STOCHASTIC MODELS
Vol. 20, No. 1, pp. 71–101, 2004

Transient Analysis of Fluid Flow Models via Stochastic Coupling to a Queue

Soohan Ahn and V. Ramaswami*

AT&T Labs, Research, Florham Park, New Jersey, USA

ABSTRACT

Markovian fluid flow models are used extensively in performance analysis of communication networks. They are also instances of Markov reward models that find applications in several areas like storage theory, insurance risk and financial models, and inventory control. This paper deals with the transient (time dependent) analysis of such models. Given a Markovian fluid flow, we construct on the same probability space a sequence of queues that are stochastically coupled to the fluid flow in the sense that at certain selected random epochs, the distribution of the fluid level and the phase (the state of the modulating Markov chain) is identical to that of the work in the queue and the phase. The fluid flow is realized as a stochastic process limit of the processes of work in the system for the queues, and the latter are analyzed using the matrix-geometric method. These in turn provide the needed characterization of transient results for the fluid model.

Key Words: Fluid-flow; Queues; QBD; Transient results; Matrix-geometric method; Stochastic coupling.

*Correspondence: V. Ramaswami, AT&T Labs, Research, 180 Park Avenue, E-233 Florham Park, NJ 07932, USA; E-mail: vramaswami@att.com.

DOI: 10.1081/STM-120028392 1532-6349 (Print); 1532-4214 (Online)

www.dekker.com

1. INTRODUCTION

This work is an extension to transient (time dependent) results of those of Ramaswami[12] and of Ahn and Ramaswami[3] on steady state results for stochastic fluid flows. In the models considered, the net rate of change of the fluid at any instant depends on the state (called phase) of a finite state, continuous time Markov chain (CTMC).

A direct computation of the transient results by numerical integration of the partial differential equations (pde's) for the transient probabilities is both difficult and hazardous due to possible build up of errors. Spectral methods are not very successful either; see Refs.[6,16]. An effort based on ad-hoc series expansions by Sericola[14] is highly successful, but it does not lend itself to generalizations or to probabilistic interpretations. We wish to provide a powerful alternative that also holds some hope of extensions in the future to more challenging models (such as those modulated by countable or continuous state space processes which are needed to model heavy tails and self similarity.) Our work is rooted in matrix-geometric methods.[7,9,13] The fact that those have been generalized to operator settings, see e.g., Refs.[10,17] gives us the hope that generalizations of our approach to the infinite dimensional case would indeed become possible. We emphasize that such a generalization has not been made in this paper.

Our approach is based on an approximation of the fluid model by the amounts of work in a sequence of Markov modulated queues of the quasi birth and death (QBD) type. That enables the use of matrix-geometric methods for queues. Our main result, Theorem 13, obtains the Laplace transforms (with respect to time) of the time dependent joint distribution of the fluid level and the phase in a form that lends itself to stable computations. The characterization uses certain kernels that appear in the matrix-geometric method and enable one to "jump" directly to a "last epoch before time t," and thereby avoids building the story over many small intervals up to time t as in a numerical solution of the pde's. Following is an overview of the approach and the organization of the paper.

A first step in our analysis is a "spatial uniformization" of the phase process. Under that scheme described in Sec. 2, the phase process is modeled as a Markovian point process such that potential increases to the fluid buffer between two successive epochs of that point process have an exponential distribution with a *common* rate, say θ. That allows one to replace the upward trajectories of the fluid process between the spatial uniformization epochs by jumps occurring at those epochs, and to consider the resulting path as a path of the work in the system for a queue with Markovian arrivals and independent, identically distributed (*iid*) work amounts; herein lies the value of spatial uniformization.

Unfortunately however, the resulting queue does not have a simple structure, because in it each service time is also directly proportional to the interarrival time; we want to avoid that inconvenience. Therefore, we consider a queue in which arrivals come at the spatial uniformization epochs but bring amounts of work that are not only *iid* exponential random variables, but are also independent of the arrival process. Now, paths of the resulting work process do not coincide with those of the fluid process anymore. But nevertheless, we can show (see Sec. 4) that a stochastic coupling still results in a distributional sense between the fluid flow and the work in the queue at points of the chosen Markovian point process.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Unfortunately, distributional coupling would only provide weak limits at individual embedded points and is not adequate for studying the behavior at an arbitrary point t or of functionals of the paths such as the busy period. Therefore, in Sec. 3, we construct a sequence of nested spatial uniformizations on a common probability space defined by a set of θ_n with $\theta_n \to \infty$, as $n \to \infty$, with the hope that, since the mean quanta sizes $1/\theta_n \to 0$, we shall, in the limit, realize the fluid flow as a pathwise limit of the work processes.

That such a hope is not misplaced is illustrated by Fig. 1 where we consider the fluid flow modulated by the Markov process with generator

$$Q = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix},$$

in which the fluid level increases at rate 1 in state 1 and decreases at rate 1 in state 2. The figure depicts the results of a simulation experiment with spatial uniformizations given by $\theta_n = n$, for various values of n. The left hand side plots give the sample paths of the fluid model and of the work in the queues, and the right hand side plots give the paths of their differences. Note that as $n \to \infty$, the differences do become negligible.

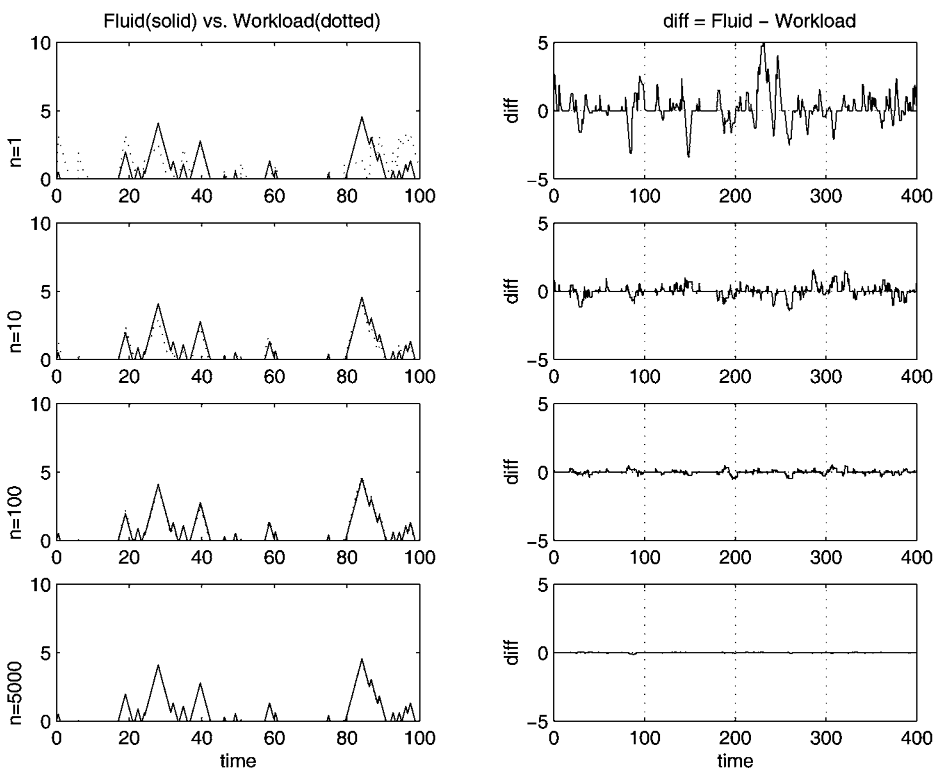

Figure 1. Comparison of the paths of the fluid flow process and workload in the queue $\mathcal{Q}^{(n)}$ for $n = 1, 10, 100, 5000$.

At a high-level, this article is but an implementation of the above simple ideas. As noted, the details of the construction of the needed queues is given in Sec. 3. In Sec. 4, we establish stochastic coupling at the spatial uniformization epochs and also show that each queue is of the QBD type; see Theorem 3 and Theorem 5. In Sec. 5, Theorem 8 we establish the stochastic process limit result that forms the basis of our approach. To let the main ideas flow uninterrupted, the technical details of its proof are moved to a later section, viz., Sec. 9.

Lest the reader should miss a major subtlety of our construction, we note that to compare the sample path increments of the spatial discretizations, it is necessary to make their construction such that all of them are modulated by a *common* phase process and are nested. Our construction in Sec. 3 satisfies this important condition, and that is exploited many times, and particularly so to establish the stronger (pathwise) limit theorems.

In Sec. 6, we develop the transient analysis of the queues using matrix-geometric techniques. These are then used in Sec. 7 to obtain the transient results for the fluid flow model through a limit process. Specifically, Sec. 7 introduces three fundamental kernels for the fluid model that hold the key to its transient analysis. In Sec. 8, numerical computations are performed for a set of examples, and the results are compared to those of Sericola.[14] Finally, in Sec. 10, we provide some concluding remarks.

2. THE MODEL AND SPATIAL UNIFORMIZATION

We assume as given an irreducible, CTMC of "phases" with a finite state space $S = S_1 \cup S_2 \cup S_3$ and infinitesimal generator Q, such that: during sojourn of the CTMC in state $i \in S_1$, the fluid level increases at rate $c_i > 0$; during sojourn of the CTMC in state $j \in S_2$, the fluid level decreases at rate $c_j > 0$; and during sojourn of the CTMC in S_3, the fluid level remains constant. Throughout this article, I will denote an identity matrix and $\mathbf{1}$ a column vector of 1's both of whose dimensions will be determined by the context in which they appear. Where it is necessary to indicate the dimension explicitly, we will write I_n to denote the $n \times n$ identity matrix.

For later use, we define the diagonal matrices

$$C_j = \mathrm{diag}\{c_i, i \in S_j\}, \quad j = 1, 2, 3, \tag{1}$$

where we set $c_i = 1$ for all $i \in S_3$, and let $C = \mathrm{diag}(C_1, C_2, C_3)$ and $c = \sum_{i \in S} c_i$. We partition the states of the Markov chain in conformity with the three sets S_i identified above and denote its infinitesimal generator in partitioned form as

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix}. \tag{2}$$

Throughout, to avoid confusion between submatrices in a partitioned structure and elements of a matrix, the (i, j)th element of a matrix A will be denoted by $[A]_{ij}$ or as $A(i, j)$ instead of as A_{ij} as is often customary.

A spatial uniformization (for the fluid flow) is effected by modeling the Markov process of phases as a Markov renewal process (MRP) with exponential sojourn times such that the potential increases to the fluid level between epochs of that MRP are identically distributed. To that end, we let $\{(J_n, t_n) : n \geq 0\}$ be such a MRP, with successive states $J_n \in S$, transition epochs $0 = t_0 < t_1 < t_2 < \cdots$, and with semi-Markov kernel $H(\cdot)$ defined such that $H(i, j; t)$, the (i, j)th element of $H(t)$, is given by

$$H(i, j; t) = P\{J_{n+1} = j, t_{n+1} - t_n \leq t | J_n = i\} = \left(1 - e^{-\theta c_i t}\right) [P_\theta]_{ij}, \tag{3}$$

where

$$P_\theta = \theta^{-1} C^{-1} Q + I, \quad \text{and} \quad \theta \geq \max_{i \in S} \left\{-[C^{-1}Q]_{ii}\right\}. \tag{4}$$

The semi-Markov process (SMP) $\mathcal{J} = \{J(t) : t \geq 0\}$ is specified such that it takes the value J_n in the interval $t_n \leq t < t_{n+1}$. The following result shows that $\mathcal{J}$ is indeed a realization of the phase process.

