Effective Bandwidths for TES Processes

Jiangbin Yang, Ioannis Lambadaris[*]

Broadband Networks Laboratory

Dept. of Systems & Computer Engineering

Carleton University

Ottawa, Ontario K1S 5B6, Canada Email: (ioannis,yang)@sce.carleton.ca

February 12, 2001

Abstract

In this paper, we derive the effective bandwidth functions of TES (Transform-Expand-Sample) processes. These processes are particularly important for modeling of video and other multimedia traffic patterns found in broadband networks. Such processes are highly autocorrelated and have general marginal distributions. To examine the impact of autocorrelation on effective bandwidths and illustrate the computational procedure, we compute the effective bandwidth functions of some specific TES processes with different autocorrelation structures. We simulate some queues with TES processes as input, numerically study the exponential decay of their queue length tail probabilities, and show how well the derived effective bandwidth functions can characterize the exponential decay rates.

Keywords: Traffic Modeling, Transform Expand Sample (TES), Effective Bandwidth

[*]Corresponding author

1

1 Introduction

TES (Transform-Expand-Sample) is a versatile methodology for modeling time series with general marginal distributions and a broad range of autocorrelation functions [7]. The marginal distributions and autocorrelation functions may be estimated from empirical data, or specified theoretically. For any given marginal distribution and autocorrelation function, TES can produce a sequence of random variates exactly matching the marginal distribution and closely approximating the autocorrelation function. A sequence of random variates generated by TES is called a TES process.

TES has been applied to video traffic modeling, where the objective is to find good mathematical models for video frame size sequences. Generally, video frame sizes from a source are autocorrelated. Autocorrelation has significant impact on the performance of a queueing system. Therefore, accurate modeling of video traffic should not only capture the marginal distribution of the frame sizes, but also more importantly the characteristics of the autocorrelation function. TES is a convenient tool capable of doing both simultaneously. [6] and [10] studied TES modeling of variable bit rate (VBR) compressed video traffic. [12] and [3] studied a composite TES model for MPEG coded video traffic. [9] studied a Markov-Renewal-Modulated TES model for modeling a full-length JPEG video traffic trace.

Effective bandwidth is a traffic descriptor, which characterizes how much bandwidth should be allocated to an input traffic process in order to satisfy certain quality of service (QoS) requirements for buffer overflow probability, queueing delay, etc.. A systematic review of effective bandwidth was given by [4] and references therein. Effective bandwidth has been applied to admission control and resource management in works such as [1] and [13].

Effective bandwidth of a traffic process is a function of two free parameters representing a space and time scaling respectively. Usually, we consider the limiting form of an effective bandwidth function for the time parameter going to infinity, and leave it as a function of the space parameter only. Effective bandwidth functions of some traffic processes have been derived by [5], [1] and [4].

For a stationary and ergodic process, the inverse of the effective bandwidth function (i.e., the value of the space parameter for given effective bandwidth) is the asymptotic exponential decay rate of the queue length tail probability of a queue with the process as input and the effective bandwidth as service rate (a discussion and references on this can be found in [4]). The space parameter is called so because it characterizes the asymptotic queue length distribution. By determining the exponential decay rate using the effective bandwidth function, we can fairly evaluate the queueing performance.

Application of effective bandwidth is not limited to a single traffic process. More importantly, effective bandwidth is a convenient tool for queueing performance analysis of multiplexed or aggregated traffic processes. The effective bandwidth function of aggregated independent traffic processes is the sum of the effective bandwidth functions of the individual traffic processes, which may be of common or different traffic classes such as voice, video or data [4]. Therefore, knowing the effective bandwidth function for each class of traffic processes, we can easily obtain the effective bandwidth function of the aggregated traffic processes, and use it to study the queueing performance of the multiplexed or aggregated traffic processes.

The purpose of this paper is to derive the effective bandwidth function of general TES processes. The rest of the paper is organized as follows. In Section 2, by approximating TES with its quantized version, we derive the effective bandwidth function for TES when the traffic process is considered over a finite time horizon (i.e., the time parameter is finite). In Section 3, we prove that the effective bandwidth exists as the time parameter goes to infinity and derive the limiting form. In Section 4, in examples, we compute the effective bandwidth functions of some specific TES processes with different autocorrelation structures. In Section 5, we simulate some queues with TES processes as input, numerically study the exponential decay of the queue length tail probabilities, and compare the exponential decay rate with the result of effective bandwidth. In Section 6, we summarize our results and conclude.

2 Derivation of the Effective Bandwidths over Finite Time Horizon

A TES process may be a TES$^+$, TES$^-$, or extended TES process with a Markov chain innovation sequence [7] [8]. A TES$^+$ process gives rise primarily to a positive lag-1 autocorrelation, and a TES$^-$ process a negative lag-1 autocorrelation. When there is a structure in the burstiness of a traffic process, a Markov chain innovation sequence can be used to better capture the structure. In this paper, we focus our derivation and analysis on TES$^+$ processes. Derivation of effective bandwidth functions for TES$^-$ and the extended TES with a Markov chain innovation sequence can be done in a similar but more complicated way, and will be presented in future works.

Let $\{Z_n : n = 0, 1, 2, \cdots\}$ denote a TES$^+$ process. This process can represent a sequence of video frame sizes for video traffic modeling (for example, [6], [12] [3], and [9]). Using similar notations of [7],

$$Z_n = D(U_n^+),$$

$$D(\cdot) = F_Z^{-1}(S_\xi(\cdot)),$$

$$U_n^+ = \begin{cases} U_0, & n = 0, \\ < U_{n-1}^+ + V_n >, & n \geq 1. \end{cases}$$

$F_Z(\cdot)$ is the marginal probability distribution function of Z_n, $< \cdot >$ is the modulo-1 operator[1], and $S_\xi(\cdot)$ is a stitching transformation[2]. U_0 is a continuous uniform random variable in $[0, 1)$, and $\{V_n : n = 1, 2, \cdots\}$ is an innovation sequence of independent and identically distributed (i.i.d.) random variables, which can have an arbitrary distribution in $(-\infty, \infty)$. Note that no matter what distribution V_n has, the marginal distribution of $\{U_n^+\}$ is always uniform on $[0, 1)$ by the construc-

---

[1]For any real x, $< x >= x - \max\{\text{integer } N : N \leq x\}$. Thus, $< 1.4 >= 0.4$, and $< -1.4 >= 0.6$.

[2]For $\xi \in [0, 1]$ and $y \in [0, 1)$, $S_\xi(y)$ is defined to be y/ξ if $y \in [0, \xi)$ and $(1 - y)/(1 - \xi)$ if $y \in [\xi, 1)$. $S_\xi(\cdot)$ transforms a uniformly distributed random variable on $[0, 1)$ to another uniformly distributed random variable on $[0, 1)$. For $\xi \in (0, 1)$, $S_\xi(\cdot)$ is continuous on the unit circle and smoothes the sample path discontinuities due to modulo-1 crossings of point 0.

tion. However, the distribution of V_n together with the stitching transformation determines the autocorrelation structure of $\{U_n^+\}$ and $\{Z_n\}$.

