tilde{w}_{t+\tau}\tilde{w}_t]$ increases in proportion to time t . Hence, we have

$$\Lambda_A(t+\tau, t) \sim H(0)^2 t, \quad \text{for large } t. \quad (2.13)$$

This implies that

$$A_t \sim mt + H(0)\tilde{w}_t, \quad \text{for large } t. \quad (2.14)$$

Here we use the symbol $\sim$ to indicate that as $t \rightarrow \infty$ the joint distribution of n intervals $(A_{t+t_1} - A_{t+s_1}, A_{t+t_2} - A_{t+s_2}, \dots, A_{t+t_n} - A_{t+s_n})$ with $s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n$ converges to the corresponding one for the Gaussian process on the right hand side of (2.14).

The approximation (2.14) can be understood also from the following reasoning: From (2.7) and (2.9), $\int_{-\infty}^t H(t-s)dw_s$ is a stationary process with zero mean and a finite variance. Hence, Z_t is a Gaussian process with zero mean and a bounded variance. On the other hand, the mean and the variance of $mt + H(0)\tilde{w}_t$ increases in proportion to time t . Consequently, $mt + H(0)\tilde{w}_t$ becomes dominant in (2.12) for large t .

From (2.10), $H(0)$ in (2.14) can be represented by the autocovariance function $\Lambda_R(t)$ of the input rate process R_t as follows:

$$\begin{aligned} H(0)^2 &= \int_0^\infty \int_0^\infty h(t)h(s)dt ds \\ &= 2 \int_0^\infty \int_0^\infty h(t+s)h(s)ds dt \\ &= 2 \int_0^\infty \Lambda_R(t)dt. \end{aligned} \tag{2.15}$$

Remark 2.1. If $h(t)$ satisfies the condition

$$|h(t)| < M \text{ and } \int_0^\infty t|h(t)| < \infty \tag{2.16}$$

for some positive constant M , it satisfies conditions (2.6) and (2.7). The followings are special cases of $h(t)$ satisfying (2.16).

- *Case 1.*

$$|h(t)| < M, \quad 0 \leq t \leq T, \quad \text{and} \quad h(t) = 0, \quad t > T, \tag{2.17}$$

where M and T are positive constants.

- *Case 2.*

$$h(t) = \sum_{i=1}^N \sigma_i t^{n_i} e^{-\kappa_i t} \cos(\omega_i t + \varphi_i), \quad t > 0, \tag{2.18}$$

where N is a positive integer, n_i 's are non-negative integers, κ_i 's are positive constants, and σ_i 's, ω_i 's and φ_i 's are real constans.

- *Case 3.*

$$h(t) = \int_0^t g(t-s)f(s)ds, \tag{2.19}$$

where $g(t)$ is a function of Case 1 and $f(t)$ is a function of Case 2.

2.3 Examples of input rate processes

In this section we present various examples of variable input rate processes and give derivations for the corresponding approximations (2.14) of A_t for large t .

Example 2.1. Ornstein-Uhlenbeck process

A. Simonian [34] uses this process as an input rate process to obtain the stationary probability distribution of the buffer content. This process is defined by the stochastic differential equation

$$dR_t = \kappa(m - R_t)dt + \sigma dw_t,$$

where κ and σ are positive constants. The solution is given by

$$R_t = m + \int_{-\infty}^t \sigma e^{-\kappa(t-s)} dw_s.$$

This is the case that $h(t) = \sigma e^{-\kappa t}$ in (2.5). Then, $H(t) = \sigma e^{-\kappa t}/\kappa$. Since $h(t)$ satisfies conditions (2.6) and (2.7), we have

$$A_t \sim m t + \frac{\sigma}{\kappa} \tilde{w}_t.$$

Example 2.2. Superposition of Ornstein-Uhlenbeck processes

A fluid queue model with this input rate process is discussed by H. Kobayashi and Q. Ren [16] though their starting point is an analysis for a fluid queue model with multiple On-Off types of traffic. The latter model will be discussed in Example 2.5.

A superposition of independent Ornstein-Uhlenbeck processes is defined as

$$R_t = \sum_{i=1}^N R_t^{(i)},$$

where

$$\begin{aligned} dR_t^{(1)} &= \kappa_1(m_1 - R_t^{(1)})dt + \sigma_1 dw_t^{(1)}, \\ dR_t^{(2)} &= \kappa_2(m_2 - R_t^{(2)})dt + \sigma_2 dw_t^{(2)}, \\ &\dots \\ dR_t^{(N)} &= \kappa_N(m_N - R_t^{(N)})dt + \sigma_N dw_t^{(N)}. \end{aligned}$$

Here, $w_t^{(i)}, i = 1, \dots, N$, are mutually independent Wiener processes. The process R_t represents an input rate model in which fluid is poured into the buffer from N independent information sources with different input rates. The solution is given by

$$R_t = m + \sum_{i=1}^N \int_{-\infty}^t \sigma_i e^{-\kappa_i(t-s)} dw_s^{(i)},$$

where $m = \sum_{i=1}^N m_i$. Since $w_s^{(i)}, i = 1, \dots, N$, are mutually independent Wiener processes, the autocovariance function is given by

$$\Lambda_R(t) = \sum_{i=1}^N \frac{\sigma_i^2}{2\kappa_i} e^{-\kappa_i|t|}, \quad (2.20)$$

and the spectral-density function by

$$\mathcal{L}_R(\omega) = \sum_{i=1}^N \frac{\sigma_i^2}{2\kappa_i} \cdot \frac{1}{\kappa_i^2 + \omega^2}.$$

From (2.10), the Fourier transforms $\mathcal{L}_R(\omega)$ and $\mathcal{H}(i\omega)$ of $\Lambda_R(t)$ and $h(t)$ respectively are related as

$$\mathcal{L}_R(\omega) = |\mathcal{H}(i\omega)|^2. \quad (2.21)$$

Hence, we can easily see that $h(t)$ takes the form

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

with some unknown constants α_i . If we want to know the values of α_i 's, we have to solve a polynomial equation. It is clear that $h(t)$ satisfies conditions (2.6) and (2.7). From (2.15) and (2.20), we have

$$H(0)^2 = \sum_{i=1}^N \frac{\sigma_i^2}{\kappa_i^2}.$$

Consequently, we obtain the approximation

$$A_t \sim m t + \sqrt{\sum_{i=1}^N \frac{\sigma_i^2}{\kappa_i^2}} \tilde{w}_t.$$

We notice that the approximation of A_t has been obtained without representing $h(t)$ explicitly, or equivalently without calculating constants α_i 's.

Example 2.3. *Linear differential equation model*

The input rate process for this model is defined as

$$R_t = m + \tilde{R}_t,$$

where $\tilde{R}_t$ is determined from the stable linear differential equation

$$a_n \frac{d^n}{dt^n} \tilde{R}_t + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} \tilde{R}_t + \cdots + a_0 \tilde{R}_t = v_t, \quad (2.22)$$

denoting by v_t a white noise, namely, $v_t = dw_t/dt$.

This is a continuous version of a linear autoregressive process which is used for representing the input rate process in simulation [25, 14]. Then R_t is written as

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

where $h(t)$ is the inverse Laplace transform of

$$\mathcal{H}(p) = \frac{1}{a_n p^n + a_{n-1} p^{n-1} + \cdots + a_0}.$$

If the differential equation of (2.22) is stable, $h(t)$ is represented in the form (2.18) and satisfies conditions (2.6) and (2.7). The Fourier transform of $h(t)$ is given by $\mathcal{H}(i\omega)$, and from (2.21), $\Lambda_R(t)$ is given by the the inverse Fourier transform of $|\mathcal{H}(i\omega)|^2$. Then from (2.14) and (2.15) we have

$$A_t \sim m t + \sqrt{2 \int_0^\infty \Lambda_R(s) ds} \cdot \tilde{w}_t. \quad (2.23)$$

In case of continuous version of autoregressive moving average processes, $v_t = dw/dt$ in (2.22) is replaced with

$$v_t = \int_{t-T}^t g(t-s) dw_s,$$

where T is a positive constant and $g(t)$ is a function satisfying (2.17). Then, R_t is given by

$$\begin{aligned} R_t &= m + \int_{-\infty}^t \int_{\tau-T}^{\tau} h(t-\tau) g(\tau-s) dw_s d\tau \\ &= m + \int_{-\infty}^t \hat{h}(t-s) dw_s, \end{aligned}$$

where

$$\hat{h}(t) = \int_0^t h(t-s)g(s)ds.$$

Since $\hat{h}(t)$ is a function satisfying (2.19), conditions (2.6) and (2.7) are satisfied. At this time, $\Lambda_R(t)$ in (2.23) is the inverse Fourier transform of $|\mathcal{H}(i\omega)\mathcal{G}(i\omega)|^2$ where $\mathcal{G}(i\omega)$ is the Fourier transform of $g(t)$.

Example 2.4. *Estimate from measured autocovariance function*

We consider a situation where the mean m and the autocovariance function $\Lambda_R(t)$ are given in some way or are estimated from an actual measurement of the input rate process R_t . In such a case, even if no other information on the structure of R_t is available, we can construct an approximate Gaussian process as usual;

$$A_t \sim m t + \sqrt{2 \int_0^\infty \Lambda_R(s)ds} \cdot \tilde{w}_t.$$

Hence the approximate formula for the buffer content derived in the next section is also available.

In some cases, however, it might be necessary to check whether or not there exists a function $h(t)$ such that $h(t) = 0$ for $t < 0$ and that $h(t)$ satisfies conditions (2.6) and (2.7). The author thinks that such a function does exist in most cases for coded video traffic since the input rate process usually behaves like an autoregressive process or an autoregressive moving average process discussed in Example 2.3.

Example 2.5. *An On-Off model with multiple types of traffic*

In this model, there are N_k On-Off sources of type k , $k = 1, 2, \dots, K$, which independently and asynchronously alternate between On-states and Off-states. Let On- and Off-periods of type k sources be exponentially distributed with parameters μ_k and λ_k , respectively;

$$\begin{aligned} \mu_k^{-1} &= \text{the mean length of an On-period of a type } k \text{ source,} \\ \lambda_k^{-1} &= \text{the mean length of an Off-period of a type } k \text{ source,} \end{aligned}$$

and r_k be the input rate of a type k source at the On-state. H. Kobayashi and Q. Ren [16] approximate this input rate process R_t by a superposition of Ornstein-Uhlenbeck processes which is discussed in Example 2.2.

Let $R_t^{(k)}$ be the aggregated input rate process of type k sources and $A_t^{(k)}$ be the corresponding input process. Then $R_t^{(k)}$ is a birth-and-death process on $\{0, 1, \dots, N_k\}$. As N_k tends to infinity and r_k tends to 0 in such a way that $N_k r_k$ is constant, the birth-and-death process converges to an Ornstein-Uhlenbeck process. In the sense, $R_t^{(k)}$ can be approximated by an Ornstein-Uhlenbeck process discussed in Example 2.1. The mean m_k and the autocovariance function $\Lambda_k(t)$ of $R_t^{(k)}$ are given by

$$m_k = \frac{N_k r_k \lambda_k}{\lambda_k + \mu_k} \quad \text{and} \quad \Lambda_k(t) = \sigma_k^2 e^{-(\lambda_k + \mu_k)|t|},$$

where

$$\sigma_k^2 = \frac{N_k r_k^2 \lambda_k \mu_k}{(\lambda_k + \mu_k)^2}.$$

By using formula (2.15), $A_t^{(k)}$ is approximated as

$$A_t^{(k)} \sim \frac{N_k r_k \lambda_k}{\lambda_k + \mu_k} t + \sqrt{\frac{2 N_k r_k^2 \lambda_k \mu_k}{(\lambda_k + \mu_k)^3}} d\tilde{w}_t.$$

Since the total input rate process A_t is given by $A_t = \sum_{k=1}^K A_t^{(k)}$, we have

$$A_t \sim \sum_{k=1}^K \frac{N_k r_k \lambda_k}{\lambda_k + \mu_k} t + \sqrt{\sum_{k=1}^K \frac{2 N_k r_k^2 \lambda_k \mu_k}{(\lambda_k + \mu_k)^3}} d\tilde{w}_t.$$

Example 2.6. A generalized On-Off model

In H. Kobayashi and Q. Ren [16], the input rate process is approximated by a superposition of Ornstein-Uhlenbeck processes under the condition that $N_k, k = 1, \dots, K$, are sufficiently large and that Off- and On-periods are exponentially distributed. We here consider an input rate process given as a superposition of many On-Off types of input rate processes with possibly different parameters and different On- and Off-period distributions. Although we cannot obtain $h(t)$ explicitly, we may derive an approximation of the input process A_t .

Let there be N mutually independent sources numbered 1 to N , and for source i we denote by r_i the input rate at On-state and by $R_t^{(i)}$ the On-Off type input rate process. Let $X_l^{(i)}$ and $Y_l^{(i)}$ be the lengths of the l -th Off- and On-periods for source i . Then $\{X_l^{(i)}, Y_l^{(i)}\}$ forms an alternate renewal process. We assume that the distributions of On- and Off-periods have finite means and variances:

$$\begin{aligned} 1/\lambda_i = E[X_l^{(i)}] < \infty, & \quad 1/\mu_i = E[Y_l^{(i)}] < \infty, \\ \xi_i^2 = Var[X_l^{(i)}] < \infty, & \quad \zeta_i^2 = Var[Y_l^{(i)}] < \infty. \end{aligned}$$

In order to avoid nuisance discussion, we also assume that $\xi_i^2 > 0$ and $\zeta_i^2 > 0$, and r_i 's are all sufficiently small.

If N is large enough, R_t is approximated by a stationary Gaussian process having the stationary distribution $N(m, \sigma_R^2)$ with

$$m = \sum_{i=1}^N \frac{r_i \lambda_i}{\lambda_i + \mu_i}, \quad \sigma_R^2 = \sum_{i=1}^N \frac{r_i^2 \lambda_i \mu_i}{(\lambda_i + \mu_i)^2}. \quad (2.24)$$

Hence, although $h(t)$ in (2.5) can't be given explicitly, $h(t)$ exists and is expected to satisfy conditions (2.6) and (2.7) on the analogy of the discussion in Examples 2.4 and 2.5. Then, we may use the approximation of A_t .

Now let us consider the individual input process. For simplicity, we assume that the input rate process $R_t^{(i)}$ starts from Off-state. Then, the input process $A_t^{(i)}$ of source i , namely, $A_t^{(i)} = \int_0^t R^{(i)}(u) du$, satisfies the following inequalities:

$$r_i \sum_{l=1}^{L_t^{(i)}} Y_l^{(i)} \leq A_t^{(i)} < r_i \sum_{l=1}^{L_t^{(i)}+1} Y_l^{(i)}, \quad (2.25)$$

where

$$L_t^{(i)} = \max \left[L \mid \sum_{l=1}^L \{X_l^{(i)} + Y_l^{(i)}\} \leq t \right].$$

Since the ratio between both extreme sides in (2.25) approaches 1 as t becomes large, we may use $r_i \sum_{l=1}^{L_t^{(i)}} Y_l^{(i)}$ as an approximation of $A_t^{(i)}$ for large t . Since $Y_l^{(i)}$, $l = 1, 2, \dots$, are independent and identically distributed random variables, $\sum_{l=1}^{L_t^{(i)}} Y_l^{(i)}$ is

approximately normally distributed with the mean

$$\begin{aligned} E\left[\sum_{l=1}^{L_t^{(i)}} Y_l^{(i)}\right] &= \sum_{n=1}^{\infty} E\left[\sum_{l=1}^{L_t^{(i)}} Y_l^{(i)} \mid L_t^{(i)} = n\right] P(L_t^{(i)} = n) \\ &\sim \frac{\lambda_i}{\lambda_i + \mu_i} t, \end{aligned}$$

and the variance

$$\begin{aligned} Var\left[\sum_{l=1}^{L_t^{(i)}} Y_l^{(i)}\right] &= \sum_{n=1}^{\infty} E\left[\sum_{l=1}^{L_t^{(i)}} (Y_l^{(i)})^2 \mid L_t^{(i)} = n\right] P(L_t^{(i)} = n) - E\left[\sum_{l=1}^{L_t^{(i)}} Y_l^{(i)}\right]^2 \\ &\sim \frac{\lambda_i \mu_i (D_i^2 + E_i^2)}{(\lambda_i + \mu_i)^3} t, \end{aligned}$$

where D_i, E_i are *coefficients of variations* defined by $D_i = \xi_i \lambda_i$ and $E_i = \zeta_i \mu_i$. Then, we have

$$E[A_t^{(i)}] \sim \frac{r_i \lambda_i}{\lambda_i + \mu_i} t, \quad Var[A_t^{(i)}] \sim \frac{r_i^2 \lambda_i \mu_i (D_i^2 + E_i^2)}{(\lambda_i + \mu_i)^3} t.$$

Since $A_t^{(i)}, i = 1, \dots, N$, are assumed to be mutually independent, the distribution of the total input process $A_t = \sum_{i=1}^N A_t^{(i)}$ for large t is approximated by a normal distribution with the mean and the variance given by

$$\begin{aligned} E[A_t] &\sim \sum_{i=1}^N \frac{r_i \lambda_i}{\lambda_i + \mu_i} t, \\ Var[A_t] &\sim \sum_{i=1}^N \frac{r_i^2 \lambda_i \mu_i (D_i^2 + E_i^2)}{(\lambda_i + \mu_i)^3} t \end{aligned}$$

Consequently, we have

$$A_t \sim \sum_{i=1}^N \frac{r_i \lambda_i}{\lambda_i + \mu_i} t + \sqrt{\sum_{i=1}^N \frac{r_i^2 \lambda_i \mu_i (D_i^2 + E_i^2)}{(\lambda_i + \mu_i)^3}} dw_t. \quad (2.26)$$

2.4 A diffusion approximation for the fluid queue model

If we want to get an exact solution for the steady-state distribution of the fluid queue model with a Gaussian input rate process, we have to deal with a two-dimensional stochastic process $(Q_t, R_t)_{t \geq 0}$ on $\mathbf{R}[0, \infty) \times \mathbf{R}(-\infty, \infty)$. However, it seems difficult and practically impossible to do this even for the simplest case in which the variable input rate process is given as an Ornstein-Uhlenbeck process [34]. In this section, then, we shall approximate the process Q_t on $\mathbf{R}[0, \infty)$ by a one-dimensional diffusion process x_t defined on $\mathbf{R}(-\infty, \infty)$. Namely we intend to reduce the two-dimensional problem into a one-dimensional problem by considering the marginal process Q_t with diffusion approximation. Notice that Q_t may not be a Markov process any longer.

The approximating process x_t has no reflection barrier and is allowed to take negative values. We consider $Q_t = 0$ when $x_t \leq 0$ and the probability that $Q_t = 0$ is approximated by the probability that $x_t \leq 0$.

We define the one-dimensional diffusion process x_t by a stochastic differential equation in the form

$$dx_t = a(x_t) dt + \sqrt{b(x_t)} dw_t, \quad (2.27)$$

where w_t is a standard Wiener process and the coefficient of drift $a(x)$ and the coefficient of diffusion $b(x)$ are dependent on the position x at the moment [38]. The stochastic differential equation above should be interpreted as

$$x_t = x_0 + \int_0^t a(x_s) ds + \int_0^t \sqrt{b(x_s)} dw_s,$$

where the second integral is understood as Ito's integral.

Let A_t be the *input process* to the buffer and V_t be the *output process* from the buffer. Then the buffer content Q_t is written as

$$Q_t = Q_0 + A_t - V_t, \quad (2.28)$$

and from (2.1) A_t and V_t are represented in terms of the input rate process R_t as

$$A_t = \int_0^t R_s ds, \quad V_t = \int_0^t U_s ds, \quad (2.29)$$

where

$$U_t = \begin{cases} C, & \text{if } Q_t > 0 \text{ or } R_t \geq C, \\ R_t, & \text{if } Q_t = 0 \text{ and } R_t < C. \end{cases} \quad (2.30)$$

Now we approximate A_t and V_t by one-dimensional diffusion processes separately.

As was seen in Section 2.2, the input process A_t is approximated by

$$dy_t = a dt + \sqrt{b} dw_t, \quad (2.31)$$

for large t , where $a = m$ and $b = 2 \int_0^\infty \Lambda_R(s) ds$. This approximation is valid only for large t , but we shall use the approximation for all $t > 0$.

For the approximation of V_t , we have to take the value of Q_t into consideration.

Denoting by $I(\cdot)$ the identity function, from (2.30), V_t is written as

$$\begin{aligned} V_t &= \int_0^t I(Q_s > 0 \text{ or } R_s \geq C) C ds + \int_0^t I(Q_s = 0 \text{ and } R_s < C) R_s ds \\ &= \int_0^t I(Q_s > 0) C ds + \int_0^t I(Q_s = 0 \text{ and } R_s \geq C) C ds \\ &\quad + \int_0^t I(Q_s = 0 \text{ and } R_s < C) R_s ds. \end{aligned}$$

Since $Q > 0$ whenever $R_t > C$,

$$I(Q_s = 0 \text{ and } R_s \geq C) C + I(Q_s = 0 \text{ and } R_s < C) R_s = I(Q_s = 0) R_s.$$

Hence V_t is represented as

$$V_t = \int_0^t I(Q_s > 0) C ds + \int_0^t I(Q_s = 0) R_s ds. \quad (2.32)$$

In order to evaluate the second term, we shall approximate $R_t dt$ on $\{Q_t = 0\}$ by a one-dimensional diffusion process

$$\eta dt + \sqrt{b'} d\hat{w}_t, \quad (2.33)$$

where $\hat{w}_t$ is a standard Wiener process. Then by replacing $I(Q_t = 0)$ with $I(x_t \leq 0)$ and $I(Q_t > 0)$ with $I(x_t > 0)$ in (2.32), we have an approximating process z_t of V_t as

$$z_t = \int_0^t I(x_s > 0) C ds + \int_0^t I(x_s \leq 0) d(\eta s + \sqrt{b'} \hat{w}_s). \quad (2.34)$$

The parameters η and b' in (2.33) might have to be estimated. From the fact that $R_t < C$ whenever $Q_t = 0$ and that the stationary distribution of R_t is normal distribution $N(m, \sigma_R^2)$, it is reasonable to estimate the mean η by the conditional expectation

$$\eta = E[R | R < C], \quad (2.35)$$

where R is a generic random variable subjecting to $N(m, \sigma_R^2)$. On the other hand, the value of b' need not be estimated here because it is not relevant to the approximation of the steady-state probabilities of Q_t as will be discussed later.