Theorem 1. *The process $\mathcal{J} = \{J(t), t \geq 0\}$ is a CTMC with infinitesimal generator Q.*

Proof. It is easy to verify that P_θ is indeed a nonnegative, stochastic matrix. If we assume $J(0) = i \in S$, then the SMP stays in i exactly n steps with probability $[P_\theta]_{ii}^{n-1}(1 - [P_\theta]_{ii})$. Since $t_1 - t_0, t_2 - t_1, \ldots, t_n - t_{n-1}$ are conditionally independent given $J_0, J_1, \ldots, J_{n-1}$, with probability $[P_\theta]_{ii}^{n-1}(1 - [P_\theta]_{ii})$, a sojourn time of $\mathcal{J}$ in state i is the sum of n *iid* exponentially distributed random variables with rate θc_i. Since $[P_\theta]_{ii} = 1 + [Q]_{ii}/(\theta c_i)$, we have for the Laplace-Stieltjes transform (LST) $f_i(s)$ of the sojourn time in state i,

$$f_i(s) = \sum_{n=1}^{\infty} \left(\frac{\theta c_i}{s + \theta c_i}\right)^n [P_\theta]_{ii}^{n-1}(1 - [P_\theta]_{ii}) = \frac{-[Q]_{ii}}{s - [Q]_{ii}},$$

showing that the sojourn time in i is exponentially distributed with parameter $-[Q]_{ii}$. Also, it can be seen that, given that $\mathcal{J}$ transits out of i into a different state, the probability that it moves into j is $[P_\theta]_{ij}/(1 - [P_\theta]_{ii}) = -[Q]_{ij}/[Q]_{ii}$. Hence the result. □

Consider a sojourn interval of the SMP in i for some $i \in S_1$; that is distributed as $\exp(\theta c_i)$, and during that interval, fluid accumulates at rate c_i per unit time. Thus the total additional fluid accumulation in that interval is distributed as $\exp(\theta)$. This underlies our reason for using the nomenclature "spatial uniformization." We note that this is similar to the "stochastic discretization" Adan and Resing[2] used in an operational manner, but formulated more formally.

Later in the paper (see Sec. 4), we will consider a sequence of spatial uniformizations given by the values $\theta_n = n\lambda$, $n = 1, 2, \ldots$, where $\lambda = \max_{i \in S} \left\{-[C^{-1}Q]_{ii}\right\}$. In those contexts, the matrix $P_{n\lambda} = \frac{1}{n\lambda} C^{-1} Q + I$ will be denoted simply as P_n, and the matrix P_1 simply as P. Also, we shall assume P_n to be partitioned in conformity with the partitioning of the state space as

$$P_n = \begin{pmatrix} P_{n11} & P_{n12} & P_{n13} \\ P_{n21} & P_{n22} & P_{n23} \\ P_{n31} & P_{n32} & P_{n33} \end{pmatrix}. \tag{5}$$

Finally, for later use, we note the following equations which are easy to verify:

$$P_{nii} = \frac{n-1}{n} I + \frac{1}{n} P_{ii}, \quad i = 1, 2, 3; \tag{6}$$

and

$$P_{nij} = \frac{1}{n} P_{ij} \quad \text{for } i \neq j, \ \ i, j \in \{1, 2, 3\}. \tag{7}$$

3. COUPLED QUEUES

3.1. Preliminaries

We assume the following as given on a *common* probability space $(\Omega, \mathscr{A}, \mathscr{P})$: A collection of mutually independent Poisson processes, say, $\mathscr{M}_{n,k}$, and $\mathscr{N}_{n,k}$ with rates λc_k respectively for $k \in S$, and $n \geq 1$; and a discrete time Markov chain $\mathscr{L} = \{L_n : n \geq 0\}$ of phases which has transition matrix P and is independent of all the Poisson processes $\mathscr{M}_{n,k}$, and $\mathscr{N}_{n,k}$, $n \geq 1$, $k \in S$. Without loss of generality, we shall assume that $L_0 = i$ for some $i \in S$.

We use $\oplus$ for denoting superposition of processes; thus, $\mathscr{M}_{n,j} \oplus \mathscr{N}_{n,j}$ denotes the superposition of $\mathscr{M}_{n,j}$ and $\mathscr{N}_{n,j}$, and $\bigoplus_{k=1}^{n} \mathscr{M}_{k,j}$ denotes the superposition of the processes $\mathscr{M}_{1,j}, \ldots, \mathscr{M}_{n,j}$. Also, the epochs of the process $\bigoplus_{r=1}^{n} \bigoplus_{j \in S} (\mathscr{M}_{r,j} \oplus \mathscr{N}_{r,j})$ will be denoted by the sequence $\{s_k^n : k \geq 0\}$.

With these as building blocks, we construct for almost all sample points in Ω: (a) a phase process $\mathscr{J} = \{J(t) : t \geq 0\}$ which is a CTMC with generator Q but realized through a spatial uniformization construct as described in Sec. 2; (b) a process $\mathscr{F} = \{F(t) : t \geq 0\}$ such that $F(t)$ increases at rate c_j while $J(t) = j \in S_1$, decreases at rate c_j while $J(t) = j \in S_2$, and remains constant while $J(t) \in S_3$ — i.e., $(F(t), J(t))$ is the fluid flow process of interest; and (c) for each $n \geq 1$, a queue $\mathscr{Q}^{(n)} = \{Q^{(n)}(t) : t \geq 0\}$ and its associated work process $\mathscr{W}^{(n)} = \{W^{(n)}(t) : t \geq 0\}$, where $Q^{(n)}(t)$ is the queue length and $W^{(n)}(t)$ is the amount of work in the queue $\mathscr{Q}^{(n)}$ at time $t+$.

3.2. The Construction

To avoid pedantry and to save notations, we shall suppress the sample point in the ensuing discussion which is indeed a sample point by sample point construction. We denote the set of epochs of the Poisson process $\bigoplus_{r=1}^{n} \mathscr{N}_{r,j}$ by $A_{n,j}$, $j \in S$ and let $A_n = \bigcup_{j \in S} A_{n,j}$; the arrival epochs to the queue $\mathscr{Q}^{(n)}$ will be a subset of A_n as we shall see later. Similarly, we denote the epochs of the Poisson process $\bigoplus_{r=1}^{n} \mathscr{M}_{r,j}$ by $D_{n,j}$ and let $D_n = \bigcup_{j \in S} D_{n,j}$; the departure epochs of $\mathscr{Q}^{(n)}$ will form a subset of them as we shall see later.

Construction of the Phase Process $\mathcal{J}$. Let $a_0 = 0$, and let a_1 be the first epoch of $\mathcal{N}_{1,L_0}$ that occurs after time 0. In general, set a_{n+1} to be the first epoch of $\mathcal{N}_{1,L_n}$ to occur after the epoch a_n. Let $J(t) = L_n$ in the interval $a_n \le t < a_{n+1}$. We note first of all that $\{(L_n, a_n) : n \ge 0\}$ is a MRP of the type discussed in Sec. 2. Also, from Theorem 1, $\mathcal{J} = \{J(t) : t \ge 0\}$ is a CTMC with infinitesimal generator Q. Indeed the epochs $\{a_n\}$ form a set of spatial uniformization epochs for the fluid process modulated by the phase process, with respect to the parameter λ; see Sec. 2.

Construction of the Fluid Flow $\mathcal{F}$. Without loss of generality, we assume the initial condition $F(0) = 0$ and define the process $\{F(t)\}$ such that for $t \in (a_n, a_{n+1})$, $F(t) = F(a_n) + c_j(t - a_n)$ if $J(t) = j \in S_1$, $F(t) = \max[0, F(a_n) - c_j(t - a_n)]$ if $J(t) = j \in S_2$, and finally $F(t) = F(a_n)$ if $J(t) \in S_3$. Defined thus, clearly the joint process $\{(F(t), J(t))\}$ is stochastically equivalent to the fluid model of interest.

Construction of the Queues. For each n, the queue $\mathcal{Q}^{(n)}$ will be defined in terms of the successive embedded epochs $t_0^n = 0$, and $\{t_k^n : k \ge 1\}$ where there is an arrival, departure or phase transition; we emphasize that some phase transitions may be from a phase to itself, as necessitated for instance in the uniformization process. These epochs will be such that $\{t_k^n\} \subset \{s_k^n\}$, where s_k^n were introduced earlier, and $\{t_k^n\}$ yield a spatial uniformization with respect to $n\lambda$. It will be assumed that service is rendered by the server only when the phase is in S_2. Also, for all queues, the queue size at time 0 will be defined to be 0 to match our initial condition $F(0) = 0$. In what follows we shall denote by Q_k^n and J_k^n the queue length (number of customers in the system $\mathcal{Q}^{(n)}$) and the phase at the epoch t_k^n+ respectively.

(a) Let $t_0^n = 0$, $Q_0^n = 0$ and $J_0^n = i$.
(b) Having defined t_k^n and (Q_k^n, J_k^n), we first specify the next time point t_{k+1}^n and then the value of the queue size and phase at that epoch. The queue size in $\mathcal{Q}^{(n)}$ is assumed to remain constant over intervals of the form $[t_k^n, t_{k+1}^n)$; that is, we shall set $Q^{(n)}(t) = Q_k^n$ for all $t \in [t_k^n, t_{k+1}^n)$. There are several cases to consider.

Case 1. If $J_k^n \in S_1$, then t_{k+1}^n is defined to be the first epoch in A_{n,J_k^n} to come after t_k^n, and Q_{k+1}^n is set to $1 + Q_k^n$ — that is, t_{k+1}^n is defined to be an arrival epoch to the queue $\mathcal{Q}^{(n)}$. The phase at J_{k+1}^n is set to $J(t_{k+1}^n)$; note that a phase change occurs at the newly defined epoch iff that epoch $t_{k+1}^n \in A_{1,J_k^n}$ and a different phase is entered; otherwise, that epoch will constitute a self-transition for the phase in the queue $\mathcal{Q}^{(n)}$.

Case 2. If $J_k^n \in S_3$, then t_{k+1}^n is once again defined to be the first epoch in A_{n,J_k^n} to come after t_k^n, but Q_{k+1}^n is set to the same value as Q_k^n. — that is, the queue length remains constant just as the fluid level. The phase J_{k+1}^n is set to $J(t_{k+1}^n)$.

Case 3. If $J_k^n = j \in S_2$, then we set t_{k+1}^n to be the first epoch in $A_{n,j} \cup D_{n,j}$ to come after t_k^n. The queue length at that epoch is set depending on whether that epoch comes from $A_{n,j}$ or from $D_{n,j}$. Specifically, Q_{k+1}^n is set to the same value as Q_k^n if $t_{k+1}^n \in A_{n,j}$; it is changed to $\max(0, Q_k^n - 1)$ if the new epoch $t_{k+1}^n \in D_{n,j}$. Thus, the next epoch is just a phase transition epoch (with no effect on queue size) if it is in $A_{n,j}$, and a departure epoch (with no phase change) if it is in $D_{n,j}$ and a departure is indeed possible; note that except when the epoch is in $A_{1,j}$ and the new phase entered is different, the new epoch is a dummy phase change transition epoch.

Note that the queue $\mathcal{Q}^{(n)}$ is specified above only in terms of its arrival, departure and phase change epochs, and we did not assume any order of service. Now, we assume services are done in the FIFO order. With this assumption, it is easy to see that the construction also specifies the work process $W^{(n)}(t)$ at each point t for almost all $\omega \in \Omega$.

Construction of the Process $\mathcal{Y}^{(n)} = \{Y^{(n)}(t)\}$. Associated with the work in the constructed queue $\mathcal{Q}^{(n)}$, let us now define the process $Y^{(n)}(t)$ such that for $t_k^n \leq t < t_{k+1}^n$,

$$Y^{(n)}(t) = \begin{cases} W^{(n)}(t_k^n) + c_j(t - t_k^n), & \text{if } J(t_k^n) = j \in S_1 \\ \max(0, W^{(n)}(t_k^n) - c_j(t - t_k^n)), & \text{if } J(t_k^n) = j \in S_2 \\ W^{(n)}(t_k^n), & \text{if } J(t_k^n) \in S_3; \end{cases} \tag{8}$$

note that in the intervals $[t_k^n\, t_{k+1}^n)$, the phase process remains constant and the rate of growth of $Y^{(n)}(t)$ mimics that of $F(t)$. Indeed, it will be proved later, that $\{F(t) : t \geq 0\}$ can be realized a.s., as the pathwise stochastic process limit of $\{Y^{(n)}(t) : t \geq 0\}$.

4. PROPERTIES OF $\mathcal{Q}^{(n)}$ AND COUPLING

We shall now show that each arrival in the queue $\mathcal{Q}^{(n)}$ brings in a random amount of work distributed as $\exp(n\lambda)$ independently of the history of the process $(Q^{(n)}(t), J(t))$ up to its arrival epoch. (Since the server serves at variable rates c_j, $j \in S_2$, we avoid using the term "service time" so that the amount of work brought in by a customer is clearly distinguished from his "time in service.")

Theorem 2.