By definition [4], the effective bandwidth of $\{Z_n : n = 0, 1, 2, \cdots\}$ over a finite time horizon of k time units is

$$\alpha(s, k) = \frac{1}{sk} \log E(e^{s \sum_{n=0}^{k-1} Z_n}), \tag{1}$$

where $s > 0$ is the space parameter and integer $k \geq 1$. Generally, direct evaluation of (1) will result in formulation of recursive non-tractable integral equations. To overcome this difficulty, we will first discretize the innovation density function of V_n, and then approximate the original TES model with a so-called quantized TES (QTES) model described below. QTES has been shown as a good approximation of TES in [11]. By using QTES, we will be able to derive a tractable solution of (1) in matrix form.

In order to get a QTES process approximating the original TES process, we first discretize the unit interval $[0, 1)$ into (integer) M equal subintervals $[0, \frac{1}{M})$, $[\frac{1}{M}, \frac{2}{M})$, $\cdots$, and $[\frac{M-1}{M}, 1)$. Then, we discretize the innovation probability density (mass) function of the original TES process. For $m = 0, 1, \cdots, M - 1$, let

$$\phi_m = \sum_{i=0}^{\pm\infty} P\{V_1 \in [i + \frac{m}{M}, i + \frac{m+1}{M})\}, \tag{2}$$

with $\sum_{m=0}^{M-1} \phi_m = 1$. Let $\{J_n : n = 1, 2, \cdots\}$ be a sequence of i.i.d. discrete random variables such that

$$P(J_n = \frac{m}{M}) = \phi_m$$

for $m = 0, 1, \cdots, M - 1$, which will be the QTES innovation sequence.

We define another sequence of discrete random variables $\{C_n^+ : n = 0, 1, 2, \cdots\}$ as follows:

$$C_n^+ = \begin{cases} C_0, & n = 0, \\ < C_{n-1}^+ + J_n >, & n \geq 1, \end{cases}$$

where C_0 is a discrete uniform random variable in $\{0, \frac{1}{M}, \cdots, \frac{M-1}{M}\}$, and $< \cdot >$ is the Modulo-1 operator. Then, the QTES process $\{Z_n : n = 0, 1, 2, \cdots\}$ approximating the original TES process is

such that

$$Z_n = D(Q_n^+),$$

where

$$D(\cdot) = F_Z^{-1}(S_\xi(\cdot)) \text{ and } Q_n^+ = \sum_{m=0}^{M-1} \mathbf{1}_{\{C_n^+ = \frac{m}{M}\}} \frac{m + W_n}{M}.$$

$F_Z(\cdot)$ is the marginal probability distribution function of the original TES process, $S_\xi(\cdot)$ is a stitching transformation as defined in the TES model, $\mathbf{1}$ is the indicator function, and $\{W_n : n = 0, 1, 2, \cdots\}$ is a sequence of i.i.d. continuous uniform random variables in $[0, 1)$. Note that the marginal distribution of $\{Q_n^+\}$ is a continuous uniform distribution on $[0, 1)$.

For $m = 0, 1, \cdots, M - 1$ and integer $k \geq 1$, define the conditional effective bandwidth of the QTES process as

$$\alpha_m(s, k) = \frac{1}{sk} \log E(e^{s \sum_{n=1}^{k} Z_n} | C_0^+ = \frac{m}{M}). \tag{3}$$

By the stationarity of $\{C_n^+ : n = 0, 1, \cdots\}$, for any integer $l \geq 0$.

$$\alpha_m(s, k) = \frac{1}{sk} \log E(e^{s \sum_{n=l+1}^{l+k} Z_n} | C_l^+ = \frac{m}{M}). \tag{4}$$

Let $< \cdot | M >$ denote the modulo-M operator[3] for positive integer M. Then, by the Markovian property of $\{C_n^+ : n = 0, 1, \cdots\}$,

$$\alpha_m(s, k + 1)$$

$$= \frac{1}{s(k+1)} \log E(e^{s \sum_{n=1}^{k+1} Z_n} | C_0^+ = \frac{m}{M})$$

$$= \frac{1}{s(k+1)} \log \sum_{l=0}^{M-1} E(e^{s \sum_{n=1}^{k+1} Z_n} | C_1^+ = \frac{l}{M}) \cdot P(C_1^+ = \frac{l}{M} | C_0^+ = \frac{m}{M})$$

$$= \frac{1}{s(k+1)} \log \sum_{l=0}^{M-1} E(e^{sD(Q_1^+)} | C_1^+ = \frac{l}{M}) \cdot E(e^{s \sum_{n=2}^{k+1} Z_n} | C_1^+ = \frac{l}{M}) \cdot \phi_{<l-m|M>}.$$

For $l = 0, 1, \cdots, M - 1$, let

$$g_{s,D}(l) = E(e^{sD(Q_1^+)} | C_1^+ = \frac{l}{M})$$

---

[3] For any real x and positive integer M, $< x | M > = x - M \cdot \max\{\text{integer } N : N \leq \frac{x}{M}\}$. Thus, $< 36 | 10 > = 6$, and $< -36 | 10 > = 4$.

$$= \int_0^1 e^{sD(\frac{l+u}{M})} du. \tag{5}$$

Then, because of (4),

$$\alpha_m(s, k+1) = \frac{1}{s(k+1)} \log \sum_{l=0}^{M-1} g_{s,D}(l) \cdot e^{sk\alpha_l(s,k)} \cdot \phi_{<l-m|M>},$$

or equivalently

$$e^{s(k+1)\alpha_m(s,k+1)} = \sum_{l=0}^{M-1} g_{s,D}(l) \cdot e^{sk\alpha_l(s,k)} \cdot \phi_{<l-m|M>}, \tag{6}$$

for $k \geq 1$. Similarly, for $k = 0$ we have

$$
\begin{aligned}
\alpha_m(s, 1) &= \frac{1}{s} \log E(e^{sZ_1} | C_0^+ = \frac{m}{M}) \\
&= \frac{1}{s} \log \sum_{l=0}^{M-1} g_{s,D}(l) \cdot \phi_{<l-m|M>},
\end{aligned}
$$

or equivalently

$$e^{s\alpha_m(s,1)} = \sum_{l=0}^{M-1} g_{s,D}(l) \cdot \phi_{<l-m|M>}. \tag{7}$$

For $k \geq 1$, we define M-element column vectors

$$\overline{\beta}(s, k) = (e^{sk\alpha_0(s,k)}, e^{sk\alpha_1(s,k)}, \cdots, e^{sk\alpha_{M-1}(s,k)})^T.$$