From (2.31) and (2.34) the approximating process x_t of Q_t is written as

$$\begin{aligned} x_t &= x_0 + y_t - z_t \\ &= x_0 + \int_0^t d(a s + \sqrt{b} w_s) \\ &\quad - \int_0^t I(x_s > 0) C ds - \int_0^t I(x_s \leq 0) d(\eta s + \sqrt{b'} \hat{w}_s) \\ &= x_0 + \int_0^t I(x_s > 0) d\{(a - C)s + \sqrt{b} w_s\} \\ &\quad + \int_0^t I(x_s \leq 0) d\{(a - \eta)s + \sqrt{\beta} \tilde{w}_s\}, \end{aligned}$$

assuming that $\sqrt{b} dw_t + \sqrt{b'} d\hat{w}_t$ is represented as $\sqrt{\beta} d\tilde{w}_t$ with a standard Wiener process $\tilde{w}_t$. Since the sets $\{x_s > 0\}$ and $\{x_s \leq 0\}$ are disjoint, we may use the same standard Wiener process in both of the last two integrals in the above equation. Thus we have

$$\begin{aligned} x_t &= x_0 + \int_0^t I(x_s > 0) d\{(a - C)s + \sqrt{b} w_s\} \\ &\quad + \int_0^t I(x_s \leq 0) d\{(a - \eta)s + \sqrt{\beta} w_s\}. \end{aligned} \quad (2.36)$$

Rewriting the above equation in the form of (2.27), we have

$$dx_t = \begin{cases} (a - C) dt + \sqrt{b} dw_t, & x_t > 0, \\ (a - \eta) dt + \sqrt{\beta} dw_t, & x_t \leq 0. \end{cases} \quad (2.37)$$

This stochastic differential equation is characterized by the jump drift coefficient

$$a(x) = \begin{cases} a - C, & x_t > 0, \\ a - \eta, & x_t \leq 0, \end{cases}$$

and the jump diffusion coefficient

$$b(x) = \begin{cases} b, & x_t > 0, \\ \beta, & x_t \leq 0. \end{cases}$$

The approximating process x_t has been derived in several steps of approximations. Therefore theoretically there is no guarantee that the process x_t simulates the behavior of the original process Q_t well when Q_t is small. However, we may expect that it approximates Q_t well when Q_t is large for the following reasons. Suppose that $Q_{t_0} > 0$ and let τ be the last departure time from level 0 of the process Q_t before time t_0 . During the time interval $(\tau, t_0]$, from (2.30), $Q_t > 0$ and $U_t = C$. Thus from (2.28) we have

$$Q_t = A_t - A_\tau + C(t - \tau), \quad \text{for } t \in (\tau, t_0].$$

In our approximation scheme, if we assume that $x_\tau = 0$ and $x_t > 0$ for $t \in (\tau, t_0]$, then from (2.37) we have

$$x_t = y_t - y_\tau + C(t - \tau), \quad \text{for } t \in (\tau, t_0].$$

We see that here we just approximate $A_t - A_\tau$ by $y_t - y_\tau$. This approximation is justified from (2.14) for large $t - \tau$. If Q_{t_0} , or its approximation x_{t_0} , is large, then $t_0 - \tau$ would be large, and we may expect that the behavior of Q_{t_0} is well approximated by that of x_{t_0} .

Stationary solution for the fluid queue model

We now proceed to derive the steady-state distribution of the diffusion process x_t described by the stochastic differential equation (2.37) under the conditions $b > 0$, $\beta > 0$, $a - C < 0$ and $a - \eta < 0$. The equation implies that the behavior of the diffusion process differs between the region $x > 0$ and the region $x < 0$. Then it is necessary to joint the stationary solutions in these two regions by setting up a suitable boundary condition at $x = 0$. We will show that an appropriate boundary condition at $x = 0$ is $bp(+0) = \beta p(-0)$, where $p(x)$ is the stationary density function of x_t . This implies that $p(x)$ is discontinuous at $x = 0$ in case of $b \neq \beta$.

In order to discuss the boundary condition and the stationary distribution, we shall approximate the process x_t by a sequence of diffusion processes with smooth drift and diffusion coefficients. Let $a^{(n)}(x)$ and $b^{(n)}(x)$ be monotone and sufficiently smooth functions such that $a^{(n)}(x) = a(x)$ and $b^{(n)}(x) = b(x)$ for x outside the interval $(-n^{-1}, n^{-1})$. Then the stochastic differential equation

$$dx_t^{(n)} = a^{(n)}(x_t^{(n)}) dt + \sqrt{b^{(n)}(x_t^{(n)})} dw_t$$

has a unique solution $x_t^{(n)}$ [38]. Since $a^{(n)}(x) \rightarrow a(x)$ and $b^{(n)}(x) \rightarrow b(x)$ as $n \rightarrow \infty$ except for $x = 0$, the sequence of stochastic processes $\{x_t^{(n)}\}$ converges to a limiting process x_t . It is clear that x_t satisfies the original stochastic differential equation (2.37). So, we shall take this limit x_t as our solution of the equation (2.37).

Let $p^{(n)}(x, t)$ be the probability density function of $x_t^{(n)}$ at time t . Then it satisfies the Fokker-Planck equation [38]

$$\frac{\partial p^{(n)}(x, t)}{\partial t} = -\frac{\partial}{\partial x} a^{(n)}(x) p^{(n)}(x, t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} b^{(n)}(x) p^{(n)}(x, t).$$

Clearly the distribution of $x_t^{(n)}$ at t converges to the stationary distribution as $t \rightarrow \infty$. Let us denote by $p^{(n)}(x)$ the stationary probability density function. Then, as $t \rightarrow \infty$

$$p^{(n)}(x, t) \rightarrow p^{(n)}(x), \quad \frac{\partial p^{(n)}(x, t)}{\partial t} \rightarrow 0,$$

and as $x \rightarrow \pm\infty$

$$p^{(n)}(x) \rightarrow 0, \quad \frac{\partial p^{(n)}(x)}{\partial x} \rightarrow 0.$$

Hence $p^{(n)}(x)$ satisfies the differential equation

$$-a^{(n)}(x)p^{(n)}(x) + \frac{b^{(n)}(x)}{2} \frac{d}{dx} p^{(n)}(x) = 0,$$

and by solving the equation we have

$$p^{(n)}(x) = \frac{A^{(n)}}{b^{(n)}(x)} \exp\left[2 \int_0^x \frac{a^{(n)}(y)}{b^{(n)}(y)} dy\right].$$

The multiplicative constant $A^{(n)}$ is determined from the total probability condition, namely, $\int_{-\infty}^{\infty} p^{(n)}(x) dx = 1$ as

$$(A^{(n)})^{-1} = \frac{1}{b^{(n)}(x)} \int_{-\infty}^{\infty} \exp\left[2 \int_0^x \frac{a^{(n)}(y)}{b^{(n)}(y)} dy\right] dx.$$

Since $P[x_t^{(n)} \leq x] \rightarrow P[x_t \leq x]$ as $n \rightarrow \infty$, the stationary probability density function $p(x)$ of x_t is given by

$$\begin{aligned} p(x) &= \lim_{n \rightarrow \infty} p^{(n)}(x) \\ &= \begin{cases} p_+(x), & x > 0, \\ p_-(x), & x \leq 0, \end{cases} \end{aligned}$$

where

$$\begin{aligned} p_+(x) &= \frac{A}{b} \exp\left[2 \frac{a-C}{b} x\right], \\ p_-(x) &= \frac{A}{\beta} \exp\left[2 \frac{a-\eta}{\beta} x\right]. \end{aligned} \tag{2.38}$$

The multiplicative constant A is given by

$$A^{-1} = \frac{1}{\beta} \int_{-\infty}^0 \exp\left[2 \frac{a-\eta}{\beta} x\right] dx + \frac{1}{b} \int_0^{\infty} \exp\left[2 \frac{a-C}{b} x\right] dx = \frac{C-\eta}{2(a-\eta)(C-a)}.$$

From this solution, we see that $p(x)$ is independent of β on $x > 0$ and that the boundary condition at $x = 0$ should be set as $b p(+0) = \beta p(-0)$.

This stationary distribution of x_t provides an approximation of the stationary distribution of Q_t . Let us denote by Q a generic random variable subjected to the stationary distribution of Q_t . Then from the above result, $P[Q > x]$ for $x > 0$ is approximated as

$$\begin{aligned} P[Q > x] &\approx \int_x^\infty p(u) du \\ &= B \exp\left[-2\frac{C-a}{b}x\right] \end{aligned} \quad (2.39)$$

and the mass at $x = 0$ is approximated as $P[Q = 0] \approx 1 - B$, where

$$B = \frac{a - \eta}{C - \eta}.$$

From (2.35), η can be calculated as

$$\begin{aligned} \eta &= \int_{-\infty}^C \frac{r}{\sqrt{2\pi}\sigma_R} e^{-\frac{(r-a)^2}{2\sigma_R^2}} dr / \int_{-\infty}^C \frac{1}{\sqrt{2\pi}\sigma_R} e^{-\frac{(r-a)^2}{2\sigma_R^2}} dr \\ &= \frac{a\Phi(\theta) - \sigma_R\phi(\theta)}{\Phi(\theta)}, \end{aligned}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $\Phi(x) = \int_{-\infty}^x \phi(u)du$ and $\theta = (C - a)/\sigma_R$. Consequently, B is also written as

$$B = \frac{\phi(\theta)}{\theta\Phi(\theta) + \phi(\theta)}. \quad (2.40)$$

Remark 2.2. A. Simonian [34] obtains a similar approximation formula for the fluid queue model with input rate varying as the Ornstein-Uhlenbeck process:

$$P[Q > x] \approx B_S \exp\left[-2\frac{C-a}{b}x\right],$$

where

$$B_S = \frac{e^{-\theta^2}}{\theta\sqrt{2\pi}}, \quad \theta = \frac{C-a}{\sigma_R}. \quad (2.41)$$

H. Kobayashi and Q. Ren [16] also propose an approximation formula for the fluid queue model with multiple types of On-Off input traffic:

$$P[Q > x] \approx B_K \exp\left[-2\frac{C-a}{b}x\right],$$

where

$$B_K = \frac{1}{\theta^* \sqrt{2\pi\sigma_R^2}} e^{-\frac{\sigma_R^2}{2}\theta^{*2}}, \quad \theta^* = \frac{C-a}{\sigma_R^2}.$$

Changing the above equation into a comparable form, B_K is given by

$$B_K = \frac{1}{\theta \sqrt{2\pi}} e^{-\frac{\theta^2}{2}} = \frac{\phi(\theta)}{\theta}, \quad \theta = \frac{C-a}{\sigma_R}. \quad (2.42)$$

We notice that these approximation formulas including ours are of a common form:

$$\tilde{B} \exp[-2\frac{C-a}{b}x].$$

This is not a coincidence, but is the consequence of the fact that the input process A_t can be approximated by $at + \sqrt{b}w_t$ for large t . With respect to the coefficient $\tilde{B}$, however, they are different. We will evaluate them in the next section by a numerical experiment.

Remark 2.3. The same approximation for $P[Q > 0]$ as in (2.39) can be derived without using the Wiener process approximation in the negative region. Consider R_t and Q_t in a sufficiently long time interval $[0, T]$, and divide the interval into three sets according to their values so that

$$\begin{aligned} \mathcal{T}^+ &= \{t \in [0, T] : R_t > C\}, \\ \mathcal{T}^{-+} &= \{t \in [0, T] : R_t \leq C \text{ and } Q_t > 0\}, \\ \mathcal{T}^{-0} &= \{t \in [0, T] : R_t \leq C \text{ and } Q_t = 0\}. \end{aligned}$$

Note that $Q_t > 0$ on $\mathcal{T}^+$, and that $\{t : Q_t > 0\} = \mathcal{T}^+ \cup \mathcal{T}^{-+}$. The total variation of Q_t in $[0, T]$ is written as

$$\begin{aligned} Q_T - Q_0 &= \int_{\{t: Q_t > 0\}} (R_t - C) dt \\ &= \int_{\mathcal{T}^+} (R_t - C) dt - \int_{\mathcal{T}^{-+}} (C - R_t) dt. \end{aligned}$$

Compared with two integrals in the right hand side, $Q_T - Q_0$ is negligibly small. So we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{\mathcal{T}^+} (R_t - C) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{\mathcal{T}^{-+}} (C - R_t) dt. \quad (2.43)$$

We shall evaluate these two limits separately. Since R_t is ergodic, the Borel measure of the set $\mathcal{T}^+$ (i.e. the sum of lengths of intervals composing $\mathcal{T}^+$) is almost equal to $P[R > C] T$ and the average of R_t on $\mathcal{T}^+$ is almost equal to $E[R | R > C]$, where R is a random variable subjecting to the stationary distribution $N(m, \sigma_R^2)$ of R_t . Hence, if we write $\alpha^+ = P[R > C]$ and $\beta^+ = E[R | R > C]$, the limit in the left hand side of (2.43) is given by $(\beta^+ - C)\alpha^+$.

In order to evaluate the right hand side of (2.43) we need the average of R_t on $\mathcal{T}^{-+}$. This value is difficult to obtain because the basic processes R_t and Q_t defining $\mathcal{T}^{-+}$ are dependent. The average of R_t on $\mathcal{T}^{-+} \cup \mathcal{T}^{-0}$ is equal to $\beta^- = E[R | R < C]$ for large T . We assume that the average of R_t on $\mathcal{T}^{-+}$ is approximated by β^- . This is equivalent to assume that the average of R_t on $\mathcal{T}^{-0}$ is approximated by β^- . The same assumption was used to derive our approximate Wiener process in (2.37). Let α^{-+} be the ratio of the Borel measure of the set $\mathcal{T}^{-+}$ to T . Then the limit in the right hand side of (2.43) is approximated by $(C - \beta^-)\alpha^{-+}$.

By equating these evaluations of limits in (2.43), we have

$$\alpha^{-+} \approx \frac{\beta^+ - C}{C - \beta^-} \alpha^+.$$

Hence the stationary probability that the buffer is positive is given by

$$P[Q > 0] \approx \alpha^+ + \alpha^{-+} = \frac{\beta^+ - \beta^-}{C - \beta^-} \alpha^+.$$

Since $\alpha^+ \beta^+ + (1 - \alpha^+) \beta^- = a$ and $\beta^- = \eta$, the right hand side is rewritten as $(a - \eta)/(C - \eta)$. This leads to the approximation formula (2.39).

2.5 Numerical results and discussions

Exponentially distributed On-Off model

To check the accuracy of our approximation, we will compare the approximate values with simulation results. The model is the On-Off model with multiple types of traffic, discussed in H. Kobayashi and Q. Ren [16]. There are two types of sources with exponential distributions for On- and Off-periods. The parameters are as follows.

- Type 1 sources: $\lambda_1 = 0.4$, $\mu_1 = 1.5$, $r_1 = 2$,
- Type 2 sources: $\lambda_2 = 0.6$, $\mu_2 = 0.75$, $r_2 = 1$,
- The distributions of periods of On-state and Off-state are exponential, namely, $D_i = E_i = 1$ for $i = 1, 2$,
- The number of type 1 sources: $N_1 = 25$,
- The number of type 2 sources: $N_2 = 50$, and
- $C = 38$.

We also test the cases in which $N_1 = 18$, $N_2 = 36$ and $N_1 = 24$, $N_2 = 48$.

Figure 2.1 shows approximate values and simulation results for the tail probability $P[Q > x]$. The approximate values are very close to the simulation results. The approximation seems to overestimate the tail probability except for a very large x . However, as indicated in Table 1, it underestimates the probability $P[Q > 0]$ at $x = 0$. We cannot recognize this from Figure 2, but if we look closer, the curves of the approximation and the simulation intersect near $x = 0$.

Table 2.1 compares the coefficient B in (2.40) with B_S and B_K in Remark 2. First we see that B_S is not accurate. A. Simonian says in his paper [34] that his result for low load significantly differs from the result reported by D. Anick [2].

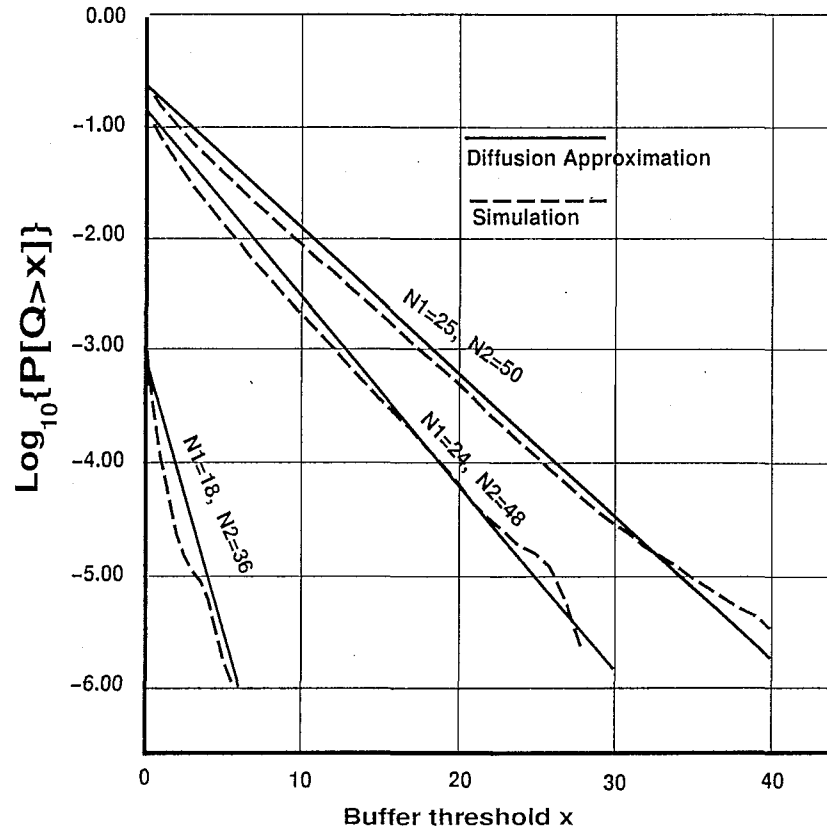


Figure 2.1: Tail probability $P[Q > x]$ of exponentially distributed On-Off model

Number of sources	θ	$\log_{10}\{P[Q > 0]\}$			
		Simulation	$\log\{B\}$	$\log\{B_K\}$	$\log\{B_S\}$
$N_1 = 25, N_2 = 50$	0.976	-0.615	-0.631	-0.595	-0.802
$N_1 = 24, N_2 = 48$	1.244	-0.756	-0.848	-0.830	-1.166
$N_1 = 18, N_2 = 36$	3.158	-2.821	-3.063	-3.064	-5.229

Table 2.1: Comparison of approximations for $P[Q > 0]$

Next we see that B and B_K provide values very close to each other. Since

$$B = \frac{\phi(\theta)}{\theta\Phi(\theta) + \phi(\theta)} \quad \text{and} \quad B_K = \frac{\phi(\theta)}{\theta},$$

$B_K \approx B$ for large θ . Figure 2.2 shows the graphs for B and B_K for various θ . We see that B_K is very close to B over wide range. However, B_K exceeds 1 significantly as θ becomes small. Since B_K is an approximation of the probability $P[Q > 0]$, this is an undesirable property. On the other hand, B approaches 1 as $\theta \rightarrow 0$. The author thinks that B gives a more preferable approximation.

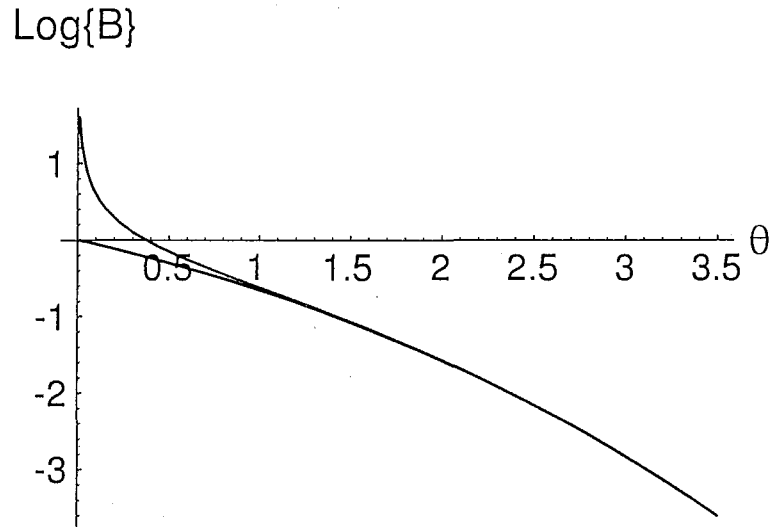


Figure 2.2: Comparison between B and B_K : The lower curve is B and the upper curve is B_K .

Non-exponentially distributed On-Off model

We also test our approximation for the On-Off model with phase-type On-period distributions. The model is the same as the previous one except for the On-period distributions are changed to Erlangian or hyper-exponential distributions. Figure 4 compares our approximation with simulation results. E in the figure indicates a common coefficient of variation of the On-period distributions for type 1 and type 2 sources.

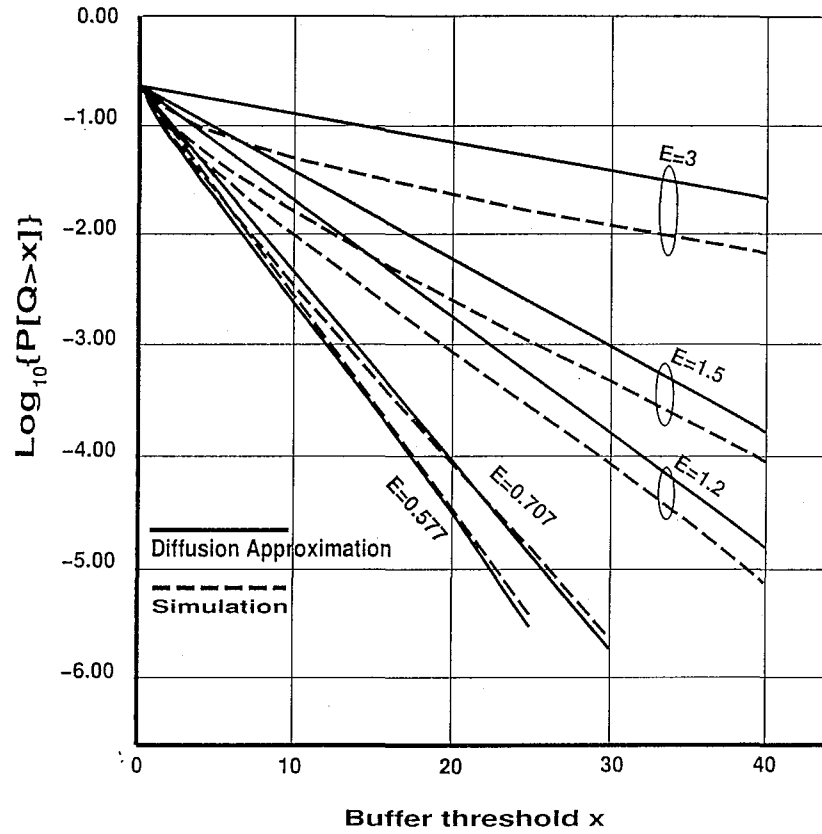


Figure 2.3: Tail probability $P[Q > x]$ of non-exponentially distributed On-Off model

The lines with $E = 0.707$ and 0.557 are for the cases with Erlangian distributions of phase 2 and phase 3. In these cases our approximation is very good. They seem to show the accuracy of our approximation in case of $E \leq 1$.

The lines with $E = 1.2, 1.5$ and 3 are for the cases with balanced mean hyper-exponential distributions with the indicated coefficient of variation. In these cases our approximation overestimates the tail probabilities, though the decay rate (the slope of the logarithm of the tail probability) and the probability of positive buffer content are very close to the results of the simulation. Our approximation was derived from two main assumptions that $A_t \sim a dt + \sqrt{b} dw_t$ for all t and that $I(Q_t = 0)R_t dt \sim E[R|R < C]dt + \sqrt{b'} dw_t$. The decay rate of $P[Q > x]$ is related to the former for large x and $P[Q > 0]$ to the latter. Hence the author thinks that the discrepancy in these cases comes from the assumption $A_t \sim a dt + \sqrt{b} dw_t$ near $x = 0$.

Chapter 3

Tail behavior of stationary distribution

3.1 Introduction

In the previous chapter, we introduce *Gaussian-type input rate processes* 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. Moreover, applying a diffusion approximation to a fluid queue model with a Gaussian-type input rate process, we derive an approximation formula

$$P[Q_t > x] \approx B \exp\left[-\frac{C-m}{H(0)^2/2}x\right] \quad (3.1)$$

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, diffusion approximations are said to require heavy traffic assumptions in general. That is, there is no guarantees except for heavy traffic.

In this chapter, we give a theoretical guarantee for the decay rate of the approx-

imation formula (3.1). Specifically we prove that for sufficiently large x ,

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

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

This chapter is organized as follows. In Section 3.2, we show a relation between input rate processes proposed so far for coded video traffic and Gaussian-type input rate processes, and discuss a fluid queue model with a Gaussian-type input rate process. Section 3.3 is devoted to our main theorem which gives an upper bound and a lower bound for the tail probability of the buffer content distribution.

