THE QTES/PH/1 QUEUE

Benjamin Melamed
Rutgers University
RUTCOR
P.O. Box 5062
New Brunswick, NJ 08903, USA

Qiang Ren and Bhaskar Sengupta
NEC USA Inc.
C&C Research Laboratories
4 Independence Way
Princeton, NJ 08540, USA

Abstract

In performance analysis of computer and communication systems, the designer is often faced with two challenges: modelling an input process accurately and then solving a resulting queueing problem. In this paper, we present a modelling methodology called QTES (Quantized Transform-Expand-Sample) which can be used to model the interarrival distribution (histogram) accurately and to capture the effect of autocorrelations approximately. We then show that a difficult queueing problem (using a QTES input with autocorrelations) can be solved. Our method is analytically tractable and numerically robust to the extent of the examples presented (and those tried but not presented).

Keywords: TES processes, QTES processes, semi-Markov queues, single-server queues

1 Introduction

The single-server queue plays a pivotal role in the performance evaluation of many communication and computer systems. A relevant example in today's emerging technologies is the analysis of a buffer (or multiplexer) which receives aggregate traffic (input) from many sources. To analyze the performance of such a buffer in the presence of empirical data, the designer often goes through the following steps:

1. A model of the input process is constructed from the empirical data. Such a model may be either fitted to an empirical aggregate input process, or to individual traffic sources which are then combined into the aggregate process through superposition (multiplexing). In either case, issues of paramount interest are the validity of the model and parameter estimation.

2. An analytical server-system model (multiplexer, single-server queue, etc.) is constructed, and methods are devised to compute performance measures of interest (e.g., queue-length and waiting-time distributions). The issues of paramount interest here center on analytical tractability and computational feasibility.

The literature on performance evaluation has usually focussed on just one of these steps, but not both. For example, a vast body of queueing literature exists which can be used to solve Item 2 above (fluid models, Markov modulated Poisson processes, semi-Markov processes, etc.). However, there does not exist any systematic way in which a designer can take empirical data and fit it into one of these processes (however, *ad hoc* methods for some specific problems do exist). Similarly, some researchers have focussed on accurate modelling of source traffic, but these methods have not as yet led to analytically tractable models, which can be used to predict multiplexer behavior. Consequently, when a designer is given an arbitrary input process, the only way of predicting queueing behavior is by simulation (which is often inadequate in predicting long tail behavior accurately).

Traditional models are not well-suited to modelling traffic on emerging high-speed networks. A salient feature of projected major traffic sources there is a high degree of burstiness (e.g., compressed video and file transfer). The mathematical underpinning of burstiness is complex; two main factors contributing to traffic burstiness are the marginal distribution and autocorrelation function of traffic interarrival times, particularly strong positive short-term autocorrelations [11]. For additional discussion of traffic modeling and burstiness, see [3]. The impact of autocorrelation in traffic processes on queueing measures such as queue-length and waiting-time distributions can be very dramatic. As an example we cite a simulation study [15] which investigated the effect of introducing increasing autocorrelations into a Poisson process without altering the (exponential) marginal distribution. When offered to an exponential server, the two traffic streams give rise to mean waiting times differing by orders of magnitude. The same phenomenon has been observed in analytical queueing models, for instance [5, 6, 7, 13, 25, 28]. These studies support the view that modeling an autocorrelated traffic stream as a renewal process is fraught with modeling risk. Worse still, a renewal traffic model is bound to err on the optimistic side, if the modeled traffic has appreciable short-term positive autocorrelations. Consequently, in order to obtain good predictions of performance measures resulting from offering a bursty traffic stream to a server system, it is important to faithfully capture both first-order and second-order properties of the traffic process.

TES (Transform-Expand-Sample) is a new class of models [8, 9, 10, 16, 17], which can accurately model empirical sequences, and in particular, can effectively capture traffic burstiness [11]. The TES modeling methodology (to be reviewed later) can accurately capture both first-order and second-order statistics of empirical data; more specifically, the model exactly captures the empirical marginal distribution (histogram) and approximately captures the empirical autocorrelation function (equivalently, spectral density), *simultaneously.* In addition, the fitting of parameters to a given dataset can be carried out in a systematic manner [4, 12]. While TES models have proven to be accurate in a variety of application domains [19], TES-based queueing models, however, have been analytically intractable. To date, queueing models with TES-based traffic have only been studied via Monte Carlo simulation. The analysis of such models is difficult, due to the fact that TES processes are essentially transformed Markov sequences, over an uncountable state space.

In this paper, we present a new class of stochastic processes, called QTES (Quantized TES), which can be used both to model a large class of stochastic processes and also to solve numerically a broad class of queueing problems, in which a QTES arrival process is offered to a single-server queue (see also [18]). The QTES modelling methodology approximates a continuous-state TES process by a discrete-state QTES counterpart. In principle, one may study the corresponding *QTES/PH/1* queue directly, without regard to the underlying TES process. In practice, however, one is better off first using known TES theory and software to fit a TES model to empirical data, and then proceeding to approximate it by a QTES model; the analysis is carried out on the *QTES/PH/1* queue rather than the underlying *TES/PH/1* queue, which is as yet untractable. This way we have the best of both worlds: a tractable queueing model which is still quite accurate in the sense that the traffic model captures first-order and second-order statistics of the empirical traffic data.

The method of solving the *QTES/PH/1* queues is based on the theory developed by Sengupta [26, 27]. This theory uses a continuous-state analog of the well-known matrix-geometric method of Neuts [23]. There are two advantages of using this theory. First, it is usually simpler to find the transition rates in the continuous-state problem when compared to its discrete-state counterpart; and second, the theory gives (in a very straightforward manner) *all* performance measures of interest (e.g., queue-length distributions at arrival epochs and at arbitrary time, as well as waiting-time, sojourn-time and virtual waiting-time distributions). This is to be contrasted with other techniques in queueing theory which usually yield either the queue-length distribution or the waiting-time distribution but not both. The method itself is based on first characterizing the distribution of time spent in the system by the customer in service (see [26, 27] for details). This method differs significantly from most other methods, which are usually based on characterizing Markov chains through appropriate imbeddings. Finally, the theory allows the use of semi-Markov arrival processes, which take into account autocorrelations in the arrival stream to a queueing system. An additional step is required to map the QTES process into a semi-Markov processes.

The rest of this paper is organized as follows. In section 2, we give a brief overview of TES processes and in section 3, we describe its discretized and quantized counterparts (DTES and QTES respectively). In section 4, we develop the analytical model for the *QTES/PH/1* queue and in section 5, we provide some numerical results. Note that most of the focus in the numerical results is on addressing the issues surrounding Item 2 above, since the issues of Item 1 have been reported elsewhere.

2 TES Processes

This section provides a brief overview of TES processes and the TES modeling methodology. For detailed discussion see [16, 8, 9, 10]. TES modeling differs from other modeling approaches in that it aims to fit a prescribed marginal distribution and a prescribed autocorrelation function *simultaneously*, via a stationary stochastic process. Thus, TES models are specifically designed to capture the statistical signature of empirical time series vis-a-vis both first-order and second-order statistical properties.