For the case of $k = 0$, we pose

$$\overline{\beta}(s, 0) = \overline{\mathbf{1}} = (1, 1, \cdots, 1)^T.$$

We also define the M-dimensional square matrices

$$
G_{s,D} = \begin{bmatrix}
g_{s,D}(0) & 0 & \cdots & 0 & 0 \\
0 & g_{s,D}(1) & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & g_{s,D}(M-2) & 0 \\
0 & 0 & \cdots & 0 & g_{s,D}(M-1)
\end{bmatrix},
$$

and

$$H^+ = \begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_{M-1} \\ \phi_{M-1} & \phi_0 & \phi_1 & \cdots & \phi_{M-2} \\ \phi_{M-2} & \phi_{M-1} & \phi_0 & \cdots & \phi_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_0 \end{bmatrix}.$$

Then, from (6) and (7), for $k \geq 1$,

$$\overline{\beta}(s,k) = H^+ G_{s,D} \overline{\beta}(s, k-1) = (H^+ G_{s,D})^k \overline{\mathbf{1}}. \tag{8}$$

Then, the effective bandwidth

$$
\begin{aligned}
& \alpha(s,k) \\
= \ & \frac{1}{sk} \log E(e^{s \sum_{n=0}^{k-1} Z_n}) \\
= \ & \frac{1}{sk} \log \sum_{m=0}^{M-1} \frac{1}{M} E(e^{s \sum_{n=0}^{k-1} Z_n} | C_0^+ = \frac{m}{M}) \\
= \ & \frac{1}{sk} \log \sum_{m=0}^{M-1} \frac{1}{M} g_{s,D}(m) e^{s(k-1)\alpha_m(s,k-1)} \\
= \ & \frac{1}{sk} \log(\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} \overline{\beta}(s, k-1)) \\
= \ & \frac{1}{sk} \log(\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} (H^+ G_{s,D})^{k-1} \overline{\mathbf{1}}).
\end{aligned}
\tag{9}
$$

It should be noted that a TES process is entirely specified by the innovation density (mass) function, the desired marginal distribution, the stitching transformation, and the initial distribution of the innovation. In turn, for given space parameter s and time parameter k, the effective bandwidth of a TES process is totally determined by these four factors. This can be seen in (9), where actually the vector $\frac{1}{M}\overline{\mathbf{1}}$ is the initial distribution of the innovation, H^+ is composed of the discretized values of the innovation density (mass) function, and $G_{s,D}$ for given s is determined by the desired marginal distribution and the stitching transformation.

3   Perron-Frobenius Theory for Infinite Time Horizon

In this section, by applying the Perron-Frobenius theory, we derive the effective bandwidth over an infinite time horizon, i.e.,

$$\alpha(s) = \lim_{k \to \infty} \alpha(s, k) \tag{10}$$

for $s > 0$. We will prove that

$$\alpha(s) = \frac{1}{s} \log \rho(H^+ G_{s,D}), \tag{11}$$

where matrix function $\rho(\cdot)$ takes the spectral radius of a matrix as its value. We will use the Frobenius theorem and a proposition in [2].

Theorem 1 (Frobenius, Thm.2, Ch.XIII, [2]) *An irreducible non-negative matrix A always has a positive characteristic value r that is a simple root of the characteristic equation. The moduli of all the other characteristic values do not exceed r. To the "maximal" characteristic value r, there corresponds a characteristic vector with positive coordinates.*

Proposition 1 (Proposition 2, §6, Ch.XIII, [2]) *A non-negative matrix $A \geq 0$ with the maximal positive characteristic value r and with a corresponding characteristic vector $\overline{z} = (z_1, z_2, \cdots, z_M) > 0$ is similar to the product of r and a stochastic matrix P:*

$$A = ZrPZ^{-1},$$

where matrices $Z = diag[z_i]_{i=1}^M$, and $P = \frac{1}{r} Z^{-1} A Z$.

By the Proposition,

$$A^k = r^k Z P^k Z^{-1}.$$

Since P is a stochastic matrix, P^k is a stochastic matrix as well.

We now prove (11). If the probability mass $\phi_0 = 1$, it is easy to verify that (11) holds. Suppose that $\phi_0 \in [0, 1)$. Then the matrix H^+ is irreducible and non-negative, and so is $H^+ G_{s,D}$. Let

$$r = \rho(H^+ G_{s,D})$$

be the spectral radius, i.e., the maximal characteristic value of the matrix $H^+ G_{s,D}$. Then, using the notations of matrices Z and P defined in the previous Proposition, for $k \geq 1$, from (9),

$$
\begin{aligned}
& \alpha(s, k) \\
= \quad & \frac{1}{sk} \log(\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} (H^+ G_{s,D})^{k-1} \overline{\mathbf{1}}) \\
= \quad & \frac{1}{sk} \log(\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} (r^{k-1} Z P^{k-1} Z^{-1}) \overline{\mathbf{1}}) \\
= \quad & \frac{1}{s} \log r + \frac{1}{sk} \log(\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} r^{-1} (Z P^{k-1} Z^{-1} \overline{\mathbf{1}})).
\end{aligned}
$$

$Z^{-1} \overline{\mathbf{1}}$ is a positive vector, and for any $k \geq 1$, P^{k-1} is a stochastic (transition probability) matrix. Hence, $P^{k-1} Z^{-1} \overline{\mathbf{1}}$ is a positive vector. Z is positive diagonal. Therefore, $Z P^{k-1} Z^{-1} \overline{\mathbf{1}}$ is a positive vector, and all its elements are upper bounded by $\max_{1 \leq i \leq M} \{z_i\} \cdot \max_{1 \leq i \leq M} \{z_i^{-1}\}$ independent of k. Since $G_{s,D}$ is a positive diagonal matrix, all the elements of $G_{s,D} r^{-1} Z P^{k-1} Z^{-1} \overline{\mathbf{1}}$ are positive and upper bounded by $\frac{1}{r} \max_{0 \leq l \leq M-1} \{g_{s,D}(l)\} \cdot \max_{1 \leq i \leq M} \{z_i\} \cdot \max_{1 \leq i \leq M} \{z_i^{-1}\}$ independent of k, and so is $\frac{1}{M} \overline{\mathbf{1}}^T G_{s,D} r^{-1} Z P^{k-1} Z^{-1} \overline{\mathbf{1}}$. Therefore,

$$
\alpha(s) = \frac{1}{s} \log r = \frac{1}{s} \log \rho(H^+ G_{s,D}).
$$

This completes the proof of (11).

4 Computation of Effective Bandwidths of Specific TES Processes

In this section, taking the exponential distribution as an example of the marginal distribution, using (11) we will compute the effective bandwidth functions of three TES processes that are i.i.d., moderately autocorrelated, and highly autocorrelated, respectively. Without loss of generality, we will assume no stitching for simplicity.

For exponential distribution with mean 1, the inverse probability distribution function is

$$
F_Z^{-1}(y) = \log \frac{1}{1 - y}.
$$

Then, for $l = 0, 1, \cdots, M - 1$ and $s > 0$, by (5),

$$g_{s,D}(l) = \int_0^1 \frac{M^s}{(M - l - u)^s} du.$$

For $l = 0, 1, \cdots, M - 2$, $g_{s,D}(l)$ exists for any $s > 0$. However, for $l = M - 1$, $g_{s,D}(l)$ exists only for $0 < s < 1$. For $0 < s < 1$,

$$g_{s,D}(l) = \frac{M^s}{1 - s}[(M - l)^{1-s} - (M - l - 1)^{1-s}],$$

for any $l = 0, 1, \cdots, M - 1$. Thus, we obtain the matrix $G_{s,D}$.