3.2 A fluid queue model with a Gaussian-type input rate process

3.2.1 Coded video traffic

As noted in Chapter 1, there are some simulation studies in which coded video traffic is modeled as a discrete time Gaussian process. In [25, 14], coded 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 ε_n 's are mutually independent and identically distributed random variables with zero mean. The random variables ε_n 's are usually assumed to be normally distributed. 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 [10], 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. It is proposed in [10] to use a zero-memory nonlinearity (ZMNL), say $\max(0, X_n)$. If many video traffics are multiplexed, then the probability that X_n is negative becomes negligibly small since the standard deviation of X_n becomes relatively small compared with the mean of X_n . 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

$$X_n = m + \sum_{i=-\infty}^n h_{n-i} \varepsilon_i. \quad (3.2)$$

In [30], 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 $R_t = m + \int_{-\infty}^t h(t-s) dw_s$ are continuous versions of processes of type (3.2).

3.2.2 Model description

We discuss a fluid queue model with a Gaussian-type input rate process, in which the stochastic behavior of the *buffer content* Q_t is represented by a differential equation

$$\frac{dQ_t}{dt} = \begin{cases} R_t - C, & \text{if } Q_t > 0 \text{ or } R_t > C, \\ 0, & \text{otherwise,} \end{cases} \quad (3.3)$$

where C is a constant *output rate* and R_t is a Gaussian-type input rate process defined by

$$R_t = m + \int_{-\infty}^t h(t-s) dw_s. \quad (3.4)$$

Here, m is a positive constant, w_t is a standard Wiener process, and $h(t)$ is a real valued measurable function such that

$$h(t) = 0 \quad \text{for } t < 0, \quad (3.5)$$

$$\int_0^{\infty} |h(t)| dt < \infty, \quad (3.6)$$

$$\int_0^{\infty} h(t)^2 dt < \infty, \quad \text{and} \quad (3.7)$$

$$\int_0^{\infty} |H(t)| dt < \infty, \quad (3.8)$$

where

$$H(t) = \int_t^{\infty} h(s) ds. \quad (3.9)$$

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

$$\int_0^{\infty} H(t)^2 dt \leq \int_0^{\infty} |h(t)| dt \int_0^{\infty} |H(t)| dt < \infty. \quad (3.10)$$

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 (3.6) to (3.8) are strictly positive.

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 (3.4) is considered as the output of a linear system with a white noise input dw_t/dt and an impulse response $h(t)$ [31]. Assumption (3.6) is the stability condition for the system, and assumption (3.7) is another essential condition in order that the integration in (3.4) exists. These assumptions are unavoidable conditions to discuss or to construct a linear system with a white noise input. Here, we assume (3.8), 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 3.1 below).

Remark 3.1. In Chapter 2, we gave some examples of $h(t)$ satisfying (3.6), (3.7) and (3.8). The following is a summary of those examples.

- Ornstein-Uhlenbeck process

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

where σ is a real constant and κ 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, σ_i 's are real constants and κ_i 's are positive constants.

- Continuous version of moving average process

$$|h(t)| < M, \quad 0 \leq t \leq T, \quad \text{and} \quad h(t) = 0, \quad t > T, \quad (3.11)$$

where M and T are positive constants.

- Continuous version of autoregressive process

$$h(t) = \sum_{i=1}^N \sigma_i t^{n_i} e^{-\kappa_i t} \cos(\omega_i t + \varphi_i), \quad t > 0, \quad (3.12)$$

where N is a positive integer, n_i 's are non-negative integers, κ_i 's are positive constants, and σ_i 's, ω_i 's and φ_i 's are real constants.

- Continuous version of autoregressive moving average process

$$h(t) = \int_0^t g(t-s)f(s)ds, \quad (3.13)$$

where $g(t)$ is a function of type (3.11) and $f(t)$ is a function of type (3.12).

Remark 3.2. 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 3.3. 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:

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

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

$$\begin{aligned} H(0)^2/2 &= \frac{1}{2} \int_0^\infty \int_0^\infty h(t)h(s)dt ds \\ &= \int_0^\infty \int_0^\infty h(t+s)h(s)ds dt \\ &= \int_0^\infty \Lambda_R(t)dt, \end{aligned} \quad (3.14)$$

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.

3.3 Bounds for the tail probability $P(Q_\infty > x)$

In this section, we prove the following theorem which justifies, in a sense, the rate of the exponential decay in the approximation formula (3.1).

Let Q_∞ be a random variable subjecting to the steady-state distribution of the buffer content and A_t be the *input process* defined as

$$A_t = \int_0^t R_s ds. \quad (3.15)$$

Theorem 3.1. *There are positive constants B_0 , B_1 and B_2 and a non-negative constant B_3 such that for sufficiently large x ,*

$$\frac{1}{B_0 + B_1\sqrt{x}} \exp\left[-\frac{C-m}{H(0)^2/2}x\right] < P(Q_\infty > x) < (B_2 + B_3x) \exp\left[-\frac{C-m}{H(0)^2/2}x\right], \quad (3.16)$$

where $H(0) = \int_0^\infty h(t)dt$. If the covariance function β_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.

3.3.1 Proof of the upper bound

To get the upper bound in (3.16), we employ the following result used in Simonian and Virtamo's paper [35].

The stationary distribution of the buffer content satisfies

$$P(Q_\infty > x) \leq \int_0^{+\infty} dt \int_{-\infty}^C (C - y) p_t(x, y) dy, \quad (3.17)$$

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

In [35] (pp.1735 (2.10)), instead of $p_t(x, y)dy$, the expression $p_t(x|y)d\Phi(y)$ is used in (3.17), 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 (3.17) 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 (3.15) and (3.4), 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

$$p_t(x, y) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma_t|}} \exp\left[-\frac{1}{2}(x + \mu t, y - m)\Sigma_t^{-1}(x + \mu t, y - m)^T\right], \quad (3.18)$$

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

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

where α_t is the variance of $A_t - Ct$, β_t the covariance between $A_t - Ct$ and R_0 , and γ the variance of R_0 . Then α_t , β_t and γ are written in terms of $h(t)$ and $H(t)$ as

follows:

$$\alpha_t = H(0)^2 t - 2H(0) \int_0^t H(s) ds + 2 \int_0^\infty H(s)^2 ds - 2 \int_0^\infty H(t+s)H(s) ds, \quad (3.19)$$

$$\beta_t = - \int_0^\infty H(t+s)h(s) ds + \int_0^\infty H(s)h(s) ds, \quad (3.20)$$

$$\gamma = \int_0^\infty h(s)^2 ds. \quad (3.21)$$

Lemma 3.1. α_t , β_t and γ have the following properties:

1. α_t and β_t are absolutely continuous functions of t such that $\alpha_0 = 0$ and $\beta_0 = 0$.
2. γ and α_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 ξ_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 Δt .
6. $\alpha_{\Delta t} = \gamma \{\Delta t\}^2 + o(\{\Delta t\}^2)$ for small Δt .

Proof. 1. Since $H(t)$ is absolutely continuous, α_t and β_t are also absolutely continuous. It is obvious from (3.19) and (3.20) that $\alpha_0 = 0$ and $\beta_0 = 0$.

2. Since γ and α_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 (3.20) 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 (3.19), 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 (3.6), (3.7), (3.8) and (3.10), the right-hand side of the above inequality is finite.

5. By differentiating (3.20), we have

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

It is easily checked from (3.7) 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 (3.19), 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 (3.18). Rearranging the term in the brackets, we have

$$\begin{aligned}p_t(x, y) &= \frac{1}{\sqrt{(2\pi)^2(\gamma\alpha_t - \beta_t^2)}} \exp\left[-\frac{\gamma(x + \mu t)^2 - 2\beta_t(x + \mu t)(y - m) + \alpha_t(y - m)^2}{2(\gamma\alpha_t - \beta_t^2)}\right] \\ &= \frac{1}{\sqrt{(2\pi)^2(\gamma\alpha_t - \beta_t^2)}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t} - \frac{\{y - m - \frac{\beta_t}{\alpha_t}(x + \mu t)\}^2}{2(\gamma - \frac{\beta_t^2}{\alpha_t})}\right].\end{aligned}\quad (3.22)$$

Substituting $p_t(x, y)$ in (3.17) with (3.22), we have

$$\begin{aligned}P(Q_\infty > x) &\leq \frac{1}{2\pi} \int_0^\infty dt \cdot \frac{1}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] \\ &\quad \times \int_{-\infty}^C \frac{C - y}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}} \exp\left[-\frac{\{y - m - \frac{\beta_t}{\alpha_t}(x + \mu t)\}^2}{2(\gamma - \frac{\beta_t^2}{\alpha_t})}\right] dy.\end{aligned}\quad (3.23)$$

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 (3.23) 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

$$P(Q_\infty > x) < \sqrt{\frac{\gamma}{2\pi}} \int_0^\infty \frac{z_t}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] dt. \quad (3.24)$$

Lemma 3.2. *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 $t > 0$*

$$z_t = \frac{\mu - \frac{\beta_t}{\alpha_t}(x + \mu t)}{\sqrt{\gamma - \frac{\beta_t^2}{\alpha_t}}} \leq D_1 + D_2 x. \quad (3.25)$$

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 3.1, $\gamma - \beta_t^2/\alpha_t$ is positive for $t > 0$ and bounded from above by γ . We note that $\lim_{t \rightarrow 0}(\gamma - \frac{\beta_t^2}{\alpha_t}) = 0$, $\lim_{t \rightarrow 0}(1 - \frac{\beta_t}{\alpha_t}t) = 0$, 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 $D_1 = \sup_{t>0} f(t) - D_2 x_0$, then D_1 and D_2 are finite and the inequality (3.25) holds.

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

From the inequality (3.24) and Lemma 3.2, we have

$$P(Q > x) < (D_1 + D_2 x) \sqrt{\frac{\gamma}{2\pi}} \int_0^\infty \frac{1}{\sqrt{\alpha_t}} \exp\left[-\frac{(x + \mu t)^2}{2\alpha_t}\right] dt. \quad (3.26)$$

To evaluate the integrand in the right-hand side of (3.26), we shall replace α_t with $H(0)^2(t + \xi_t)$. From Lemma 3.1, ξ_t is a bounded function and we let $\bar{\xi}$ is an upper bound of $|\xi_t|$.

Lemma 3.3. *Let δ be an arbitrary positive number. Then, for $x > \mu\bar{\xi} + \delta/2$ and $t > 0$,*

$$\frac{1}{\sqrt{H(0)^2(t + \xi_t)}} \exp\left[-\frac{(x + \mu t)^2}{2H(0)^2(t + \xi_t)}\right] < \frac{D_3}{\sqrt{H(0)^2(t + \bar{\xi})}} \exp\left[-\frac{(x + \mu t - \delta)^2}{2H(0)^2(t + \bar{\xi})}\right], \quad (3.27)$$

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

Proof. Let $F(t)$ be the ratio between the right and the left sides of (3.27), namely,

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

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

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

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

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

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 (3.27). $\square$

Lemma 3.4.

$$\int_0^\infty \frac{1}{\sqrt{t + \bar{\xi}}} \exp\left[-\frac{(x + \mu t - \delta)^2}{2H(0)^2(t + \bar{\xi})}\right] dt < D_4 \exp\left[-\frac{\mu}{H(0)^2/2}x\right], \quad (3.28)$$

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 (3.28) 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

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

Applying Lemmas 3.3 and 3.4 to (3.26), we have for $x > \mu\bar{\xi} + \delta/2$,

$$\begin{aligned} P(Q > x) &< D_4(D_1 + D_2 x) \sqrt{\frac{\gamma}{2\pi}} \exp\left[-\frac{\mu}{H(0)^2/2} x\right] \\ &= (B_2 + B_3 x) \exp\left[-\frac{\mu}{H(0)^2/2} x\right], \end{aligned}$$

where $B_2 = D_4 D_1 \sqrt{\gamma/2\pi}$ and $B_3 = D_4 D_2 \sqrt{\gamma/2\pi}$.

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

Remark 3.4. Recall that $\beta_0 = 0$ and $\beta'_t = \Lambda_R(t)$. Then, β_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 β_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 β_t is always positive or not, by accumulating the sample autocovariance of $R_{i\Delta t}, i = 1, 2, \dots, N$.

3.3.2 Proof of the lower bound

Lemma 3.5. *The tail distribution of the buffer content Q_∞ in the steady-state satisfies*

$$P(Q_\infty > x) \geq \max_{t \geq 0} P(A_t - Ct > x). \quad (3.29)$$

Proof. According to [3, 13], 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 (3.29). $\square$

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

$$P(A_t - Ct > x) = \bar{\Phi}(y_t), \quad (3.30)$$

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 ξ_t is a function such that $\alpha_t = H(0)^2(t + \xi_t)$. From Lemma 3.1, 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 3.6.

$$P(Q_\infty > x) > \bar{\Phi}(\min_{t > \bar{\xi}} \tilde{y}_t). \quad (3.31)$$

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 (3.29) and (3.30), we have

$$P(Q_\infty > x) \geq \max_{t > \bar{\xi}} \bar{\Phi}(y_t) > \max_{t > \bar{\xi}} \bar{\Phi}(\tilde{y}_t) = \bar{\Phi}(\min_{t > \bar{\xi}} \tilde{y}_t). \quad \square \quad (3.32)$$

By using Lemma 3.6, 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 (3.32) and the inequality

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

we have

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

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

$$P(Q_\infty > x) > \frac{1}{B_0 + B_1\sqrt{x}} \exp[-\frac{C - m}{H(0)^2/2}x], \quad (3.33)$$

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

This completes the proof of Theorem 3.1.

Chapter 4

Data analysis and modeling of VBR video traffic with scene changes

4.1 Introduction

In ATM networks, coded video traffic, called *variable-bit-rate* (VBR) video traffic, is considered to become main traffic since video transmissions require larger bandwidth than voice and data ones. However, its statistical behavior is not well known partly because it is not long since VBR video codec started to develop and partly because its data consists of a huge amount of records which are mutually highly dependent.

J. Beran et al. [4] analyze various processes of video traffic data, using R/S analysis and variance-time analysis, and assert that these processes exhibit long-range dependence. The input rate processes discussed so far, for example, discrete time Gaussian processes [25, 14, 10, 27], Markovian processes [11, 2, 19, 36] and Gaussian-type input rate processes introduced in Chapter 2, are not long-range

dependent processes. *Self-similar* processes such as *fractional Brownian motions* [26] and *fractional ARIMA processes* [9] are typical long-range dependent processes. In the opinion of J. Beran et al. [4] video traffic would be modeled as a self-similar process.

In this chapter we statistically analyze a certain VBR video traffic data in detail, and based on the analysis we develop a simulation model to examine the statistical property of an aggregated traffic and the tail distribution of the buffer content of an ATM multiplexer.

In the VBR video traffic, “spikes” appear at moments of scene changes and the statistical behavior of the traffic differs from scene to scene. Hence we analyze the data by taking account of scene changes. We divide the sequence of the traffic data into subsequences corresponding to scenes, and classify all the subsequences into five types of scenes from a viewpoint of camera work and degree of the motion of subjects. Then, we analyze the subsequences for each scene type.

Based on the analysis, we develop a simulation model consisting of three sub-models, a Markovian scene transition model, a spike model and an AR model for bit rates following the spike. In our model, successive scene types are selected according to a Markov chain from the set of the five types of scenes, and the parameters of the spike model and the AR model for a scene are selected randomly according to scene-type dependent distributions estimated from the data.

A sequence of traffic is generated by the simulation model and compared with the original sequence in four aspects; the marginal distribution, the autocorrelation function, the R/S function and the spectral density function. The R/S function is used for testing whether a sequence is long-range dependent or not. In these aspects, the simulated traffic is recognized to resemble the real one very well. In particular, the simulated sequence exhibits long-range dependence in R/S analysis as well as the original sequence though any self-similar processes are not used in the model.

We examine the statistical property of an aggregated sequence generated by the

simulation model. In the sequence the effects of scene changes almost disappear and the marginal distribution is almost normal. It also exhibits long-range dependence in R/S analysis. In the aspects of the marginal distribution and the R/S function, it is similar to the input rate process from a fractional Brownian process, a typical self-similar process. However its spectral density function is much different.

We also examine the tail behavior of the steady-state queue length distribution of an ATM multiplexer with multiple sources. The simulation results indicate that the approximation formula (2.39) provides good approximate values when the number of sources is large, say more than 150. When the number of sources is 50, the formula underestimates the simulated tail distribution but the decay rate is similar. In this sense, the approximation formula provides a useful tool to analyze and estimate the congestion in the buffers of ATM multiplexers.

Finally, we discuss long-range dependence of VBR video traffic, which Beran et al. indicate by using R/S analysis and time-variance analysis. Our analysis and modeling of the real traffic data seem to suggest that the traffic is not long-range dependent in the rigorous sense.

The construction of this chapter is as follows. In Section 4.2 we explain the actual video traffic sequence and statistically analyze it by dividing into subsequences corresponding to scenes. In Subsection 4.2.1 scene type transitions are analyzed. The scene length, the average rate of scene, the spike of scene, and the body of scene are analyzed in Subsections 4.2.2, 4.2.3, 4.2.4 and 4.2.5, respectively. In Section 4.3 we present a simulation model for the VBR video traffic. In Section 4.4 we compare the statistical property of a simulated traffic with that of the real data. In Subsection 4.4.1 we simulate a single source and compare the simulated traffic with the real one in aspects of the marginal distribution, the autocorrelation function, the R/S function and the spectral density function. In Subsection 4.4.2 we examine statistical property of the aggregated traffic sequence from multiple sources. In Section 4.5 we show the behavior of the steady-state queue length distribution of ATM

multiplexers with multiple sources and compare it with the approximation formula (2.39) in Chapter 2. In Section 4.6 we discuss long-range dependence property of VBR video traffic.

4.2 Data analysis of VBR video traffic with scenes

We analyze a certain real VBR video traffic data. The video is a travel movie in USA by a car. The traffic data is generated from a video source in which the video data is compressed according to the modified H.261. The standard H.261 includes a mechanism to transmit compressed video data with a constant bit rate (CBR). In the modified H.261, the mechanism for CBR transmission is eliminated and a mechanism for controlling a maximum bit rate is added.

The real traffic data is a sequence $\{X(k), k = 1, 2, \dots, K\}$, where $X(k)$ represents the bit rate in Mbps generated at the k -th frame. The frame interval is $1/30$ second and the number of frames N is 79686. So the time length of the whole sequence is about 43 minutes. Figure 4.1 shows the bit rates of the first 10,000 frames. The value of $X(k)$ ranges from 0 to 6.144 Mbps.

In the real traffic sequence there exist a number of *spikes* (see Figures 4.1, 4.2). Observing the video and the traffic sequence, we see spikes appear at moments of scene changes and the behavior of $X(k)$ is largely different among scenes.

In our analysis, first we divide the sequence $\{X(k) : k = 1, \dots, K\}$ into subsequences corresponding to scenes. The subsequence corresponding to the n -th scene is represented as $\{X_n(k) : k = 1, \dots, N_n\}$. We call this subsequence as the *sequence of the n -th scene*. In the sequence of the n -th scene, first few values are recognized to form a *spike*. The subsequence $\{X_n(1), \dots, X_n(M_n)\}$ of the sequence of the n -th scene forming the spike will be referred as the *spike sequence* of the n -th scene. The remainder of the sequence $\{X_n(k) : k = M_n + 1, \dots, N_n\}$ will be referred as the *body sequence* of the n -th scene. The way of defining M_n in the real data will be discussed

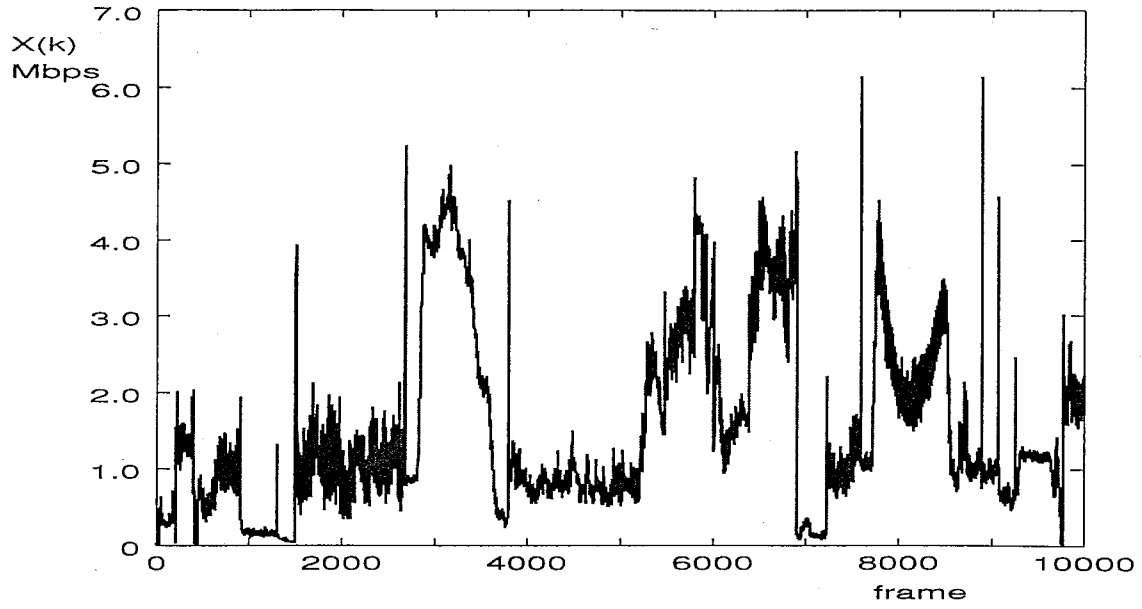
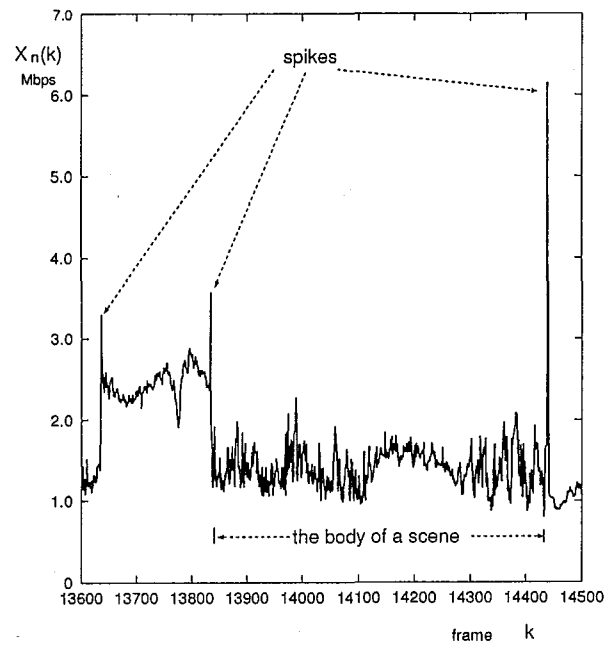


Figure 4.1: Measured VBR video traffic

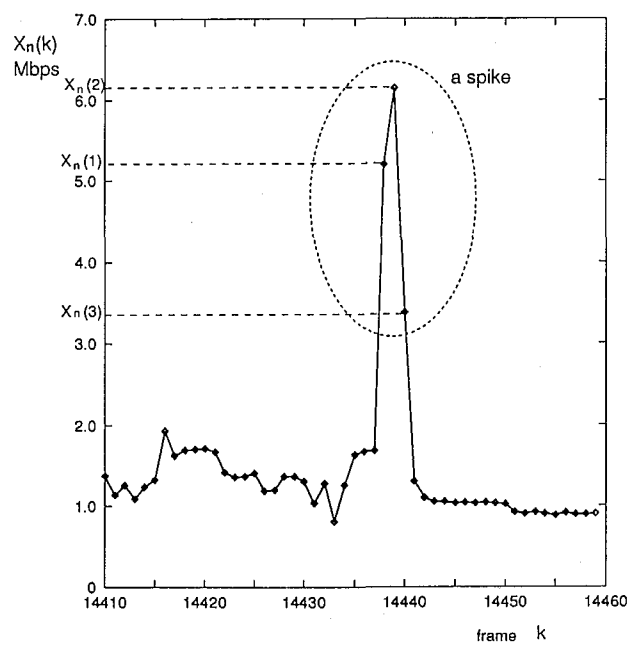
in Subsection 4.2.4. There are 163 scenes in the traffic data, but we analyze here first 100 scenes or 43526 frames.