We adopt throughout the following notational conventions. Two distinct notations will be used for the indicator function, depending on context: For a set A, $1_A(x)$ denotes a deterministic indicator function, while for an event, B, $1\{B\}$ denotes a stochastic indicator function, with the sample point argument suppressed for typographic clarity. For a function, $f(x)$, the corresponding Laplace transform is defined as

$$\tilde{f}(s) = \begin{cases} \displaystyle\int_0^\infty f(x)e^{-sx}dx, & \text{if } f \text{ is a continuous function,} \\ \displaystyle\sum_v f(v)e^{-sv}, & \text{if } f \text{ is a discrete function,} \end{cases}$$

and tilde denotes the Laplace transform operator. In particular, if $f = \{f(0), f(1), \ldots, f(M-1)\}$ for some integer $M > 0$, the corresponding discrete Fourier transform [24] is given by

$$\tilde{f}(i2\pi\nu) = \sum_{m=0}^{M-1} f(m)\, e^{-i2\pi\nu m/M}, \quad \nu = 0, 1, \ldots, M-1. \tag{1}$$

The discrete inverse Fourier transform of (1) is defined by

$$f(m) = \frac{1}{M}\sum_{\nu=0}^{M-1} \tilde{f}(i2\pi\nu)\, e^{i2\pi\nu m/M}, \quad m = 0, 1, \ldots, M-1. \tag{2}$$

The construction of TES models involves two random sequences in lockstep. The first sequence, called a *background* TES process, serves an auxiliary role, and is chosen as either $\{U_n^+\}_{n=0}^\infty$ or $\{U_n^-\}_{n=0}^\infty$; these have the form

$$U_n^+ \;=\; \begin{cases} U_0, & n = 0 \\ \langle U_{n-1}^+ \;+\; V_n \rangle, & n > 0 \end{cases} \tag{3}$$

and

$$U_n^- \;=\; \begin{cases} U_n^+, & n \text{ even} \\ 1 \;-\; U_n^+, & n \text{ odd} \end{cases} \tag{4}$$

respectively, where U_0 is distributed uniformly on $[0, 1)$, $\{V_n\}_{n=1}^\infty$ is a sequence of iid random variables, independent of U_0, called the *innovation sequence*, and angular brackets denote the modulo-1 (fractional part) operator $\langle x \rangle = x - \max\{\text{integer } n : n \le x\}$. The superscript notation in (3) and (4) is motivated by the fact that $\{U_n^+\}$ and $\{U_n^-\}$ can generate lag-1 autocorrelations in the range [0,1] and [-1,0], respectively. From now on, we always append plus or minus superscripts to other mathematical objects associated with $\{U_n^+\}$ and $\{U_n^-\}$, but the superscript is omitted when the distinction is immaterial. Intuitively, the modulo-1 arithmetic used in defining TES processes in (3) and (4) can be interpreted as random walks on a circle of circumference 1 (unit circle), with mean step size $E[V_n]$.

The second sequence, called a *foreground TES process* and denoted by $\{X_n^+\}_{n=0}^\infty$ or $\{X_n^-\}_{n=0}^\infty$, respectively, is a transformed version of a background TES process,

$$X_n^+ \;=\; D(U_n^+), \quad X_n^- \;=\; D(U_n^-), \tag{5}$$

where D is a measurable transformation from $[0, 1)$ to the reals, called a *distortion*. Eq. (5) defines two classes of TES models, denoted TES$^+$ and TES$^-$, respectively, and these foreground sequences are the end-product TES models.

An important family of distortions is based on an empirical time series, $\{Y_n\}_{n=0}^N$; these distortions consist of composite transformations of the form

$$D_{Y,\xi}(x) \;=\; \hat{H}_Y^{-1}(S_\xi(x)), \quad x \in [0, 1). \tag{6}$$

Here, the inner transformation, S_ξ, is a "smoothing" operation, called a *stitching transformation*, parameterized by $0 \le \xi \le 1$, and given by

$$S_\xi(y) \;=\; \begin{cases} y/\xi, & 0 \le y \le \xi \\ (1-y)/(1-\xi), & \xi \le y < 1. \end{cases} \tag{7}$$

The outer transformation, $\hat{H}_Y^{-1}$, is the inverse of the empirical (histogram) distribution function computed from an empirical time series, $Y = \{Y_n\}$, as

$$\hat{H}_Y^{-1}(x) \;=\; \sum_{j=1}^{J} 1_{[\hat{C}_{j-1}, \hat{C}_j)}(x)\, [l_j + (x - \hat{C}_{j-1})\, \frac{w_j}{\hat{p}_j}], \quad 0 \le x \le 1, \tag{8}$$

where J is the number of histogram cells, $[l_j, r_j)$ is the support of cell j with width $w_j = r_j - l_j > 0$, $\hat{p}_j$ is the probability estimator of cell j and $\{\hat{C}_i\}_{i=0}^J$ is the cdf of $\{p_j\}_{j=1}^J$, i.e., $\hat{C}_j = \sum_{i=1}^j \hat{p}_i$, $1 \le j \le J$ ($\hat{C}_0 = 0$ and $\hat{C}_J = 1$).

The rationale for TES processes stems from the following facts. First, all background TES sequences are stationary Markovian, and their marginal distribution is uniform on $[0,1)$, *regardless* of the probability law of the innovations $\{V_n\}$ selected [8]. It can be shown [8] that the τ-step transition density of $\{U_n^+\}$ is

$$g_\tau^+(y|x) \;=\; \begin{cases} \sum\limits_{\nu=-\infty}^{\infty} \tilde{f}_V^\tau(i2\pi\nu)\, e^{i2\pi\nu(y-x)}, & 0 \;\leq\; y,x \;<\; 1 \\ 0, & otherwise \end{cases} \tag{9}$$

and that of $\{U_n^-\}$ is

$$g_\tau^-(y|x) \;=\; \begin{cases} g_\tau^+(y|x), & 0 \leq y,x < 1,\ \tau\ even \\ \sum\limits_{\nu=-\infty}^{\infty} \tilde{f}_V^\tau(i2\pi\nu)\, e^{-i2\pi\nu(y+x)}, & 0 \leq y,x < 1,\ \tau\ odd \\ 0,\ \ otherwise & \end{cases} \tag{10}$$

(recall that tilde denotes the Laplace transform). Second, the *inversion method* [2] permits us, in principle, to transform any uniform variate to others with arbitrary distributions as follows: if U is uniform on $[0,1)$ and F is a prescribed distribution, then $X = F^{-1}(U)$ has distribution F. In particular, $F = \hat{H}_Y$ is just a special case. And third, for $0 < \xi < 1$, the effect of S_ξ is to render the sample paths of background TES sequences more "continuous-looking". Because stitching transformations preserve uniformity [16], the inversion method can still be applied to stitched background processes $\{S_\xi(U_n)\}$. It follows that any foreground TES sequence of the form

$$X_n \;=\; \hat{H}_Y^{-1}(S_\xi(U_n)), \tag{11}$$

obtained from any background sequence $\{U_n\}$, is always guaranteed to have the empirical (histogram) distribution $\hat{H}_Y$, *regardless* of the innovation density f_V and stitching parameter ξ selected. The choice of a pair (f_V, ξ) will largely determine the dependence structure of (11), and in particular, its autocorrelation function. Thus, TES modeling decouples the fitting of the empirical distribution from the fitting of the empirical autocorrelation function. Since the former is guaranteed, one can concentrate on the latter. We now note that even though the TES background sequence is Markovian, the foreground sequence is (in general) not Markovian. The reason is that the map D is not invertible in general due to the presence of the stitching function. However, for the special case of $\xi = 1$, the foreground sequence is also Markovian because the map D is one-to-one.

In practice, TES modeling has the form of a heuristic search for a suitable pair (f_V, ξ), such that the ensuing TES model approximates well the leading empirical autocorrelations and gives rise to sample path realizations which adequately resemble the empirical record. Effective TES modeling requires software support. To construct TES models that fit empirical data, we used TEStool — a visual interactive software environment [4, 20], designed to support TES modeling. TEStool allows the user to read in empirical time series and calculate their statistics (histogram, autocorrelation function and spectral density) in textual and graphical representations. It provides services to construct and modify TES models and to superimpose the corresponding TES statistics on their empirical counterparts. The search proceeds in an interactive style, guided by visual feedback: each model modification triggers a recalculation and redisplay of the corresponding statistics. TES model autocorrelations and spectral densities are calculated numerically from fast and accurate formulas developed in [8, 9]. Recently, an algorithmic modeling approach has been devised and implemented for TES modeling [12], in which the TES parameters are chosen by solving a non-linear programming problem so as to yield a weighted least-squares fit between the empirical and model autocorrelation functions.

3 Discretized and Quantized TES Processes

The rationale for introducing discretized and quantized TES processes is to gain analytical tractability by reducing the continuous state space of background TES processes (the unit circle) to a finite state space. More precisely, discrete TES processes (referred to as *DTES processes*) have a finite state space and constitute a discretized approximation of ordinary TES processes. Quantized TES processes (referred to as *QTES processes*) are obtained from DTES processes by extending the latter to the unit circle, but in such a way that the corresponding transition density has a quantized domain of conditioning, in a sense to be made precise in the sequel. The final TES approximation is obtained through a two-stage construction to be described in the two following subsections.

