Steady State Analysis of Finite Fluid Flow Models using Finite QBDs

Soohan Ahn

Department of Statistics, The University of Seoul

Jeonnong-dong, 90 Dong daeman-gu

Seoul 130-743, South Korea

J. Jeon

Department of Statistics, Seoul National University

Seoul 151-742, South Korea

V. Ramaswami

AT&T Labs-Research, 180 Park Avenue

Florham Park, NJ 07932, USA

Abstract

The Markov modulated fluid model with finite buffer of size β is analyzed using a stochastic discretization yielding a sequence of finite waiting room queueing models with iid amounts of work distributed as $exp(n\lambda)$. The n-th approximating queue's system size is bounded at a value q_n such that the corresponding expected amount of work $q_n/(n\lambda) \to \beta$ as $n \to \infty$. We demonstrate that as $n \to \infty$, we obtain the exact performance results for the finite buffer fluid model from the processes of work in the system for these queues. The necessary (strong) limit theorems are proven for both transient and steady state results. Algorithms for steady state results are developed fully and illustrated with numerical examples.

1 Introduction

This paper deals with a finite buffer, stochastic fluid flow modulated by a finite state, continuous time, Markov process which modulates the rates of (linear) change of the flow at any instant. For Markov modulated fluid flows, the infinite buffer case (see [2] and references therein) has been analyzed by many authors using a variety of methods. Just as in the infinite buffer case, for the finite buffer case also, one can – by arguments similar to those in [3] – derive a set of partial differential equations for the time-dependent state probabilities of the model by considering time points t and $t + \Delta t$ and by letting $\Delta t \to 0$. From these, by letting $t \to \infty$, it is easy to obtain a finite system of linear differential equations for the steady state probabilities of the model (see equation (62) for that equation). While the differential equations are the same in both cases, in the infinite buffer case, we get an initial value problem, while in the finite buffer case we get a two-point boundary value problem. The boundary conditions for the finite buffer case correspond to the situations when the buffer is empty and when the buffer is full. From the general theory of differential equations [5], it is well-known that if a set of linear differential equations has a solution satisfying the boundary conditions, then that solution is also unique. However, the finite buffer case also involves an additional challenge in the determination of the boundary conditions which are not easy to derive. Even if one could do that, as in the infinite buffer case, approaches based on a numerical solution of differential equations directly or using spectral methods is fraught with problems of accuracy and efficiency. Also, for the specific set of models at hand, certain modes of their general solution are unstable, and that causes considerable technical difficulties; see [11] for a discussion.

In [11], V. Ramaswami developed a matrix-geometric steady state analysis of infinite buffer fluid flow models modulated by a finite state, continuous time Markov chain, and demonstrated that it can be carried out via a discrete-time quasi birth and death process (QBD) for which efficient algorithms are well developed. These avoid the difficulties in approaches based on differential equations. Recently, Ahn & Ramaswami [1], [2] demonstrated that approach to be rooted in a stochastic coupling of the fluid flow to suitably defined Markov modulated queues. In [2], we also noted that the matrix-geometric approach allows us to "jump" to an embedded epoch prior to t and avoids having to build the story by discretization over time as done, for instance, in the numerical solution of differential equations. Furthermore, although those papers deal exclusively with finite state Markovian environments, we note that the appeal of their probabilistic approach rests in the possibility they open for generalizations using operator geometric methods [14], [8], [10] to fluid flows modulated by general state space Markov chains. Consideration of such general flows is, in turn, motivated by a desire to incorporate effects like heavy tails and self-similarity in the model that cannot be done using a finite state space chain for modulation.

This paper is an extension of the approaches in [1], [2] to the analysis of a Markov modulated fluid flow with a finite buffer of size $\beta > 0$, which is the continuous analog of the queue with finite waiting room. During those intervals when the buffer is full, incoming flow in excess of the drain rate is assumed to be lost permanently. A recent piece of work [13] by Soares & Latouche, of which we have become aware recently, obtains the steady state behavior of the finite buffer fluid flow model using level crossing arguments of the type introduced in Ramaswami [11]. While, modulo notations, this work yields the same results for the steady state distributions, through our approach based on stochastic discretization, we are able to also set the stage for the transient analysis.