We observe that the behavior of the sequence $\{X_n(k)\}$ of the n -th scene depends largely on the camera work and the motion of subjects in the video. If inactive subjects are shot with a fixed camera, $X_n(k)$'s take small values and their behavior is almost flat. If inactive subjects are shot with a moving camera or if subjects in a scene are moving, the corresponding sequence $\{X_n(k)\}$ forms a waved curve or a mountain-like curve. Here we classify them from the following viewpoints;

1. camera is moving or not,
2. subjects in the scene are moving or not, and
3. the degree of motion.



(a) Scenes and spikes



(b) A spike

Figure 4.2: Scene and spike

The 100 scenes are classified into five scene types.

Type 1 *Scene of low degree motion:* A scene of this type is shot with a fixed camera and the motion of subjects is inactive (26 scenes).

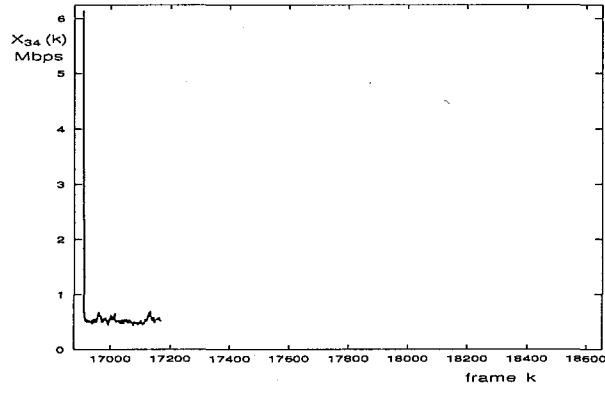
Type 2 *Scene of normal degree motion:* A scene of this type is shot with a fixed camera and the motion of subjects is active, but not so hard (31 scenes).

Type 3 *Scene of high degree motion:* A scene of this type is shot with a fixed camera and the motion of subjects is hard (16 scenes).

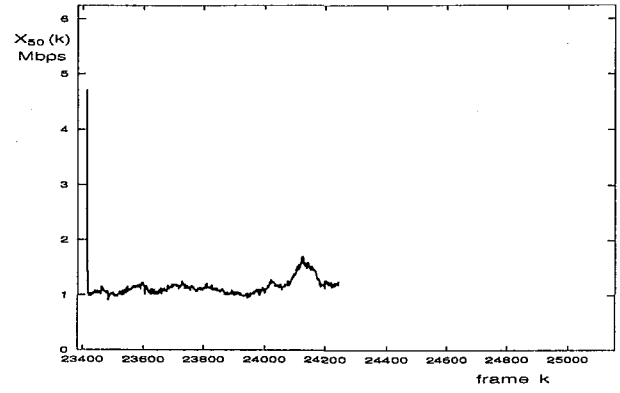
Type 4 *Scene from a car:* A scene of this type is shot from a moving car (9 scenes).

Type 5 *Scene with a panning camera:* A scene of this type is shot with a panning camera (18 scenes).

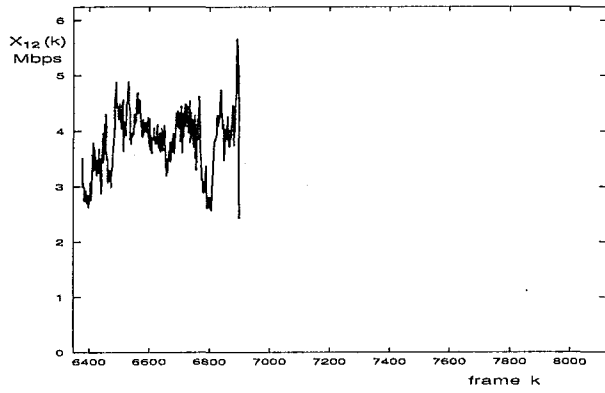
Figure 4.3 shows examples of the behavior of $X_n(k)$ for the five scene types. These graphs exhibit different patterns; a flat line in (a), a mountain-like curve in (e), and a waved curve in (b),(c),(d).



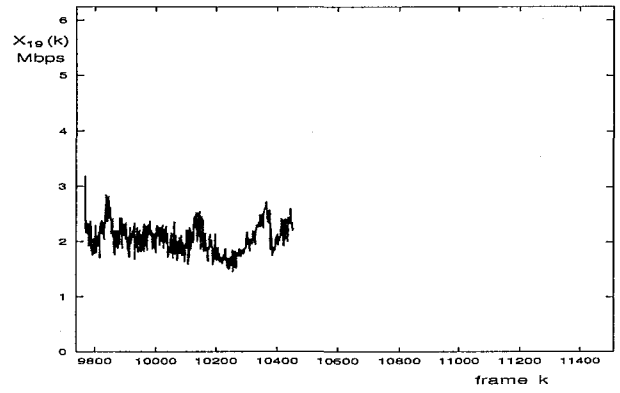
(a) #34 of Type 1 (low degree motion)



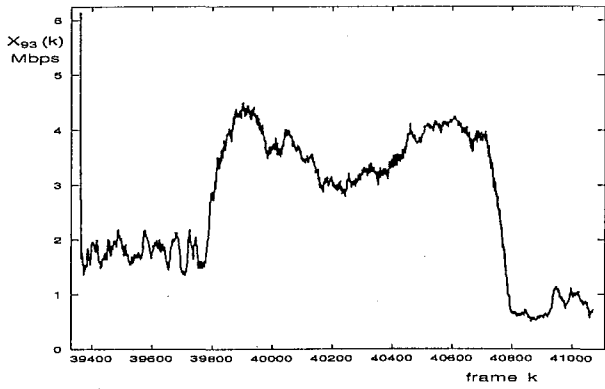
(b) #50 of Type 2 (normal degree motion)



(c) #12 of Type 3 (high degree motion)



(d) #19 of Type 4 (from a car)



(e) #93 of Type 5 (with a panning camera)

Figure 4.3: Examples of sequences $\{X_n(k)\}$ of scenes

4.2.1 Transition of scene types

First we observe transitions of scene types. Let q_{ij} be the frequency of transitions from Type i to Type j and r_{ij} be the transition rate $q_{ij}/\sum_k q_{ik}$. Then they are given in a matrix form as

$$Q = (q_{ij}) = \begin{array}{c} \nearrow \\ \begin{array}{l} \text{Type1} \\ \text{Type2} \\ \text{Type 3} \\ \text{Type 4} \\ \text{Type 5} \end{array} \end{array} \begin{array}{ccccc} \text{Type1} & \text{Type2} & \text{Type3} & \text{Type4} & \text{Type5} \\ \left[\begin{array}{ccccc} 8 & 8 & 0 & 5 & 5 \\ 8 & 11 & 4 & 1 & 7 \\ 4 & 3 & 9 & 0 & 0 \\ 2 & 1 & 1 & 2 & 3 \\ 3 & 8 & 2 & 1 & 3 \end{array} \right] \end{array}, \quad (4.1)$$

$$R = (r_{ij}) = \begin{array}{c} \nearrow \\ \begin{array}{l} \text{Type1} \\ \text{Type2} \\ \text{Type3} \\ \text{Type4} \\ \text{Type5} \end{array} \end{array} \begin{array}{ccccc} \text{Type1} & \text{Type2} & \text{Type3} & \text{Type4} & \text{Type5} \\ \left[\begin{array}{ccccc} .308 & .308 & .000 & .192 & .192 \\ .258 & .355 & .129 & .032 & .226 \\ .250 & .188 & .562 & .000 & .000 \\ .222 & .111 & .111 & .222 & .333 \\ .176 & .471 & .118 & .059 & .176 \end{array} \right] \end{array} \quad (4.2)$$

There is no transition from Type 1 (Scene of low degree motion) to Type 3 (Scene of high degree motion). Once the scene type becomes Type 3 it has a tendency to stay at Type 3. There is no transition from Type 3 to Type 4 (Scene from a car) or Type 5 (Scene with a panning camera). If the present scene type is Type 5, then the next scene type becomes Type 2 (Scene of normal degree motion) with high probability.

4.2.2 Length of scene

Table 4.1 lists the sample mean $\overline{N}^{(j)}$ and the sample standard deviation $S_N^{(j)}$ of the length N_n of the sequences for scenes of Type j ($j = 1, \dots, 5$). The histogram of N_n is given in Figure 4.4. For Type 1, $\overline{N}^{(1)}$ and $S_N^{(1)}$ are low, and for Type 5, $\overline{N}^{(5)}$ and $S_N^{(5)}$ are large. A gamma distribution is fitted to the distribution of the length N_n of Type j ($j = 1, \dots, 5$). The gamma distribution has two parameters α and β , and its density function is given by

$$f(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} & (x \geq 0), \\ 0 & (x < 0), \end{cases} \quad (4.3)$$

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ is the gamma function. The expectation of the distribution is α/β and the variance is α/β^2 .

Table 4.1: The sample mean and the deviation of N_n

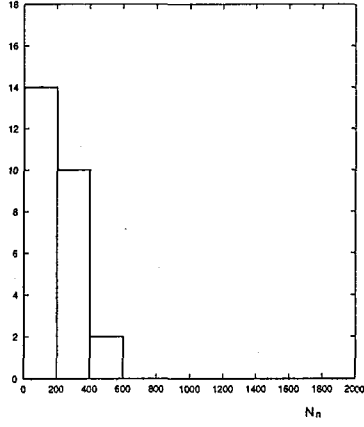
Scene type	$\overline{N}$	S_N	$\hat{\alpha}_N$	$\hat{\beta}_N$
Type 1 (Scene of low degree motion)	207	110	3.54	.0171
Type 2 (Scene of normal degree motion)	413	320	1.67	.0040
Type 3 (Scene of high degree motion)	374	295	1.61	.0043
Type 4 (Scene from a car)	434	145	8.96	.0206
Type 5 (Scene with a panning camera)	859	364	5.57	.0065

The parameters $\alpha_N^{(j)}$ and $\beta_N^{(j)}$ of the gamma distribution fitted to the distribution of N_n of scenes of Type j can be estimated using $\overline{N}^{(j)}$ and $S_N^{(j)}$ as

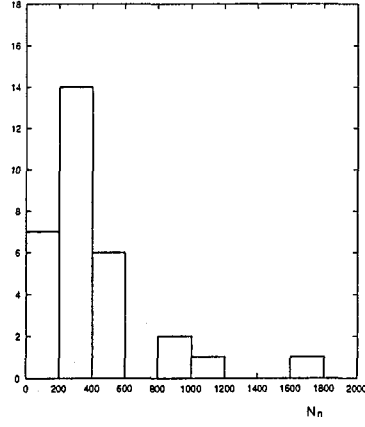
$$\hat{\alpha}_N^{(j)} = (\overline{N}^{(j)} / S_N^{(j)})^2 \quad (4.4)$$

$$\hat{\beta}_N^{(j)} = \overline{N}^{(j)} / (S_N^{(j)})^2. \quad (4.5)$$

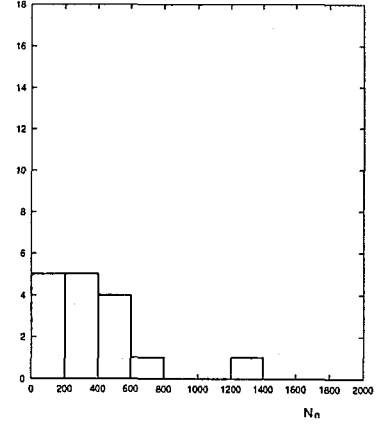
The estimated values of these parameters are also listed in Table 4.1. For each j , the gamma distribution with these estimated parameters is not rejected with 10% significance in a test of goodness of fit.



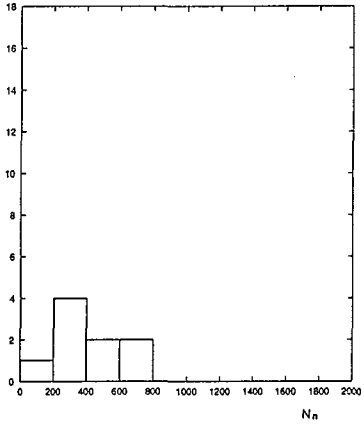
(a) Type 1 (low degree motion)



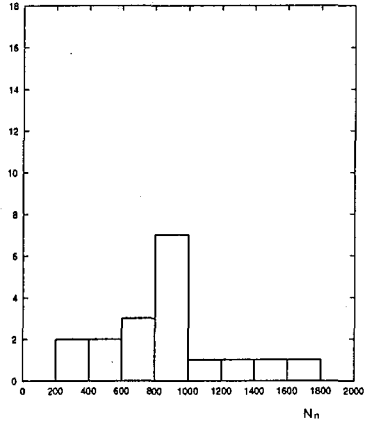
(b) Type 2 (normal degree motion)



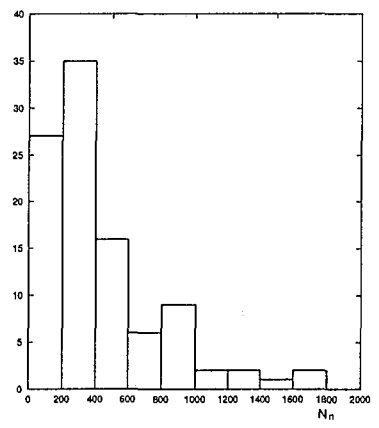
(c) Type 3 (high degree motion)



(d) Type 4 (from a car)



(e) Type 5 (panning camera)



(f) Total

Figure 4.4: Histogram of the scene length N_n

4.2.3 Average bit rate

We define the average bit rate A_n and the bit rate deviation D_n of the body sequence of the n -th scene as

$$A_n = \sum_{k=M_n+1}^N X_n(k)/(N_n - M_n), \quad (4.6)$$

$$D_n = \sqrt{\sum_{k=M_n+1}^{N_n} (X_n(k) - A_n)^2 / (N_n - M_n)}, \quad (4.7)$$

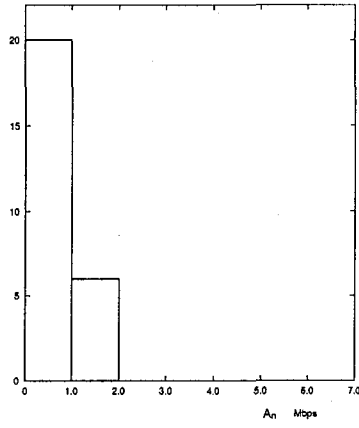
where N_n is the length of the scene and M_n is the length of the spike sequence the latter of which will be defined in Subsection 4.2.4. Table 4.2 lists the sample mean $\bar{A}^{(j)}$ and the sample standard deviation $S_A^{(j)}$ of the average rates A_n of scenes of Type j ($j = 1, \dots, 5$). The histogram of A_n is given in Figure 4.5. For Type 1, $\bar{A}^{(1)}$ and $S_A^{(1)}$ is low, for Type 3, $\bar{A}^{(3)}$ is large, and for Type 5, $S_A^{(5)}$ is large.

A gamma distribution is fitted to the distribution of the average bit rate A_n of Type j ($j = 1, \dots, 5$). The parameters $\alpha_A^{(j)}$ and $\beta_A^{(j)}$ can be estimated using $\bar{A}^{(j)}$ and $S_A^{(j)}$ as

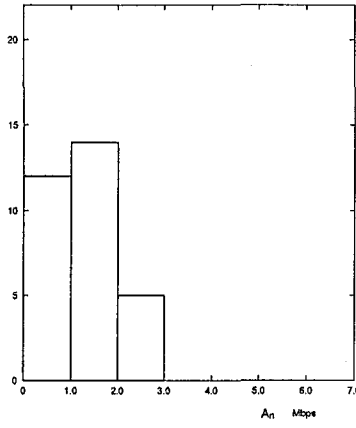
$$\hat{\alpha}_A^{(j)} = (\bar{A}^{(j)} / S_A^{(j)})^2 \quad (4.8)$$

$$\hat{\beta}_A^{(j)} = \bar{A}^{(j)} / (S_A^{(j)})^2. \quad (4.9)$$

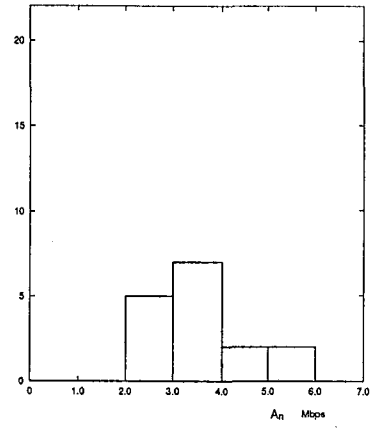
The estimated values of these parameters are also listed in Table 4.2. For each j , the gamma distribution with these estimated parameters is not rejected with 10% significance in a test of goodness of fit.



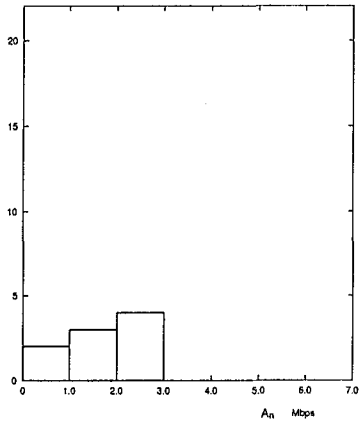
(a) Type 1 (low degree motion)



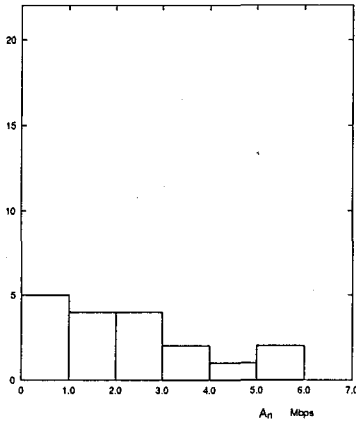
(b) Type 2 (normal degree motion)



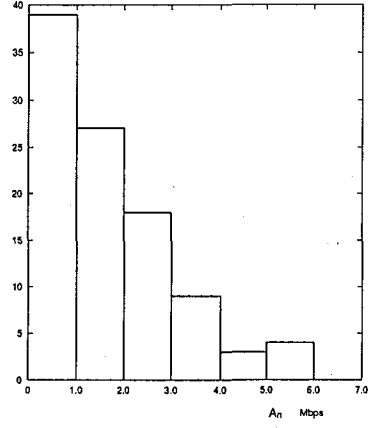
(c) Type 3 (high degree motion)



(d) Type 4 (from a car)



(e) Type 5 (panning camera)



(f) Total

Figure 4.5: Histogram of the average bit rate A_n

Table 4.2: The sample mean and the deviation of A_n

Scene type	$\bar{A}$	S_A	$\hat{\alpha}_A$	$\hat{\beta}_A$
Type 1 (Scene of low degree motion)	0.573	0.446	1.65	2.88
Type 2 (Scene of normal degree motion)	1.161	0.596	3.79	3.27
Type 3 (Scene of high degree motion)	3.616	0.934	14.9	4.15
Type 4 (Scene from a car)	1.856	0.788	5.55	2.99
Type 5 (Scene with a panning camera)	2.364	1.517	2.43	1.02

4.2.4 Spike of scene

Let A'_n and D'_n be the average bit rate and bit rate deviation of the sequence including the spike sequence of the n -th scene. They are determined by equations (4.6) and (4.7) with 0 in the places of M_n . We define length of the spike sequence M_n by the following rule: M_n is the maximum number $h(\leq 5)$ such that $X_n(k) > A'_n + 2D'_n$ for all k ($0 \leq k \leq h$). If $X_n(k) \leq A'_n + 2D'_n$ for all k ($0 \leq k \leq 5$), then $M_n = 0$. In scenes of Type 3, $A'_n + 2D'_n$ is sometimes larger than the peak value of spike. At this time M_n takes zero.

The spike length M_n ranges from 0 to 5. The most frequent length is different among scene types. For Type 1 and Type 2 it is three, and it is two in other types of scenes. The distributions of M_n is shown in Figure 4.6.

The maximum value X_n^{max} of the bit rate in a spike sequence seems to distribute uniformly between A_n and 6.144Mbps (see Figure 4.7). This phenomenon is common among all scene types. If the spike length M_n is two, then $X_n(2)$ is larger than $X_n(1)$ in most cases. If M_n is three, then $X_n(2)$ mostly takes the largest value and $X_n(3)$ takes the smallest value among the three. Thus in most cases $X_n^{max} = X_n(2)$.

When the spike length M_n is three, $X_n(1)$ and $X_n(2)$ are positively correlated and $X_n(1) - A_n$ is near to $0.83(X_n(2) - A_n)$. Also $X_n(2)$ and $X_n(3)$ are positively

correlated and $X_n(3) - A_n$ is near to $0.42(X_n(2) - A_n)$ (see Figure 4.8).