3.1 DTES Processes

In the first stage we introduce a discrete variant of TES over a finite state space. To this end, we partition the unit circle into a finite number of subintervals (called *slices*), and effectively "collapse" each slice subinterval into a new state. Intuitively, this finite-state process only tracks the slice index, on the assumption that if the slice sub intervals are "small enough", then the exact location of the original process within a slice is "well-approximated" by the slice index. Specifically, let $\mathcal{Z}_M = \{0, 1, \ldots, M-1\}$, for some integer $M > 0$. A quantization of the unit circle, $[0, 1)$, is a partition,

$$\Gamma(M) \;=\; \{\Gamma_m = [m\delta, (m+1)\delta) :\; m \in \mathcal{Z}_M\},$$

of the interval $[0, 1)$ into M subintervals (slices), Γ_m, each of length $\delta = \delta(M) = 1/M$. The integer-valued modulo-M operator from the integers onto $\mathcal{Z}_M$ is denoted by $\langle\cdot\rangle_M$ and defined by

$$\langle m\rangle_M \;=\; m \;-\; M\lfloor\frac{m}{M}\rfloor, \tag{12}$$

where $\lfloor x\rfloor = \max\{n :\; n \text{ is an integer and } n \leq x\}$.

Let C_0 be an integer-valued random variable, distributed uniformly over $\mathcal{Z}_M$, that is, $P\{C_0 = m\} = 1/M$, for all $m \in \mathcal{Z}_M$. Let, further, $\{J_n\}_{n=1}^{\infty}$ be an iid sequence of integer-valued random variables, independent of C_0, called the *innovation sequence*, analogously to the setting of TES processes in Section 2. The common probability distribution of the J_n is denoted by

$$f_J(j) \;=\; P\{J_n = j\} \;=\; \phi_j, \quad j \in \mathcal{Z}_M.$$

In analogy to Eqs. (3) and (4)), *background DTES processes* are defined by

$$C_n^+ \;=\; \begin{cases} C_0, & n = 0 \\ \\ \langle C_{n-1}^+ \;+\; J_n \rangle_M, & n > 0 \end{cases} \tag{13}$$

and

$$C_n^- \;=\; \begin{cases} C_n^+, & n \text{ even} \\ \\ M - 1 - C_n^+, & n \text{ odd} \end{cases} \tag{14}$$

Intuitively, a background DTES process, $\{C_n\}$, is a random walk on a finite circular lattice. Moreover, when C_n assumes a value $k \in \mathcal{Z}_M$, we may interpret state k to stand for slice Γ_k.

The following facts hold for DTES processes of the forms (13) and (14). Background DTES processes are stationary Markovian, the former with a stationary transition structure and the latter with a time-dependent transition structure. Their marginal distributions are all uniform on $\mathcal{Z}_M$ (see e.g., [14] and a continuous-state case in [8]). Their τ-step transition probabilities, denoted $h_\tau^+(k|j) = P\{C_{n+\tau}^+ = k | C_n^+ = j\}$ and $h_\tau^-(k|j) = P\{C_{n+\tau}^- = k | C_n^- = j\}$, respectively, have the representations as given in the following lemma.

Lemma 1. For $k, j \in \mathcal{Z}_M$ and $\tau = 0, 1, \ldots$,

$$h_\tau^+(k|j) \;=\; \frac{1}{M} \sum_{\nu=0}^{M-1} \tilde{f}_J^\tau(i2\pi\nu)\, e^{i2\pi\nu(k-j)/M}, \tag{15}$$

and

$$h_\tau^-(k|j) \;=\; \begin{cases} \frac{1}{M} \sum\limits_{\nu=0}^{M-1} \tilde{f}_J^\tau(i2\pi\nu)\, e^{i2\pi\nu(k-j)/M}, & \tau \text{ even} \\ \\ \frac{1}{M} \sum\limits_{\nu=0}^{M-1} \tilde{f}_J^\tau(i2\pi\nu)\, e^{-i2\pi\nu(k+j+1)/M}, & \tau \text{ odd} \end{cases} \tag{16}$$

where

$$\tilde{f}_J(i2\pi\nu) \;=\; \sum_{m=0}^{M-1} f_J(m)\, e^{-i2\pi\nu m/M}, \qquad \nu \in \mathcal{Z}_M, \tag{17}$$

is the discrete Fourier transform of the discrete probability distribution function, $f_J \equiv \{\phi_m : \; m \in \mathcal{Z}_M\}$, with $\phi_m = f_J(m), \; m \in \mathcal{Z}_M$; see [24] for details.

Proof: The proof for h^+ can be carried out analogously to Theorem 1 in [8] for

ordinary TES processes, but here we outline a simpler version.
¿From (13),

$$C^+_{n+\tau} \;=\; \langle C^+_n \;+\; J_{n+1} \;+\; \cdots \;+\; J_{n+\tau}\rangle_M,$$

which implies by the modulo-M algebra,

$$\langle C^+_{n+\tau} - C^+_n\rangle_M \;=\; \langle J_{n+1} + \cdots + J_{n+\tau}\rangle_M.$$

Thus, $h^+_\tau(k|j)$ depends on k and j, only through $\langle k-j\rangle_M$. Since $J_{n+1},\ldots,J_{n+\tau}$ are iid random variates, $h^+_\tau(k|j)$ is the τ-fold circular convolution (with period M) of f_J, evaluated at point $\langle k-j\rangle_M$; see [24]. Hence, the discrete Fourier transform, $\tilde{h}^+_\tau$, of h^+_τ coincides with $\tilde{f}^\tau_J = \left(\tilde{f}_J\right)^\tau$, namely, for all $j \in \mathcal{Z}_M$,

$$\tilde{h}^+_\tau(i2\pi\nu|j) \;=\; \tilde{f}^\tau_J(i2\pi\nu), \quad \nu \in \mathcal{Z}_M. \tag{18}$$

¿From (2), inverting (18) yields

$$\begin{aligned} h^+_\tau(k|j) \;&=\; \frac{1}{M}\sum_{\nu=0}^{M-1} \tilde{h}^+_\tau(i2\pi\nu|j)\, e^{i2\pi\nu\langle k-j\rangle_M/M} \\ &=\; \frac{1}{M}\sum_{\nu=0}^{M-1} \tilde{f}^\tau_J(i2\pi\nu)\, e^{i2\pi\nu\langle k-j\rangle_M/M} \end{aligned}$$

The right-hand side above reduces to (15), since $e^{i2\pi\nu\langle k-j\rangle_M/M} \;=\; e^{i2\pi\nu(k-j)/M}$. To see that, note that for $k-j \;\geq\; 0$ this is immediate, while for $k-j \;<\; 0$, we have $\langle k-j\rangle_M = M+k-j$ from (12).
The case for h^- follows in a similar way. □

We remark that from (13), it can be seen that a simple alternative representation for the 1-step transition probabilities, $h^+(k|j)$, is

$$h^+(k|j) \;=\; \phi_{\langle k-j\rangle_M},$$

or, equivalently, the corresponding $M \times M$ 1-step transition matrix, $\mathcal{H}^+ = [h^+(k|j)]$, is

$$\mathcal{H}^+ \;=\; \begin{bmatrix} \phi_0, & \phi_1, & \phi_2, & \cdots, & \phi_{M-1} \\ \phi_{M-1}, & \phi_0, & \phi_1, & \cdots, & \phi_{M-2} \\ \vdots & \vdots & & & \vdots \\ \phi_1, & \phi_2, & \cdots, & \cdots, & \phi_0 \end{bmatrix}$$

which is a so-called *circulant* matrix. Nevertheless, we shall use the discrete Fourier representation in (16), as it leads to numerical calculation of other statistics (e.g., autocorrelation functions) with favorable computational properties.