In [2], we considered coupled queues defined on a common probability space through a sequence of nested "spatial uniformizations." A spatial uniformization amounts to a discretization of the time axis through a Markovian point process such that in the inter-event intervals of that process, the potential increment to the (unrestricted) flow process are identically distributed as exponential random variables. That allowed us to measure the fluid level approximately in terms of the number of "exponentially distributed" chunks, and thereby obtain a queueing model for consideration that is also represented by a quasi birth and death process (QBD). By letting the parameter of uniformization to give progressively finer and finer discretizations, the fluid process was obtained as the (strong) stochastic process limit of the work in the queues thus generated. The matrix-geometric analysis of the QBDs gave, through a limit process, a complete characterization of the transient and steady state characteristics of the fluid flow.

Unlike in [2], here we assume, however, that each associated queue has a finite bound on the number of customers that can be present simultaneously. These queues and their buffer sizes (in terms of the number of "customers") are to be constructed on a common probability space such that the fluid flow with finite buffer of size β can be realized as the pathwise stochastic process limit of the work processes of the associated queues. By analyzing the associated queues using matrix-geometric methods [8], [6], the results for the finite buffer fluid model are then obtained through a limit process.

Our approach is motivated by Figure 1 and Figure 2 showing the results of a pair of simulation experiments for two finite buffer fluid models, of which the first has "traffic intensity" less than unity and the second a traffic intensity larger than unity. (For specific details of the parameters etc., refer to Section 3.) Stochastic discretization of the models is governed by a parameter n, and successive discretizations get finer and finer as $n \to \infty$. In each figure, along with the fluid level and the work in the queue which were simulated for different values of n, we have also presented the graphs of the difference between these. Note that in both cases, as $n \to \infty$, the processes of work seem to converge to a limit which is the fluid flow. Indeed, one can prove this formally as a stochastic process limit theorem in the sense of [15]. The formal proof of that limit theorem is arduous, but proceeds along the same lines as in [2], and will therefore be only sketched here since readers familiar with [2] should be able to complete the details. We, however, wish to emphasize that certain aspects of the construction of the queues such as their having *iid* work amounts and being modulated by a common phase process are critical to our approach.

Our limit theorems in Section 4 concern both the transient and steady state distributions of the finite buffer fluid flow. However, to keep the length of this paper within reasonable bounds, in the sections following that we shall limit ourselves only to steady state results which are simpler due to reasons to be stated later.

2 The Model and Spatial Uniformization

We begin with a specification of the fluid model to be considered and a review of spatial uniformization used in Ahn & Ramaswami [2].

We assume as given an irreducible, continuous time Markov chain (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 $c_i > 0$ as long as it is below the given buffer size β; during sojourn of the CTMC in state $j \in S_2$, the fluid level decreases at rate $c_j > 0$ if the buffer is nonempty; and during sojourn of the CTMC in S_3, the fluid level remains constant. Numerous special cases of this canonical model have been considered in the literature, although for the finite buffer case no exact results are known.

For non-triviality, we assume throughout that the sets S_1 and S_2 are nonempty. Also, for any set A, we let $|A|$ denote the number of elements in A. Also, 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

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

where we set $c_i = 1$ for all $i \in S_3$, and let $C = diag(C_1, C_2, C_3)$. 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

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

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

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

where

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

The associated continuous time process $\mathcal{X} = \{\chi(t) : \ t \geq 0\}$ specified such that it takes the value χ_n in the interval $t_n \leq t < t_{n+1}$ is the SMP of interest to us. Note that the MRP under consideration may make a self-transition from a state to itself in one step. Nevertheless, it can be shown that the continuous time process $\mathcal{X}$ is indeed a realization of the phase process, namely the Markov chain with infinitesimal generator Q. (This is accomplished by verifying that the sojourn time in each state i, i.e., the time before it leaves i to a *different* state, is exponentially distributed with parameter $-Q_{ii}$ and the probability of transiting to $j \neq i$ from i is $-Q_{ij}/Q_{ii}$; see [2], Theorem 1.) The consideration of χ as a SMP or as a CTMC depends on whether self-transitions to states that occur in the underlying MRP are considered or not.