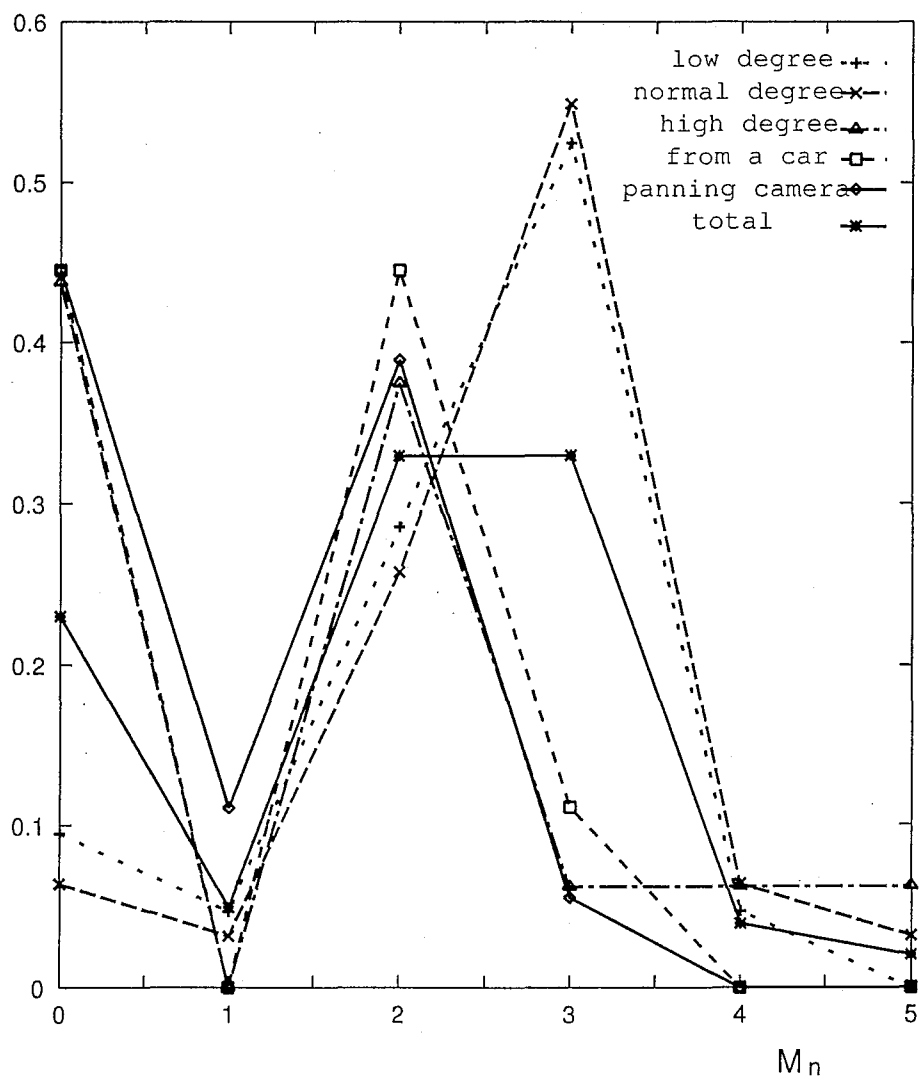
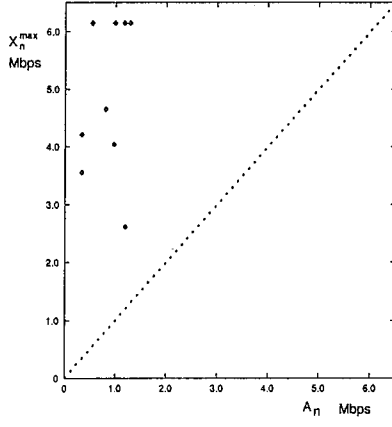
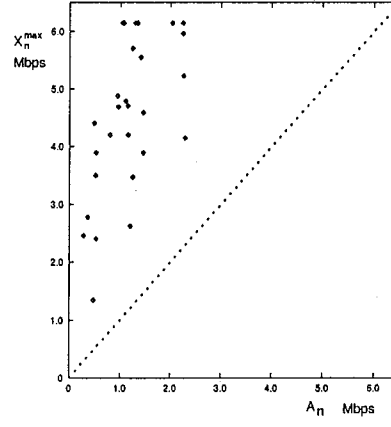


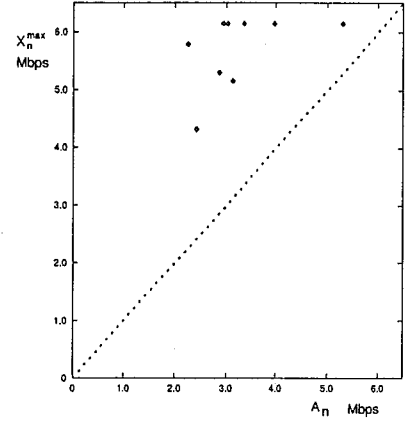
Figure 4.6: Distribution of the spike length M_n



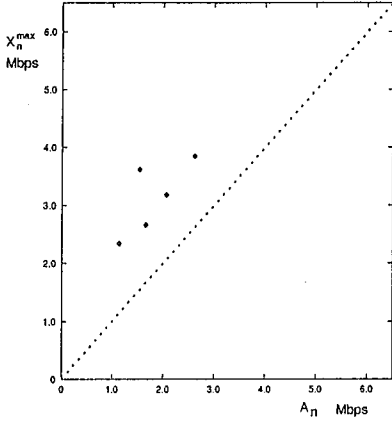
(a) Type 1 (low degree motion)



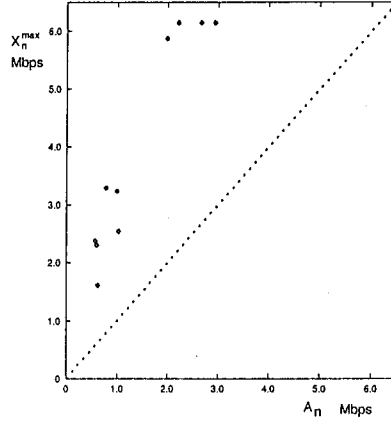
(b) Type 2 (normal degree motion)



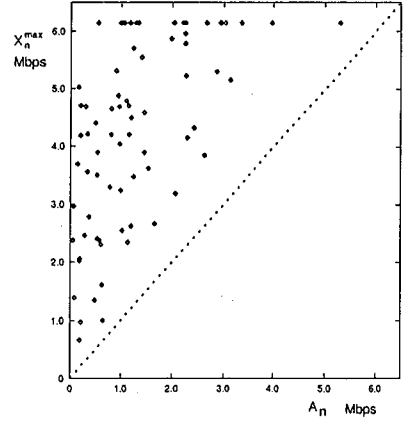
(c) Type 3 (high degree motion)



(d) Type 4 (from a car)

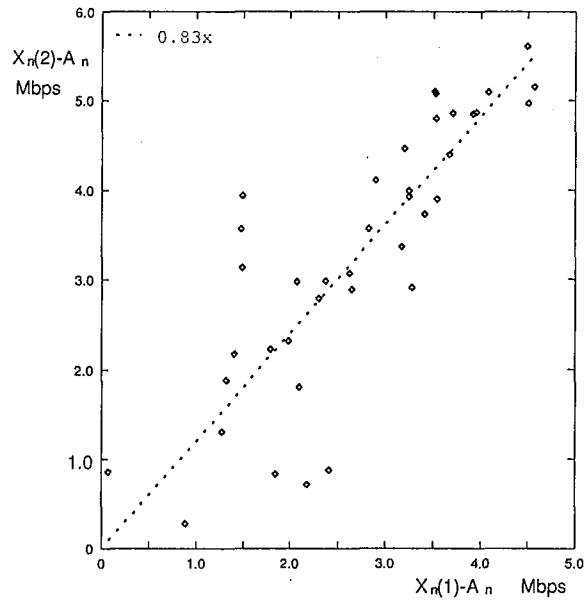


(e) Type 5 (panning camera)

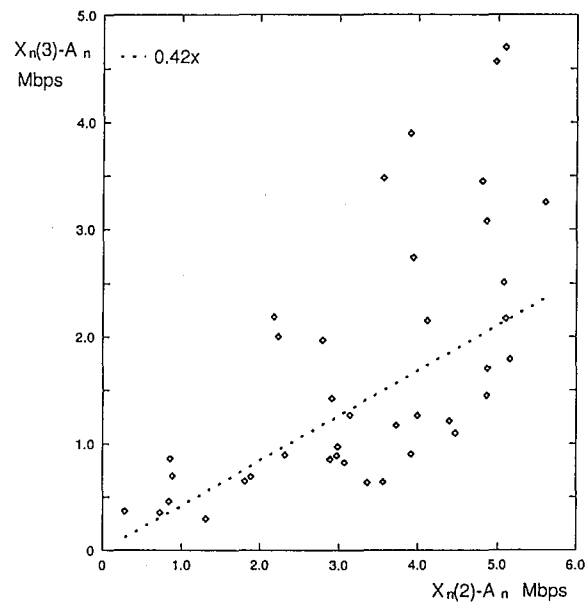


(f) Total

Figure 4.7: Scatter diagram of average bit rate A_n and the max value X_n^{max} of the spike sequence



(a) Scatter diagram of $X_n(1) - A_n$ and $X_n(2) - A_n$



(b) Scatter diagram of $X_n(2) - A_n$ and $X_n(3) - A_n$

Figure 4.8: Scatter diagram of bit rates in a spike

4.2.5 Body sequence of scene

In Figure 4.3, we saw that most of body sequences of scenes $\{X_n(k) : k = M_n + 1, \dots, N_n\}$ behave like autoregressive (AR) processes. The only exception is the graph in Figure 4.3(e), i.e. a scene of Type 5. Here we analyze 82 scenes which are of types other than Type 5. We will call these scenes as *AR scenes*.

An AR process $\{W(k) : k = M + 1, \dots, N\}$ is represented in a form;

$$\begin{aligned} Z(k) &= W(k) - \mu, \\ Z(k) &= \sum_{i=1}^m a_i Z(k-i) + \epsilon(k), \end{aligned} \quad (4.10)$$

where m is the order of the process, μ is the average rate, a_i 's are coefficients, and $\epsilon(k)$'s are mutually independent and identically normally distributed random variables with zero mean and variance σ^2 .

We estimate the order m of the AR process together with parameters μ , σ and a_i 's by using the MDL (Minimum Description Length) principle [33]. The principle selects the most appropriate probability model from several candidates.

For the average rate μ , we estimate it by the average bit rate $\hat{\mu} = A = \sum_{k=M+1}^N W(k)/(N-M)$, and set $Z(k) = W(k) - A$ ($k = M+1, \dots, N$). For coefficients a_i 's, let m be a candidate order and $\hat{a}_i(m)$'s be the solution of the equations

$$\sum_{k=M+1}^N Z(k-h) \left(Z(k) - \sum_{i=M+1}^m \hat{a}_i(m) Z(k-i) \right) = 0, \quad h = 1, \dots, m. \quad (4.11)$$

$\hat{a}_i(m)$'s are maximum likelihood estimators in the sense that they make the estimate

$$\hat{\sigma}(m)^2 = \sum_{k=1}^N \left(Z(k) - \sum_{i=1}^m \hat{a}_i(m) Z(k-i) \right)^2 \quad (4.12)$$

of σ^2 minimum.

Then the value of MDL for the candidate value m is defined by

$$MDL(m) = (N-M) \log \hat{\sigma}^2(m) + \frac{m+1}{2} \log (N-M). \quad (4.13)$$

An m which makes $MDL(m)$ smallest is chosen as the estimate $\hat{m}$ of the order for the process. We write $\hat{\sigma}$ and $\hat{a}_i$'s instead of $\hat{a}_i(\hat{m})$ and $\hat{\sigma}_n^2(\hat{m})$.

We estimate the parameters for the AR process fitted to the body sequence of an AR scene in the way stated above. We will denote the estimates for the n -th scene as $\hat{m}_n$, $\hat{\mu}_n$, $\hat{a}_{i,n}$ and $\hat{\sigma}_n$.

In our case, the estimated order $\hat{m}_n$ ranges from 0 to 9. We also estimate the order $\hat{m}_n$ by using AIC (Akaike's Information Criterion) [1], but it tends to be larger than the one estimated with MDL.

Unfortunately, we can't find any specific tendency for $\hat{a}_{n,i}$'s and $\hat{\sigma}_n$. Probably this is because the number of parameters is too large. So we make an estimation by limiting the order less than or equal to two. Let $\hat{m}'_n$ be the value of m which makes $MDL(m)$ in (4.13) smallest among $m = 0, 1$ or 2 . Then there are 64 scenes for which $\hat{m}'_n = 2$, and 17 scenes for which $\hat{m}'_n = 1$. There is only 1 scene for which $\hat{m}'_n = 0$. We will denote the estimates $\hat{a}_{n,i}(\hat{m}'_n)$ and $\sigma_n(\hat{m}'_n)$ as $\hat{\mu}'_n$, $\hat{a}'_{n,i}$ and $\hat{\sigma}'_n$. Note that $\hat{\mu}'_n = \hat{\mu}_n = A_n$ since it is not dependent on the order.

AR coefficients

For the AR coefficients $\hat{a}'_{n,i}$'s, we don't find any specific difference among distributions for four scene types. However the distributions are much different in the cases with $\hat{m}'_n = 1$ and $\hat{m}'_n = 2$. Hence we discuss here the distribution of AR coefficients with classification by $\hat{m}'_n$ but without classification by scene type.

Figure 4.9 shows the distribution of the coefficient $\hat{a}'_{n,1}$ when $\hat{m}'_n = 1$. From this figure, the distribution of $1 - \hat{a}'_{n,1}$ seems to be exponential. Its sample average is 0.141.

Figure 4.10 is the scatter diagram of $\hat{a}'_{n,1}$ and $\hat{a}'_{n,2}$ when $\hat{m}'_n = 2$. The distribution of $\hat{a}'_{n,2}$ is shown in Figure 4.11 and the distribution of $\hat{a}'_{n,1} + \hat{a}'_{n,2}$ is shown in Figure 4.12. From Figure 4.12, the distribution of $1 - (\hat{a}'_{n,1} + \hat{a}'_{n,2})$ seems to be

exponential. Its expectation is estimated as 0.216 from the sample mean. These exponential distributions for $\hat{m}'_n = 1$ and 2 are not rejected with 10% significance by a test of goodness of fit.

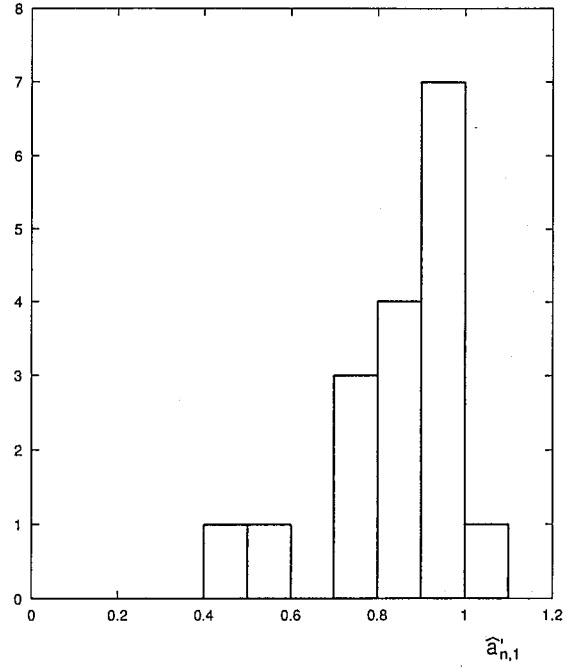


Figure 4.9: Histogram of $\hat{a}'_{n,1}$ when $\hat{m}'_n = 1$

The distribution of $\hat{a}'_{n,2}$ in Figure 4.11 seems to be bimodal with modes near to 0.15 and -0.30 . Its sample mean and sample deviation are 0.010 and 0.096, respectively.

As shown in Figures 4.9 and 4.10, almost all of the estimated coefficients satisfy the stable condition of AR processes;

$$\begin{aligned}
 & -1 < a'_{n,1} < 1 \quad \text{when } \hat{m}'_n = 1, \\
 & a'_{n,1} + a'_{n,2} < 1, \quad a'_{n,1} - a'_{n,2} < 1, \quad -1 < a'_{n,2} < 1 \quad \text{when } \hat{m}'_n = 2.
 \end{aligned}$$

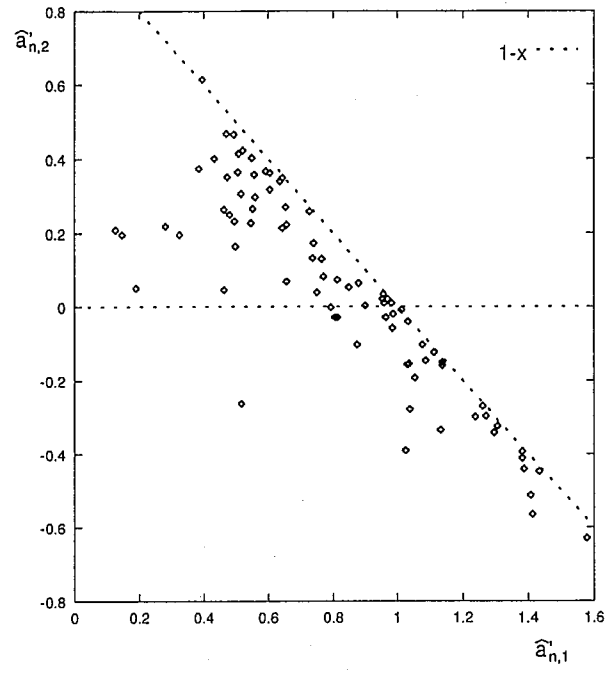


Figure 4.10: Scatter diagram of $\hat{a}'_{n,1}$ and $\hat{a}'_{n,2}$ when $\hat{m}'_n = 2$

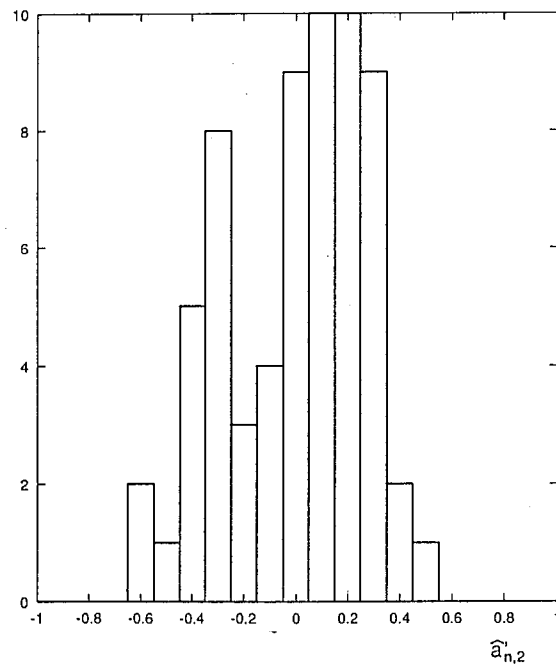


Figure 4.11: Histogram of $\hat{a}'_{n,2}$ when $\hat{m}'_n = 2$

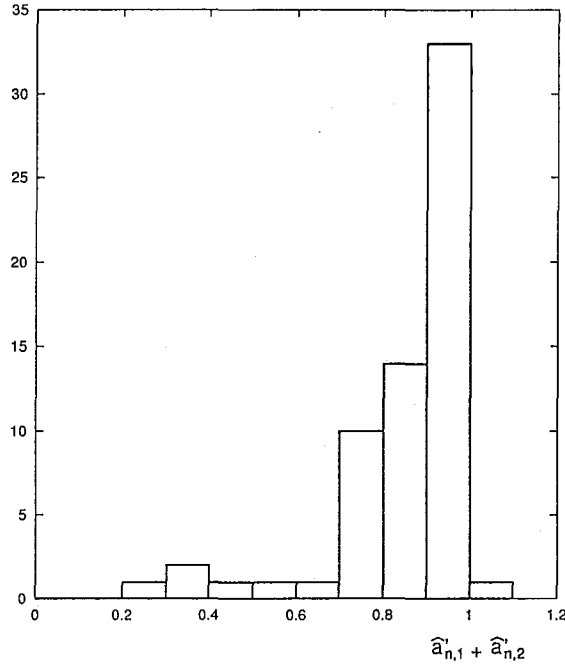


Figure 4.12: Histogram of $\hat{a}'_{n,1} + \hat{a}'_{n,2}$ when $\hat{m}'_n = 2$

Standard deviation $\hat{\sigma}'_n$ of $\epsilon(k)$'s

Figure 4.13 is the scatter diagram of the average rate A_n and the standard deviation $\hat{\sigma}'_n$ of $\epsilon(k)$ for the AR process. There exists a weak positive correlation between A_n and $\hat{\sigma}'_n$.

The standard deviation $\hat{\sigma}'_n$ seems to range between $0.02A_n$ and $0.16\sqrt{A_n} + 0.02A_n$ (see Figure 4.13(e)). Table 4.3 lists the average $\bar{v}^{(j)}$ and the standard deviation $S_v^{(j)}$ of the statistic

$$v_n = (\hat{\sigma}'_n - 0.02A_n)/\sqrt{A_n} \quad (4.14)$$

of Type j ($j = 1, \dots, 4$). The histogram of v_n is given in Figure 4.14. A gamma distribution is fitted to the distribution of v_n of Type j ($j = 1, \dots, 4$).

The parameters $\alpha_v^{(j)}$ and $\beta_v^{(j)}$ can be estimated using $\bar{v}^{(j)}$ and $S_v^{(j)}$ as

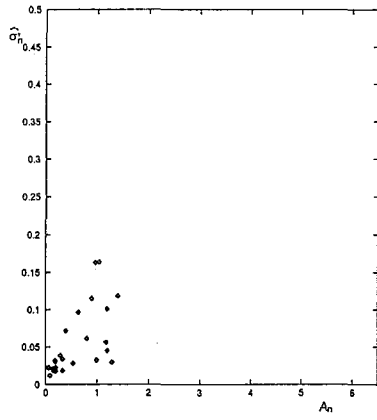
$$\hat{\alpha}_v^{(j)} = (\bar{v}^{(j)}/S_v^{(j)})^2 \quad (4.15)$$

$$\hat{\beta}_v^{(j)} = \bar{v}^{(j)} / (S_v^{(j)})^2. \quad (4.16)$$

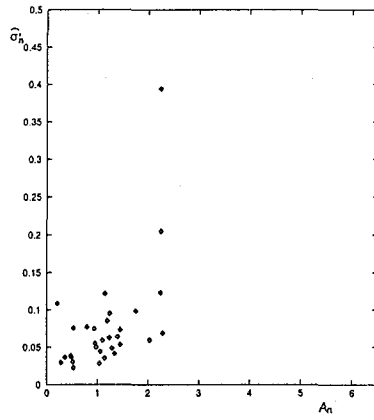
The estimated values of these parameters are also listed in Table 4.3. For each j , the gamma distribution with these estimated parameters is not rejected with 10% significance in a test of goodness of fit.

Table 4.3: The sample mean and the deviation of v_n

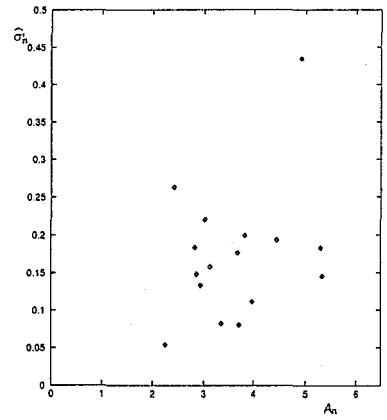
Scene type	$\bar{v}$	S_v	$\hat{\alpha}_v$	$\hat{\beta}_v$
Type 1 (Scene of low degree motion)	.0595	.0366	2.64	44.4
Type 2 (Scene of normal degree motion)	.0528	.0523	1.02	19.3
Type 3 (Scene of high degree motion)	.0535	.0425	1.58	29.6
Type 4 (Scene from a car)	.0799	.0307	6.77	84.8



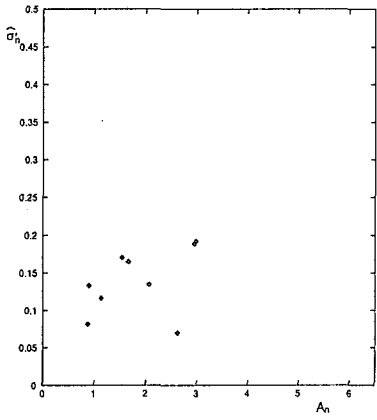
(a) Type 1 (low degree motion)



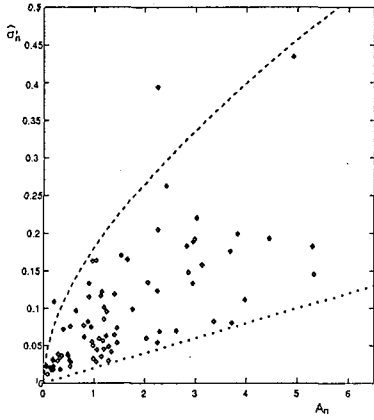
(b) Type 2 (normal degree motion)



(c) Type 3 (high degree motion)

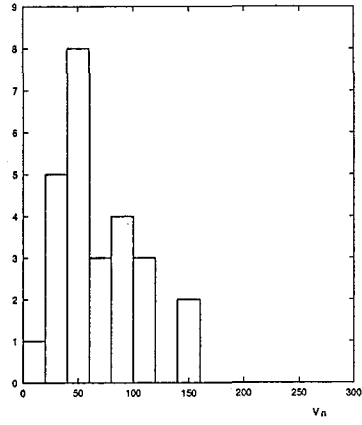


(d) Type 4 (from a car)

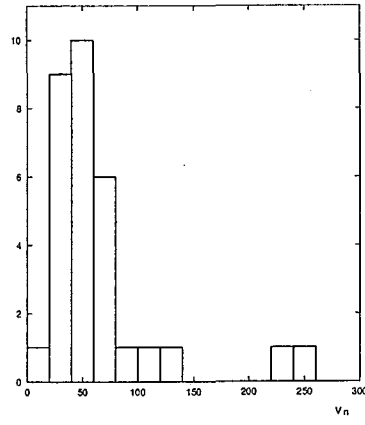


(e) Total

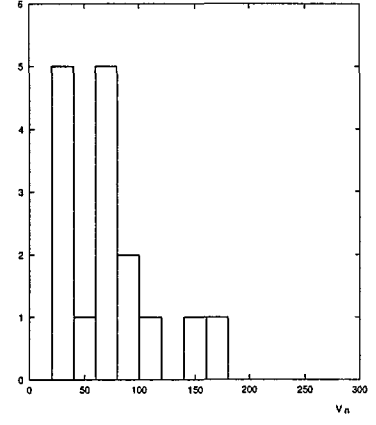
Figure 4.13: Scatter diagram of A_n and $\hat{\sigma}'_n$



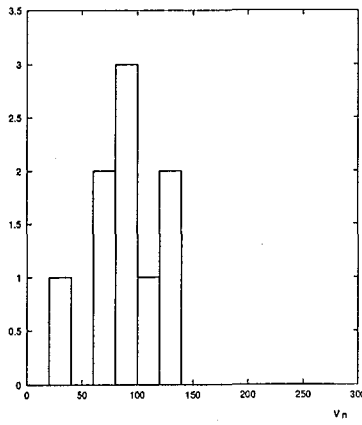
(a) Type 1 (low degree motion)



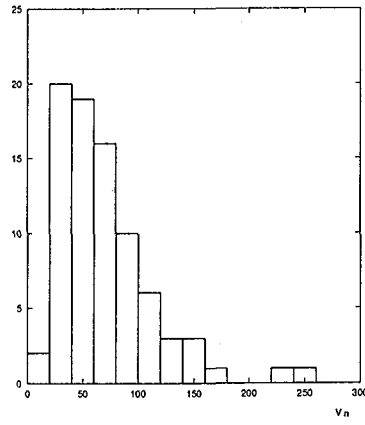
(b) Type 2 (normal degree motion)



(c) Type 3 (high degree motion)



(d) Type 4 (from a car)



(e) Total

Figure 4.14: Histogram of v_n

4.2.6 Dependence between parameters

In the previous subsection, we showed the correlation between the average rate A_n and the standard deviation $\hat{\sigma}'_n$ of the AR process. In this subsection we analyze the dependence between other parameters. Figure 4.15(a) is the scatter diagram of the length N_n and the average bit rate A_n . Figure 4.15(b) is the scatter diagram of N_n and $\hat{\sigma}'_n$. We can't find any clear dependence from these diagrams.