(a) *Arrivals to the queue occur only at those epochs t_k^n for which $J(t_k^n -) \in S_1$; that is, the epoch is a phase transition epoch in A_n from S_1 for that queue (which may very well be a phase self transition).*
(b) *Departures to the queue can occur at t_k^n only if $J(t_k^n -) \in S_2$, $Q_{k-1}^n > 0$ and $t_k^n \in D_n$. Also, the phase immediately after each departure epoch is the same as that immediately prior to that epoch.*
(c) *Assume that work gets depleted at rate c_j while $J(t) = j \in S_2$, and that no service is rendered while $J(t) \in S_1 \cup S_3$. Then the amounts of work done in $\mathcal{Q}^{(n)}$ between successive departure epochs are iid random variables distributed as* $\exp(n\lambda)$.

Proof. The first two assertions are obvious, and we only need to prove (c). Its proof is very similar to that of Theorem 2 in Ref.[3] but given below for completeness.

The service of a customer in $\mathcal{Q}^{(n)}$ begins at an epoch t_k^n with $J_k^n \in S_2$. The service completion epoch is the first epoch t_{k+r}^n, $r \geq 1$ to come after t_k^n for which $J_{k+r}^n \in S_2$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

and $t^n_{k+r} \in D_n$. We let $f_i(s)$ denote the transform of the amount of work to be done by the server until the next departure epoch, given that service starts in phase $i \in S_2$.

From our construction of t^n_k, clearly the interval of time to the next epoch of transition is the minimum of $2n$ exponential distributions with the same parameter λc_i. Since during that interval, work is being depleted at rate c_i, the total work rendered during that interval is also exponentially distributed but with parameter $2n\lambda$ as one may easily verify. Also, the probability that the customer in service departs at that epoch is $\frac{1}{2}$.

Now, if the work of the customer in service is not completed at that epoch, then it is an epoch of phase change which may or may not be in S_2. If it is not in S_2, note that the service of the customer is halted until the phase returns to the set S_2; and such a return will happen w.p. 1 since the phase process is irreducible and hence recurrent non-null (we have a finite number of phases).

Using all these facts and denoting by $b(i, j), i, j \in S_2$ the probability that a service interrupted from phase i gets restarted in phase j, we can now write the equations

$$f_i(s) = \frac{1}{2}\frac{2n\lambda}{s + 2n\lambda} + \frac{1}{2}\frac{2n\lambda}{s + 2n\lambda}\sum_{j \in S_2} b(i, j) f_j(s), \quad i \in S_2.$$

If we set $\mathbf{f}(s)$ to be the column vector of $f_i(s)$'s and B to be the matrix whose elements are $b(i, j), i, j \in S_2$, it is now easy to verify that the above set of linear equations has the solution

$$\begin{aligned}\mathbf{f}(s) &= \left[I - \frac{n\lambda}{s + 2n\lambda}B\right]^{-1}\frac{n\lambda}{s + 2n\lambda}\mathbf{1}\\ &= \sum_{k=0}^{\infty}\left(\frac{n\lambda}{s + 2n\lambda}\right)^{k+1} B^k\mathbf{1}.\end{aligned}$$

Since B is a stochastic matrix, we have

$$\mathbf{f}(s)\mathbf{1} = \frac{n\lambda}{s + n\lambda}\mathbf{1}. \tag{9}$$

Thus, the marginal distribution of the amount of service rendered is exponential with parameter $n\lambda$ and is independent of the starting phase, and the proof is complete. □

The process $\{(Q^{(n)}(t), J(t) : t \geq 0\}$ has a simple structure which lends to an analysis by the matrix-geometric method. That is the main import of our next result.

Theorem 3. *The queues $\mathbb{Q}^{(n)}$ have the following properties*:

(a) *All the queues $\mathbb{Q}^{(n)}$ are modulated by the same continuous time phase process $J(\cdot)$ which is a CTMC with infinitesimal generator Q on the state space S.*
(b) *For each n, $\{(Q^{(n)}(t), J(t)) : t \geq 0\}$ is a continuous time QBD.*
(c) *The embedded sequence $\{(Q^{(n)}(t^n_k), J(t^n_k)) : k \geq 0\}$ is a discrete time QBD.*

Proof. Using the construction, it is elementary to verify that $\{(Q(t_k^n), J(t_k^n))\}$ is a discrete time QBD with transition matrix

$$\begin{pmatrix} M_n & B_{n0} & 0 & 0 & \cdots \\ B_{n2} & B_{n1} & B_{n0} & 0 & \cdots \\ 0 & B_{n2} & B_{n1} & B_{n0} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \tag{10}$$

where $M_n = B_{n2} + B_{n1}$, and

$$B_{n0} = \begin{pmatrix} P_{n11} & P_{n12} & P_{n13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$B_{n1} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2}P_{n21} & \frac{1}{2}P_{n22} & \frac{1}{2}P_{n23} \\ P_{n31} & P_{n32} & P_{n33} \end{pmatrix},$$

$$B_{n2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2}I & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{11}$$

where the matrices P_{nij} are defined by (5)–(7). Furthermore, in state (n, j), $n \geq 0$, $j \in S$, the sojourn time (i.e., the time to the next epoch t_k^n) is exponentially distributed with mean $(n\lambda c_j)^{-1}$ for $j \in S_1$, $(2n\lambda c_j)^{-1}$ for $j \in S_2$, and $(n\lambda)^{-1}$ for $j \in S_3$. This completes the proof. □

Let us denote the queues $\mathcal{Q}^{(n)}$ constructed above by $\mathcal{Q}^{(n)}(\lambda)$ to explicitly note that $\mathcal{Q}^{(n)}$ is the nth queue arising from a spatial uniformization of the phase process, with respect to λ. Then it follows from the above theorem that the laws of both $\mathcal{Q}^{(1)}(n\lambda)$ and $\mathcal{Q}^{(n)}(\lambda)$ are identical. At this point, it is worth noting that the epochs $\{t_k^n : k \geq 0\}$ used in the construction provide a spatial uniformization of the phase process with respect to the parameter $n\lambda$, and that these spatial uniformization epochs are nested in the sense that

$$\{t_k^n : k \geq 0\} \subset \{t_k^{n+1} : k \geq 0\}.$$

From these facts, we record a noteworthy result whose importance stems from its immediate implication that once we have analyzed the queue $\mathcal{Q}^{(1)}(\lambda)$, the results for $\mathcal{Q}^{(n)}(\lambda)$, can be obtained by just changing λ to $n\lambda$ in the final formulae for the queue $\mathcal{Q}^{(1)}(\lambda)$. We state the result formally as a theorem.

Theorem 4. *The queue $\mathcal{Q}^{(n)}(\lambda)$ (which is the nth queue constructed from a spatial uniformization of the phase process with respect to λ) has the same law as the queue $\mathcal{Q}^{(1)}(n\lambda)$ (which is the first queue constructed from a spatial uniformization with respect to $n\lambda$).*

The next result is a coupling theorem and gives the primary reason for our interest in the constructed queueing models.

Theorem 5.

(a) *The process* $\{(Y^{(n)}(t_k^n), J(t_k^n)) : k \geq 0\}$ *has the same probability law as the process* $\{(F(t_k^n), J(t_k^n)) : k \geq 0\}$.
(b) *The process* $\{(Y^{(n)}(t_k^n-), J(t_k^n)) : k \geq 1\}$ *has the same probability law as the process* $\{(Y^n(t_k^n+), J(t_k^n)) : k \geq 1\}$.

Proof. We will show that the partial subsequences up to m have identical distributions for all m. Note that Part (a) is clearly true for $m = 0$ and similarly Part (b) is clearly true for $m = 1$. The proof is completed by mathematical induction by observing that all sequences considered are Markov, and furthermore due to their construction, for all (x, j), the conditional distribution of $(F(t_{k+1}^n-), J(t_{k+1}^n))$ given $(F(t_k^n+), J(t_k^n)) = (x, j)$ is the same as the conditional distribution of $(Y(t_{k+1}^n-), J(t_{k+1}^n))$ and of $(Y(t_{k+1}^n+), J(t_{k+1}^n))$ given $(Y(t_k^n+), J(t_k^n)) = (x, j)$. □

5. FLUID FLOW AS A LIMIT PROCESS

We have noted that in the nth queue $\mathcal{Q}^{(n)}$, the amount of work brought by each customer is distributed as $\exp(n\lambda)$. As $n \to \infty$, these quanta become smaller and smaller. Given that the rates of change of the processes $Y^{(n)}(\cdot)$ and $F(\cdot)$ in the intervals $[s_k^n, s_{k+1}^n)$ are identical and that these have the same distribution at the end points of these intervals, it appears reasonable to expect that as $n \to \infty$, the distribution of $(Y^{(n)}(t), J(t))$ should tend to that of $(F(t), J(t))$. We actually prove a stronger result yielding convergence in probability.

Theorem 6.

(a) *For all* $t \geq 0$, *the following convergence in probability holds*:

$$Y^{(n)}(t) \to F(t) \text{ in pr. as } n \to \infty.$$

(b) *For all* $t > 0, x \geq 0$, *we have*

$$\mathcal{P}_{0i}[F(t) \leq x, J(t) = j] = \lim_{n\to\infty} \mathcal{P}_{0i}[Y^{(n)}(t) \leq x, J(t) = j],$$

where, by $\mathcal{P}_{0i}$ *we denote the conditional probability given that the respective process starts in the state* $(0, i)$.

Proof. Part (b) is a direct consequence of (a). So as not to impede the flow of the paper, the proof of (a) is presented in Sec. 9 separately. □

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

In an entirely analogous manner, we can actually prove the following result; we omit the details of the proof which are along the same lines as that of Theorem 6.

Theorem 7. *For all $u_r > 0$, $r = 1, \ldots, m$, $x_r \geq 0$ and $j_r \in S$, we have*

$$\begin{aligned}&\mathscr{P}_{0i}[F(u_r) \leq x_r, J(u_r) = j_r, 1 \leq r \leq m]\\&\quad = \lim_{n\to\infty} \mathscr{P}_{0i}[Y^{(n)}(u_r) \leq x_r, J(u_r) = j_r, 1 \leq r \leq m].\end{aligned} \tag{12}$$

Due to Theorem 11.3.1v of Ref.,[18] we have that the process $(Y^{(n)}(\cdot), J(\cdot))$ converges as $n \to \infty$ to the process $(F(\cdot), J(\cdot))$ in the sense that

$$\lim_{n\to\infty} \int_\Omega f d\mathscr{P}_n = \int_\Omega f d\mathscr{P}, \tag{13}$$

for all continuous, bounded real-valued functions f on Ω. Under the assumption that the probability space is a separable, metric space, we can now invoke the Skorohod Representation Theorem (See Ref.[15] and also Chapter 3 of Ref.[18]) to assert that there exist versions $\widetilde{Y}^{(n)}(\cdot)$ and $\widetilde{F}(\cdot)$ of the processes $Y^{(n)}(\cdot)$ and $F(\cdot)$ respectively such that

$$\mathscr{P}[\widetilde{Y}^{(n)}(t) \to \widetilde{F}(t) \text{ as } n \to \infty \forall t] = 1.$$

In view of its importance, we state the result as a theorem.

Theorem 8. *There exist on a common probability space versions $\widetilde{Y}^{(n)}(t)$ and $\widetilde{F}(t)$ of the processes $Y^{(n)}(t)$ and $F(t)$ such that*

$$\mathscr{P}[\widetilde{Y}^{(n)}(t) \to \widetilde{F}(t) \textit{ as } n \to \infty] = 1.$$

Throughout the rest of the paper, we shall (by an abuse of notations) assume that $Y^{(n)}(\cdot)$ and $F(\cdot)$ are such versions so that where necessary we may take sample path limits. This also allows us to assert the following "continuous mapping principle," namely, that for any continuous functional f of the paths (see Ref.[18] for applicable definitions of the topology),

$$f(Y^{(n)}, J) \Rightarrow f(F, J), \quad \text{as } n \to \infty,$$

where $\Rightarrow$ denotes convergence in the sense of Eq. (13); we refer to Ref.[18] for details.