Consider now an unrestricted fluid flow, i.e., one in which the fluid level ranges over $(-\infty, \infty)$ and is modulated by the SMP $\mathcal{X}$ such that the net input rate to the fluid buffer is c_i when $\chi(t) = i \in S_1$, is $-c_i$ when $\chi(t) = i \in S_2$, and is 0 when $\chi(t) \in S_3$. Assume that the SMP starts in $i \in S_1$, and consider its first transition interval. This is distributed as $exp(\theta c_i)$, and during this, fluid accumulates at rate c_i per unit time. Thus the total additional fluid accumulation in that interval is distributed as $exp(\theta)$. Similarly, for $i \in S_2$, the total depletion of the fluid level in the interval of sojourn of the SMP in i would be distributed as $exp(\theta)$. This underlies our reason for calling our construction by the name "spatial uniformization." Unlike standard uniformization which uniformizes the inter-event time distributions (i.e., distributions of time intervals), the procedure adopted uniformizes the distributions over space (i.e., the fluid level). It can be seen that for the finite queues we shall construct later, spatial uniformization plays a crucial role in simplifying their structure by allowing the queues to have iid amounts of work for successive customers.

We define

$$n_0 = \inf\{m \ : \ m \in \mathcal{N}^+, \ m/\beta \geq \max_{i \in S} \{-[C^{-1}Q]_{ii}\}\}, \text{ and } \lambda = n_0/\beta, \quad (5)$$

where $\mathcal{N}^+$ is the set of positive integers. We will consider in this paper, a sequence of spatial uniformizations of the fluid flow given by

$$\theta_n = n \, \frac{n_0}{\beta} = n\lambda, \ n = 1, 2, \cdots , \quad (6)$$

and a corresponding set of queues with finite buffer sizes $q_n - 1$, where

$$q_n = n \, n_0, \quad n = 1, 2, \cdots . \quad (7)$$

The queue corresponding to θ_n will be such that each customer brings in an exponentially distributed amount of work with mean $1/(n\lambda)$. See Section 3 for details of the construction of the queues. Note that defined thus, the n-th queue is such that when it is saturated it has q_n customers in the system and their total expected amount of work is $q_n/(n\lambda) = n_0/\lambda = \beta$. Our main thrust will be to show that the work processes of these queues converge as $n \to \infty$ to the fluid flow with finite buffer size β in both an a.s., sense and in the sense of stochastic process convergence as defined in [15]. This is what will allow us to analyze the finite fluid flow in terms of matrix-geometric methods for the queues we construct.

In the following, the nonnegative, stochastic matrix

$$P(\theta_n) = \frac{1}{\theta_n} C^{-1} Q + I \quad (8)$$

will be denoted simply as P_n, and the matrix P_1 simply as P. We shall assume P_n to be partitioned in conformity with the partitioning of the state space as

$$P_n = \begin{pmatrix} P_{n11} & P_{n12} & P_{n13} \\ P_{n21} & P_{n22} & P_{n23} \\ P_{n31} & P_{n32} & P_{n33} \end{pmatrix}. \quad (9)$$

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

$$P_{nii} = \frac{n-1}{n} I + \frac{1}{n} P_{ii}, \quad i = 1, 2, 3; \quad (10)$$

and

$$P_{nij} = \frac{1}{n} P_{ij} \ \text{ for } \ i \neq j, \quad i, j \in \{1, 2, 3\}. \quad (11)$$