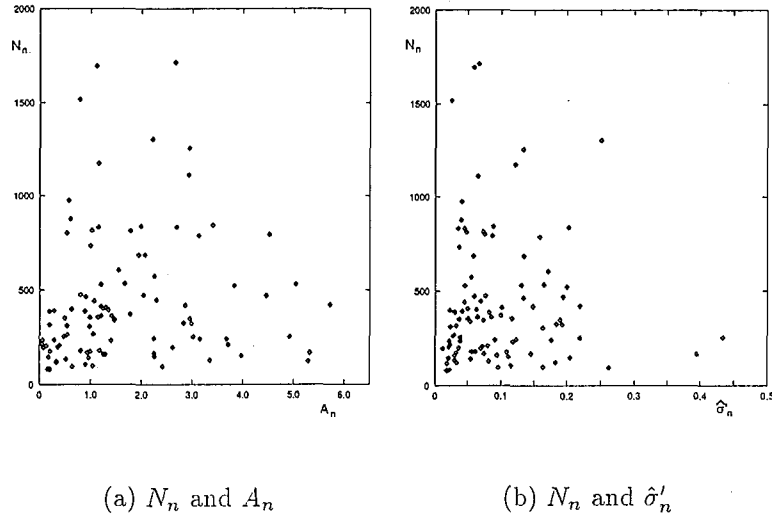
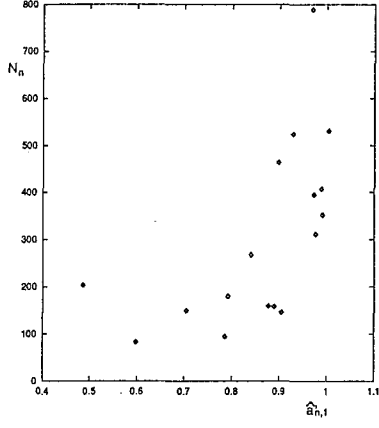
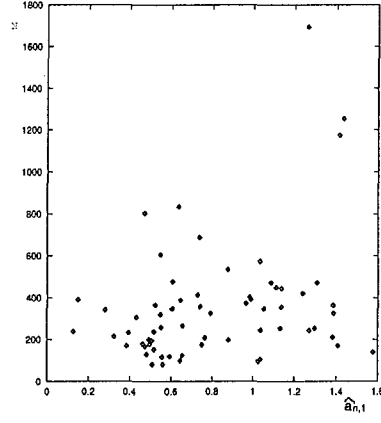


Figure 4.15: Scatter diagram (1)

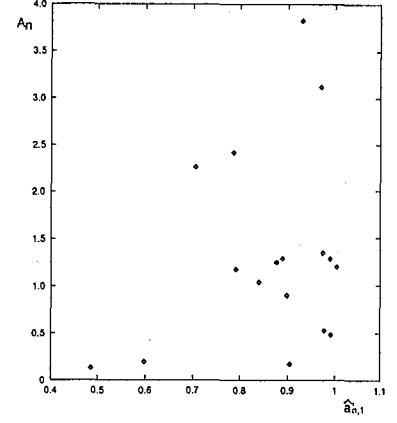
Figure 4.16 shows scatter diagrams for other combinations of estimated parameters. Figures 4.16(a) and (b) are the scatter diagrams of N_n and $\hat{a}'_{n,1}$ when $\hat{m}'_n = 1$ and $\hat{m}'_n = 2$. Figures 4.16(c) and (d) are the scatter diagrams of $\hat{a}'_{n,1}$ and A_n when $\hat{m}'_n = 1$ and $\hat{m}'_n = 2$. Figures 4.16(e) and (f) are the scatter diagrams of $\hat{a}'_{n,1}$ and $\hat{\sigma}'_n$ when $\hat{m}'_n = 1$ and $\hat{m}'_n = 2$. From these figures we can't find any clear correlations among these estimated parameters.



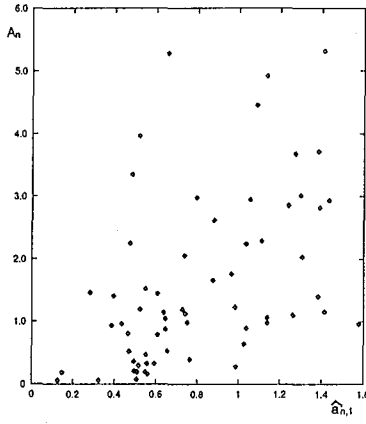
(a) $\hat{a}'_{n,1}$ and N_n at $\hat{m}'_n = 1$



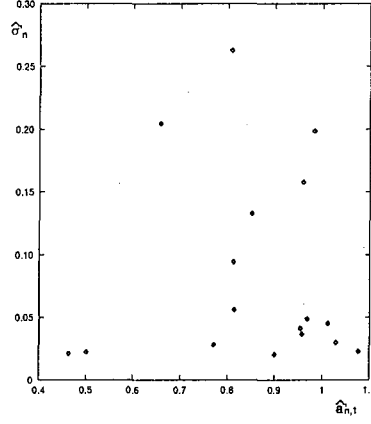
(b) Scatter diagram of $\hat{a}'_{n,1}$ and N_n when $\hat{m}'_n = 2$



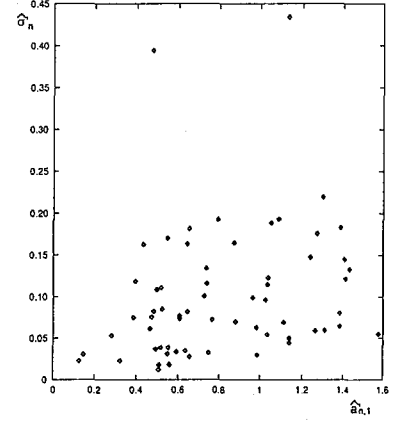
(c) $\hat{a}'_{n,1}$ and A_n at $\hat{m}'_n = 1$



(d) $\hat{a}'_{n,1}$ and A_n at $\hat{m}'_n = 2$



(e) $\hat{a}'_{n,1}$ and $\hat{\sigma}'_n$ at $\hat{m}'_n = 1$



(f) $\hat{a}'_{n,1}$ and $\hat{\sigma}'_n$ at $\hat{m}'_n = 2$

Figure 4.16: Scatter diagram (2)

4.3 Modeling of VBR traffic

To achieve a simulation for an ATM multiplexer, we introduce a model for VBR bit rates based on the statistical analysis in the previous section. We denote by $\{Y(k) : k = 1, \dots, K\}$ the generated whole sequence by the model and by $\{Y_n(k) : k = 1, \dots, N_n\}$ the subsequence corresponding to the n -th scene.

Transition of scene types

We assume the scene type changes according to a Markov chain with transition matrix R in (4.2). The initial scene type is randomly selected according to the stationary distribution.

Length of scene

The length N_n of the sequence of the n -th scene is modeled as an independent and identically distributed random variables whose distribution depends on the scene type. The random variable N_n is generated according to the gamma distribution indicated in Table 4.1.

Average bit rate

The average bit rate A_n of the sequence of the n -th scene is modeled as an independent and identically distributed random variable whose distribution depends on the scene type. The random variable A_n is generated according to the gamma distribution indicated in Table 4.2.

Spike of scene

To simplify the model, we assume that the spike length M_n to be constant and equal to three. We generate the spike sequence $\{Y_n(1), Y_n(2), Y_n(3)\}$ in the following manner. The second value of a spike sequence $Y_n(2)$ is randomly selected according to the uniform distribution over the interval $(A_n, 6.144)$. Further we set $Y_n(1) = 0.83(Y_n(2) - A_n) + A_n$ and $Y_n(3) = 0.42(Y_n(2) - A_n) + A_n$.

Body sequence of scene

The body sequences of AR scenes are modeled with AR processes. As said in Subsection 4.2.5, scenes of Type 5 can't be modeled with AR processes. But even so we model them with AR processes here. Parameters of the AR process of the n -th scene other than N_n and A_n , i.e. the order m_n , the coefficients $a_{n,i}$'s, the standard deviation σ_n and the initial values $Y_n(M_n + 1), \dots, Y_n(M_n + m_n)$, are selected randomly according to scene-type dependent distributions stated below. Parameters of scenes of Type 5 are selected by using the same distributions as those for Type 2.

We randomly choose the order m_n of the AR process from two candidates, 1 and 2, with probabilities 17/81 and 64/81, respectively. These probabilities are common for all scene types.

The AR coefficients $a_{n,i}$'s are generated as follows.

1. When $m_n = 1$, we generate $a_{n,1}$ so that $1 - a_{n,1}$ is subjecting to the exponential distribution with mean 0.141.
2. When $m_n = 2$, we first generate $a_{n,2}$ according to a normal distribution with mean 0.010 and variance 0.096² (the distribution in Figure 4.11 seems to be bimodal, but we use a normal distribution here to keep the model simple). Next we generate $a_{n,1}$ so that $1 - (a_{n,1} + a_{n,2})$ is subjecting to the exponential distribution with mean 0.216.

Again, these probability distributions for $a_{n,i}$'s are common for all scene types.

To generate the standard deviation σ_n of the error term $\epsilon(k)$'s of the AR process, we first generate a value v_n according to the gamma distribution in Table 4.3. Then we set $\sigma_n = v_n \sqrt{A_n} + 0.02A_n$. This gamma distribution depends on the scene type.

For initial values $Y_n(M_n + 1), \dots, Y_n(M_n + m_n)$, we generate m_n random numbers according to the normal distribution with mean A_n and variance σ_n^2 .

4.4 Simulated sequence of coding bit rates

4.4.1 Simulation of a single source

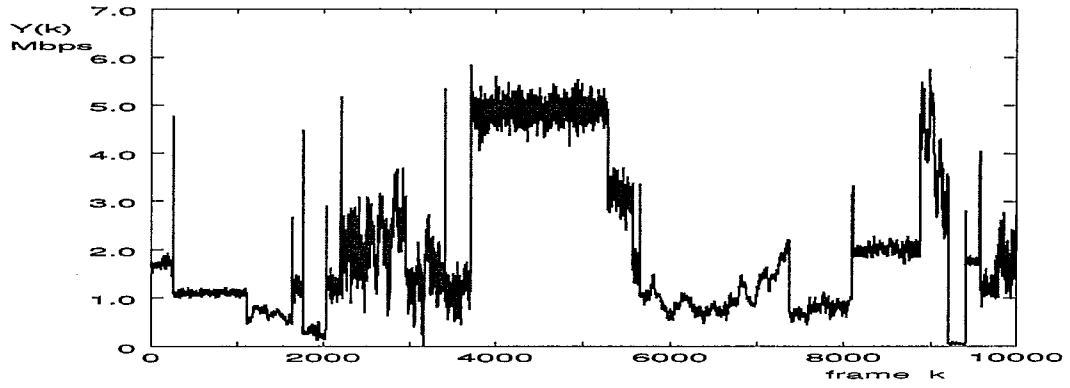


Figure 4.17: Simulated traffic of a single source

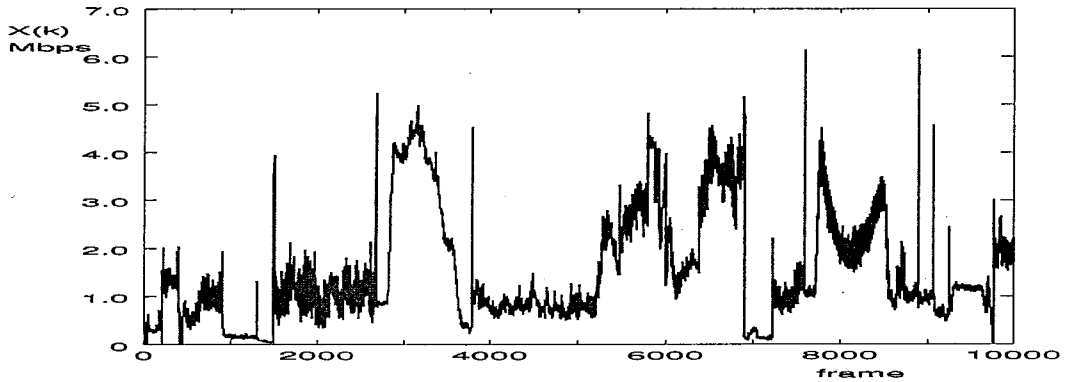


Figure 4.18: Real VBR video traffic

To check whether our model is adequate or not, we generate a sequence $\{Y(k) : k = 1, \dots, 2000000\}$ and observe its statistical property. Figure 4.17 shows the first 10,000 frames of the generated sequence. For the reference, the first 10,000 frames of the real VBR video traffic is shown in Figure 4.18. Comparing Figures 4.18 and 4.17, there is a little difference. Specifically there is no scene whose bit rate sequence shows a mountain-like curve in Figure 4.17. This is because we use AR models even

for scenes of Type 5 in our simulation. However, the author has an impression that basic behaviors of these two sequences are similar.

To make more rigorous check, we compare their marginal distributions, autocorrelation functions, R/S functions and spectral density functions.

Marginal distributions

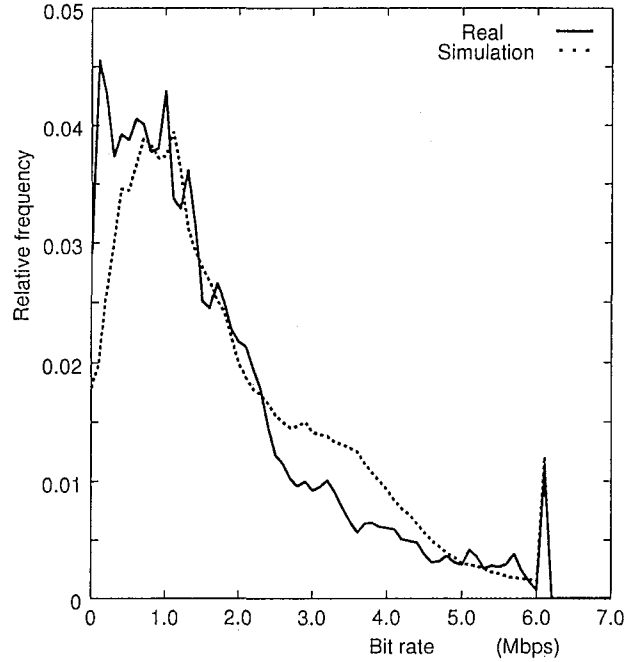


Figure 4.19: Marginal distributions of $\{X(k)\}$ and $\{Y(k)\}$

Figure 4.19 shows graphs of relative frequencies of the bit rates $\{X(k) : k = 1, \dots, 79686\}$ and $\{Y(k) : k = 1, \dots, 2000000\}$. Note that we take a far longer sequence for $\{Y(k)\}$ than $\{X(k)\}$ to estimate the marginal distribution more accurately. These two curves have similar shape. However the relative frequencies of smaller bit rates of $\{Y(k)\}$ are less than those of $\{X(k)\}$ and the relative frequencies of larger ones are a little more. This may be caused by the replacement of the model of Type 5 scenes with AR models. In scenes of Type 5, $X(k)$ is apt to start from low bit rate, increase and go over the average bit rate, and then go down to low bit

rate. Hence $X(k)$ ranges from low values to high values, but we assume that $Y(k)$ ranges near the average value by putting the process as an AR process in our model. Another reason might be the length of $\{X(k)\}$ which is not long enough to estimate the true marginal distribution.

Autocorrelation functions

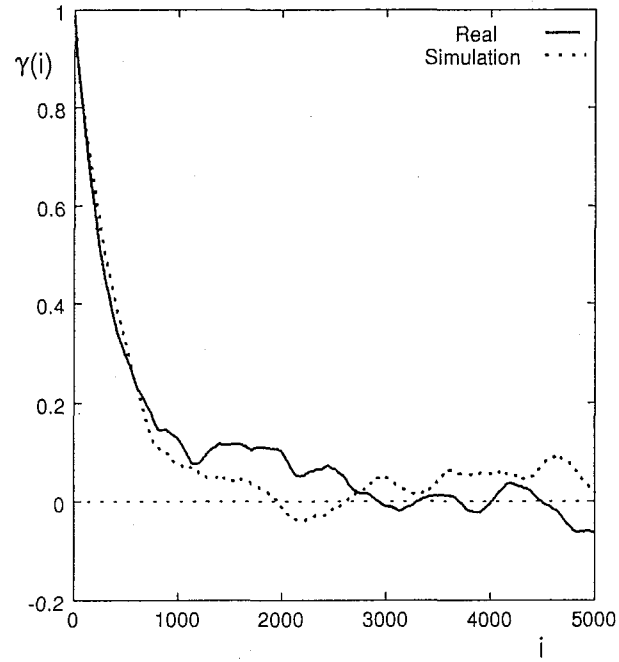


Figure 4.20: Autocorrelation functions of $\{X(k)\}$ and $\{Y(k)\}$

Autocorrelation function of a sequence $\{W(k)\}$ is defined as

$$\gamma(i) = \sum_{k=1}^{N-i} \frac{(W(k) - \mu)(W(k-i) - \mu)}{NS^2} \quad (i = 0, 1, \dots, K), \quad (4.17)$$

where

$$\mu = \sum_{k=1}^K \frac{W(k)}{K}$$

$$S^2 = \sum_{k=1}^K \frac{(W(k) - \mu)^2}{K}.$$

Figure 4.20 shows the autocorrelation functions of $\{X(k) : k = 1, \dots, 79686\}$ and $\{Y(k) : k = 1, \dots, 79686\}$. They are similar in the region of i less than 800, but in the region $i > 800$ they look different. Since these sequences have very strong dependence, it is difficult to judge whether the autocorrelation functions are similar or not in the entire region.

R/S analysis

Let the sequence $\{W(k)\}$ be a wide-sense stationary stochastic process with autocorrelation function $\gamma(i)$. The sequence $\{W(k)\}$ is said to exhibit *long-range dependence* [5] if

$$\gamma(i) \sim i^{-D} L(i), \text{ as } i \rightarrow \infty, \quad (4.18)$$

where $0 < D < 1$ and $L(i)$ is a slowly varying function in the sense

$$\lim_{t \rightarrow \infty} L(tx)/L(t) = 1, \text{ for all } x > 0. \quad (4.19)$$

From the definition, for a long-range dependent process the autocovariance function is non-summable, i. e.,

$$\sum_{i=0}^{\infty} \gamma(i) = \infty.$$

In AR processes or more general Markov processes, tail of the autocovariance function decays as $\gamma(i) \sim c\rho^{-i}$, $0 < \rho < 1$, and hence $\sum_{i=0}^{\infty} \gamma(i) < \infty$. They are called *short-range dependent processes*.

R/S analysis is used to check the long-range or short-range dependence of processes. Let $\bar{W}(u)$ be the average and $S(u)$ the standard deviation of the first u values, $W(1), \dots, W(u)$, of the sequence $\{W(k)\}$, that is

$$\begin{aligned} \bar{W}(u) &= \sum_{i=1}^u W(i)/u \\ S^2(u) &= \sum_{i=1}^u (W(i) - \bar{W}(u))^2/u. \end{aligned}$$

Then we introduce a statistic

$$\frac{R(u)}{S(u)} = \frac{1}{S(u)} [\max(0, V_1(u), V_2(u), \dots, V_u(u)) - \min(0, V_1(u), V_2(u), \dots, V_u(u))] \quad (4.20)$$

where $V_k(u) = \sum_{i=1}^k (W(i) - \bar{W}(u))$, $1 \leq k \leq u$. We refer the function $R(u)/S(u)$ in (4.20) as the *R/S function*. It is empirically known that many time series satisfy the relation

$$E[R(u)/S(u)] \sim cu^H, \text{ as } u \rightarrow \infty \quad (4.21)$$

with *Hurst parameter* H . If the process is long-range dependent, H takes a value between 0.5 and 1 as $u \rightarrow \infty$. If the process is short-range dependent like a Markov process or an AR process, H is equal to 0.5 as $u \rightarrow \infty$.

Figure 4.21 shows *R/S* functions of $\{X(k) : k = 1, \dots, 79686\}$, $\{Y(k) : k = 1, \dots, 79686\}$ and an AR process of order 2. In the graph the abscissa is for $\log_{10} u$ and the ordinate is for $\log_{10} (R(u)/S(u))$. So the gradient represents the Hurst parameter H .

For both the real traffic sequence $\{X(k)\}$ and the simulated sequence $\{Y(k)\}$ the Hurst parameter H is about 0.8, and for the AR process of order 2, H is about 0.5. Hence both the sequences $\{X(k)\}$ and $\{Y(k)\}$ seem to exhibit long-range dependence in the *R/S* analysis. On this respect the generated sequence $\{Y(k)\}$ shows a similar property to the real sequence $\{X(k)\}$.