For any given innovation density (mass) function of TES process, ϕ_m for $m = 0, 1, \cdots, M - 1$ can be obtained according to (2). Then, H^+ can be obtained. In this section, we consider three continuous uniform innovation density functions symmetric around 0 with different lengths of support: 1, 1/2, and 1/8, respectively. Figure 1 shows the three innovation density functions. These three innovation density functions together with the exponential marginal distribution will generate three TES processes that are i.i.d., moderately autocorrelated, and highly autocorrelated, respectively. Computation of the spectral radius of $G_{s,D}H^+$, and thus the effective bandwidth functions (11), can be performed by using software such as Matlab. Numerical results (e.g., Figure 2) show that the computed effective bandwidths converge as M gets larger. Table 1 and Figure 3 present some of the computed results with $M = 20$, 160 and 1280 for the i.i.d., moderately autocorrelated and highly autocorrelated TES processes, respectively.

For the i.i.d. TES process with exponential marginal distribution of mean 1, explicit form of the effective bandwidth function can be derived directly from the definition (1):

$$\alpha(s) = \frac{1}{s} \log \frac{1}{1 - s}, \tag{12}$$

for $s \in (0, 1)$. As a check formula, this verifies the effective bandwidth function values for the i.i.d. TES process computed from the matrix form solution (11), such as the values shown in the second column of Table 1.
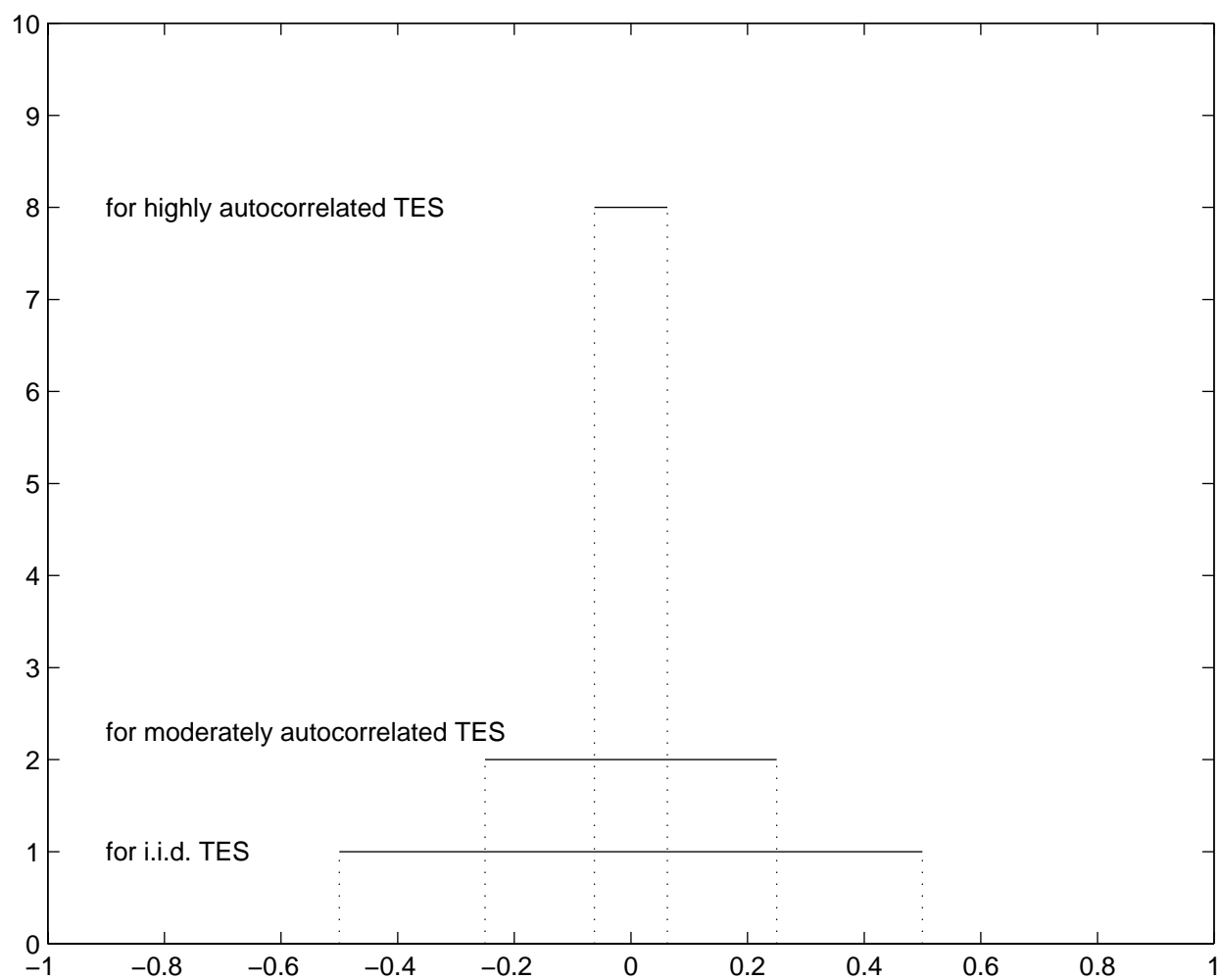
Figure 1: The three innovation probability density functions for generating (1) i.i.d., (2) moderately autocorrelated, and (3) highly autocorrelated TES processes.
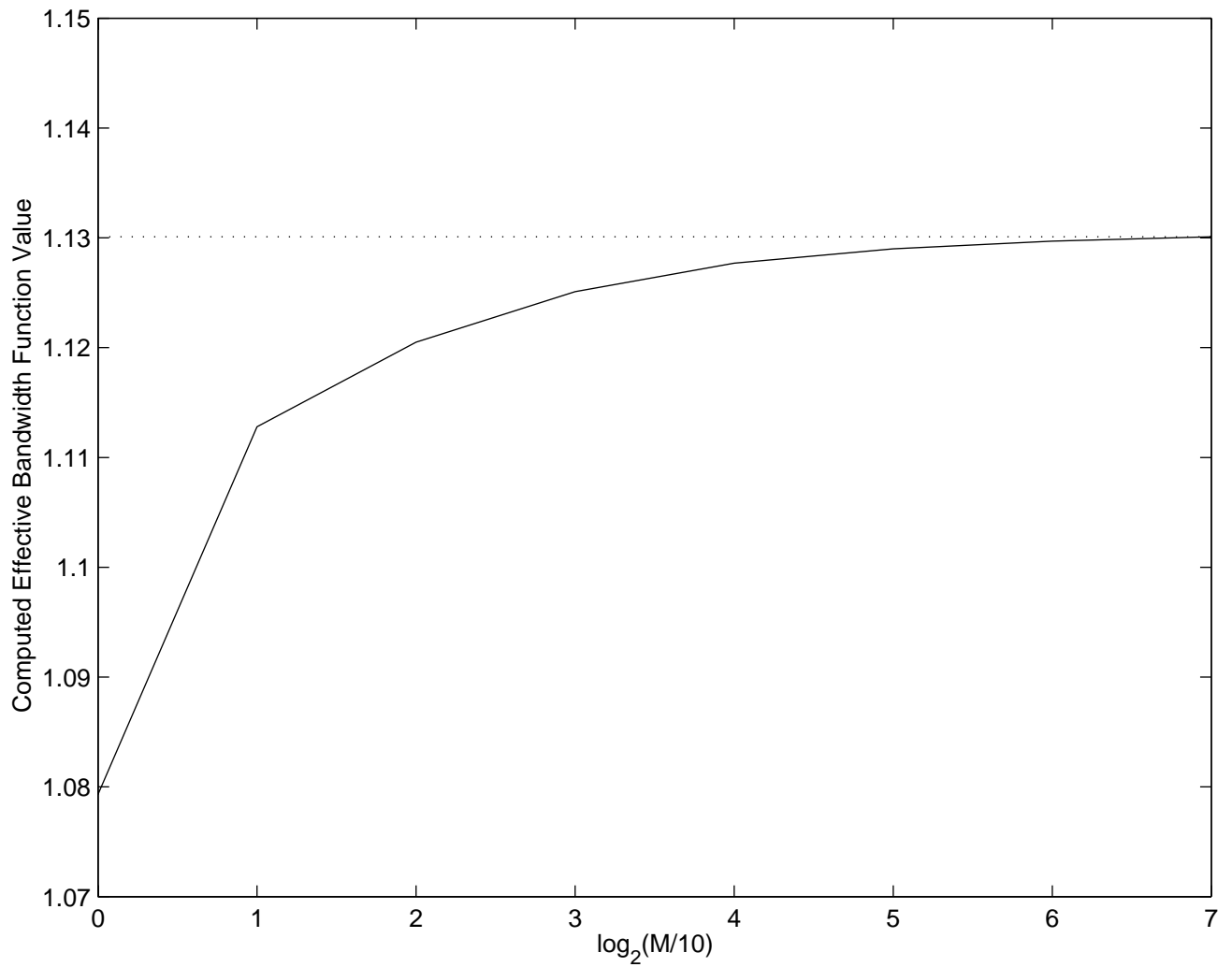
Figure 2: Convergence of the computed effective bandwidths at $s = 0.1$ with respect to M for the moderately autocorrelated TES process.
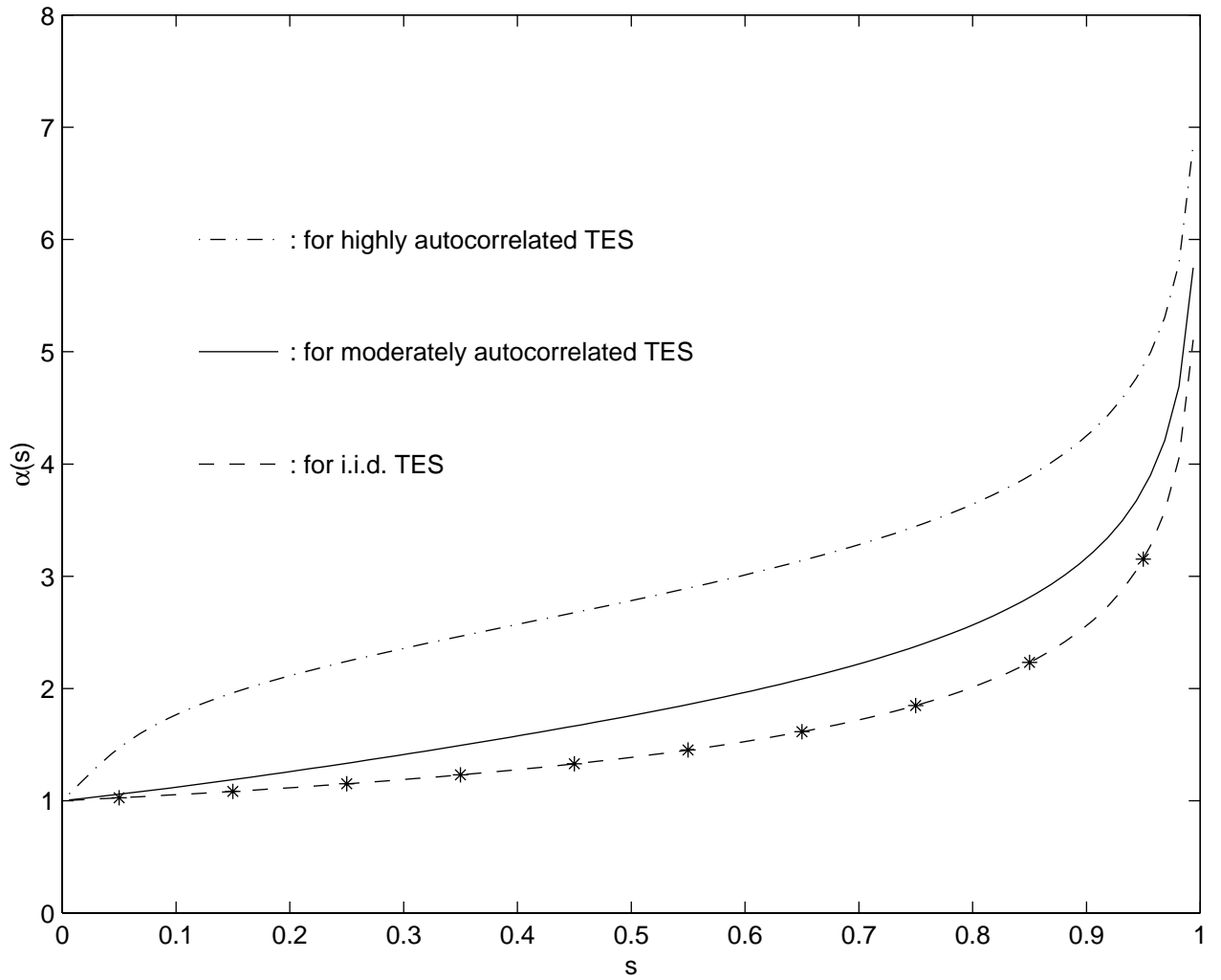
Figure 3: Effective bandwidth functions of the three TES processes with the same marginal exponential distribution of mean 1 but different autocorrelation structures. "*" plots the analytical effective bandwidth function (12) for the i.i.d. TES process.

	i.i.d.	moderately autocorrelated	highly autocorrelated
$s = 0.01$	1.0050	1.012	1.18
$s = 0.1$	1.0536	1.130	2.04
$s = 0.2$	1.1157	1.278	2.38
$s = 0.5$	1.3863	1.791	3.05
$s = 0.9$	2.5584	3.203	4.56

Table 1: Effective bandwidths at different values of the space parameter s of the three TES processes.