6. TRANSIENT ANALYSIS OF $\mathscr{Y}^{(n)}$

In this section, we present the busy-period and idle-period analysis of the process $(Y^{(n)}, J)$. The limit results as $n \to \infty$ will give, by the continuous mapping principle, the corresponding results for (F, J). For the analysis, we need certain results on the nth coupled queue process $(\mathscr{Q}^{(n)}, J)$. A key ingredient of the analysis is the matrix transform $\widehat{R}_n(s)$ which was introduced by Ramaswami[11] as a generalization of Neuts' R-matrix. First, we recall its definition and characterizing equations.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

6.1. The Kernel $\widehat{R}_n(\cdot)$

Recall that $\{(Q_k^n, J_k^n, t_k^n) : k \geq 0\}$ is the embedded MRP in the nth queue $\mathcal{Q}^{(n)}$ arising from a spatial uniformization of the phase process with respect to λ. We showed in Sec. 4 that this MRP has exponential sojourn times depending on phases, and that its embedded sequence $\{(Q_k^n, J_k^n) : k \geq 0\}$ is a discrete time QBD process defined by the matrices B_{ni}, $i = 0, 1, 2$ (see Eq. (11)). Furthermore, the mean sojourn time in state (n, j) is given by $1/(n\lambda c_j)$ if $j \in S_1 \cup S_3$, and by $1/(2n\lambda c_j)$ if $j \in S_2$. Defining the matrix

$$\Lambda = \text{diag}(\lambda I_{|S_1|}, 2\lambda I_{|S_2|}, \lambda I_{|S_3|}),$$

we can therefore write for $Re(s) \geq 0$, the LST of the kernel of the MRP under consideration as being given by

$$\begin{pmatrix} \widehat{M}_n(s) & \widehat{B}_{n0}(s) & 0 & 0 & \cdots \\ \widehat{B}_{n2}(s) & \widehat{B}_{n1}(s) & \widehat{B}_{n0}(s) & 0 & \cdots \\ 0 & \widehat{B}_{n2}(s) & \widehat{B}_{n1}(s) & \widehat{B}_{n0}(s) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \tag{14}$$

where

$$\widehat{B}_{ni}(s) = n\Lambda C[sI + n\Lambda C]^{-1} B_{ni}, \quad i = 0, 1, 2,$$

and $B_{ni} = \widehat{B}_{ni}(0)$ are given by (11). The matrix $\widehat{M}_n(s) = \widehat{B}_{n1}(s) + \widehat{B}_{n2}(s)$.

Following Ref.[11], for $Re(s) > 0$, let the matrix $\widehat{R}_n(s)$ be the solution in the class of matrices of spectral radius at most one of the matrix quadratic equation

$$\widehat{R}_n(s) = \widehat{B}_{n0}(s) + \widehat{R}_n(s)\widehat{B}_{n1}(s) + \widehat{R}_n^2(s)\widehat{B}_{n2}(s), \tag{15}$$

where by an abuse of notations, we let $\widehat{R}_n^k(s) = [\widehat{R}_n(s)]^k$.

It is well-known that in the set of LSTs, such a kernel $\widehat{R}_n(\cdot)$ solving (15) exists uniquely, has spectral radius strictly less than unity for $Re(s) > 0$, and furthermore that the matrix $\widehat{R}_n(0+)$ is indeed Neuts' R-matrix for the embedded Markov chain which is a QBD. It has also been shown in Ref.[11] that for all $k \geq 0$, the kernel $\widehat{R}_n^k(\cdot)$ is the LST of the k-fold convolution $R_n^k(\cdot)$ of a matrix distribution function $R_n(\cdot)$ on $[0, \infty)$. Furthermore, for $k \geq 0$, $i, j \in S$, we can interpret $[R_n^k(t)]_{ij}$, the (i, j)th element of the convolution $R_n^k(t)$, as the expected number of visits to the state $(m + k, j)$ in the time interval $\min(t, \tau(m))$, where $\tau(m)$ is the return time to "level" m in the embedded MRP, given that the MRP starts in the state (m, i). We can interpret $[R_n^k(dt)]_{ij}$ as the elementary probability that a visit to (k, j) is made in the interval $(t, t + dt)$ and that no visits are made by the MRP to states of the form $(0, j)$ in $(0, t]$, given that it starts in state $(0, i)$.

We note that due to the structure of $\widehat{B}_{n0}$ (zero rows), the matrices $\widehat{R}_n(s)$ have the structure

$$\widehat{R}_n(s) = \begin{pmatrix} \widehat{R}_{n,11}(s) & \widehat{R}_{n,12}(s) & \widehat{R}_{n,13}(s) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{16}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The zero rows of $\widehat{R}_n(s)$ are inherited by $\widehat{R}_n^v(s)$ for $v \geq 1$, and from the above equation, we also have the following equation for the latter:

$$\widehat{R}_n^v(s) = \begin{pmatrix} [\widehat{R}_{n,11}(s)]^v & [\widehat{R}_{n,11}(s)]^{v-1}\widehat{R}_{n,12}(s) & [\widehat{R}_{n,11}(s)]^{v-1}\widehat{R}_{n,13}(s) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{17}$$

In what follows, we shall again abuse the notations slightly and write $\widehat{R}_{n11}^v(s)$ in place of $[\widehat{R}_{n,11}(s)]^v$.

The computation of $\widehat{R}(s)$ is efficiently accomplished using the L-R algorithm of Ref.[8] (with the matrices $\widehat{B}_{ni}(s)$ in place of A_i there), and it is trivial to note from the probabilistic interpretations of that algorithm that one obtains quadratic convergence; this is so since exactly as in the case of R, the termination of the algorithm in its kth step corresponds to ignoring paths that go beyond level (queue length) 2^k.

6.2. Transient Analysis of $\mathscr{Y}^{(n)}$ in a Busy Period

Assume that at time 0, the fluid process and the queue start empty and that the phase is i for some $i \in S_1$. Then note that $Y^{(n)}(0) = 0$. We shall call time 0 the start of a busy period for the fluid flow as well as for the process $\mathscr{Y}^{(n)}$. In what follows, to save space we sometimes write $Y_u^{(n)}$ in place of $Y^{(n)}(u)$. Define $[\mathbf{V}_n]_{i,j}(t, x)$ such that

$$[\mathbf{V}_n]_{i,j}(t, x) = \mathscr{P}_{0i}[Y_t^{(n)} \leq x, J(t) = j, Y_u^{(n)} > 0 \quad \text{for all } 0 < u \leq t].$$

Then, for $x > 0$, $[\mathbf{v}_n]_{i,j}(t, x) = \frac{\partial}{\partial x}[\mathbf{V}_n]_{i,j}(t, x)$ is the joint density of $Y^{(n)}(t)$ and the phase at time t. In the following, we let $\mathbf{V}_n(t, x)$ and $\mathbf{v}_n(t, x)$ denote respectively the matrices with elements $[\mathbf{V}_n]_{i,j}(t, x), i \in S_1, j \in S$ and $[\mathbf{v}_n]_{i,j}(t, x), i \in S_1, j \in S$. Also, we represent each matrix in partitioned form in accordance with the partition of the state space and denote their limits w.r.t n by $\mathbf{V}_\infty$ and $\mathbf{v}_\infty$; the existence of the limits is a consequence of the continuous mapping principle in Sec. 5. Thus,

$$\begin{aligned}
\mathbf{V}_n(t, x) &= (\mathbf{V}_{n,1}(t, x), \mathbf{V}_{n,2}(t, x), \mathbf{V}_{n,3}(t, x)),\\
\mathbf{V}_\infty(t, x) &= (\mathbf{V}_{\infty,1}(t, x), \mathbf{V}_{\infty,2}(t, x), \mathbf{V}_{\infty,3}(t, x)),\\
\mathbf{v}_n(t, x) &= (\mathbf{v}_{n,1}(t, x), \mathbf{v}_{n,2}(t, x), \mathbf{v}_{n,3}(t, x))\\
\mathbf{v}_\infty(t, x) &= (\mathbf{v}_{\infty,1}(t, x), \mathbf{v}_{\infty,2}(t, x), \mathbf{v}_{\infty,3}(t, x)).
\end{aligned}$$

Now, we can prove the following theorem:

Theorem 9. *For complex variables α and s with $Re(s) > 0$, $Re(\alpha) \geq 0$, let*

$$\widehat{\mathbf{V}}_{n,l}(t, \alpha) = \int_{[0,\infty)} e^{-\alpha x} V_{n,l}(t, dx), \quad l = 1, 2, 3,$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

be the LST of $\mathbf{V}_{n,l}(t, x)$ *with respect to x, and let*

$$\widetilde{\widehat{\mathbf{V}}}_{n,l}(s, \alpha) = \int_0^\infty e^{-st}\widehat{V}_{nl}(t, \alpha)dt$$

denote the Laplace transform with respect to t of $\widehat{V}_{nl}(t, \alpha)$*. Then*

$$\widetilde{\widehat{\mathbf{V}}}_{n,1}(s, \alpha) = (\alpha + n\lambda)\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}[sI + (\alpha + n\lambda)C_1]^{-1} \tag{18}$$

$$\widetilde{\widehat{\mathbf{V}}}_{n,2}(s, \alpha) = n\lambda\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}\widehat{R}_{n,12}(s)[sI + 2n\lambda C_2]^{-1} \tag{19}$$

$$\widetilde{\widehat{\mathbf{V}}}_{n,3}(s, \alpha) = \frac{n\lambda}{s + n\lambda}\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}\widehat{R}_{n,13}(s) \tag{20}$$

Proof. (a) For $i, j \in S_1$, we can see from the definition of $\mathscr{Y}^{(n)}$ that

$$\begin{aligned}&[\mathbf{V}_n]_{i,j}(t, x)\\ &\quad = \delta_{i,j}e^{-n\lambda c_j t}I(c_j t \le x) + \sum_{k=1}^{\infty}\int_0^t [R_n^k(du)]_{i,j}e^{-n\lambda c_j(t-u)}I(x - c_j(t-u) \ge 0)\\ &\qquad \times \int_0^{x-c_j(t-u)} \frac{(n\lambda)^k}{\Gamma_k}y^{k-1}e^{-n\lambda y}dy.\end{aligned}$$

In the above, the first term on the right hand side above corresponds to the case where no transition of the embedded MRP occurs in $(0, t]$. In the second term, the variable u in the integral is the last epoch of transition of the embedded MRP before time t; see the interpretation of $R_n^k(\cdot)$ given earlier. In computing the distribution, we have used the facts that: (a) $Y^{(n)}(t) = W^{(n)}(u) + c_j(t-u)$; and (b) given $Q^{(n)}(u) = k$, $W^{(n)}(u)$ is distributed as a sum of k *iid* $\exp(n\lambda)$ random variables.

Taking LST with respect to x, for $i, j \in S_1$, we can now write

$$\begin{aligned}[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha) &= \int_0^\infty e^{-\alpha x}[V_n]_{i,j}(t, dx)\\ &= \delta_{i,j}e^{-\alpha c_j t}e^{-n\lambda c_j t}\\ &\quad + \sum_{k=1}^{\infty}\int_0^t [R_n^k(du)]_{i,j}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k e^{-(\alpha+n\lambda)c_j(t-u)}.\end{aligned}$$

If we also take Laplace transformation of the above function on t w.r.t. s, $Re(s) > 0$, then the transform has the following form

$$\begin{aligned}[\widetilde{\widehat{\mathbf{V}}}_n]_{i,j}(s, \alpha) &= \int_0^\infty e^{-st}[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha)dt\\ &= \delta_{ij}(s + (\alpha + n\lambda)c_j)^{-1}\\ &\quad + \sum_{k=1}^{\infty}\left[\widehat{R}_n^k(s)\right]_{i,j}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k \frac{1}{s + (\alpha + n\lambda)c_j}.\end{aligned}$$

It follows that

$$\begin{aligned}
&\widetilde{\widetilde{\mathbf{V}}}_{n,1}(s, \alpha) \\
&\quad = (sI + (\alpha + n\lambda)C_1)^{-1} \\
&\qquad + n\lambda\hat{R}_{n,11}(s)\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}[sI + (\alpha + n\lambda)C_1]^{-1} \\
&\quad = (\alpha + n\lambda)\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}[sI + (\alpha + n\lambda)C_1]^{-1}.
\end{aligned}$$

(b) Similarly, for $i \in S_1$, $j \in S_2$, we get

$$\begin{aligned}
[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha) &= \int_0^\infty e^{-\alpha x}[V_n]_{i,j}(t, dx) \\
&= \sum_{k=1}^{\infty}\int_0^t [R_n^k(du)]_{i,j} e^{-2n\lambda c_j(t-u)}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k;
\end{aligned}$$

we have used the following facts: at the last epoch u of the MRP before time t, the phase visited should be j and some $k \geq 1$ customers should be in the system at $u+$ (since the busy period has not ended before t); the total amount of remaining work at time t is indeed the value of $Y_t^{(n)}$; due to the exponential distribution of the amount of work of each customer, the remaining amounts of work at t for customers have the same distribution $\exp(n\lambda)$ as at the time point u as one may easily verify; finally, given $j \in S_2$, $P(t_{k+1}^n - t_k^n > u) = \exp(-2n\lambda c_j u)$.