3 Approximating queues

3.1 Preliminaries

We assume the following as given on a *common* probability space $(\Omega, \mathcal{A}, \mathcal{P})$: A collection of mutually independent Poisson processes, say, $\mathcal{M}_{n,i}$, and $\mathcal{N}_{n,i}$ with rates λc_i respectively for $i \in S$, and $n \geq 1$; and a discrete time Markov chain $\mathcal{L} = \{L_n : n \geq 0\}$ of phases which has transition matrix P and is independent of all the Poisson processes $\mathcal{M}_{n,i}$, and $\mathcal{N}_{n,i}$, $n \geq 1, i \in S$.

For notational convenience, we use $\oplus$ for denoting superposition of processes; thus, $\mathcal{M}_{n,j} \oplus \mathcal{N}_{n,j}$ denotes the superposition of $\mathcal{M}_{n,j}$ and $\mathcal{N}_{n,j}$, and $\oplus_{i=1}^{n} \mathcal{M}_{i,j}$ denotes the superposition of the processes $\mathcal{M}_{1,j}, \cdots, \mathcal{M}_{n,j}$. Without loss of generality, we shall assume that $L_0 = i$ for some $i \in S$.

With these as building blocks, we will construct for almost all sample points in Ω: (a) a phase process $\mathcal{J} = \{J(t) : t \geq 0\}$ which is a CTMC with generator Q but realized through a spatial uniformization construct as described in Section 2; (b) a process $\mathcal{F}_\beta = \{F_\beta(t) : t \geq 0\}$ such that $F_\beta(t)$ increases at rate c_j while $J(t) = j \in S_1$ as long as the level of the process is below the finite buffer size β, decreases at rate c_j while $J(t) = j \in S_2$ and $F_\beta(\cdot)$ is positive, and remains constant while $J(t) \in S_3$ — i.e., $(F_\beta(t), J(t))$ is the finite buffer fluid flow process of interest; and (c) for each $n \geq 1$, a queue length process $\mathcal{Q}^{(n)} = \{Q^{(n)}(t) : t \geq 0\}$ with bounded queue size q_n (the size q_n is the upper bound on the number of customers that can be present simultaneously) and modulated by the phase process, and its associated work process $\mathcal{W}^{(n)} = \{W^{(n)}(t) : t \geq 0\}$. (All processes are defined to have right continuous sample paths.) Moreover, the queues $\mathcal{Q}^{(n)}$ are to be constructed such that; all queues $\mathcal{Q}^{(n)}$ are modulated by a common phase process; the amounts of work brought in by successive customers in $\mathcal{Q}^{(n)}$ are independent and $exp(n\lambda)$ distributed. Throughout, unless otherwise stated, for any process under consideration "state at a time point t" will always denote the state at $t+$.

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. This construction is almost identical to that of [2] except for the introduction of the finite buffers. We denote the set of epochs of the Poisson process $\oplus_{i=1}^{n}\mathcal{N}_{i,j}$ by $A_{n,j},\ j \in S$ and let $A_n = \cup_{j\in S} A_{n,j}$; the arrival epochs to the queue $\mathcal{Q}^{(n)}$ will be a subset of A_n as we shall see later. Similarly, we denote the epochs of the Poisson process $\oplus_{i=1}^{n}\mathcal{M}_{i,j}$ by $D_{n,j}$ and let $D_n = \cup_{j\in S} D_{n,j}$; the departure epochs of $\mathcal{Q}^{(n)}$ will form a subset of them as we shall see later.

Construction of the Phase Process $\mathcal{J}$: Note that $\mathcal{L} = \{L_n, n \geq 0\}$ is a discrete time Markov chain of phases which has the transition matrix P and is independent of all Poisson processes $\mathcal{N}_{n,j}, \mathcal{M}_{n,j}, n \geq 1, j \in S$. 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 $a_n \leq t < a_{n+1}$. We note first of all that $\{(L_n, a_n) : n \geq 0\}$ is a SMP of the type discussed in Section 2. Also, from Theorem 1 in [2], $\mathcal{J} = \{J(t) : t \geq 0\}$ is a continuous time Markov chain 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 Section 2.