3.2 QTES Processes

Recall that any background DTES process, $\{C_n\}$, of the form (13) or (14), is marginally uniform on $\mathcal{Z}_M$, i.e.,

$$P\{C_n = m\} \;=\; \frac{1}{M} \;=\; \delta, \quad m \in \mathcal{Z}_M. \tag{19}$$

The corresponding QTES process, $\{Q_n\}$, is a continuous extension of $\{C_n\}$, defined by

$$Q_n \;=\; \sum_{m \in \mathcal{Z}_M} 1\{C_n = m\}\,\delta[m + W_n], \tag{20}$$

where the sequence $\{W_n\}$ is iid with uniform marginals on $[0,1]$. In other words, Q_n is conditionally uniform on Γ_m, given that $C_n = m$. Combining (19) and (20), it immediately follows that any QTES background sequence is uniform on $[0,1)$. The τ-step transition density, $q_\tau(y|x)$, of $\{Q_n\}$, is given by

$$q_\tau(y|x) \;=\; h_\tau(I_\Gamma(y)|I_\Gamma(x))\,\frac{1}{\delta} \;=\; h_\tau(I_\Gamma(y)|I_\Gamma(x))\,M, \tag{21}$$

where $I_\Gamma(z)$ is the (unique) slice index, $k = k(z)$, in the partition, Γ, such that $z \in \Gamma_k$. Thus, $q_\tau(y|x)$ is only a function of the partition, Γ, in the sense that it does not depend on x and y, but only on the partition indices in which they lie. In other words, q_τ is constant in each variable over partition slices, implying that it maps a continuous domain to a finite range. Thus, we obtain a reduction from a transition function, g_τ, of two continuous variables (in a TES$^+$ background process) to a corresponding q_τ (in a QTES$^+$ background process), which only requires a finite number of values for its representation. It is this reduction in complexity that motivated the term "QTES processes". In fact, q_τ converges to g_τ, when the discretization level, M, tends to infinity, i.e.,

$$\lim_{\delta \to 0} q_\tau^+ \;=\; g_\tau^+, \quad \lim_{\delta \to 0} q_\tau^- \;=\; g_\tau^-, \tag{22}$$

as can be seen from (9) and (15), and from (10) and (16), respectively.

Let D be a distortion, and let $\{Z_n\}$ be the corresponding foreground sequence, with elements $Z_n = D(Q_n)$. Thus, the QTES foreground sequence, $\{Z_n\}$, has the same marginal distribution, $\hat{H}_Y$, as the TES foreground sequence $\{X_n\}$, with $X_n = \hat{H}_Y^{-1}(S_\xi(U_n))$; recall that $\hat{H}_Y$ is the empirical distribution of the underlying empirical time series, $Y = \{Y_n\}$. Let further $\bar{D}_m$ be the (partial) mean of Z_n on slice Γ_m, namely,

$$\bar{D}_m \;=\; \frac{1}{\delta}\int_{m\delta}^{(m+1)\delta} D(x)\,dx,$$

and notice that

$$\mu_Z \;=\; \int_0^1 D(x)\,dx \;=\; \delta\tilde{D}(0) \;=\; \delta\sum_{j=0}^{M-1}\bar{D}_j. \tag{23}$$

The autocorrelation functions, $\rho_Z^+(\tau)$ of $\{Z_n^+\}$ and $\rho_Z^+(\tau)$ of $\{Z_n^+\}$, are computed in the following lemma.

Lemma 2. For $\tau = 0,\ 1, \ldots,$

$$\rho_Z^+(\tau) \;=\; \frac{\delta^2}{\sigma_Z^2}\sum_{\nu=1}^{M-1}\tilde{f}_J^\tau(i2\pi\nu)\,|\tilde{D}(i2\pi\nu)|^2, \tag{24}$$

and

$$
\rho_Z^-(\tau) \;=\; \begin{cases} \rho_Z^+(\tau), & \tau \text{ even}, \\[2ex] \dfrac{\delta^2}{\sigma_Z^2} \displaystyle\sum_{\nu=1}^{M-1} \tilde{f}_J^\tau(i2\pi\nu)\, \tilde{D}^2(i2\pi\nu)\, e^{-i2\pi\nu\delta}, & \tau \text{ odd}. \end{cases} \tag{25}
$$

Proof: Analogously to Theorem 3 in [8], the autocovariance function, ψ_Z^+, of $\{Z_n^+\}$ evaluates to

$$
\begin{aligned}
\psi_Z^+(\tau) \;&=\; E[Z_0^+\, Z_\tau^+] \;-\; \mu_Z^2 \;=\; \int_0^1 \int_0^1 D(x)\, D(y)\, q_\tau^+(y|x)\, dx\, dy \;-\; \mu_Z^2 \\
&=\; \sum_{k=0}^{M-1} \sum_{j=0}^{M-1} \frac{1}{\delta}\, h_\tau^+(k|j) \int_{k\delta}^{(k+1)\delta} \int_{j\delta}^{(j+1)\delta} D(x)\, D(y)\, dx\, dy \;-\; \mu_Z^2 \quad \text{by (21)} \\
&=\; \delta^2 \sum_{\nu=0}^{M-1} \tilde{f}_J^\tau(i2\pi\nu) \sum_{k=0}^{M-1} \sum_{j=0}^{M-1} \bar{D}_k\, \bar{D}_j\, e^{i2\pi\nu\delta(k-j)} \;-\; \delta^2\, |\tilde{D}(0)|^2 \qquad \text{by (15), (23)} \\
&=\; \delta^2 \sum_{\nu=0}^{M-1} \tilde{f}_J^\tau(i2\pi\nu)\, |\tilde{D}(i2\pi\nu)|^2 \;-\; \delta^2\, |\tilde{D}(0)|^2,
\end{aligned}
$$

which agrees with (24).
The proof for (25) is analogous to that of Theorem 4 in [8], and is not reproduced here. □

3.3 Construction of DTES Innovation Densities

Since the DTES innovation sequence, $\{J_n\}_{n=1}^\infty$, is an iid sequence of integer-valued random variables, the sequence is fully characterized by its common density, $f_J \equiv \{\phi_m : \; m \in \mathcal{Z}_M\}$, or equivalently, by the ϕ_m. The relations between TES and QTES suggest that we may use a "good" TES innovation density, f_V, to construct the desired DTES innovation density, f_J. A procedure to construct the latter from the former is as follows:

Step 1: Find a satisfactory TES innovation density, f_V, on the unit circle $[0, 1)$, using either a heuristic search (see [4]) or an automated search (see [12]).

Step 2: Select a discretization level, M, for the DTES process.

Step 3: Partition the unit circle, $[0, 1)$, into M equal slices, yielding

$$
\Gamma(M) \;=\; \{\Gamma_m = [m\delta, (m+1)\delta) : \; m \in \mathcal{Z}_M\},
$$

and then construct the distribution of the DTES innovation variates as

$$
\phi_m \;=\; \int_{\Gamma_m} f_V(v) dv. \tag{26}
$$

For greater accuracy of the QTES model, we can always increase the discretization level, M, in Step 2 (at the expense of an increased computation time). But for practical purposes, we only need increase M to a point that the performance measures of the QTES model are sufficiently close to their TES counterparts.

4 Analysis of the *QTES/PH/1* Queue

In this section, we first show that the QTES process is a semi-Markov process. We then derive some preliminary results about the computation of a key matrix and finally, we show how to compute the performance measures of interest.