If we take Laplace transformation of the above function on t w.r.t. s, $Re(s) > 0$, then the transform has the following form

$$\begin{aligned}
[\widetilde{\widetilde{\mathbf{V}}}_n]_{i,j}(s, \alpha) &= \int_0^\infty e^{-st}[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha)dt \\
&= \sum_{k=1}^{\infty}\left[\widehat{R}_n^k(s)\right]_{i,j}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k \frac{1}{s + 2n\lambda c_j}.
\end{aligned}$$

It follows from (17) that

$$\begin{aligned}
\widetilde{\widetilde{\mathbf{V}}}_{n,2}(s, \alpha) &= \sum_{k=1}^{\infty}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k \widehat{R}_{n,11}^{k-1}(s)\widehat{R}_{n,12}(s)\,[sI + 2n\lambda C_2]^{-1} \\
&= n\lambda\left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1}\widehat{R}_{n,12}(s)\,[sI + 2n\lambda C_2]^{-1}.
\end{aligned}$$

Hence, we have (19).

(c) Finally, we can write for $i \in S_1$ and $j \in S_3$,

$$\begin{aligned}
[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha) &= \int_0^\infty e^{-\alpha x}[V_n]_{i,j}(t, dx) \\
&= \sum_{k=1}^{\infty}\int_0^t [R_n^k(du)]_{i,j} e^{-n\lambda(t-u)}\left(\frac{n\lambda}{\alpha + n\lambda}\right)^k.
\end{aligned}$$

If we take Laplace transformation of the above function on t w.r.t. s, $Re(s) > 0$, then the transform has the following form:

$$
\begin{aligned}
[\widetilde{\widetilde{\mathbf{V}}}_n]_{i,j}(s, \alpha) &= \int_0^\infty e^{-st}[\widehat{\mathbf{V}}_n]_{i,j}(t, \alpha)dt \\
&= \sum_{k=1}^{\infty} \left[\widehat{R}_n^k(s)\right]_{i,j} \left(\frac{n\lambda}{\alpha + n\lambda}\right)^k \frac{1}{s + n\lambda}.
\end{aligned}
$$

From this, it follows using (17) that

$$
\begin{aligned}
\widetilde{\widetilde{\mathbf{V}}}_{n,3}(s, \alpha) &= [s + n\lambda]^{-1} \sum_{k=1}^{\infty} \left(\frac{n\lambda}{\alpha + n\lambda}\right)^k \widehat{R}_{n,11}^{k-1}(s)\widehat{R}_{n,13}(s) \\
&= \frac{n\lambda}{s + n\lambda} \left[\alpha I - n\lambda(\widehat{R}_{n,11}(s) - I)\right]^{-1} \widehat{R}_{n,13}(s). \qquad \square
\end{aligned}
$$

6.3. Idle Period and Busy Cycle of $\mathscr{Y}^{(n)}$

Assume that a busy period of $\mathscr{Y}^{(n)}$ (i.e., a period when the process remains continuously positive) starts at time 0. Let β_n and ι_n denote respectively the duration of the busy period starting at time 0 and the duration of the idle period (the period where $Y^{(n)} = 0$) following that busy period. Let $\zeta_n = \beta_n + \iota_n$, and note that it represents the length of a busy cycle. We also note that a busy period can start only with phase in S_1, and it always ends with phase in S_2; similarly, an idle period can start only with phase in S_2 and must end with phase in S_1. Bearing these in mind, for $t \geq 0$, let us define the kernels $B_n(t)$, $U_n(t)$ and $W_n(t)$ such that

$$
\begin{aligned}
&[B_n]_{i,j}(t) = \mathscr{P}_{0i}\left[\beta_n \leq t, J(\beta_n) = j\right], \quad i \in S_1, j \in S_2, \\
&[U_n]_{j,k}(t) = \mathscr{P}_{0j}\left[\iota_n \leq t, J(\iota_n) = k\right], \quad j \in S_2, k \in S_1, \\
&[W_n]_{i,k}(t) = \mathscr{P}_{0i}\left[\zeta_n \leq t, J(\zeta_n) = k\right], \quad i, k \in S_1.
\end{aligned}
$$

and also denote their LSTs respectively by $\widehat{B}_n(s)$, $\widehat{U}_n(s)$ and $\widehat{W}_n(s)$.

We note that ι_n is a dwell time of the underlying phase process $\mathscr{J}$ in $S_2 \cup S_3$ which starts in a phase in S_2 and ends upon entering the set S_1. Thus, ι_n can be determined completely by using the underlying phase process, and it does not depend on n. From these facts, it follows that $\widehat{U}_n(s) = \widehat{U}(s)$, where

$$
\begin{aligned}
\widehat{U}(s) &= (sI - Q_{22})^{-1}[Q_{21} + Q_{23}(sI - Q_{33})^{-1}Q_{31}] \\
&\quad + (sI - Q_{22})^{-1}Q_{23}(sI - Q_{33})^{-1}Q_{32}\widehat{U}(s),
\end{aligned}
$$

and we can write

$$
\widehat{U}(s) = [I - \widehat{D}(s)]^{-1}(sI - Q_{22})^{-1}[Q_{21} + Q_{23}(sI - Q_{33})^{-1}Q_{31}], \tag{21}
$$

where

$$\widehat{D}(s) = (sI - Q_{22})^{-1}Q_{23}(sI - Q_{33})^{-1}Q_{32}. \tag{22}$$

Finally, we also have

$$\widehat{W}_n(s) = \widehat{B}_n(s)\widehat{U}(s).$$

We can also calculate $\widehat{U}(s)$ as the transform of an absorption time distribution in a Markov process. To that end, consider an absorbing Markov process defined by the generator

$$\begin{bmatrix} Q_{-1} & Q_{\cdot 1} \\ 0 & 0 \end{bmatrix},$$

where

$$Q_{-1} = \begin{pmatrix} Q_{22} & Q_{23} \\ Q_{32} & Q_{33} \end{pmatrix}, \quad Q_{\cdot 1} = \begin{pmatrix} Q_{21} \\ Q_{31} \end{pmatrix}.$$

If we denote the LSTs of the distributions of the absorption times given different initial states in $S_2 \cup S_3$ by

$$[sI - Q_{-1}]^{-1}Q_{\cdot 1} = \begin{pmatrix} \widehat{U}_{21}(s) \\ \widehat{U}_{31}(s) \end{pmatrix},$$

then $\widehat{U}(s) = \widehat{U}_{21}(s)$. This also shows that $U(\cdot)$ is a phase type distribution.

7. TRANSIENT ANALYSIS OF (*F*, *J*)

7.1. Preliminaries

For $Re(s) > 0$, we now define the following matrices which will play a key role in our analysis.

$$K_n^*(s) = (I - \widehat{R}_{n,11}(s))^{-1}\,[sI + n\lambda C_1]^{-1}\,, \tag{23}$$

$$\Psi_n(s) = \widehat{R}_{n,12}(s)n\lambda C_2[sI + 2n\lambda C_2]^{-1}, \tag{24}$$

$$\Theta_n(s) = \frac{n\lambda}{s + n\lambda}\widehat{R}_{n,13}(s). \tag{25}$$

We first give some probabilistic interpretations.

Theorem 10. *Let $[H_n(t)]_{ij}$ denote the probability that a busy period of $\mathcal{Y}^{(n)}$ ends at or before time t and in phase $j \in S_2$, given that it starts at time 0 in phase $i \in S_1$. The matrix $\Psi_n(s)$ is such that its (i, j)th element is the Laplace-Stieltjes transform of $[H_n(\cdot)]_{ij}$.*

Proof. The busy period ends at t in phase j iff at the last epoch u of the MRP before t, $Q^{(n)}(u) = 1$ and $J(u) = j$ and then the next epoch of the MRP occurs at t and marks a departure of that customer. That gives the density,

$$h_n(t) = \int_0^t R_{n,12}(du)\frac{1}{2}2n\lambda C_2 e^{-2n\lambda C_2(t-u)};$$

here, we have used the fact that the sojourn time in $(1, j)$ for $j \in S_2$ is distributed as $\exp(2n\lambda c_j)$ and that the epoch t is a departure epoch iff the epoch marking t comes from $\bigoplus_{r=1}^{n} \mathcal{M}_{rj}$ which has probability $1/2$. Taking the Laplace transform of the above expression, we get the required result. □

Corollary 1. *The matrix*

$$\Psi(s) = \lim_{n\to\infty} \Psi_n(s)$$

exists, and its (i, j)th element is the LST of the busy period duration and ending phase of the fluid process given that the busy period starts in phase $i \in S_1$.

Proof. This is an immediate consequence of the continuous mapping principle. □

Corollary 2. *Let $\widehat{G}_n(s)$ be a matrix such that its (i, j)th element is the LST of the duration of a busy period of the queue $\mathcal{Q}^{(n)}$ that starts in $(1, i)$ and ends in phase j. Then*

$$[\Psi_n(s)]_{ij} = [\widehat{B}_{n0}(s)\widehat{G}_n(s)]_{ij}, \quad \textit{for } i \in S_1, j \in S_2,$$

where $\widehat{B}_{n0}(s)$ is the matrix appearing in Eq. (14).

Proof. If the process $(Y^{(n)}, J)$ starts in $(0, i)$ with $i \in S_1$, then the queue is starting with an idle period. The first transition of the MRP then starts off a busy period of the queue in state $(1, r)$ for some $r \in S$, and the first transition is governed by $\widehat{B}_{n0}(s)$. Hence the result. □

Theorem 11. *The matrix $K_n^*(s)$ is a matrix of Laplace transforms such that its (i, j)th element is the transform*

$$\int_0^\infty e^{-st}\mathcal{P}_{0i}[J(t) = j \ \ \textit{and} \ \ Y^{(n)}(u) > 0 \ \ \textit{for all } 0 < u \le t]dt, \quad i, j \in S_1.$$

Therefore, the limit

$$K^*(s) = \lim_{n\to\infty} K_n^*(s)$$

exists, and its (i, j)th element is the Laplace transform

$$\int_0^\infty e^{-st}\mathcal{P}_{0i}[J(t) = j \ \ \textit{and} \ \ F(u) > 0 \ \ \textit{for all } 0 < u \le t]dt, \quad i, j \in S_1.$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Proof. The first assertion is immediate upon noting that $K_n^*(s) = \widetilde{\widehat{\mathbf{V}}}_{n,1}(s, 0)$; see Eq. (18). The second follows from the continuous mapping principle. □

Before we discuss the next theorem which gives a set of equations that are satisfied by the matrices $K^*(s)$, $\Psi(s)$ and $\Theta(s)$, we prove two lemmas.

Lemma 1. *For $Re(s) > 0$, the matrix $K^*(s)$ is invertible.*

Proof. Consider the matrix $\widetilde{\widetilde{V}}_{n,1}(s, 0)$ in Eq. (18). By the continuous mapping principle, it converges as $n \to \infty$ to the double transform $\widetilde{\widetilde{V}}_{\infty,1}(s, 0)$ for the fluid model. Now, looking at (18), since

$$n\lambda\{sI + n\lambda C_1\}^{-1} \to C_1^{-1} \text{ as } n \to \infty,$$

the matrix $[n\lambda(\widehat{R}_{n,11}(s) - I)]^{-1}$ must have a limit as $n \to \infty$. This also entails immediately, upon taking limits in the equation

$$\left[n\lambda\{\widehat{R}_{n,11}(s) - I\}\right]\left[n\lambda\{\widehat{R}_{n,11}(s) - I\}\right]^{-1} = I$$

defining the inverses, that the matrix $n\lambda(\widehat{R}_{n,11}(s) - I)$ also has a (nonsingular) limit as $n \to \infty$. Let

$$K(s) = \lim_{n\to\infty} n\lambda(\widehat{R}_{n,11}(s) - I), \text{ and } \widetilde{K}(s) = C_1 K(s) C_1^{-1}.$$

Letting $n \to \infty$ in (18), we can now write

$$\widetilde{\widetilde{V}}_{\infty,1}(s, \alpha) = C_1^{-1}[\alpha I - \widetilde{K}(s)]^{-1}.$$

Inverting the LST with respect to α, it is then easy to see that

$$\widetilde{V}_{\infty,1}(s, x) = C_1^{-1} \int_0^x e^{\widetilde{K}(s)u} du, \quad x > 0. \tag{26}$$

Letting $x \to \infty$ in the above, due to the interpretation of $K^*(s)$ in Theorem 11, we get

$$\widetilde{\widetilde{V}}_{\infty,1}(s, 0) = K^*(s) = -C_1^{-1}[\widetilde{K}(s)]^{-1}.$$

Thus, we have the existence of $[\widetilde{K}(s)]^{-1}$ as well as the relationship

$$K^*(s) = -C_1^{-1}[\widetilde{K}(s)]^{-1}, \tag{27}$$

and the proof of the lemma is also complete. □

Lemma 2. *The sequence $\Theta_n(s)$ has a limit as $n \to \infty$.*

Proof. By the continuous mapping principle, the left side of (20) converges to $\widetilde{\widetilde{\mathbf{V}}}_{\infty,3}(s, \alpha)$. Setting $\alpha = 0$ in (20), we see upon using (25) that

$$\widetilde{\widetilde{\mathbf{V}}}_{n,3}(s, 0) = [-n\lambda(\widehat{R}_{n,11}(s) - I)]^{-1}\Theta_n(s).$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

We have shown in the proof of Lemma 1 that the inverse on the right side of the above has a nonsingular limit as $n \to \infty$. Therefore, since the left side converges, there should exist a matrix

$$\Theta(s) = \lim_{n\to\infty} \Theta_n(s)$$

for all $Re(s) > 0$. □

We now define $\widetilde{\Psi}(s)$ and $\widetilde{\Theta}(s)$ by

$$\widetilde{\Psi}(s) = C_1 \Psi(s) C_2^{-1} \tag{28}$$

$$\widetilde{\Theta}(s) = C_1 \Theta(s), \tag{29}$$

and prove the following theorem.