Construction of the Fluid Flow $\mathcal{F}_\beta$: Without loss of generality, we will assume the initial condition $F(0) = 0$. We let $F_\beta(0) = 0$ and define the process $\{F_\beta(t)\}$ such that for $t \in [a_n,\ a_{n+1})$, $F_\beta(t) = \min\{\beta,\ F_\beta(a_n) + c_j(t - a_n)\}$ if $J(t) = j \in S_1$, $F_\beta(t) = \max[0,\ F_\beta(a_n) - c_j(t - a_n)]$ if $J(t) = j \in S_2$, and finally $F_\beta(t) = F_\beta(a_n)$ if $J(t) \in S_3$. Defined thus, clearly $F_\beta(\cdot)$ increases at rate c_j when the phase process is in $j \in S_1$ and the level of the process is below β, decreases at rate c_j in $j \in S_2$ while the level is positive, and remains constant in S_3. Clearly, the joint process $\{(F_\beta(t),\ J(t))\}$ is stochastically equivalent to the fluid model with finite buffer β starting empty and in phase i, and is modulated by the underlying Markov process $\mathcal{J}$ with a generator Q. (Other initial conditions for the fluid model are handled in an obvious manner.)

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 $\{t_k^n : k \geq 1\}$ 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. It will be assumed that service is rendered by the server only when the phase is in S_2; specifically, no service is rendered when the phase is in $S_1 \cup S_3$. Also, for all queues, the queue size at time 0 will be defined to be 0 to match our initial condition $F_\beta(0) = 0$ (other initial conditions can be accommodated with minor changes in the construction.) In the following, 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+. The specific details of the definition of these quantities are now described below.

(a) Let $t_0^n = 0$ and $Q_0^n = 0$; this initializes the queue size at time 0 to match our initial state specification $F(0) = 0$ for the fluid model. Note that we have $J_0^n = J(0) = i$ from the construction of the phase process.

(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 immediately after 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 the next queue length value Q_{k+1}^n is set to $\max\{q_n, Q_k^n+1\}$ — that is, in this case, the epoch t_{k+1}^n is defined to be an arrival epoch to the queue $\mathcal{Q}^{(n)}$; note that the arrival is lost if the system is saturated. The phase 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 at that epoch in the uniformization scheme; 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 the next epoch 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 the queue length value Q_{k+1}^n is set to the same value as Q_k^n — that is, a construct is made that makes the queue length remain constant just as the fluid level would remain constant over the interval under consideration (note that we are assuming that no work is being done in S_3.) The phase J_{k+1}^n is set to $J(t_{k+1}^n)$; note that a phase transition to a different phase occurs at the newly defined epoch iff that epoch $t_{k+1}^n \in A_{1,J_k^n}$ and the new phase entered is indeed different; otherwise, the epoch is to be treated as a self-transition epoch of the phase for the queue $\mathcal{Q}^{(n)}$.

Case 3: If $J_k^n = j \in S_2$, then we set the next epoch 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, the next queue length value 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 (i.e., with a phase self transition).

Construction of the process $\mathcal{Y}^{(n)} = \{Y^{(n)}(t)\}$: We assume that the queue $\mathcal{Q}^{(n)}$ operates under the FIFO discipline and that in it work gets done only when the phase is in the set S_2, and specifically that when the phase is $i \in S_2$, work does get depleted at rate c_i per unit time. Given a path of our construction, namely the epochs $t_k^{(n)}$, $k \geq 0$ and the phases J_k^n at those epochs, we can construct from the amount of work done between the n-th and $(n+1)$-st departure epochs, the amount of work brought in by the n_1-st customer (similarly that of the first customer by considering the path up to the first departure) and from these define the work process $W^n(t)$ associated with the queue $\mathcal{Q}^{(n)}$; the formal equations governing this construction can be written down but are not germane to our analysis.