5   Comparison with Tail Probabilities of Simulated Queues

According to the effective bandwidth theory, for a stationary and ergodic process, if $\alpha(s) = \lim_{k \to \infty} \alpha(s, k)$ exists, then for a given value of effective bandwidth or service rate c, $s = \alpha^{-1}(c)$ is the asymptotic exponential decay rate of the tail probability of queue length distribution for an infinite-size buffer, or that of the buffer overflow probability for a finite-size buffer.

Suppose a TES process of video frame sizes. The frames are arriving at equally-spaced time instants to a queue of constant service rate $c = \alpha(s)$. Then, we have a $D/\frac{1}{c}TES/1$ queueing system, where the interarrival times of frames are constant (deterministic) and the service times of frames are the frame sizes divided by the service rate c. Suppose that the buffer size is infinite, and we will focus on the tail probabilities of the queue length distributions. (Buffer overflow probabilities for finite-size buffer can be thus approximated.) We will evaluate the tail of the queue length distributions, and compare the exponential decay rate, if there is, with $s = \alpha^{-1}(c)$.

We have simulated the queueing system in OPNET. The three TES processes of different auto-correlation structures in Section 4 are considered. In the simulation, frames arrive at every second and the frame sizes have an exponential distribution with mean 1000 bits. Therefore, the "time unit"

(constant interarrival time) is 1 second, and the "workload unit" is 1000 bits. (The 1 second time unit is assumed for simplicity of presentation. For MPEG video, the time unit would be 1/30 second.) The simulation time is 10 hours. Therefore, 36000 frames are generated and 36000 observations of queue length upon arrival are collected for each of the simulated queueing system. Based on the queue length observations, empirical cumulative distribution functions (CDF) can be computed in OPNET. Then, the empirical tail probability is one minus the corresponding empirical CDF value.

For fixed service rate $c = 1.130Kb/s$, the theoretical exponential decay rates from the effective bandwidth functions $s = \alpha^{-1}(c)$ can be computed as 0.2225 (workload unit)$^{-1}$, 0.1 (workload unit)$^{-1}$, and 0.008 (workload unit)$^{-1}$ for the i.i.d., moderately autocorrelated, and highly autocorrelated TES processes, respectively. Figure 4 presents the results for this case.

On the other hand, we can choose the same $s = 0.1(1000bits)^{-1}$ for all the three TES processes. The effective bandwidths corresponding to the row of $s = 0.1$ in Table 1 are therefore determined to be the service rates $c = \alpha(s)$ for the three TES processes, which are 1.054 Kb/s, 1.130 Kb/s, and 2.040 Kb/s, respectively. Figure 5 presents results for this case.

From either Figure 4 or Figure 5, the following points can be made regarding the empirical queue length distribution tail behaviors.