Theorem 12. *The matrices* $\widetilde{K}(s)$, $\widetilde{\Psi}(s)$ *and* $\widetilde{\Theta}(s)$ *also satisfy the following equations*:

$$\widetilde{K}(s) = \left[(Q_{11} - sI) + \widetilde{\Psi}(s)Q_{21} + \widetilde{\Theta}(s)Q_{31}\right] C_1^{-1} \tag{30}$$

$$0 = Q_{12} + \widetilde{\Psi}(s)\,(Q_{22} - sI) + \widetilde{\Theta}(s)Q_{32} + \widetilde{K}(s)\widetilde{\Psi}(s)C_2 \tag{31}$$

$$0 = Q_{13} + \widetilde{\Psi}(s)Q_{23} + \widetilde{\Theta}(s)\,(Q_{33} - sI)\,. \tag{32}$$

Proof. Using (16), (11), (6) and (7), we can rewrite Eq. (15) as:

$$\begin{aligned}
n\lambda &\left(\widehat{R}_{n,11}(s) - I\right) \\
&= n\lambda[sI + n\lambda C_1]^{-1} Q_{11} - n\lambda s[sI + n\lambda C_1]^{-1} \\
&\quad + \frac{1}{2}\widehat{R}_{n,12}(s)(2n\lambda)[sI + 2n\lambda C_2]^{-1} Q_{21} + \frac{n\lambda}{s + n\lambda}\widehat{R}_{n,13}(s) Q_{31}
\end{aligned}$$

$$\begin{aligned}
0 &= n\lambda[sI + n\lambda C_1]^{-1} Q_{12} + \frac{1}{2}\widehat{R}_{n,12}(s)(2n\lambda)[sI + 2n\lambda C_2]^{-1}[Q_{22} - sI] \\
&\quad + \frac{n\lambda}{s + n\lambda}\widehat{R}_{n,13}(s) Q_{32} + n\lambda\left(\widehat{R}_{n,11}(s) - I\right) \\
&\quad \times \frac{1}{2}\widehat{R}_{n,12}(s)(2n\lambda C_2)[sI + 2n\lambda C_2]^{-1}
\end{aligned}$$

$$\begin{aligned}
0 &= n\lambda[sI + n\lambda C_1]^{-1} Q_{13} + \frac{1}{2}\widehat{R}_{n,12}(s)(2n\lambda)[sI + 2n\lambda C_2]^{-1} Q_{23} \\
&\quad + \frac{n\lambda}{s + n\lambda}\widehat{R}_{n,13}(s)[Q_{33} - sI].
\end{aligned}$$

The asserted equations follow by using (23)–(25) and (27) to simplify the above and then taking the limits as $n \to \infty$. □

Remark 1. With $s = 0+$ and $S_3 = \phi$, the equations in the above theorem reduce to the Eqs. (2.9) and (2.11) obtained by Ramaswami[12] for $\widetilde{K}(0+)$ and $\widetilde{\Psi}(0+)$ which were denoted by the symbols K and Ψ in that article. As noted there, the solutions to these equations for a fixed s are not unique, and further work remains in developing a direct algorithm based on them.

7.2. Final Results

We are now ready to state and prove the main result of this article which, among others, gives the Laplace transform (with respect to time t) of the distribution of the fluid level at time t, given that a busy period of the fluid process starts at time 0 from the empty level in a phase of S_1. In what follows, we denote by $\mathcal{P}_{(0,S_1)}(A)$ the vector whose elements are the conditional probabilities $\mathcal{P}_{0i}(A)$ which is the conditional probability of the event A given that $F(0) = 0$ and $J(0) = i \in S_1$.

Theorem 13.

(a) *For $x > 0$, let $\tilde{\mathbf{v}}_\infty(s, x)$ denote the Laplace transform of the density function $\mathbf{v}_\infty(t, x)$ introduced in Subsec. 6.2. Then*

$$\begin{aligned} \tilde{\mathbf{v}}_{\infty,1}(s, x) &= C_1^{-1} e^{\widetilde{K}(s)x}, \\ \tilde{\mathbf{v}}_{\infty,2}(s, x) &= C_1^{-1} e^{\widetilde{K}(s)x} \widetilde{\Psi}(s), \\ \tilde{\mathbf{v}}_{\infty,3}(s, x) &= C_1^{-1} e^{\widetilde{K}(s)x} \widetilde{\Theta}(s). \end{aligned} \tag{33}$$

(b) *Let $[\widehat{W}(s)]_{ij}$ denote the Laplace-Stieltjes transform of the distribution of the length of a busy cycle (time until the start of the second busy period) which ends in phase $j \in S_1$ given that a busy period of the fluid process starts at time 0 in phase $i \in S_1$. Then, the matrix of these elements,*

$$\widehat{W}(s) = \Psi(s)\widehat{U}(s) = C_1^{-1}\widetilde{\Psi}(s)C_2\widehat{U}(s), \tag{34}$$

where $\widehat{U}(s)$ is defined in Subsec. 6.3.

(c) *For $x > 0$,*

$$\begin{aligned} &\int_0^\infty e^{-st}\, \mathcal{P}_{(0,S_1)}[F(t) > x]\, dt \\ &\quad = (I - \widehat{W}(s))^{-1} C_1^{-1} (-\widetilde{K}(s))^{-1} e^{\widetilde{K}(s)x} [I + \widetilde{\Psi}(s) + \widetilde{\Theta}(s)]\mathbf{1}, \end{aligned} \tag{35}$$

where $\mathbf{1}$ is a column vector of 1's.

(d)

$$\begin{aligned} &\int_0^\infty e^{-st} \mathcal{P}_{(0,S_1)}[F(t) = 0]\, dt \\ &\quad = (I - \widehat{W}(s))^{-1} [\Psi(s) \vdots \mathbf{0}][sI - Q_{-1}]^{-1}\mathbf{1} \\ &\quad = (I - \widehat{W}(s))^{-1} [C_1^{-1}\widetilde{\Psi}(s)C_2 \vdots \mathbf{0}][sI - Q_{-1}]^{-1}\mathbf{1}, \end{aligned} \tag{36}$$

where

$$Q_{-1} = \begin{pmatrix} Q_{22} & Q_{23} \\ Q_{32} & Q_{33} \end{pmatrix}.$$

Proof. (a) Taking limits as $n \to \infty$ in Eqs. (18)–(20), one easily gets

$$\widetilde{\widetilde{V}}_{\infty,1}(s, \alpha) = C_1^{-1}\left[\alpha I - \widetilde{K}(s)\right]^{-1},$$

$$\widetilde{\widetilde{V}}_{\infty,2}(s, \alpha) = C_1^{-1}\left[\alpha I - \widetilde{K}(s)\right]^{-1}\widetilde{\Psi}(s),$$

$$\widetilde{\widetilde{V}}_{\infty,3}(s, \alpha) = C_1^{-1}\left[\alpha I - \widetilde{K}(s)\right]^{-1}\widetilde{\Theta}(s).$$

The asserted results follow by noting that $[\alpha I - \widetilde{K}(s)]^{-1}$ is indeed the Laplace transform of $\exp[\widetilde{K}(s)x]$.

(b) We note from the fact that the busy cycle consists of the first busy period (governed by Ψ) and the immediately following idle period (governed by $\widehat{U}$) that $\widehat{W}(s) = \Psi(s)\widehat{U}(s)$. The rest is a change of notation only.

(c) This part is obtained by noting that $F(t) > x$ happens iff t is in the kth busy period of the fluid process for some k, and then using the formulae for the busy cycle distribution and the transient distribution given in (a) within a busy period. We omit the details.

(d) The proof of this is similar to (c) and uses the fact that t should now be a point within some idle period of the fluid process. □

Let us denote $\widetilde{K}(0+)$, $\widetilde{\Psi}(0+)$, $\widetilde{\Theta}(0+)$ respectively by $\widetilde{K}$, $\widetilde{\Psi}$ and $\widetilde{\Theta}$. The transient results then yield the following theorem whose formula coincides with that obtained (using much simpler arguments) in Ref.,[3] Theorem 18, for the steady state joint distribution of the fluid and phase, whenever it exists.

Theorem 14. *Assume that the fluid process is stable. Let $\overline{F}(x, j)$ denote the steady state probability that the fluid level is greater than x and the phase is j, and let $\overline{F}(x)$ denote the vector with elements $\overline{F}(x, j)$, $j \in S$. Then*

$$\overline{F}(x) = \xi_1 e^{\widetilde{K}x}[I_{|S_1|} \vdots \widetilde{\Psi} \vdots \widetilde{\Theta}],$$

where ξ is the stationary probability vector of the phase process, and ξ_1 comprises of the steady state probabilities for the states in S_1.

Proof. Multiplying the Laplace transforms in (33) by s, we note that for $i \in S_1$, $j \in S$, the (i, j)th element of

$$s(I - \widehat{W}(s))^{-1}C_1^{-1}[-\widetilde{K}(s)]^{-1}e^{\widetilde{K}(s)x}[I \vdots \widetilde{\Psi}(s) \vdots \widetilde{\Theta}(s)],$$

is the LST of the complementary distribution function

$$\mathcal{P}_i[F(t) > x, \quad J(t) = j].$$

In particular, setting $x = 0$ and letting $s \to 0$ in the above, we note that the jth element of

$$\lim_{s\to 0} s(I - \widehat{W}(s))^{-1}C_1^{-1}[-\widetilde{K}(s)]^{-1}$$

equals

$$\lim_{t\to\infty} \mathcal{P}_i[J(t) = j] = \xi_j,$$

provided the Markov renewal process governed by $\widehat{W}(\cdot)$ is recurrent nonnull. (To evaluate the limit distribution via the Markov renewal process at the start of busy periods, we need the stability condition.) Hence the result. □

8. NUMERICAL RESULTS

For comparison with the results of Sericola,[14] we take some of the examples in that paper. Specifically, we consider a system with a set of m on-off Markov sources as defining the underlying phase process. The state space $S = \{0, 1, \dots, m\}$ is such that the state i denotes that i sources are busy. We also set $S_1 = \{1, \dots, m\}$ and $S_2 = \{0\}$. (In this case S_3 is empty, and the applicable formulae are obtained by dropping the terms corresponding to S_3). For each source, the duration of on and off periods are exponentially distributed with means 1 and $1/\gamma$, respectively. The $m+1$ dimensional infinitesimal generator Q of this Markov chain is given by

$$[Q]_{i,i+1} = (m-i)\gamma, \quad \text{for } 0 \le i \le m-1,$$
$$[Q]_{i,i-1} = i, \quad \text{for } 1 \le i \le m,$$

with the diagonal elements defined such that row sums of Q are zero.

The fluid rates are determined such that a source that is on inputs at rate 1, while the system drains the fluid at rate 0.8. Thus, $c_i = |i - 0.8|$. So, if we denote the traffic intensity of the system by ρ, then γ can be determined from ρ by the equation $\rho = (m\gamma)/[0.8(1+\gamma)]$.