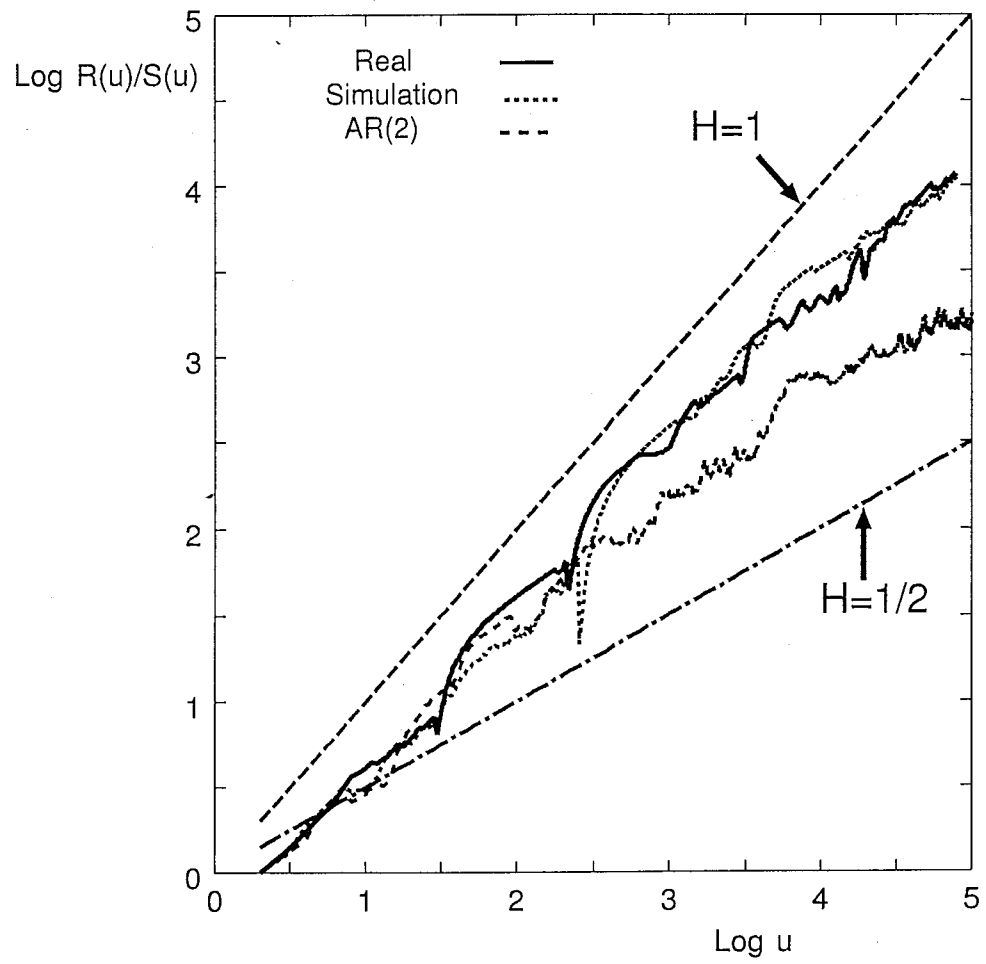


Figure 4.21: R/S functions of $\{X(k)\}$, $\{Y(k)\}$ and AR(2)

Beran et al. [4] state that VBR video traffic would be modeled as a self-similar process which is a typical long-range dependent process. A self-similar process $\{W_H(t)\}$ with parameter $0.5 < H < 1$ is a process such that

$$\begin{aligned} & \{W_H(at_1), W_H(at_2), \dots, W_H(at_m)\} \\ & \stackrel{d}{=} \{a^H W_H(t_1), a^H W_H(t_2), \dots, a^H W_H(t_m)\} \end{aligned}$$

for any $a > 0$ and for any integer m . Here “ $\stackrel{d}{=}$ ” indicates equality in distribution. If the input sequence is $\{W_H(k)\}$, then its input rate sequence $\{G_H(k)\}$ is given by $\{W_H(k) - W_H(k-1)\}$. A trite calculation shows that its autocorrelation function $\gamma(i)$ of $\{G_H(k)\}$ is given by

$$\gamma(i) = \frac{1}{2}(|i+1|^{2H} - 2|i|^{2H} + |i-1|^{2H}).$$

Since $\gamma(i) \sim H(2H-1)i^{2H-2}$ for $0.5 < H < 1$, $\{G_H(k)\}$ is a long-range dependent sequence.

One of the most popular self-similar processes is a *fractional* Brownian motion [26] defined by

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^t K(s, t) dB(s),$$

where $B(t)$ is a Brownian motion (a Wiener process), $\Gamma(x)$ is the gamma distribution and $K(s, t)$ is a function defined by

$$K(s, t) = \begin{cases} (t-s)^{H-0.5}, & 0 \leq s \leq t, \\ (t-s)^{H-0.5} - (-s)^{H-0.5}, & s < 0. \end{cases}$$

Figures 4.22 and 4.23 show a sample sequence of $\{G_{0.8}(k)\}$ given by $\{B_{0.8}(k) - B_{0.8}(k-1)\}$ and its R/S function, respectively. The R/S function is similar to that of the real data given in Figure 4.21. However, the sample sequence is much different from that of the real data presented in Figure 4.18.

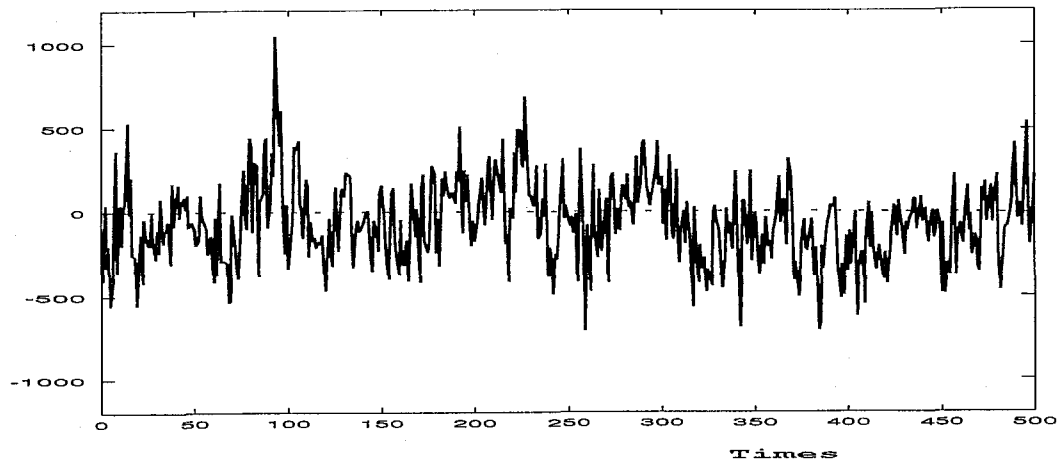


Figure 4.22: Sample sequence of $\{G_{0.8}(k)\}$

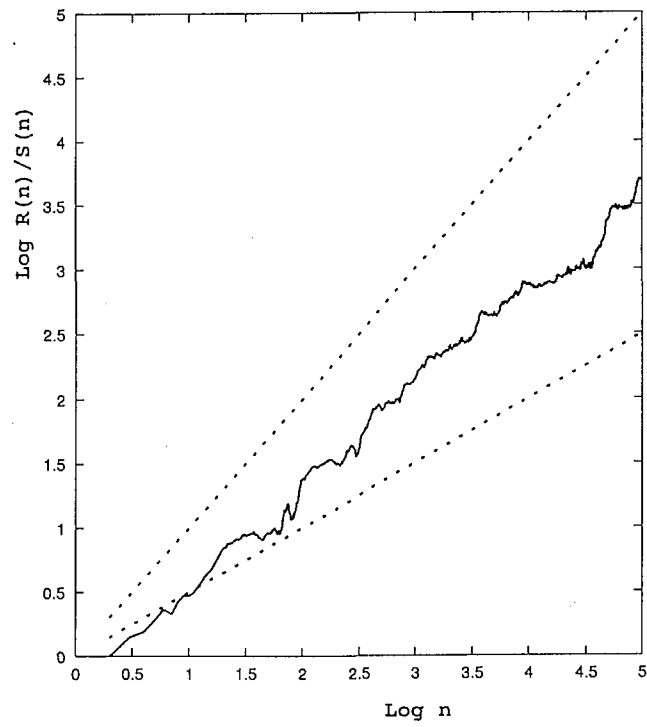


Figure 4.23: R/S function of $\{G_{0.8}(k)\}$

Spectrum analysis

Let X_t be a wide-sense stationary process and $\Lambda(t)$ be the autocovariance function of X_t . The spectral density function $S_p(f)$ of X_t is defined by the Fourier transform of $\Lambda(t)$. However, we rarely estimate a spectral density function from an empirical autocovariance function. We usually estimate it through the *periodogram* defined as

$$\hat{S}_T(f) = \frac{|\mathcal{X}_T(f)|^2}{T} \quad \text{where} \quad \mathcal{X}_T(f) = \int_0^T X_t e^{-i2\pi f t} dt.$$

There exists a relation between $\hat{S}_T(f)$ and $S_p(f)$,

$$\lim_{T \rightarrow \infty} \int_{f_0-\delta/2}^{f_0+\delta/2} \hat{S}_T(f) df = \int_{f_0-\delta/2}^{f_0+\delta/2} S_p(f) df,$$

so we can estimate $S_p(f_0)$ for a frequency f_0 as

$$S_p(f_0) \approx \frac{1}{\delta} \int_{f_0-\delta/2}^{f_0+\delta/2} \hat{S}_T(f) df.$$

The constant δ is called *bandwidth*.

In this section, we also estimate the spectral density function of a sequence $\{x(n)\}$ of length N through the periodogram as follows:

1. Using Fast Fourier Transform (FFT), obtain the Fourier transform $\hat{x}(f_k) = \sum_{n=1}^N x(n) e^{i2\pi f_k(n-1)}$, $k = 1, \dots, N/2$, of the sequence where $f_k = k/N$.
2. Obtain the periodogram $I(f_k) = \frac{|\hat{x}(f_k)|^2}{N}$, $k = 1, \dots, N/2$.
3. Average $I(f_k)$ with bandwidth of l/N .

In subsequent spectrum analyses we take $N = 65536 (= 2^{16})$ and $l = 32$.

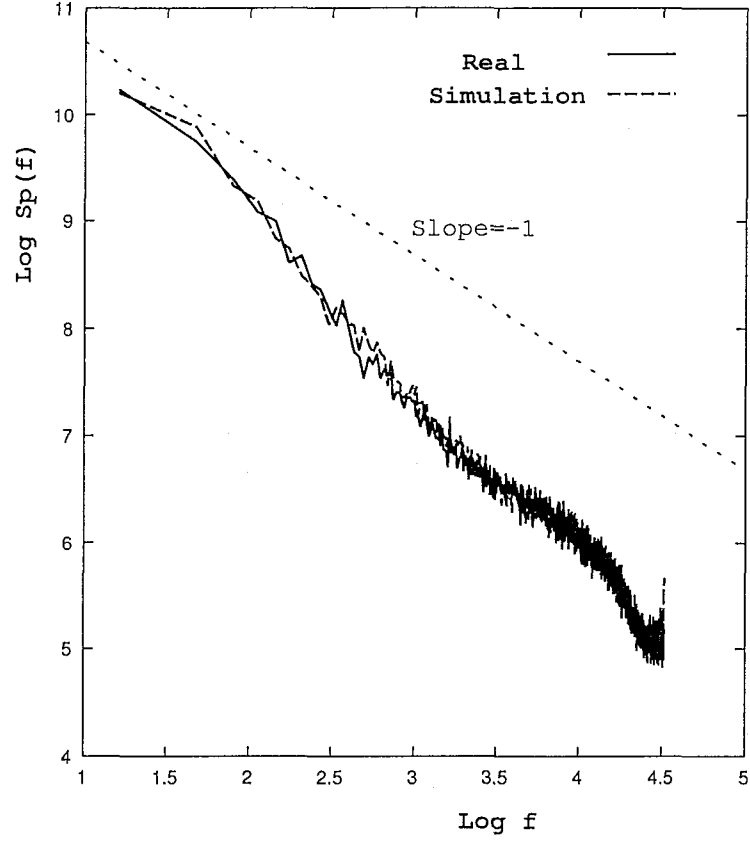


Figure 4.24: Spectral density functions of $\{X(k)\}$ and $\{Y(k)\}$

Figure 4.24 shows the estimated spectral density functions of $\{X(k)\}$ and $\{Y(k)\}$. In the figure, the abscissa presents $\log f$ and the ordinate $\log S_p(f)$. The spectral density functions of $\{X(k)\}$ and $\{Y(k)\}$ are very similar and $\log S_p(f)$ decrease at the rate of less than -1 in terms of $\log f$.

Figure 4.25 shows the estimated spectral density function of an input rate sequence $\{G_{0.8}(k)\}$ from a fractional Brownian motion $\{B_{0.8}(k)\}$. Clearly, it is different from that of the real data $\{X(k)\}$ since the decay rate is more than -1 . Theoretically, its decay rate is -0.6 when $H = 0.8$. The figure indicates that the gradient of $\log S_p(f)$ is about -0.6 .

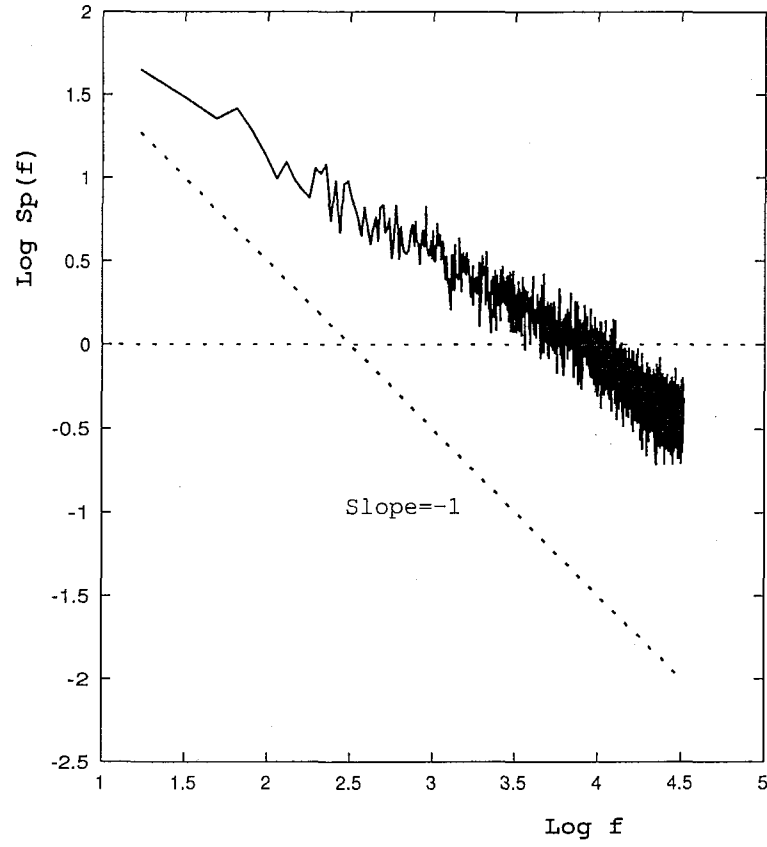


Figure 4.25: Spectral density function of $\{G_{0.8}(k)\}$

We have compared marginal distributions, autocorrelation functions, R/S functions and spectral density functions. Their results indicate that the simulated sequence $\{Y(k)\}$ has a similar statistical property to the real sequence $\{X(k)\}$ in all of these aspects. Hence our simulation model can be used for simulation studies of ATM multiplexers.

4.4.2 Simulation of multiple sources

In an ATM multiplexer, traffics from multiple sources are poured into the buffer. We will examine statistical property of the aggregated bit rate sequence. We generate 50 sequences $\{Y^h(k) : k = 1, \dots, 82000\}$, $h = 1, \dots, 50$, and aggregate them. To normalize it, we divide the aggregated bit rate by 50 as

$$Y^{(50)}(k) = \frac{\sum_{h=1}^{50} Y^h(k)}{50}, k = 2001, \dots, 82000. \quad (4.22)$$

The first 2000 frames of $\{Y^{(50)}(k)\}$ are not used to eliminate initial bias.

Figure 4.26 shows a sample sequence of $\{Y^{(50)}(k) : k = 2001, \dots, 12000\}$. In the figure, spikes at scene changes disappear and $\{Y^{(50)}(k)\}$ looks like a stationary Gaussian process. Checking the marginal distribution of $\{Y^{(50)}(k)\}$, it becomes close to a normal distribution as is shown in Figure 4.27.

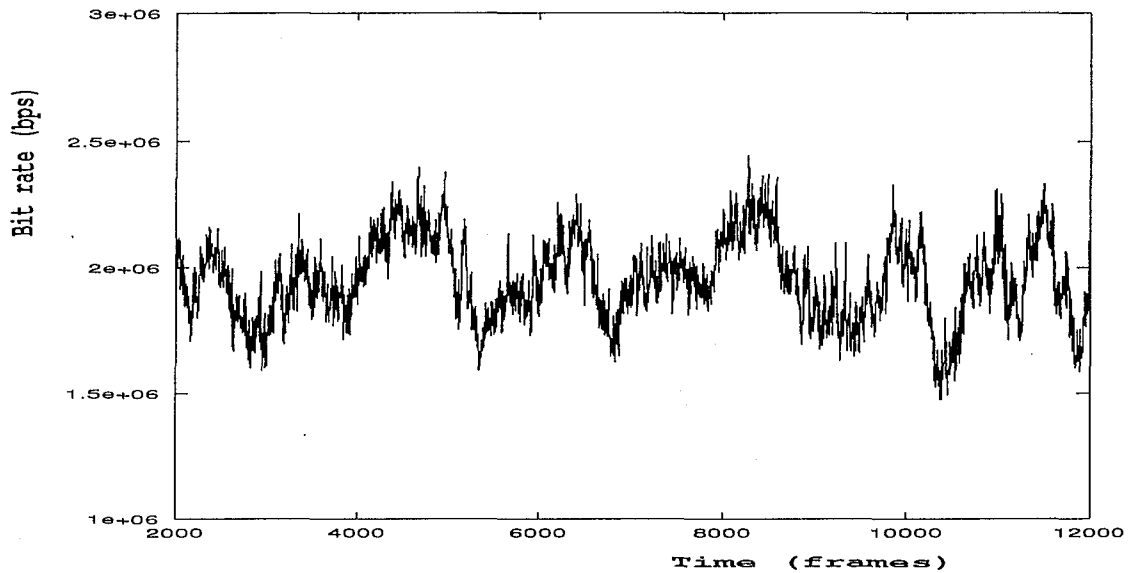


Figure 4.26: Aggregated bit rate from 50 sources

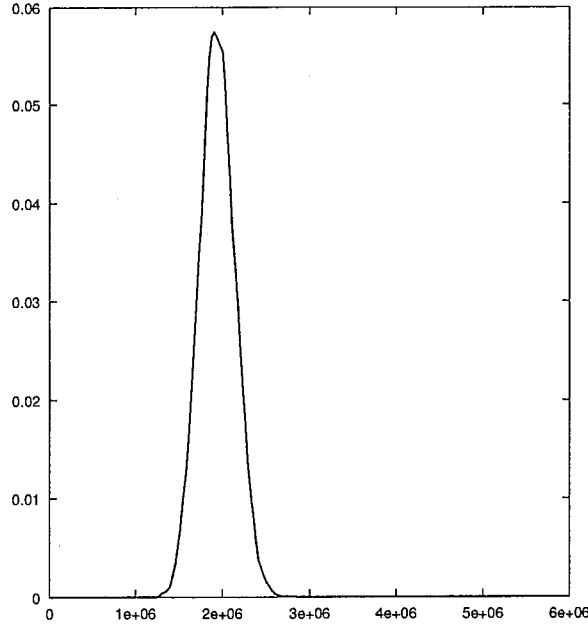


Figure 4.27: Marginal distribution of $\{Y^{(50)}(k)\}$

R/S analysis

We make *R/S* analysis to the sequence $\{Y^{(50)}(k)\}$. The Hurst parameter H for the sequence $\{Y^{(50)}(k)\}$ is about 0.8 again, as seen in Figure 4.28. Thus, it also shows long-range dependence in *R/S* analysis.

Spectrum analysis

The sequence $\{Y^{(50)}(k)\}$ is close to a stationary Gaussian process and exhibits long-range dependence in *R/S* analysis. The input rate sequence $\{G_H(k)\}$ from a fractional Brownian motion $\{B_H(k)\}$ also has the same properties. Thus, we examine whether $\{Y^{(50)}(k)\}$ is close to $\{G_H(k)\}$ or not, by comparing the spectral density functions of $\{Y^{(50)}(k)\}$ and $\{G_{0.8}(k)\}$. In Figure 4.29, $\log S_p(f)$ of $\{Y^{(50)}(k)\}$ decreases at a rate less than -1 as well as that of $\{Y(k)\}$. On the other hand, that of $\{G_{0.8}(k)\}$ decreases at a rate more than -1 . This result shows that $\{Y^{(50)}(k)\}$ is not similar to $\{G_{0.8}(k)\}$.