4.1 The semi-Markov Kernel of the QTES Process

We first show that the QTES sequence $\{Z_n\}_{n=0}^{\infty}$ forms a semi-Markov process. This semi-Markovian process is characterized by its kernel, $\mathcal{K}(t) \stackrel{def}{=} [K_{ij}(t)]_{M\times M}$, given by

$$\begin{aligned} K_{ij}(t) &\stackrel{def}{=} P[Z_n \le t,\ C_n^+ = j | C_{n-1}^- = i] \\ &= h^+(j|i) \cdot P[D(\delta(j+W_n)) \le t], \end{aligned} \tag{27}$$

where the second equality is due to (20), and $D(\cdot) = \hat{H}_Y^{-1}(S_\xi(\cdot))$. Now,

$$\begin{aligned} P[D(\delta(j+W_n)) \le t] &= P[S_\xi(\delta(j+W_n)) \le \hat{H}_Y(t)] \\ &= P[\delta(j+W_n) \ \le\ \xi\, \hat{H}_Y(t)] \\ &\quad + P[\delta(j+W_n) \ \ge\ 1-(1-\xi)\, \hat{H}_Y(t)] \\ &= \frac{1}{\delta}\, [\xi\, \hat{H}_Y(t) - j\delta]\, 1_{\Gamma_j}(\xi \hat{H}_Y(t)) \\ &\quad + \frac{1}{\delta}\, [(j+1)\delta - (1-(1-\xi)\hat{H}_Y(t))]\, 1_{\Gamma_j}(1-(1-\xi)\, \hat{H}_Y(t)), \end{aligned}$$

and so, the derivative of the kernel, $K_{ij}(t)$, can be written as

$$\begin{aligned} \frac{dK_{ij}(t)}{dt} &= \frac{h^+(j|i)}{\delta}\, \frac{d\hat{H}_Y(t)}{dt}\, [\xi\, 1_{\Gamma_j}(\xi \hat{H}_Y(t)) \ +\ (1-\xi)\, 1_{\Gamma_j}(1-(1-\xi)\hat{H}_Y(t))] \\ &= \frac{h^+(j|i)}{\delta}\, \frac{d\hat{H}_Y(t)}{dt}\, [\xi\, 1_{[a_j,b_j)}(t) \ +\ (1-\xi)\, 1_{[c_j,d_j)}(t)], \end{aligned} \tag{28}$$

where

$$\begin{aligned} a_j &= \hat{H}_Y^{-1}\Big(\tfrac{j\delta}{\xi}\Big), & b_j &= \hat{H}_Y^{-1}\Big(\tfrac{(j+1)\delta}{\xi}\Big), \\ c_j &= \hat{H}_Y^{-1}\Big(\tfrac{1-(j+1)\delta}{1-\xi}\Big), & d_j &= \hat{H}_Y^{-1}\Big(\tfrac{1-j\delta}{1-\xi}\Big). \end{aligned} \tag{29}$$

Note that intervals $[a_j, b_j)$ and $[c_j, d_j)$ do not overlap. Thus, if $1\{Q_n = j\} = 1\{Z_n \in \Gamma_j\} = 1$, then either $1_{[a_j,b_j)}(t) = 1$ (when $(j+1)\delta \ \le\ \xi$), or $1_{[c_j,d_j)}(t) = 1$ (when $(j+1)\delta \ >\ \xi$), but not both.

4.2 Some Preliminary Results

We assume that the time for service has a phase-type distribution represented by (β, R), in which β is a $1 \times K$ vector, R is a $K \times K$ rate matrix, and K is the number of phases [23].

We now show how to analyze the behavior of the *QTES/PH/1* queue by applying the theory and numerical algorithm developed in [26, 27]. The key quantity for computing the performance measures of the *QTES/PH/1* queue (*e.g.*, waiting-time distributions, queue-length distributions, etc.) is an $MK \times MK$ matrix T which is the minimal solution of the following nonlinear matrix equation

$$\begin{aligned} T \;&=\; I \otimes R \,+\, \int_0^\infty exp(Tt)\, d(\mathcal{K}(t) \otimes I)(I \otimes R^* e_K \beta) \\ &\stackrel{def}{=}\; I \otimes R \,+\, \tilde{K}(T)(I \otimes R^* e_K \beta), \end{aligned} \tag{30}$$

where e_K is a $K \times 1$ vector of ones, R^* is a diagonal matrix satisfying $-R^* e_K = R e_K$, and the symbol $\otimes$ denotes the Kronecker product operator [1]. By Theorem 2.1 of [26], T can be found by the recursion

$$T_n \;=\; \begin{cases} I \otimes R, & n = 0 \\ I \otimes R \,+\, \displaystyle\int_0^\infty exp(T_{n-1}t)\, d(\mathcal{K}(t) \otimes I)(I \otimes R^* e_K \beta), & n > 0 \end{cases} \tag{31}$$

This iteration is carried out until the maximum norm of $T_n - T_{n-1}$ falls below a given error criterion (say 10^{-7}).

In order to solve for T via (31), we need to calculate the matrix of integrals

$$\tilde{K}(T) \;=\; \int_0^\infty exp(Tt)\, d(\mathcal{K}(t) \otimes I)$$

in an efficient manner, and if possible, in a closed form. To this end, let us define

$$L_{ij}(T) \;\stackrel{def}{=}\; \int_0^\infty exp(Tt)\, dK_{ij}(t) \tag{32}$$

for $i, j \in \mathcal{Z}_M$. Note that $L_{ij}(T)$ is of dimension MK while $K_{ij}(t)$ is a scalar function. The empirical (histogram) distribution function, $\hat{H}_Y$, is a piecewise linear function (see (8)) of the form

$$\hat{H}_Y(t) \;=\; \sum_{j=1}^{J} 1_{[l_j, r_j)}(t)[\hat{C}_{j-1} + (t - l_j)\, \frac{\hat{p}_j}{w_j}], \quad -\infty < t < \infty. \tag{33}$$

Therefore,

$$\frac{d\hat{H}_Y(t)}{dt} \;=\; \sum_{j=1}^{J} 1_{[l_j, r_j)}(t)\, \frac{\hat{p}_j}{w_j}. \tag{34}$$

We now have two cases from (28), one in which $(j+1)\delta \,\le\, \xi$ and the other in which $(j+1)\delta \,>\, \xi$. For the first case, we have

$$\begin{aligned} L_{ij}(T) \;&=\; \int_0^\infty exp(Tt)\, dK_{ij}(t) \\ &=\; \frac{h^+(j|i)}{\delta}\, \xi \int_{a_j}^{b_j} exp(Tt) \left(\sum_{j=1}^{J} 1_{[l_j, r_j)}(t)\, \frac{\hat{p}_j}{w_j} \right) dt \\ &=\; \frac{h^+(j|i)}{\delta}\, \xi\, T^{-1} \left[\frac{\hat{p}_n}{w_n}\, (e^{Tb_j} - e^{Tl_n}) \right. \\ &\quad \left. +\; \frac{\hat{p}_{n+1}}{w_{n+1}}\, (e^{Tr_{n+1}} - e^{Tl_{n+1}}) \,+\, \cdots +\, \frac{\hat{p}_m}{w_m}\, (e^{Tr_m} - e^{Ta_j}) \right]. \end{aligned} \tag{35}$$

For the second case, we have (details omitted)

$$
\begin{aligned}
L_{ij}(T) \;&=\; \frac{h^+(j|i)}{\delta}\,(1-\xi)\,T^{-1}\left[\frac{\hat{p}_n}{w_n}\,(e^{Td_j}-e^{Tl_n})\right. \\
&\left.+\;\frac{\hat{p}_{n+1}}{w_{n+1}}\,(e^{Tr_{n+1}}-e^{Tl_{n+1}})\;+\cdots+\;\frac{\hat{p}_m}{w_m}\,(e^{Tr_m}-e^{Tc_j})\right]. \qquad (36)
\end{aligned}
$$

With $L_{ij}(T)$ in hand, for all i and j, one has

$$
\begin{aligned}
\left[\tilde{K}(T)\right]_{ij} \;&=\; \left[\int_0^\infty exp(Tt)\,d(\mathcal{K}(t)\otimes I)\right]_{ij} \\
&=\; \int_0^\infty \sum_{k=0}^{MK-1} [exp(Tt)]_{ik}\;d\,[\mathcal{K}(t)\otimes I]_{kj} \\
&=\; \sum_{k=0}^{MK-1}\int_0^\infty [exp(Tt)]_{ik}\;dK_{\lfloor\frac{k}{K}\rfloor\lfloor\frac{j}{K}\rfloor}(t) \\
&=\; \sum_{k=0}^{MK-1}\left[L_{\lfloor\frac{k}{K}\rfloor\lfloor\frac{j}{K}\rfloor}(T)\right]_{ik}. \qquad (37)
\end{aligned}
$$

This completes the description of the algorithm.