To effect computations, we took various values of n in increments of 1000 until the successive matrices $\Psi_n(s)$ differed by at most 1e-12 in their elements. Each $\Psi_n(s)$ was itself computed using Corollary 2 from $\widehat{G}_n(s)$ for computing which we used the logarithmic reduction algorithm.[8] The value of $\Psi_n(s)$ at the termination of the iteration process was taken as the value of $\Psi(s)$, the limiting value of the sequence $\Psi_n(s)$. The other required matrices were all derived from this. From these were computed the transforms in (35) and (36), and these transforms were inverted using the Laplace transform inversion method of Abate and Whitt.[1] Table 1 contains the values of the complementary distribution of the amount of fluid in the system and the corresponding values reported in Ref.[14].

Note that our results are consistently larger than those of Sericola. Our numerical results show a match of up to 5 decimal places with Sericola's results in Ref.[14] most of the time. Recently, in the case $m = 2$, using a significantly smaller value for the error threshold ϵ in his algorithm, Sericola has verified that our results agree to about 10 decimal places; unfortunately, due to long run times with small ϵ values, we could not verify the other cases, but we do conjecture that our values are the more accurate ones.

Table 1. Comparison of $\mathcal{P}(F(100) > x)$ values with those of Sericola[14] at time $t = 100$.

	$m = 2, \rho = 0.75$		$m = 2, \rho = 1.25$		$m = 50, \rho = 5/6$	
x	Ours	Sericola	Ours	Sericola	Ours	Sericola
0	0.749803	0.749797	0.999942	0.999933	0.823479	0.823470
2	0.186047	0.186045	0.998769	0.998760	0.485221	0.485216
4	0.056276	0.056276	0.995548	0.995539	0.326178	0.326174
6	0.016943	0.016943	0.988034	0.988025	0.219216	0.219213
8	0.005064	0.005064	0.972873	0.972865	0.146241	0.146239
10	0.001498	0.001498	0.945740	0.945732	0.096634	0.096632
12	0.000437	0.000437	0.902078	0.902070	0.063190	0.063189
14	0.000125	0.000125	0.838395	0.838388	0.040871	0.040870
16	0.000035	0.000035	0.753781	0.753774	0.026139	0.026137
18	0.000010	0.000010	0.651010	0.651004	0.016526	0.016525
20	0.000003	0.000003	0.536633	0.536629	0.010328	0.010326
22	0.000001	0.000001	0.419800	0.419797	0.006379	0.006378
24	0.000000	0.000000	0.310136	0.310133	0.003894	0.003893
26	0.000000	0.000000	0.215474	0.215472	0.002350	0.002349
28	0.000000	0.000000	0.140293	0.140292	0.001401	0.001400
30	0.000000	0.000000	0.085342	0.085342	0.000826	0.000825
32	0.000000	0.000000	0.048378	0.048378	0.000481	0.000481
34	0.000000	0.000000	0.025496	0.025496	0.000277	0.000277
36	0.000000	0.000000	0.012467	0.012467	0.000158	0.000158
38	0.000000	0.000000	0.005645	0.005645	0.000089	0.000089
40	0.000000	0.000000	0.002362	0.002362	0.000050	0.000049

9. PROOF OF THEOREM 6

9.1. Preliminaries

First, we define some notations to be used in the proof and recall that all the processes under consideration are defined on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$.

Definition 1.

(a) *For $l \in S$, $\mathcal{P}_{0l}$ and $\mathcal{E}_{0l}$ denote respectively the conditional probability and the conditional expectation given the initial state $(0, l)$ of work and phase or fluid and phase.*

(b) *Recall that $\{a_0, a_1, \cdots\}$ mark the spatial uniformization epochs of the phase process with respect to rate λ (see Sec. 3.2). Let the index $\kappa(t)$ be such that $a_{\kappa(t)}$ is the last of these epochs to occur before time t. Similarly, let $\kappa_n(t)$ denote the index of the last epoch before time t in the set $\{t_0^n, t_1^n, \cdots\}$. Note that $\{0 = a_0 < a_1 < \cdots\} \subset \{0 = t_0^n < t_1^n < \cdots\}$ and $a_{\kappa(t)} \leq t^n_{\kappa_n(t)}$ for all $n \geq 1$.*

(c) *For a given time set G, let $A^{(n)}(G)$ denote the total number of arrivals to the queue $\mathcal{Q}^{(n)}$ which occurs in G. Thus, $A^{(n)}(G)$ denotes the number of epochs t_k^n in G such that t_k^n is an epoch of $\bigoplus_{i=1}^n \mathcal{N}_{i,j}$ in G and $J(t_k^n-) = j$ for some*

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

$j \in S_1$. Clearly, $A^{(n)}(G)$ is a Poisson distributed random variable with mean $n\lambda g$ where

$$g = \int_G \sum_{j \in S_1} c_j I(J(u) = j) du.$$

(d) *Given a time set G, define $\alpha_{W^{(n)}}(G)$ and $\alpha_F(G)$ to be the total amount of work that arrives to the queue $\mathcal{Q}^{(n)}$ and to the amount of increase to the fluid buffer that occurs in G. For notational convenience, we also define*

$$\alpha_{W^{(n)},i} = \alpha_{W^{(n)}}((a_{i-1}, a_i]),$$
$$\alpha_{F,i} = \alpha_F((a_{i-1}, a_i]).$$

(e) *Given a time set G, let $\delta_{W^{(n)}}(G)$ and $\delta_F(G)$ denote respectively the maximum amount of work and fluid which can be depleted in G assuming enough exists at the beginning. From the construction, we can see that*

$$\forall G, \delta_{W^{(n)}}(G) = \delta_F(G) = \int_G \sum_{j \in S_2} c_j I(J(u) = j) du.$$

(f) *Let $X_i^{(n)}$ denote the work of the ith customer in the nth coupled queue process $\mathcal{Q}^{(n)}$. We know that $X_i^{(n)}$ has an $\exp(n\lambda)$ distribution and is independent of all other quantities.*

Remark 2.

(a) *For disjoint intervals $[a, b)$ and $[c, d)$, $A^{(n)}([a, b))$ and $A^{(n)}([c, d))$ are conditionally independent given the phases $J(a)$ and $J(c)$.*
(b) *$A^{(n)}(G)$ and $X_i^{(n)}$ are independent for all i and G.*
(c) *For all $t \geq 0$, $Y^{(n)}(t)$ and $X_j^{(n)}$, $j \geq A^{(n)}((0, t]) + 1$, are independent. This is so since $Y^{(n)}(t)$ depends only on the history up to time t.*
(d) *Given $0 \leq t_1 < t_2$, $\alpha_{W^{(n)}}((t_1, t_2])$ can be represented by*

$$\alpha_{W^{(n)}}((t_1, t_2]) = \sum_{i=A^{(n)}((0,t_1])+1}^{A^{(n)}((0,t_1])+A^{(n)}((t_1,t_2])} X_i^{(n)}$$

where $X_i^{(n)}$ is defined in (f) of Definition 1. This shows that for nonoverlapping intervals G_1 and G_2, $\alpha_{W^{(n)}}(G_1)$ and $\alpha_{W^{(n)}}(G_2)$ are conditionally independent given the phases at the beginning of these intervals.

The following is an important intermediate result in strengthening the distributional convergence that follows from the continuous mapping principle to a convergence in probability result.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Lemma 3.

(a) *For all* $i \geq 1$,

$$\mathcal{P}\left[|\alpha_{W^{(n)},i} - \alpha_{F,i}| > \epsilon \mid (a_k, J(a_k)), k \geq 0\right] \leq \frac{4n\lambda c(a_i - a_{i-1}) + 2\lambda^2 c^2 (a_i - a_{i-1})^2 + 2(n\lambda + 1)}{n^2\lambda^2\epsilon^2}, \tag{37}$$

where $c = \sum_{j\in S} c_j$.

(b) *If we let* $G_t = (a_{\kappa(t)}, t]$, *then*

$$\mathcal{P}[|\alpha_{W^{(n)}}(G_t) - \alpha_F(G_t)| > \epsilon \mid (a_k, J(a_k)), k \geq 0] \leq \frac{4n\lambda c(t - a_{\kappa(t)}) + 2\lambda^2 c^2 (t - a_{\kappa(t)})^2}{n^2\lambda^2\epsilon^2}. \tag{38}$$

(c) *As* $n \to \infty$, $t - t^n_{\kappa_n(t)}$ *converges in probability to* 0 (*uniformly*) *in* t.

Proof. (a) Let $G_i = (a_{i-1}, a_i)$. Note that the underlying phase process $\mathcal{J}$ does not change phase within G_i. Let us assume that the phase it remains in is $j \in S$.

If $j \in S_2 \cup S_3$, then there is no arrival to the queue $\mathcal{Q}^{(n)}$ and no increase to the fluid level during $(\tau_{i-1}, \tau_i]$; this follows from the definition of the fluid process and its n-th coupled queue $\mathcal{Q}^{(n)}$. Therefore, in this case the result holds trivially because

$$\alpha_{W^{(n)},i} = \alpha_{F,i} = 0.$$

Consider now the case $j \in S_1$. By Definition 1(c), $A^{(n)}(G_i)$, the number of customers arriving to $\mathcal{Q}^{(n)}$ in the interval G_i, now equals the number of epochs in G_i of the Poisson process $\bigoplus_{i=2}^{n} \mathcal{N}_{i,j}$. This is distributed as Poisson with mean $(n-1)\lambda c_j(a_i - a_{i-1})$.

Since each customer in $\mathcal{Q}^{(n)}$ brings in $1/(n\lambda)$ amount of work on the average, we get:

$$\mathcal{E}[\alpha_{W^{(n)}}(G_i) \mid (a_k, J(a_k)), k \geq 0] = (n-1)\lambda c_j(a_i - a_{i-1})/(n\lambda).$$

Now, using the fact that the rate of increase of the fluid in G_i is c_j, we get:

$$\begin{aligned}
&\mathcal{E}\left[\left(\alpha_{W^{(n)}}(G_i) - \alpha_{F,i}\right)^2 \mid (a_k, J(a_k)), k \geq 0\right] \\
&\quad = \mathcal{E}\left[\alpha^2_{W^{(n)}}(G_i) \mid (a_k, J(a_k)), k \geq 0\right] \\
&\qquad - 2c_j(a_i - a_{i-1})\mathcal{E}[\alpha_{W^{(n)}}(G_i) \mid (a_k, J(a_k)), k \geq 0] + c_j^2(a_i - a_{i-1})^2 \\
&\quad = \left[2(n-1)\lambda c_j(a_i - a_{i-1}) + \lambda^2 c_j^2 (a_i - a_{i-1})^2\right]/(n\lambda)^2.
\end{aligned} \tag{39}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Since there is another arrival at the epoch a_i, $\alpha_{W^{(n)},i}$ can be expressed as the sum of $\alpha_{W^{(n)}}(G_i)$ and an independent random variable Z which is exponentially distributed with mean $1/(n\lambda)$. From this, we get using the Chebycheff inequality,

$$\begin{aligned}
&\mathscr{P}\left[|\alpha_{W^{(n)},i} - \alpha_{F,i}| > \epsilon \mid (a_k, J(a_k)), k \geq 0\right] \\
&\quad \leq \frac{2}{\epsilon^2}\left\{\mathscr{E}\left[\left(\alpha_{W^{(n)}}(G_i) - \alpha_{F,i}\right)^2 \mid (a_k, J(a_k)), k \geq 0\right] + E[Z^2]\right\} \\
&\quad = \frac{2}{\epsilon^2}\left\{\mathscr{E}\left[\left(\alpha_{W^{(n)}}(G_i) - \alpha_{F,i}\right)^2 \mid (a_k, J(a_k)), k \geq 0\right] + \frac{n\lambda + 1}{n^2\lambda^2}.\right\}
\end{aligned} \tag{40}$$

The proof can now be completed by this inequality and Eq. (39).

(b) A little reflection will show that Eq. (39) holds for $G_i = (a_{\kappa(t)}, t]$ also. So, the proof follows along the same lines as that of (a).

(c) Note that due to the nested nature of the epochs, $t - t^n_{\kappa_n(t)}$ monotonely decreases as n increases, and $a_{\kappa(t)} \leq t^n_{\kappa_n(t)} \leq t$. Define $t^* = t - a_{\kappa(t)}$ and assume that $J(a_{\kappa(t)}) = j \in S$.