- The logarithms of empirical tail probabilities decay approximately linearly, i.e., the tail probabilities decay exponentially, except for the very large queue lengths (i.e., the very far end of the tails) observed during simulation. This is due to the increased variance of the experiment since *loss events* (i.e. events when the buffer exceeds a threshold) become more rare. This can be overcome by a) running longer simulations in order to further reduce the variance of our experiments (we show some results shortly!) or b) by running simulations based on *Importance Sampling* which is a method for variance reduction. This is however beyond the scope of the current paper and research towards this direction is in progress.
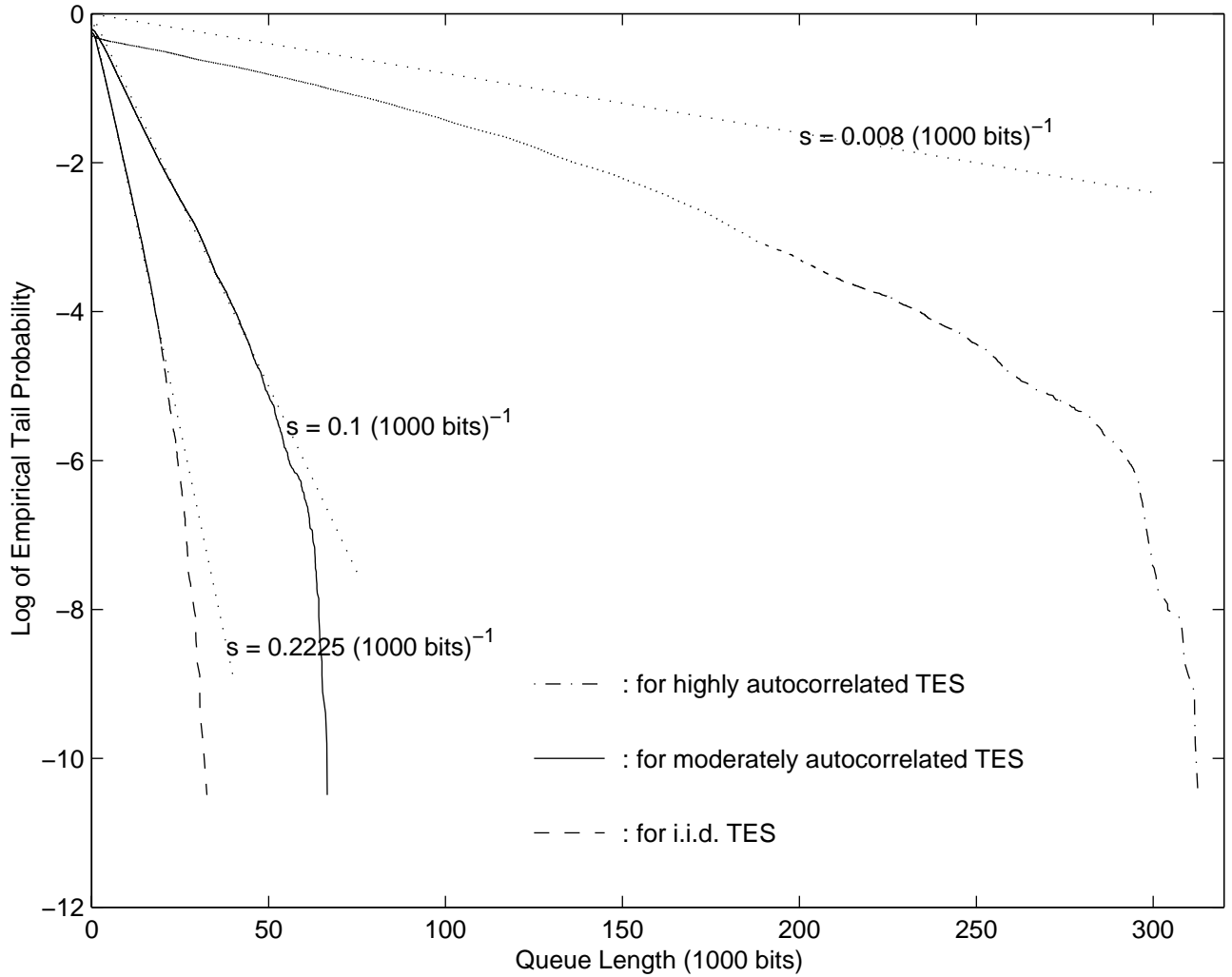
Figure 4: Plots of the logarithms of empirical tail probabilities of the queue length distributions of the simulated $D/\frac{1}{c}TES/1$ queueing systems, where the service rate is fixed as $c = 1.130Kb/s$ and the theoretical exponential decay rates from the effective bandwidth functions $s = \alpha^{-1}(c)$ vary. The dotted line is passing the origin and with the respective slope $-s$.
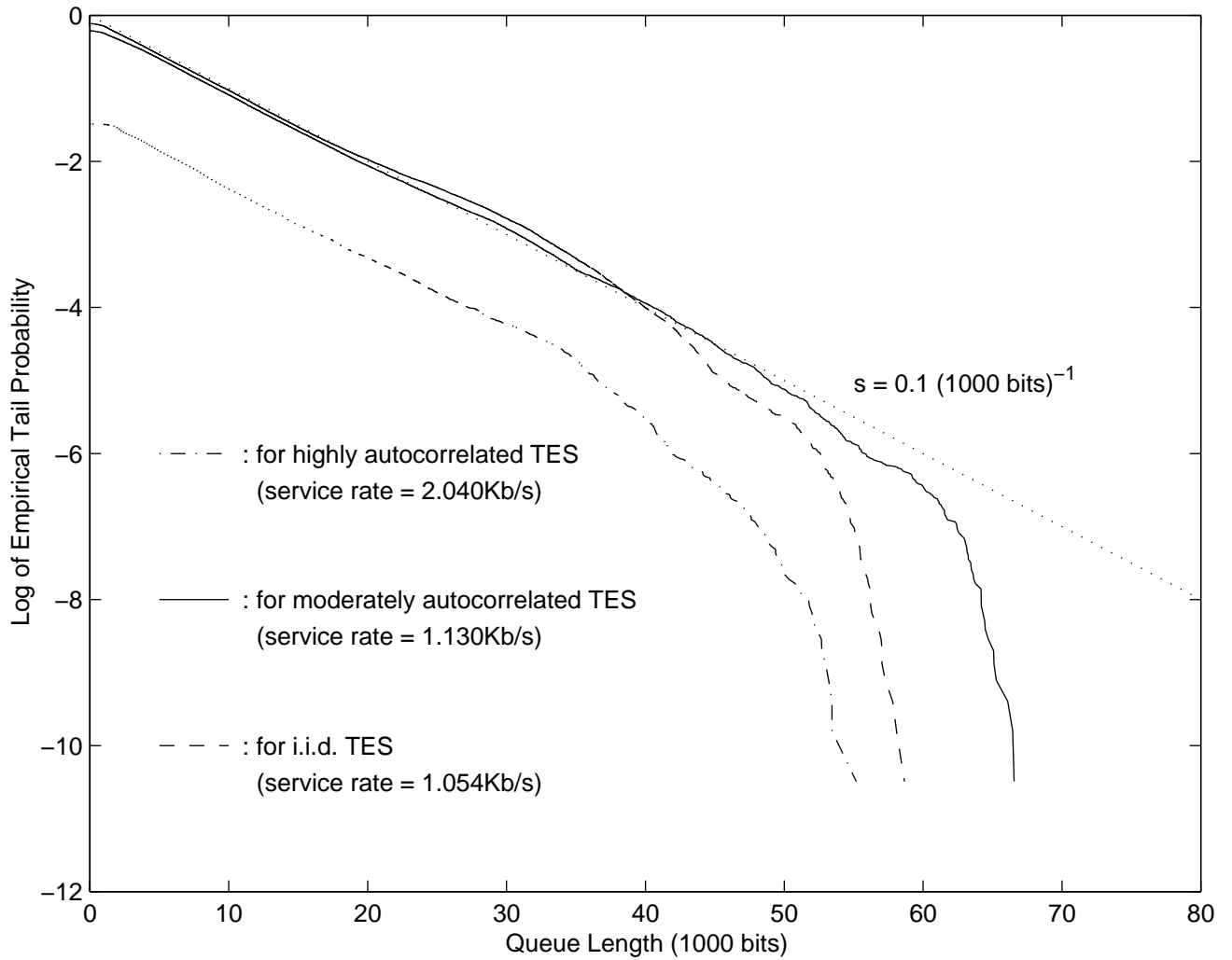
Figure 5: Comparison of the exponential decay rates $s = 0.1$ from the effective bandwidth functions with the plots of the logarithm of empirical tail probability of the queue length distributions of the simulated D/$\frac{1}{c}$TES/1 queueing systems, where the service rates $c = \alpha(s)$ vary. The dotted line is passing the origin and with slope $-s$.

- The plots of the logarithms of empirical tail probabilities are roughly parallel to the straight line of slope $-s$ passing the origin. This shows that $s = \alpha^{-1}(c)$ is very close to the actual exponential decay rate of the queue length distribution tail probabilities.

- The plots of the logarithms of empirical tail probabilities are a little bit below or very close to the straight line of slope $-s$ passing the origin. That is, $\exp(-sB)$ is roughly an upper bound of the tail probability at queue length B.

In order to reduce the variance of the simulations and get more accurate numerical results we have run longer simulations. The longer the simulation, the more observations of large values can be obtained. In Figures 6, we present results on the empirical queue length distribution tail probabilities computed from the original (10 hour simulation time) simulation, and a 50 time longer simulation of the queueing system fed by the moderately autocorrelated TES process. For the 50 time longer simulation, the plot of logarithms of empirical tail probabilities for queue lengths from 0 to 85000 bits is roughly linear, while there is a dropping end before 67000 bits for the original simulation. This tells us that the logarithms of the "real" tail probabilities decay linearly, which can be verified by longer and longer simulation.