4.3 Performance Measures of the *QTES/PH/1* Queue

Having evaluated the matrix T (and then the matrix $\tilde{K}(T)$ via (37)), one may proceed to calculate various performance measures for the *QTES/PH/1* queue. We now give these results but omit the proofs. The reader is referred to [26, 27] and the references therein.

Let α be a $1\times K$ vector such that $\alpha(R+R^*e_K\beta)=0$ and $\alpha e_K=1$. Then the mean service time, μ^{-1}, is obtained from the relation,

$$\mu \;=\; \alpha R^* e_K.$$

Let π be a $1\times M$ vector satisfying $\pi\mathcal{H}^+=\pi$ and $\pi e_M=1$. Then the mean interarrival time, λ^{-1}, is given by

$$\lambda^{-1} \;=\; \sum_{m=0}^{M-1}\pi_m\,\bar{D}_m.$$

Let $\rho=\lambda/\mu$, and assume that the queue is stable, i.e., $\rho<1$. Then the following results hold.

1. The waiting time distribution, $W_q(t)$, is given by

$$W_q(t) \;=\; 1\;-\;\mu^{-1}\,(e_M^{'}\otimes\alpha)\,exp(Tt)\,(T-I\otimes R)\,e_{MK}. \qquad (38)$$

2. The sojourn time distribution, $W(t)$, is given by

$$W(t) \;=\; 1\;-\;\mu^{-1}\,(e_M^{'}\otimes\alpha)\,exp(Tt)\,(I\otimes R^*)\,e_{MK}. \qquad (39)$$

3. The virtual waiting time distribution, $W_v(t)$, is given by

$$W_v(t) \;=\; 1 \;-\; \rho\,(e_M^{'} \otimes \alpha)\, exp(Tt)\, e_{MK}. \tag{40}$$

4. The queue length probabilities at arrival epochs, x_n, are given by

$$x_n \;=\; -\mu^{-1}\,(e_M^{'} \otimes \alpha)\, T\, \tilde{K}^n(T)\, e_{MK}, \quad n \geq 0. \tag{41}$$

5. The queue length probabilities at arbitrary time, y_n, are given by

$$y_n \;=\; \begin{cases} 1 \;-\; \rho, & n = 0 \\ \rho\,(e_M^{'} \otimes \alpha)(I - \tilde{K}(T))\, \tilde{K}^{n-1}(T)\, e_{MK}, & n > 0 \end{cases} \tag{42}$$

Remark: Note that it is possible to extend these results to the case in which the service times are from a semi-Markov process (but are phase type) and also to the case in which the service times depend on the state of the arrival process. For details, see [27].

5 Numerical Results

In this section, we illustrate the methodology with some examples. We have actually carried out similar experiments with a fairly large class of problems. Due to lack of space, we present only a part of our experience. Note that we do not discuss the modelling part of the problem here (posed as Item 1 in the first paragraph of section 1). The reason for not doing so is that it has been reported elsewhere (see [21, 22] for specific examples which include non-monotonic behavior of the autocorrelation function, and [12] for algorithmic TES modelling). So we concentrate only on how well QTES approximates TES and issues of analytical tractability and computational feasibility.

We assume that the service time has an Erlangian distribution with a scale parameter ζ and a shape parameter r, where ζ is varied to obtain different utilizations and $r = 1, 2$ or 3. We further assume that the marginal distribution of the arrival process for TES is exponential with rate $\lambda = 1$. In our algorithm, we convert this distribution into a histogram on the interval $[0, 5)$ with 25 cells. For this marginal distribution, the TES arrival process can generate different autocorrelation functions by properly choosing densities for the innovation sequence, $\{V_n\}_{n=1}^{\infty}$, over the unit circle $[0, 1)$. We consider the following four different TES innovations with the same exponential marginal distribution (thus four different autocorrelations).

Case 1: V_n is uniformly distributed on the unit circle $[0, 1)$. In this case, the corresponding TES process becomes a renewal process, so that its autocorrelation is $\rho_X^+(\tau) = 0$, for all lags $\tau \geq 1$. Thus, the queueing system is reduced to an *M/PH/1* queue.

Case 2: V_n is uniformly distributed over the two intervals $[0, 0.05)$ and $[0.95, 1)$ with probabilities of 0.375 each, and uniformly distributed over the interval $[0.15, 0.25)$ with probability 0.25.

Case 3: V_n is uniformly distributed over the two intervals, $[0, 0.1)$ and $[0.9, 1)$, with probabilities of 0.5 each.

Case 4: V_n is uniformly distributed over the two intervals, $[0, 0.05)$ and $[0.95, 1)$, with probabilities of 0.5 each.

For these four cases, we first try to figure out how well QTES approximates TES for various values of M (this is clearly an important modelling issue). The result for case 3 above is shown in Fig. 1. As can be seen, the approximation improves as M is made larger (as we should expect) and for $M = 20$ or 40, the autocorrelation function of TES is approximated extremely well by QTES. In Fig. 2, we show the autocorrelation functions for TES for all four cases mentioned above. In each of the four cases, we show the same function from QTES for an appropriate value of M. We observe that for the same value of M, the quality of the approximation actually becomes better as the input process gets close to the renewal case. Stated differently, the value of M needed for a good match of the autocorrelation function increases when the input process is highly autocorrelated. Typically, a value of $M = 40$ gives a good match for all sequences including the most highly autocorrelated one.

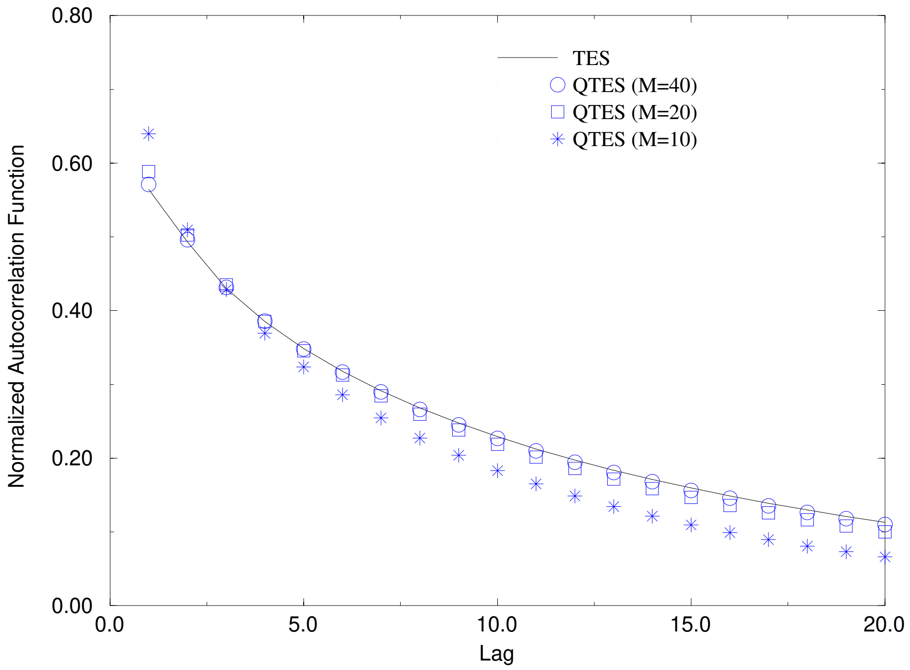

Figure 1: Normalized Autocorrelation Function of Interarrival Times for Case 3.