Now, if $j \in S_1$, $t^n_{\kappa_n(t)}$ is determined by the Poisson process $\bigoplus_{i=2}^n \mathscr{N}_{i,j}$ and it follows from renewal theory (Chapter 9 of Ref.[5]) that

$$\begin{aligned}
&\mathscr{E}\left[(t - t^n_{\kappa_n(t)})^2 \mid a_{\kappa(t)}, J(a_{\kappa(t)}) = j \in S_1\right] \\
&\quad = t^{*2}e^{-(n-1)\lambda c_j t^*} + \int_0^{t^*}(n-1)\lambda c_j(t^* - u)^2 e^{-(n-1)\lambda c_j(t^*-u)}du \\
&\quad \leq 2/((n-1)\lambda c_j)^2 \leq 2/((n-1)\lambda \underline{c})^2,
\end{aligned}$$

where $\underline{c} = \min_{j\in S} c_j$. Note that the last term in the above does not depend on t and goes to 0 as n goes to infinity.

Now, if $j \in S_2$ or $j \in S_3$, then we can get similar inequalities from the fact that if $j \in S_2$, $t^n_{\kappa_n(t)}$ is determined by the Poisson process $\bigoplus_{i=2}^n \mathscr{N}_{i,j} \oplus \bigoplus_{i=1}^n \mathscr{M}_{i,j}$; and if $j \in S_3$, $t^n_{\kappa_n(t)}$ is determined by the Poisson process $\bigoplus_{i=2}^n \mathscr{N}_{i,j}$.

Since the bounds for these conditional expectations do not depend on t or the conditioning values, and go to zero as $n \to \infty$, this allows us to show that

$$\mathscr{E}\left[(t - t^n_{\kappa_n(t)})^2\right] \to 0, \; as \; n \to \infty,$$

and that the convergence is uniform in t. Hence (c) follows. □

9.2. Proof of Theorem 6

Now, we introduce the reflection map for the proof. In the classical treatment of fluid flow models (see e.g., Ref.[18]), it is customary to consider an unrestricted fluid flow $\widetilde{F}(s) \equiv \alpha_F((0, s]) - \delta_F((0, s])$ (i.e., one without a boundary at level 0) from which the fluid process $F(t)$ with boundary at level 0 is obtained by the formula

$$F(t) = \widetilde{F}(t) - \inf_{0\leq u\leq t} \widetilde{F}(u). \tag{41}$$

The map $\widetilde{F} \to F$ will be called the reflection map and is denoted by $\mathcal{R}$.[18] Similarly, if we let $\widetilde{W}^{(n)}(s) \equiv \alpha_{W^{(n)}}((0,s]) - \delta_{W^{(n)}}((0,s])$ denote the unrestricted work flow process, we can also see that $W^{(n)}(t) = \mathcal{R}(\widetilde{W}^{(n)})(t)$.

From the fact that $Y^{(n)}(a_i) = W^{(n)}(a_i)$ for all $i \geq 0$, we can see that if $J(t) \in S_1$,

$$F(t) = F(a_{\kappa(t)}) + \alpha_F((a_{\kappa(t)}, t]),$$
$$Y^{(n)}(t) = W^{(n)}(a_{\kappa(t)}) + \alpha_{W^{(n)}}((a_{\kappa(t)}, t]) + c_{J(t)}(t - t^n_{\kappa_n(t)}), \tag{42}$$

and if $J(t) \in S_2$,

$$F(t) = \max\{0, F(a_{\kappa(t)}) - c_{J(t)}(t - a_{\kappa(t)})\},$$
$$Y^{(n)}(t) = \max\{0, W^{(n)}(a_{\kappa(t)}) - c_{J(t)}(t - a_{\kappa(t)})\}. \tag{43}$$

From Lemma 3, $t - t^n_{\kappa_n(t)}$ converges in probability to 0. Also, $\alpha_{W^{(n)}}((a_{\kappa(t)}, t])$ converges to $\alpha_F((a_{\kappa(t)}, t])$ in probability as $n \to \infty$ because by Lemma 3(b),

$$\mathcal{P}\left[|\alpha_{W^{(n)}}((a_{\kappa(t)}, t]) - \alpha_F((a_{\kappa(t)}, t])| > \epsilon\right] \leq \frac{4n\lambda ct + 2\lambda^2 c^2 t^2}{n^2\lambda^2\epsilon^2}.$$

Thus, it is sufficient to consider convergence at the epoch $a_{\kappa(t)}$ for the proof.

Finally, if $J(t) \in S_3$, then

$$F(t) = F(a_{\kappa(t)}) \quad \text{and} \quad Y^{(n)}(t) = W^{(n)}(a_{\kappa(t)}),$$

and once again, we need to only consider the epoch $a_{\kappa(t)}$. This we do next.

From the reflection map, we know that $F(a_{\kappa(t)}) = \mathcal{R}(\widetilde{F})(a_{\kappa(t)})$ and $W^{(n)}(a_{\kappa(t)}) = \mathcal{R}(\widetilde{W}^{(n)})(a_{\kappa(t)})$. Moreover, since depletion of work and fluid can occur only when the underlying phase process belongs to S_2, we can also see that

$$\inf_{0 \leq u \leq a_{\kappa(t)}} \widetilde{F}(u) = \inf_{0 \leq i \leq \kappa(t)} \widetilde{F}(a_i), \quad \text{and} \tag{44}$$

$$\inf_{0 \leq u \leq a_{\kappa(t)}} \widetilde{W}^{(n)}(u) = \inf_{0 \leq i \leq \kappa(t)} \widetilde{W}^{(n)}(a_i). \tag{45}$$

From the Eqs. (44) and (45) and the fact that $\delta_F((0,a_i]) = \delta_{W^{(n)}}((0,a_i])$ for all $i \geq 1$, it is enough to consider convergence of $\alpha_{W^{(n)}}((0,a_i])$ to $\alpha_F((0,a_i])$ for $i = 1, \ldots, \kappa(t)$ to show the convergence at $a_{\kappa(t)}$. From Lemma 3, we can show that

$$\begin{aligned}
&\mathcal{P}_{0l}\left[\max_{1 \leq i \leq \kappa(t)} |\alpha_{W^{(n)}}((0,\tau_i]) - \alpha_F((0,\tau_i])| > \epsilon \mid (a_k, J(a_k)), k \geq 1\right] \\
&\leq \mathcal{P}_{0l}\left[\bigcup_{i=1}^{\kappa(t)} \{|\alpha_{W^{(n)},i} - \alpha_{F,i}| > \epsilon/\kappa(t)\} \mid (a_k, J(a_k)), k \geq 1\right] \\
&\leq \sum_{i=1}^{\kappa(t)} \mathcal{P}_{0l}\left[|\alpha_{W^{(n)},i} - \alpha_{F,i}| > \epsilon/\kappa(t) \mid (a_k, J(a_k)), k \geq 1\right] \\
&\leq \frac{\kappa^2(t)}{\epsilon^2} \sum_{i=1}^{\kappa(t)} \frac{4n\lambda c(a_i - a_{i-1}) + 2\lambda^2 c^2 (a_i - a_{i-1})^2 + 2(n\lambda + 1)}{n^2\lambda^2} \\
&\leq \frac{\kappa^2(t)}{\epsilon^2} \times \frac{4n\lambda ct + 2\lambda^2 c^2 t^2 + 2\kappa(t)(n\lambda + 1)}{n^2\lambda^2}.
\end{aligned}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

If we let $\tau(t)$ denote the total number of epochs before time t of the Poisson process $\bigoplus_{j\in S} \mathcal{N}_{1,j}$, then $0 \le \kappa(t) \le \tau(t)$ and it follows that

$$\mathcal{P}_{0l}\left[\max_{1\le i\le\kappa(t)} |\alpha_{W^{(n)}}((0,\tau_i]) - \alpha_F((0,\tau_i])| > \epsilon\right]$$
$$\le \frac{(4n\lambda ct + 2\lambda^2c^2t^2)\mathcal{E}[\tau^2(t)] + 2\mathcal{E}[\tau^3(t)](n\lambda+1)}{n^2\lambda^2\epsilon^2}.$$

Since $\tau(t)$ is a Poisson distributed random variable with the mean $c\lambda t$, $\mathcal{E}[\tau^2(t)]$ and $\mathcal{E}[\tau^3(t)]$ are finite and do not depend on n. Therefore, the proof of Theorem 6 is complete. □

10. CONCLUDING REMARKS

We have characterized the transient distribution of stochastic fluid flows modulated by finite state CTMCs in a computable form. The characterization is in terms of three kernels obtained as limits of certain kernels arising in a sequence of matrix-geometric queues that approximate the fluid model. We demonstrated computational tractability by actually carrying out the approximation. However, at the time of preparing the final version of this paper, we have developed some direct algorithms for computing these kernels directly from Eqs. (30)–(32) that are satisfied by them. That will obviate the need to compute results for the constituent approximating queues. This new development will be discussed elsewhere as it requires several results not developed here. Among the uses of the results here are algorithms for the finite buffer case presented in Ref.[4].

ACKNOWLEDGMENTS

Soohan Ahn was supported partially by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University, Korea. The authors thank Ward Whitt for many helpful discussions and the referees for comments which improved the presentation. We also thank Benny VanHoudt whose reading of an earlier version resulted in many significant improvements.

REFERENCES

1. Abate, J.; Whitt, W. The Fourier-series method for inverting transforms of probability distributions. Queue. Sys. **1992**, *10*, 5–88.
2. Adan, I.J.B.F.; Resing, J.A.C. A two-level traffic shaper for an on-off source. Performance Evaluation **2000**, *42*, 279–298.
3. Ahn, S.; Ramaswami, V. Fluid flow models and queues - A connection by stochastic coupling. Stochastic Models **2003**, *19*, 325–348.
4. Ahn, S.; Jeon, J.; Ramaswami, V. Steady state analysis of finite fluid flow models using finite QBDs – in review.

5. Çinlar, E. *Introduction to Stochastic Processes*; Prentice-Hall, 1975.
6. Kobayashi, H.; Ren, Q. A mathematical theory for transient analysis of communication networks. IEICE Trans. Commun. **1992**, *E-75B* (12), 1266–1276.
7. Latouche, G.; Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Modeling*; SIAM and ASA: Philadelphia, 1999.
8. Latouche, G.; Ramaswami, V. A logarithmic reduction algorithm for Quasi-Birth-and-Death Processes. J. Appl. Prob. **1993**, *30*, 650–674.
9. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, 1981.
10. Nielsen, B.F.; Ramaswami, V. A computational framework for quasi birth and death processes with a continuous phase variable. In *Teletraffic Contributions for the Information Age*, Proc. of the 15th International Teletraffic Congress, Ramaswami, V., Wirth, P.E., Eds.; Elsevier, 1997, 477–486.
11. Ramaswami, V. The busy period of queues which have a matrix-geometric steady state probability vector. Opsearch **1982**, *19*, 238–261.
12. Ramaswami, V. Matrix analytic methods for stochastic fluid flows. In *Teletraffic Engineering in a Competitive World*, Proc. of the 16th International Teletraffic Congress, Smith, D., Key, P., Eds.; Elsevier, 1999, 1019–1030.
13. Ramaswami, V. Algorithmic analysis of stochastic models - the changing face of mathematics. In *Mathematics for the New Millennium*, Ramanujan Memorial Lecture, Anna University, Chennai, India, 2000; Vijayakumar, A. et.al. Eds.; 2000.
14. Sericola, B. Transient analysis of stochastic fluid models. Performance Evaluation **1998**, *32*, 245–263.
15. Skorohod, A.V. Limit theorems for stochastic processes, Theor. Probabil. Appl. **1956**, *1*, 261–290. (Republished also in *Skorohod's Ideas in Probability Theory*, V. Korolyuk, N. Portenko and H. Syta (Eds.), Institute of Mathematics of the National Academy of Sciences of the Ukraine, Kyiv, Ukraine, pp. 23–52.)
16. Tanaka, T.; Hashida, O.; Takahashi, Y. Transient analysis of fluid model for ATM statistical multiplexer. Performance Evaluation **1995**, *23*, 145–162.
17. Tweedie, R.L. Operator-geometric stationary distribution for Markov chains, with applications to queueing models. Adv. Appl. Prob. **1982**, *14*, 368–391.
18. Whitt, W. *Stochastic Process Limits - An Introduction to Stochastic-Process Limits and their Application to Queues*; Springer, 2002.

Received February 27, 2003
Accepted July 4, 2003