Discussion of the general issues on the biased-ness of empirical CDF and its impact on performance analysis is out of the scope of this paper. Nevertheless, clarification on the rapid dropping behavior of the plots of logarithms of the empirical tail probabilities, together with the first three points made previously in this section, helps us reach that (i) the tail probabilities of the queue length distributions decay exponentially, and (ii) the effective bandwidth functions can be used to characterize the exponential decay rate.
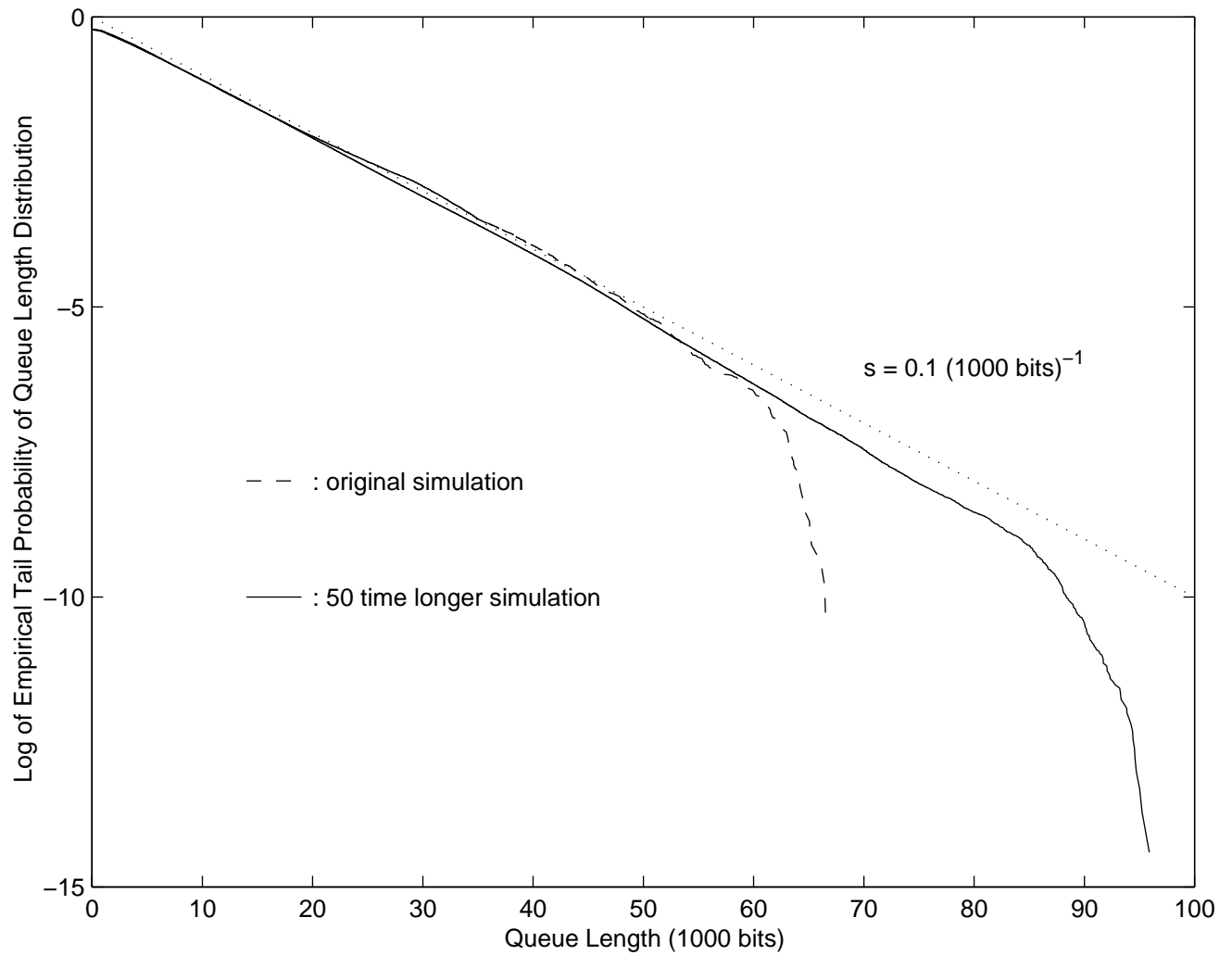
Figure 6: Comparison of the empirical queue length distribution tail probabilities computed from the original simulation to those from a 50 time longer simulation of the queueing system with the moderately autocorrelated TES process.

6   Summary and Conclusion

In this paper, we have derived the effective bandwidth functions of TES processes for video traffic modeling. We have computed the effective bandwidth functions of some specific TES processes with different autocorrelation structures as examples. The results of the effective bandwidth functions show that to meet roughly the same QoS requirement, the more autocorrelated the traffic process, the more bandwidth should be allocated to it. We have simulated queues with TES processes as input, numerically studied the exponential decay of the queue length tail probabilities, and compared the exponential decay rates with the space parameter from the effective bandwidth functions. The results show that the effective bandwidth functions can well characterize the TES traffic processes.

While TES is a convenient tool for video traffic modeling, the effective bandwidth functions of TES processes will facilitate queueing performance analysis of video traffic processes modeled by TES, efficient network design and video traffic management. Future work is on detailed analysis of exact video traces and use of effective bandwidth for buffer dimensioning. Extension to autoregressive modular processes [8] is currently in progress, which are generalized TES processes with autocorrelated innovation sequence.

References

[1] Elwalid, A. I., and Mitra, D. (1993), "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks", *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp.329-343.

[2] Gantmacher, F. R. (1974), *The Theory of Matrices*, Chelsea Publishing Company, New York.

[3] Ismail, M. R., Lambadaris, I. E., Devetsikiotis, M., and Kaye, A. R. (1995), "Modeling Prioritized MPEG Video Using TES and a Frame Spreading Strategy for Transmission in ATM Networks", *Proceedings of IEEE INFOCOM'95*, Boston, Massachusetts, pp.762-770.

[4] Kelly, F. (1996), "Notes on Effective Bandwidths", *Stochastic Networks: Theory and Applications* (ed. F. P. Kelly, S. Zachary and I. Ziedins), Vol. 4 of *Royal Statistical Society Lecture Notes Series*, Oxford University Press, pp.141-168.

[5] Kesidis, G., Walrand, J., Chang, C.-S. (1993), "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources", *IEEE/ACM Transactions on Networking*, Vol. 1, No. 4, pp.424-428.

[6] Lee, D.-S., Melamed, B., Reibman, A. R., and Sengupta, B. (1992), "TES Modeling for Analysis of a Video Multiplexer", *Performance Evaluation*, Vol. 16, pp.21-34.

[7] Melamed, B. (1993), "An Overview of TES Processes and Modeling Methodology", *Performance Evaluation of Computer and Communications Systems*, (ed. L. Donatiello and R. Nelson), Lecture Notes in Computer Science, Springer-Verlag, pp.359-393.

[8] Melamed, B. (1999), "ARM Processes and Modeling Methodology", *Stochastic Models*, Vol. 15, No. 5, pp.903-929.

[9] Melamed, B., and Pendarakis, D. (1998), "Modeling Full-Length VBR Video Using Markov-Renewal-Modulated TES Models", *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 3, pp.600-611.

[10] Melamed, B., Raychaudhuri, D., Sengupta, B., and Zdepski, J. (1994), "TES-Based Video Source Modeling for Performance Evaluation of Integrated Networks", *IEEE Transactions on Communications*, Vol. 42, No. 10, pp.2773-2777.

[11] Melamed, B., Ren, Q., and Sengupta, B. (1996), "The QTES/PH/1 Queue", *Performance Evaluation*, Vol. 26, pp.1-20.

[12] Reininger, D., Melamed, B., Raychaudhuri, D., and Sengupta, B. (1994), "Variable Bit Rate Video: Characteristics, Modeling and Multiplexing", *Proceedings of the 14-th International Teletraffic Congress*, Antibes Juan-les-Pins, France, Vol. 1a, pp.295-306.

[13] de Veciana, G., Kesidis, G., Walrand, J. (1995), "Resource Management in Wide-Area ATM Networks Using Effective Bandwidths", *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 6, pp.1081-1090.