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

$$Y^{(n)}(t) = \begin{cases} W^{(n)}(t_k^n) + c_j(t - t_k^n), & \text{if } j \in S_1 \text{ and } Q_k^n < q_n \\ W^{(n)}(t_k^n), & \text{if } j \in S_1 \text{ and } Q_k^n = q_n \\ \max(0,\ W^{(n)}(t_k^n) - c_j(t - t_k^n)), & \text{if } j \in S_2 \\ W^{(n)}(t_k^n), & \text{if } j \in S_3; \end{cases} \tag{12}$$

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 $F_\beta(t)$. Indeed, as shown for the infinite buffer case in [2], here also $\{F_\beta(t) : t \geq 0\}$ can be shown to be realized as the pathwise stochastic process limit of $\{Y^{(n)}(t) : t \geq 0\}$; see Section 4. Finally, note that since $t_{k+1}^n - t_k^n$ converges to zero a.s. as $n \to \infty$, the difference between $Y^{(n)}(t)$ and $W^{(n)}(t)$ also become negligible for large n.

Now, coming to Figure 1 and Figure 2, they provide the simulated results for the processes $Y^{(n)}(t)$ and the fluid process $F_\beta(t)$. The two cases considered respectively correspond to

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

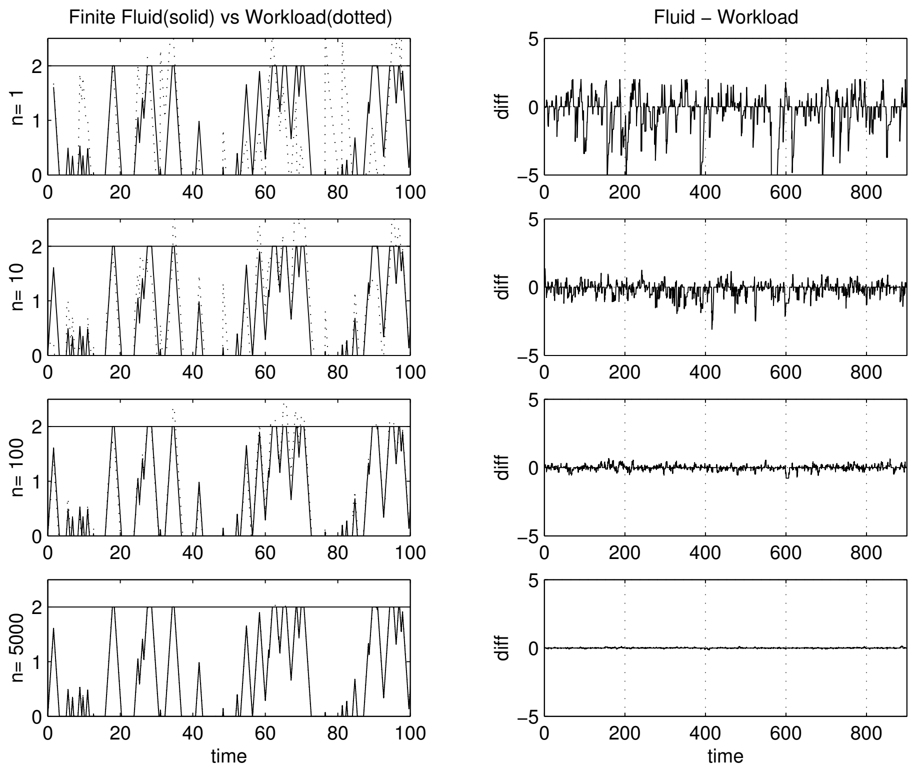

Figure 1: Comparison of the paths of a fluid flow process and workload in the queue $\mathcal{Q}^{(n)}$, for $n = 1, 10, 100, 5000$. The traffic intensity is 0.5.

and in both cases, we assume that $\beta = 2$, $C_1 = 1$, $C_2 = 1$, $S_1 = \{1\}$, $S_2 = \{2\}$ and $S_3 = \phi$. Note from Fig. 1 and Fig. 2 that as $n \to \infty$, the differences between Y^n and F_β do seem to vanish.