We now present some results from the queueing analysis of this paper, using QTES as the input process. In all the results to follow, we use QTES for the 4 cases mentioned above with an arrival rate of 1. The mean service time is varied to

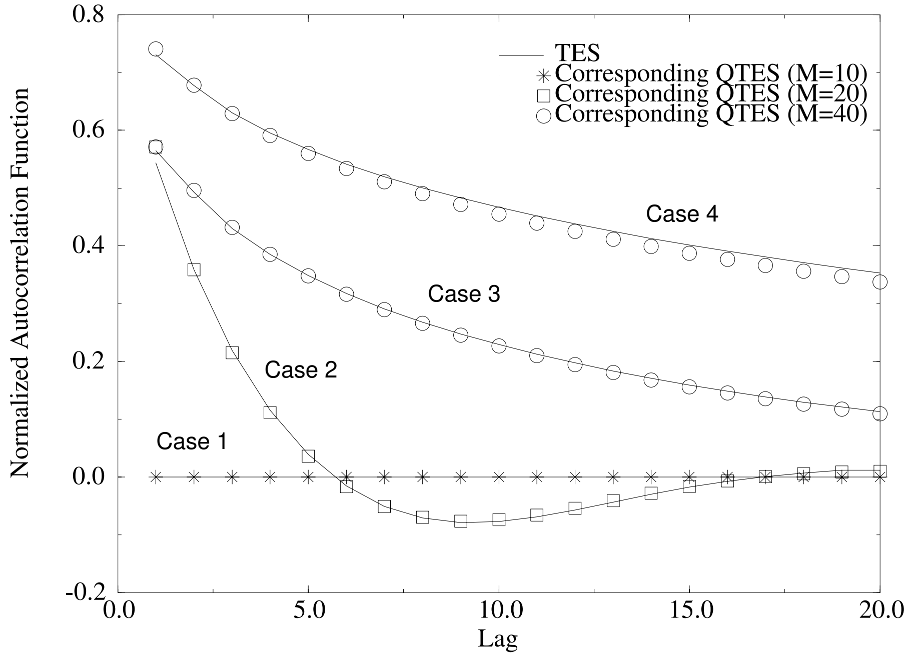

Figure 2: Normalized Autocorrelation Functions of Interarrival Times for All Four Cases.

obtain different utilizations. All results involving input TES models of arrivals are from simulations. All simulation runs were replicated to obtain the 95% confidence intervals.

In Fig. 3, we show the complementary waiting time distribution (see (38)) for case 3 for TES and for QTES with $M = 10, 20$ and 40 and a utilization of 0.8 (the service time is assumed to be exponentially distributed). As can be seen, increasing the value of M gives a better approximation to TES. Again a value such as $M = 20$ or 40 gives a good answer which would be acceptable for most applications. In Fig. 4, we show the queue-length distribution (see (42)) for case 3 for TES and QTES with $M = 10, 20$ and 40 and a utilization of 0.8 (the service time is assumed to be exponentially distributed). Again $M = 20$ or 40 give extremely good answers.

In Tables 1 and 2, we show the mean and variance of waiting time for case 3 for TES and for QTES with $M = 10, 20$ and 40, under various utilizations (the service time is assumed to be exponentially distributed). There are two entries in each case in these tables. The first corresponds to the case of 25 histogram cells and the second (in parentheses) to the case of 200 histogram cells. These examples should convince the reader that the computations can be carried out to a high degree of fidelity for the marginal distribution of the inter-arrival times. In Table 3, we show the mean waiting times for all 4 cases above for TES and for QTES for the stated values of M under various utilizations (the service times are assumed to be exponentially distributed). Finally, in Table 4, we show the mean waiting time for case 3 for TES and for QTES for $M = 40$ under various utilizations (the service time is assumed to be Erlangian with shape parameter $r = 1, 2$ and 3). For comparison purposes, we also show the

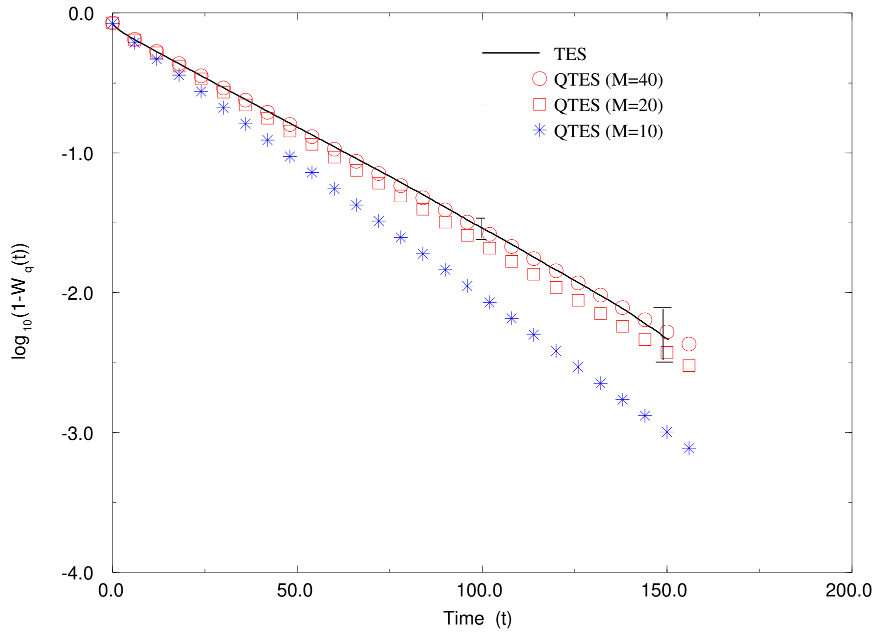

Figure 3: Logarithm of Complementary Waiting Time Distribution for Case 3.

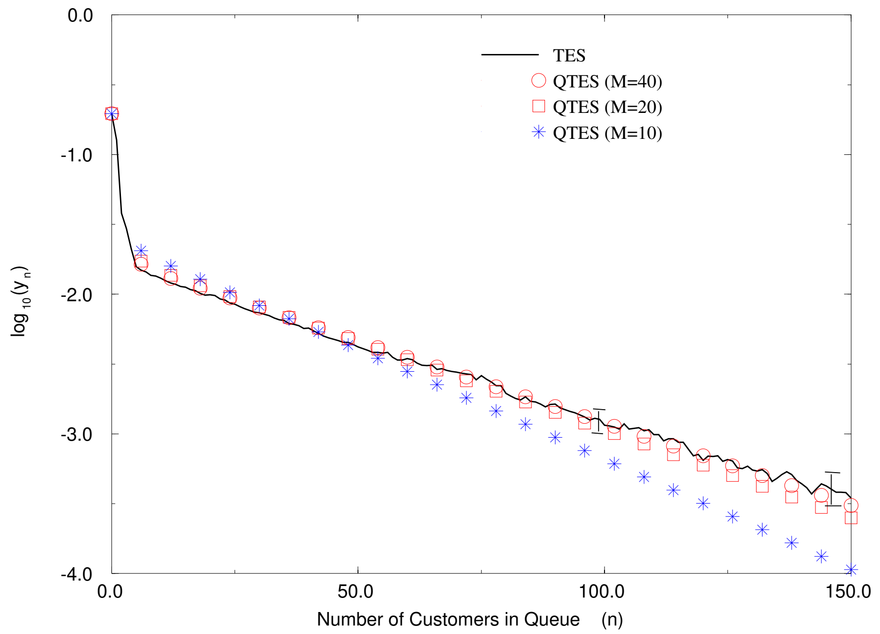

Figure 4: Logarithm of Queue Length Distribution (at arbitrary time) for Case 3.

simulation result for the *TES/D/1* queue (the deterministic service corresponds to the shape parameter $r = \infty$). Note that the results from QTES are getting closer to those from the *TES/D/1* queue as the value of the shape parameter is increased (as we should expect).

	ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8
TES	0.171 ± 0.0036 (0.178 ± 0.0053)	1.51 ± 0.033 (1.55 ± 0.018)	6.34 ± 0.15 (6.50 ± 0.18)	24.0 ± 0.4 (24.77 ± 0.22)
QTES ($M = 40$)	0.171 (0.178)	1.51 (1.54)	6.31 (6.43)	23.8 (24.35)
QTES ($M = 20$)	0.166 (0.173)	1.44 (1.47)	5.95 (6.06)	22.4 (22.84)
QTES ($M = 10$)	0.151 (0.156)	1.23 (1.25)	4.88 (4.97)	18.1 (18.5)

Table 1: Mean Waiting Time for Case 3: TES vs. QTES for $M = 10,\ 20,\ 40$ for 25 (200) histogram cells.

	ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8
TES	0.45 ± 0.01 (0.46 ± 0.01)	2.83 ± 0.074 (2.92 ± 0.03)	9.47 ± 0.24 (9.64 ± 0.34)	29.8 ± 1.24 (29.9 ± 1.1)
QTES ($M = 40$)	0.45 (0.46)	2.80 (2.85)	9.39 (9.54)	29.2 (29.77)
QTES ($M = 20$)	0.44 (0.45)	2.66 (2.71)	8.84 (8.97)	27.4 (27.88)
QTES ($M = 10$)	0.40 (0.41)	2.26 (2.30)	7.20 (7.31)	22.0 (22.41)

Table 2: Variance of Waiting Time for Case 3: TES vs. QTES for $M = 10,\ 20,\ 40$ for 25 (200) histogram cells.

		ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8
Case 1	TES ($M/M/1$)	0.05	0.267	0.9	3.2
	QTES ($M = 20$)	0.048	0.263	0.89	3.17
Case 2	TES	0.122 ± 0.0035	0.691 ± 0.015	2.08 ± 0.036	6.44 ± 0.15
	QTES ($M = 40$)	0.122	0.690	2.07	6.44
Case 3	TES	0.171 ± 0.0036	1.51 ± 0.033	6.34 ± 0.15	24.0 ± 0.4
	QTES ($M = 40$)	0.171	1.51	6.31	23.8
Case 4	TES	0.443 ± 0.053	4.95 ± 0.31	22.1 ± 1.2	85.7 ± 2.8
	QTES ($M = 40$)	0.414	4.73	21.4	82.6

Table 3: Mean Waiting Time for All Four Cases: TES vs. QTES.

	ρ=0.2	ρ=0.4	ρ=0.6	ρ=0.8
$TES/M/1$	0.171 ± 0.0036	1.51 ± 0.033	6.34 ± 0.15	24.0 ± 0.4
$QTES/M/1$	0.171	1.51	6.31	23.8
$TES/E_2/1$	0.143 ± 0.0029	1.40 ± 0.027	6.08 ± 0.12	23.4 ± 0.88
$QTES/E_2/1$	0.142	1.36	5.86	23.1
$TES/E_3/1$	0.134 ± 0.0027	1.36 ± 0.028	5.97 ± 0.13	23.2 ± 0.83
$QTES/E_3/1$	0.133	1.33	5.77	22.8
$TES/D/1$	0.117 ± 0.0027	1.28 ± 0.031	5.76 ± 0.14	22.6 ± 0.74

Table 4: Mean Waiting Time of Different Service Distributions: TES vs. QTES ($M = 40$).

The conclusions that we draw from these numerical results are:

1. TES can be approximated very well by QTES, both for the autocorrelation function and queueing behavior.

2. The quality of the approximation improves with the value of M. Typically a value of $M = 40$ gives extremely good results. Smaller values of M will suffice for lower utilizations or lower values of the autocorrelation function.

3. The analytical methods presented in the paper are computationally feasible for calculating both the waiting-time and queue-length distributions to the extent of the examples presented earlier (and also for examples that we have tried but not presented). The examples tried are limited to matrices of size 120 and 200 histogram cells. It is possible that many applications in communications, computers and manufacturing may fall within these parameters to render this method useful, provided the user is willing to accept errors of 5-10% or so.

References

[1] Bellman, R., *Introduction to Matrix Analysis*, McGraw-Hill Book Company, 1970.

[2] Devroye, L., *Non-Uniform Random Variate Generation*, Springer-Verlag, 1986.

[3] Frost, V. and Melamed, B., "Overview of Simulation and Traffic Modeling for Telecommunications Networks", *Communications Magazine*, 32(3) 70-81, 1994.

[4] Geist, D. and Melamed, B., "TEStool: An Environment for Visual Interactive Modeling of Autocorrelated Traffic", *Proceedings of the 1992 International Conference on Communications* (Chicago, Illinois), 3, 1285-1289, 1992.

[5] Heffes, H. and Lucantoni, D. M., "A Markov Modulated Characterization of Packetized Voice and Data and Related Statistical Multiplexer Performance", *IEEE J. on Selected Areas in Communications*, Vol. SAC-4, 856-868, 1986.

[6] Jacobs, P. A., "A Cyclic Queueing Network With Dependent Exponential Service Times", *J. Appl. Prob.*, 15, 573-589, 1978.

[7] Jacobs, P. A., "Heavy Traffic Results For Single-Server Queues With Dependent (EARMA) Service and Interarrival Times", *Adv. Appl. Prob.*, 12, 517-529, 1980.

[8] Jagerman, D. L. and Melamed, B., "The Transition and Autocorrelation structure of TES Processes Part I: General Theory", *Stochastic Models*, 8(2), 193-219, 1992.

[9] Jagerman, D. L. and Melamed, B., "The Transition and Autocorrelation structure of TES Processes Part II: Special Cases", *Stochastic Models*, 8(3), 499-527, 1992.

[10] Jagerman, D. L. and Melamed, B., "The Spectral Structure of TES Processes", *Stochastic Models*, Vol 10, No. 3, 599–618, 1994.

[11] Jagerman, D. L. and Melamed, B., "Burstiness Descriptors of Traffic Streams: Indices of Dispersion and Peakedness", *Proceedings of the Twenty Eighth Annual Conference on Information Sciences and Systems*, Princeton, New Jersey, March 1994.

[12] Jelenkovic, P., Melamed, B., "Automated TES Modeling of Compressed Video", *Infocom '95*, 746-752, 1995.

[13] Latouche, G., "An Exponential Semi-Markov Process, With Applications to Queueing Theory", *Stochastic Models*, 1(2), 137-169, 1985.

[14] L'Ecuyer, P., "Efficient and Portable Combined Random Number Generators", *CACM* 31, 742-749, 1988.

[15] Livny, M., Melamed, B. and Tsiolis, A. K., "The Impact of Autocorrelation on Queuing Systems", *Management Science*, 39, 322-339, 1993.

[16] Melamed, B., "TES: A Class of Methods for Generating Autocorrelated Uniform Variates". *ORSA J. on Computing*, 3(4), 317-329, 1991.

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

[18] Melamed, B., Ren, Q. and Sengupta, B., "Modeling and Analysis of a Single Server Queue with Autocorrelated Traffic", *Infocom '95*, 634-643, 1995.

[19] Melamed, B. and Hill, J., "Applications of the TES Modeling Methodology", *Proceedings of WSC '93*, 1330–1338, Los Angeles, California, 1993.

[20] Melamed, B. and Hill, J., "TEStool: A Visual Interactive Environment for Modeling Autocorrelated Time Series", *to appear in Performance Evaluation*, 1995.

[21] Melamed, B., and Sengupta, B., "TES Modeling of Video Traffic", *IEICE Transactions on Communications*, Vol. E75-B, No. 12, 1292-1300, 1992.

[22] Melamed, B., Raychaudhuri, D., Sengupta, B. and Zdepski, J., "TES-Based Video Source Modeling for Performance Evaluation of Integrated Networks", *IEEE Trans. in Comm.*, vol. 42, no. 10, 2773-2777, 1994.

[23] Neuts, M., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins, Baltimore, 1981.

[24] Oppenheim, A. and Schafer, R., *Digital Signal Processing*, Prentice-Hall, Inc., 1975.

[25] Patuwo, B. E., Disney, R. L. and McNickle, D. C. "The Effect of Correlated Arrivals on Queues", *IIE Transactions*, Vol. 25, No. 3, 105–110, 1993.

[26] Sengupta, B., "Markov Process whose Steady State Distribution is Matrix-Exponential with an Application to the *GI/PH/1* Queue", *Adv. App. Prob.*, 21, 159-180, 1989.

[27] Sengupta, B., "The Semi-Markovian Queue: Theory and Applications", *Stochastic Models*, 6(3), 383-413, 1990.

[28] Tin, P., "A Queueing System with Markov-Dependent Arrivals", *J. Appl. Prob.*, 22, 668-677, 1985.