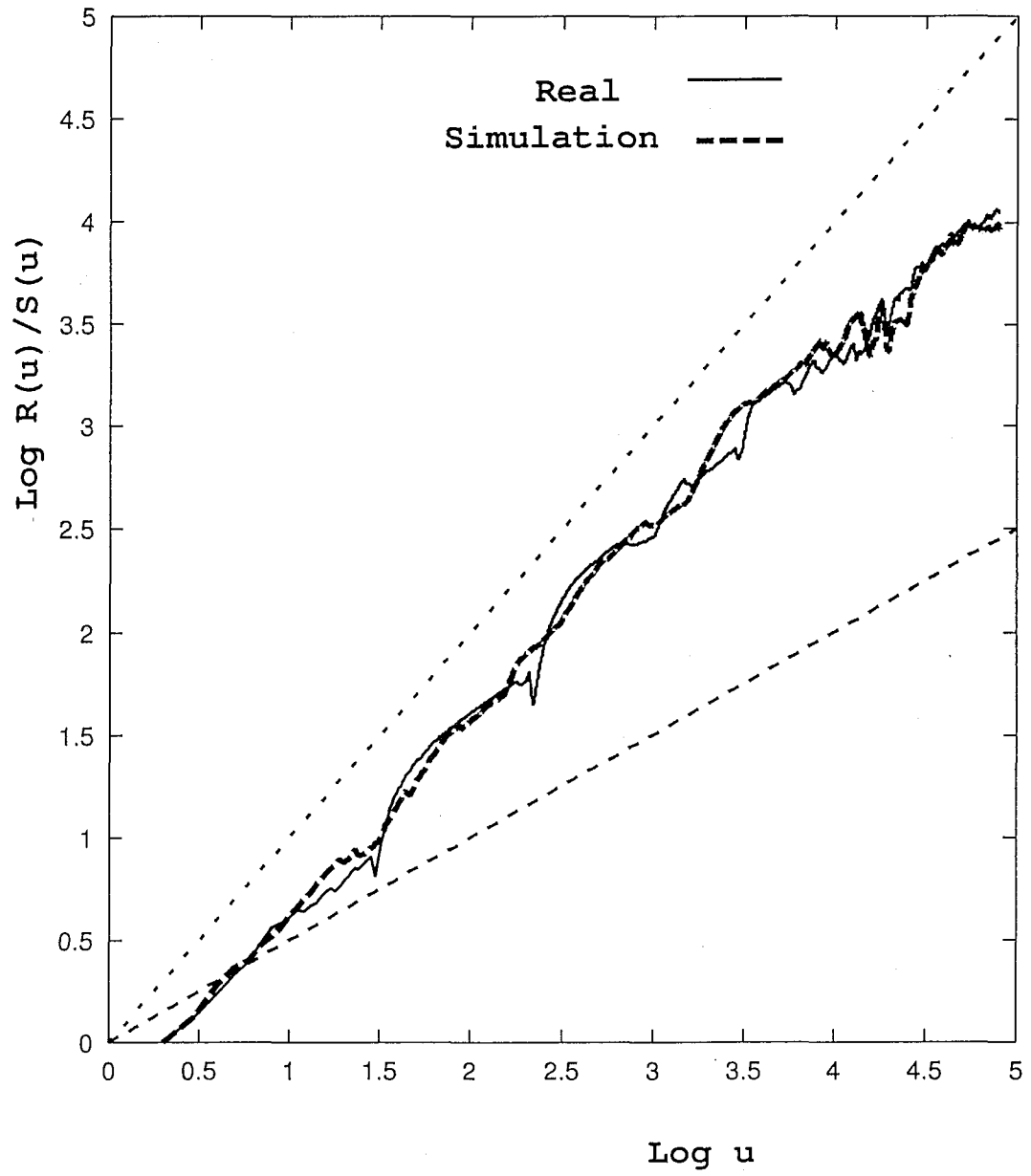


Figure 4.28: R/S functions of $\{Y^{(50)}(k)\}$ and the real data $\{X(k)\}$

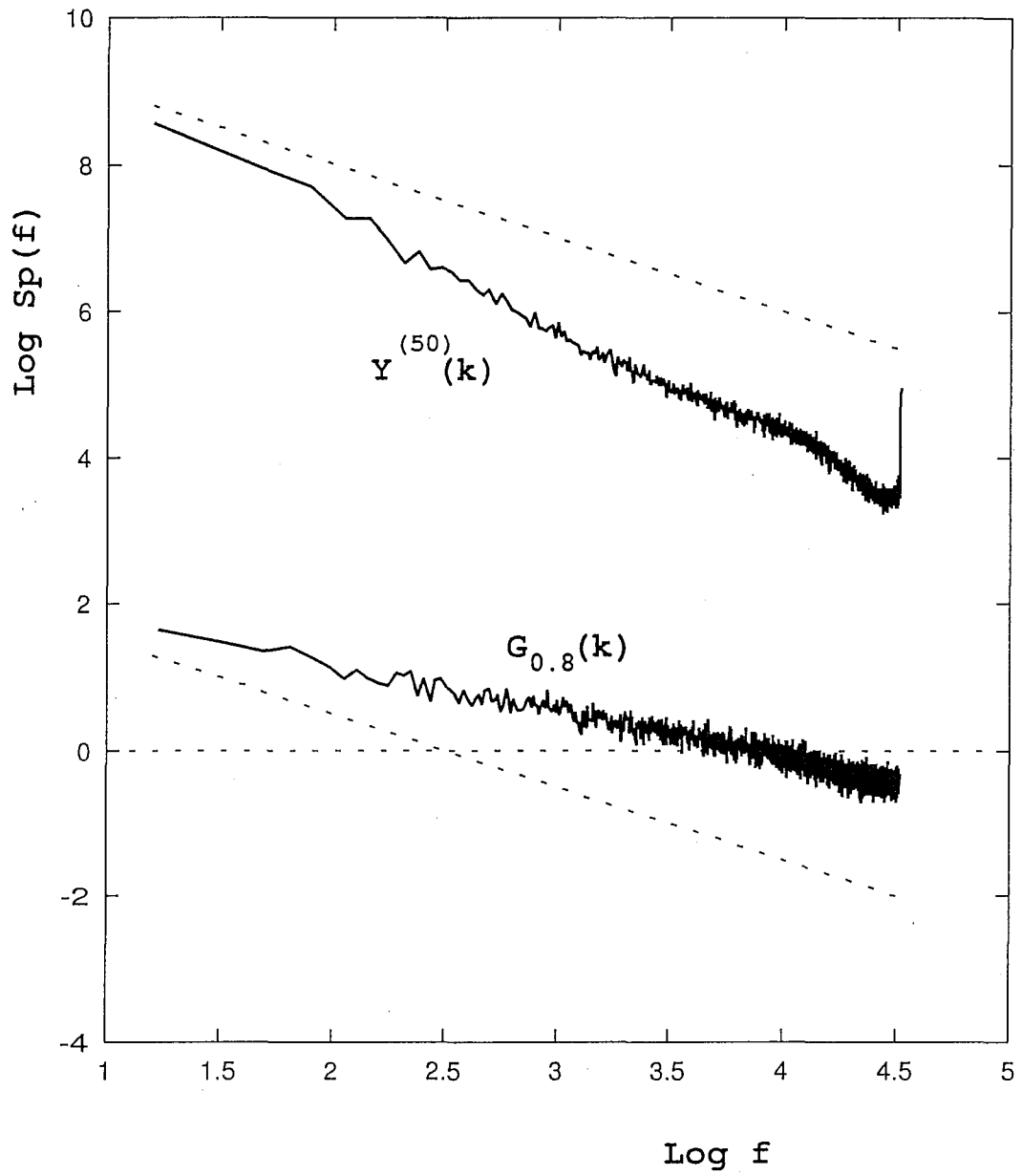


Figure 4.29: Spectral density functions of $\{Y^{(50)}(k)\}$ and $\{G_{0.8}(k)\}$

4.5 Tail distribution of an ATM multiplexer

We shall examine the tail behavior of the steady-state queue length distribution of the ATM multiplexer model introduced in Chapter 2 using a bit rate sequence generated by the simulation model. We also compare it with the approximation formula derived in Chapter 2,

$$P[Q > x] = B \exp\left[-\frac{C - m}{H(0)^2/2}x\right], \quad (4.23)$$

where

$$B = \frac{\phi(\theta)}{\theta\Phi(\theta) + \phi(\theta)}, \quad \theta = \frac{C - m}{\sigma_R}, \quad (4.24)$$

and

$$H(0)^2/2 = \int_0^\infty \Lambda(t)dt. \quad (4.25)$$

Here, C is the output rate and m , σ_R and $\Lambda(t)$ are the mean, the standard deviation and the autocovariance function of the input rate process.

If the input rate process is strictly a long-range dependent one, $H(0)^2/2$ becomes infinite and hence the approximation formula (4.23) is not applicable. In this section, we do not discuss this point. We estimate $H(0)^2/2$ from autocorrelations over a finite range as indicated below.

The parameters of approximation formula (4.23) at N sources are related to those at a single source as follows;

$$m^{(N)} = N m^{(1)}, \quad \sigma_R^{(N)} = \sqrt{N} \sigma_R^{(1)}, \quad (4.26)$$

and

$$\{H^{(N)}(0)\}^2/2 = N \{\sigma_R^{(1)}\}^2 \sum_{i=0}^{2000} \gamma^{(1)}(i)/30. \quad (4.27)$$

In our numerical experiments, we use parameter values $m^{(1)} = 1.933432 \times 10^6$, $\sigma_R^{(1)} = 1.4118186 \times 10^6$ and $\sum_{i=0}^{2000} \gamma^{(1)}(i) = 378.63999$.

Simulation results

Using a discrete version of a fluid queue model, $Q_{n+1} = \max\{0, Q_n + (R_n - C)\Delta t\}$, we simulate the steady-state queue length distributions when the number of sources N is 50, 150 and 250. We set the output rates $C = m^{(N)} + \theta \sigma_R^{(N)}$ where $\theta = 2$ and $\theta = 3$.

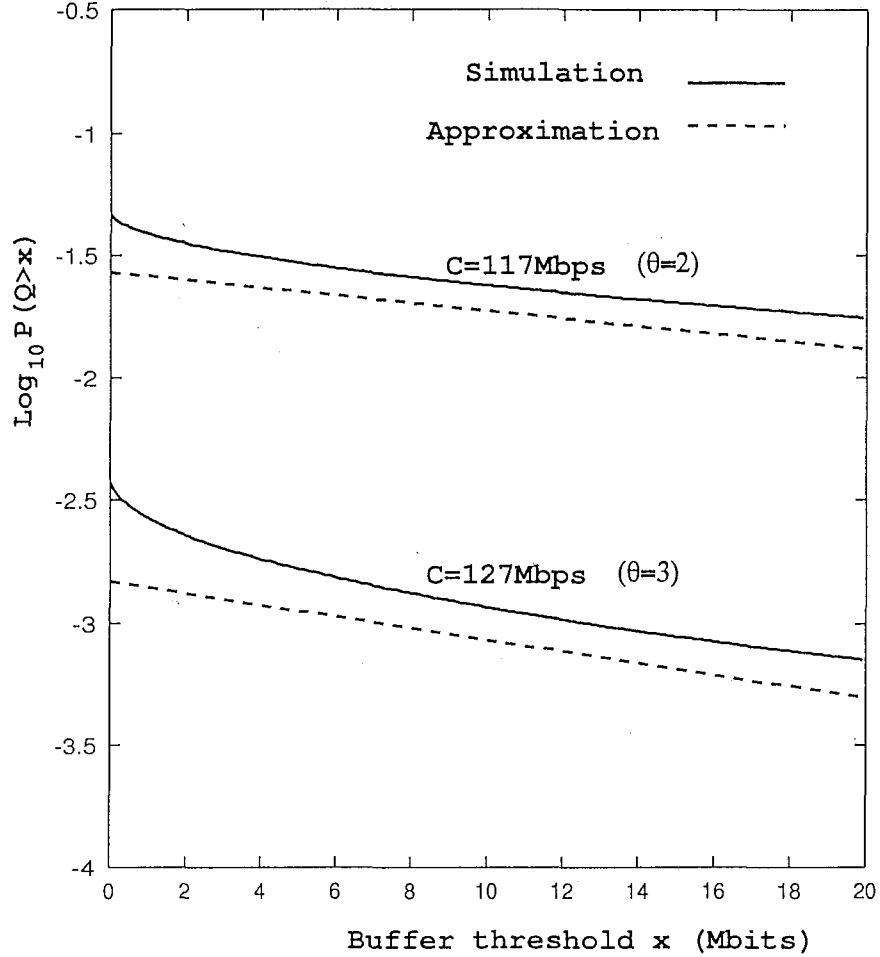


Figure 4.30: Tail distributions of a fluid queue with 50 sources

Figure 4.30 shows simulation and approximation results when $N = 50$. The ordinate presents $\log_{10} P(Q > x)$. The threshold $x = 10\text{Mbits}$ corresponds to 23500cells since one cell is 424bits. Our approximation underestimates the simulated tail dis-

tribution. However, their decay rates almost coincide with each other. The underestimation may be caused by deviating the marginal distribution of $\{Y^{(50)}(k)\}$ from a normal distribution.

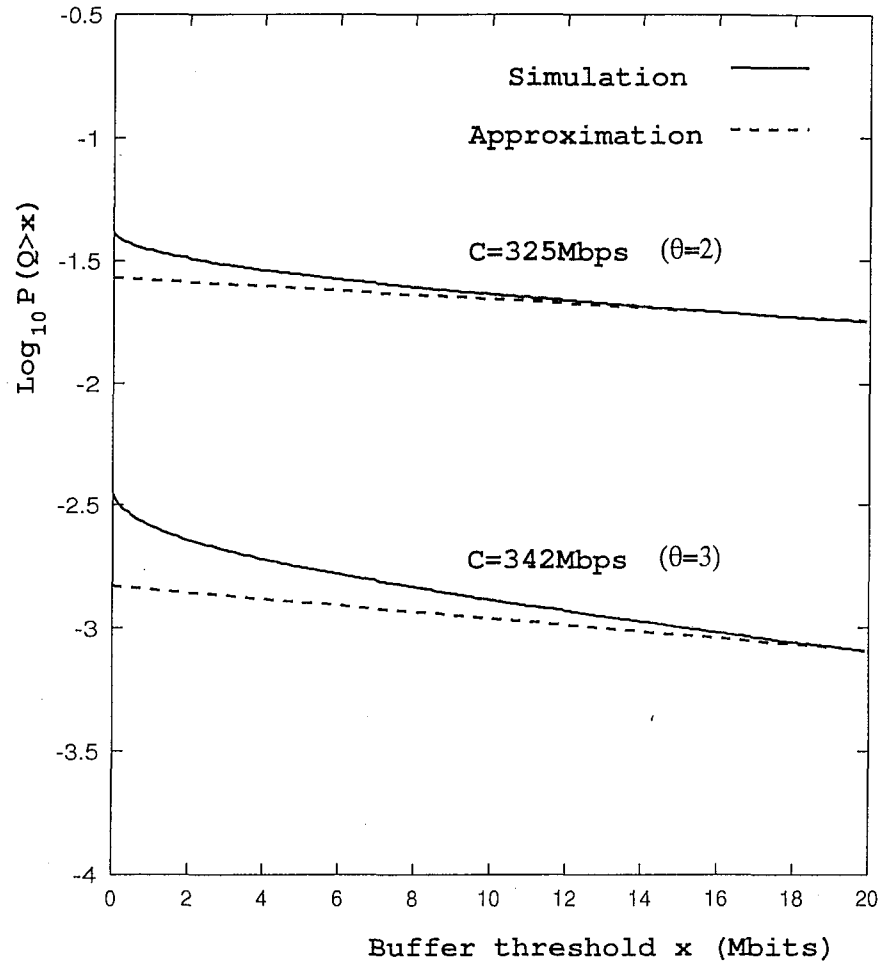


Figure 4.31: Tail distributions of a fluid queue with 150 sources

Figure 4.31 shows simulation and approximation results when $N = 150$. Compared with the cases of $N = 50$, the approximation is closer to the simulation result. The curve by the simulation approaches the line of the approximation as x becomes large.

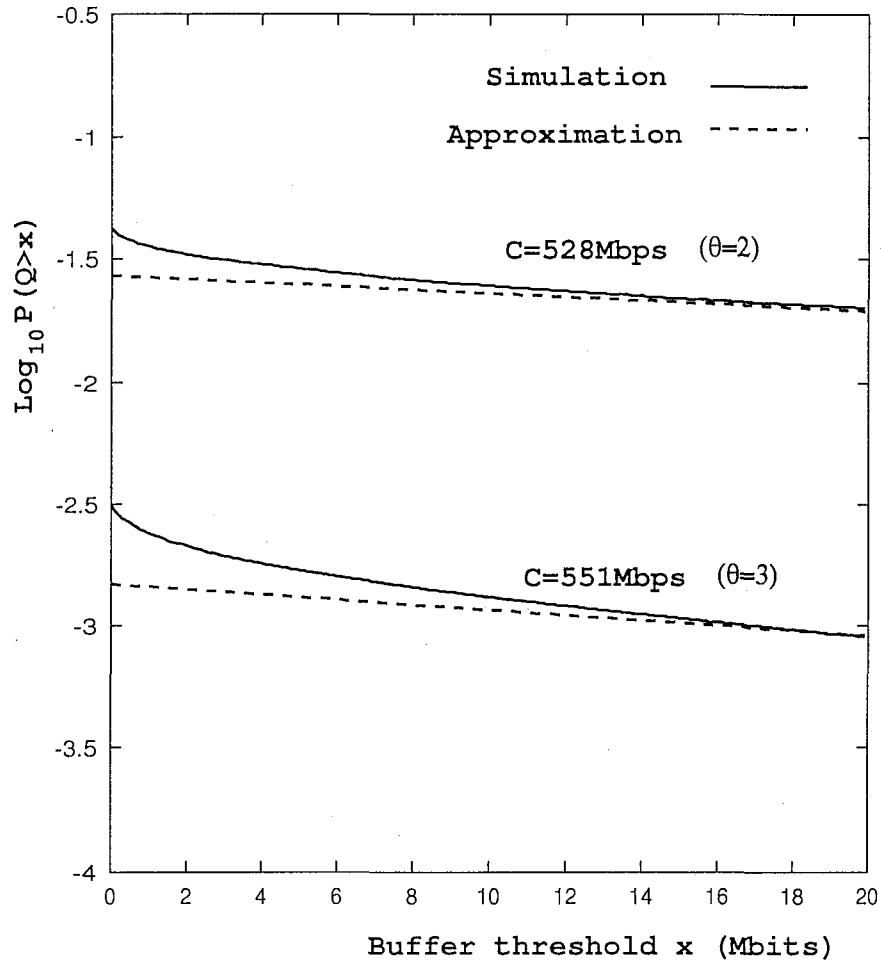


Figure 4.32: Tail distribution of a fluid queue with 250 sources

Figure 4.32 shows simulation and approximation results when $N = 250$. This figure also indicates that the curve by the simulation approaches the line of the approximation as x becomes large.

4.6 A discussion on long-range dependence

In the previous section, we showed that the approximation formula (4.23) derived in Chapter 2 provides a useful tool to analyze and estimate congestion in the buffers of ATM multiplexers, by implicitly assuming that VBR video traffic is short-range

dependent, namely, $H(0)^2/2 = \int_0^\infty \Lambda_R(s)ds < \infty$. Here we shall discuss the problem whether VBR video traffic is short-range dependent or long-range dependent.

J. Beran et al. [4] analyze 20 large sets of actual VBR video traffic data, using R/S analysis and variance-time analysis. Based on the analyses, they say “long-range dependence is an inherent characteristic of VBR video traffic” and propose to use self-similar processes such as fractional Brownian motions and fractional ARIMA processes for modeling of VBR video traffic, since self-similar processes are long-range dependent in the sense of the definition (4.18).

However, as to the Berans’ assertion, there are some questions. They are

1. Is VBR video traffic really long-range dependent in the sense of the definition (4.18)?
2. Does VBR video traffic have self-similar nature?

Concerning to the first question, generally speaking, we cannot decide rigorously whether the traffic is long-range dependent or not from a data of finite length. We can only guess whether it is long-range dependent or not by using R/S analysis or variance-time analysis. We have observed in subsection 4.4.1 that a real VBR video traffic data exhibits long-range dependence in R/S analysis. Its Hurst parameter is about 0.8. However, this result does not necessarily mean that the traffic is long-range dependent in the sense (4.18). In order to clarify whether the traffic is long-range dependent or not in the rigorous sense, we need to find causes of long-range dependence in R/S analysis. If the traffic had self-similar nature, it would be long-range dependent in the rigorous sense. But does it really has self-similar nature? This is the second question.

The real traffic data shown in Figure 4.1 does not seem to be a self-similar process. Spikes appear at moments of scene changes and the behavior of the traffic differs from scene to scene. As seen in Figure 4.21 the data exhibits long-range dependence in R/S analysis, though. In section 4.2, we analyzed the data statistically by taking account of scene changes, and in section 4.3 we developed a simulation

model consisting of three submodels, a Markovian scene transition model, a spike model and an AR model. A traffic sequence was generated using the simulation model and tested for its long-range dependence in R/S analysis. The result shows that it certainly has long-range dependence in the sense of R/S analysis as in Figure 4.21. We should notice that our simulation model consists of only short-range dependent processes and doesn't use any processes having self-similar nature.

Thus a single traffic sequence is far from a self-similar process. If we consider an aggregated sequence of multiple traffic sequences, it might be nearer to a self-similar process since the aggregation eliminates individual irregularity and makes marginal distribution close to the normal distribution.

In subsection 4.4.2, we examined the statistical property of an aggregated sequence generated by the simulation model. In the sequence the effects of scene changes almost disappear and the marginal distribution is almost normal as being expected. It also exhibits long-range dependence in R/S analysis. In these aspects, the aggregated sequence is similar to the input rate sequence $\{G_H(k)\}$ from a fractional Brownian motion $\{B_H(k)\}$, a typical self-similar process. However in the spectrum analysis in subsection 4.4.2, their spectral density functions are much different.

Further, in section 4.5, we examined the tail behavior of the steady-state queue length distribution of an ATM multiplexer using the aggregated sequence. The result shows that the approximation formula (4.23) provides good approximate values even when $H(0)^2/2$ is estimated from finite autocorrelations.

These various facts seem to suggest that VBR video traffic does not have self-similar nature and that the traffic is not long-range dependent in the sense (4.18). A key to this point is, as mentioned before, to detect what makes VBR video traffic long-range dependent in R/S analysis. By examining our model further, it might be possible to find the cause and to make clear whether the traffic is long-range dependent or not in the rigorous sense. This is a subject for future study.

Bibliography

- [1] H. Akaike, A New Look at the Statistical Model Identification, *IEEE Trans. on Automatic Control* **19** (1974) 716-723.
- [2] D. Anick, D. Mitra and M. M. Sondhi, Stochastic Theory of a Data-Handling System with Multiple Sources, *BSTJ* **61** (8) (1982) 1871-1894.
- [3] V. E. Benes, *General Stochastic Processes in the Theory of Queues* (Addison-Wesley, Reading, MA, 1963).
- [4] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, Long-Range Dependence in Variable-Bit-Rate Video Traffic, *IEEE Trans. on Communications* **43** (2/3/4) (1995) 1566-1579.
- [5] D. R. Cox, Long-Range Dependence: A Review, *Statistics: An Appraisal* H. A. David and H. T. David, Eds., (The Iowa State University Press, Ames, Iowa, 1984) 55-74.
- [6] K. Debicki and T. Rolski, A Gaussian Fluid Model, *Queueing Systems* **20** (1995) 433-452.
- [7] W. Fischer and K. Meier-Hellstern, The Markov-Modulated Poisson Process (MMPP) Cook-book, *Performance Evaluation* **18** (1992) 149-171.
- [8] H. J. Fowler and W. E. Leland, Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management, *IEEE J. on Sel. Areas in Commun.* **9** (1991) 1139-1149.

- [9] C. W. J. Granger and R. Joyeux, An Introduction to Long-Memory Times Series Models and Fractional Differencing, *J. Time Series Anal.*, **1** (1980) 15-29.
- [10] R. Grünenfelder, 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.
- [11] O. Hashida and M. Fujiki, Queueing Models for Buffer Memory in Store-and-Forward System, *Proc. ITC-7* (1973) 323/1-323/7.
- [12] H. Heffes and D. M. Lucantoni, A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance, *IEEE J. Sel. Areas of Commun.* **4** (6) (1986) 856-868.
- [13] J. M. Harrison, *Brownian Motion and Stochastic Flow Systems* (Wiley, New York, 1985).
- [14] 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.
- [15] H. Kobayashi, Performance Issues of Broad-band ISDN, Part II: Statistical Multiplexing of Multiple Types of Traffic, *Proc. the 10th ICC* (1990) 349-361.
- [16] H. Kobayashi and Q. Ren, A Diffusion Approximation Analysis of an ATM Statistical Multiplexer with Multiple Types of Traffic, Part I: Equilibrium State Solutions, *Proc. ICC'93* (2) (1993) 1047-1053.
- [17] K. Kobayashi and Y. Takahashi, Steady-State Analysis of ATM Multiplexer with Variable Input Rate through Diffusion Approximation, *Performance Evaluation* **23** (2) (1995) 145-162.

- [18] K. Kobayashi and Y. Takahashi, The Tail Behavior of the Stationary Distribution of a Fluid Queue with a Gaussian-type Input Rate Process, *J. of O. R. Society of Japan* (to appear).
- [19] L. Kosten, Stochastic Theory of Data Handling Systems with Groups of Multiple Sources, *Performance of Computer-Communication Systems* (North-Holland, 1984) 321-331.
- [20] A. Kuczura, The Interrupted Poisson Process as an Overflow Process, *BSTJ* **52** (1973) 437-450.
- [21] W. Leland, LAN Traffic Behavior from Milliseconds to Days, *7th ITC Specialists' Seminar*, Morristown (1990).
- [22] W. Leland, M. Taqqu, W. Willinger and D. Wilson, On the Self-Similar Nature of Ethernet Traffic, *IEEE/ACM Trans. on Networking* **2** (1994) 1-15.
- [23] D. Lucantoni, K. S. Meier-Hellstern and M. F. Neuts, A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes, *Adv. Appl. Prob.* **22** (2) (1990) 676-705.
- [24] F. Machihara, Completion Time of Service Unit Interrupted by PH-Markov Renewal Customers and its Applications, *Proc. ITC-12* (1988) 5.4B.
- [25] 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) 834-884.
- [26] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian Motions, Fractional Noises and Applications, *SIAM Review* **10** (4) (1968) 422-437.
- [27] M. Nomura, T. Fuji and N. Ohta, Basic Characteristics of Variable Rate Video Coding in ATM Environment, *IEEE J. Sel. Areas of Commun.* **7** (5) (1989) 752-760.

- [28] I. Norros, A Storage Model with Self-Similar Input, *Queueing Systems* **16** (1994) 387-396.
- [29] T. Okuda, H. Akimaru, A Simplified Performance Evaluation for Packetized Voice Systems, *IECE Trans.* **E73** (6) (1990) 936-941.
- [30] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, New Jersey, 1975).
- [31] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).
- [32] N. U. Prabhu, *A Study of Their Basic Stochastic Processes* (John & Sons, 1965).
- [33] J. Rissanen, Modeling by Shortest Data Description, *Automatica*, **14** (1978) 465-471.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] T. Tanaka, O. Hashida and Y. Takahashi, Transient Analysis of Fluid Model for ATM Statistical Multiplexer, *Performance Evaluation* **23** (2) (1995) 145-162.
- [38] E. Wong, *Stochastic Processes in Information and Dynamical Systems* (McGraw-Hill, New York, 1971).