Now, we summarize some basic properties of the constructed queue processes as a theorem, whose proof is essentially the same as that of Theorem 2 in [2]. We draw particular attention to Part (c) which asserts that the work brought in by successive arrivals in $\mathcal{Q}^{(n)}$ are iid $exp(n\lambda)$.

Theorem 1

(a) Arrivals to the queue $\mathcal{Q}^{(n)}$ can 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.) Note that arrivals that find a full waiting room are lost.

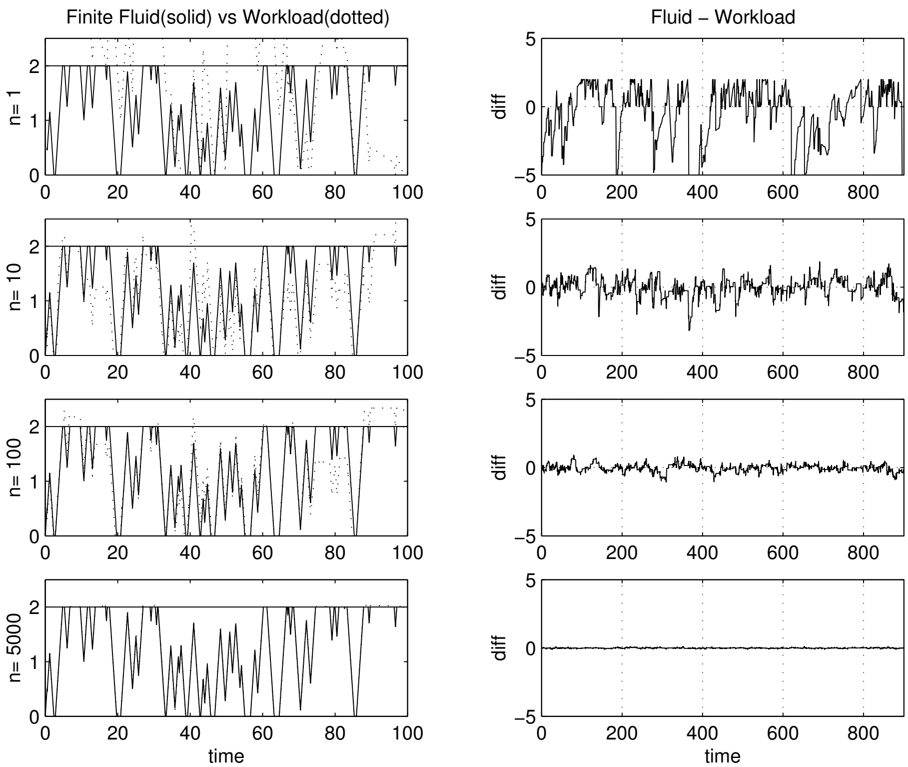

Figure 2: Comparison of the paths of a fluid flow process and workload in the queue $\mathcal{Q}^{(n)}$ for $n = 1, 10, 100, 5000$. The traffic intensity is 2.

(b) Departures from 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) The amounts of work done in $\mathcal{Q}^{(n)}$ between successive departure epochs are iid exponentially distributed random variables with mean $1/(n\lambda) = \beta/(n_0 n)$.

Remark 1

(a) The upper bound q_n on the queue length of $\mathcal{Q}^{(n)}$ corresponds to an amount of work distributed as the sum of $q_n (= n_0 n)$ i.i.d. exponentially distributed random variables each with mean $1/(n\lambda) = \beta/(n_0 n)$. This converges in probability to β as $n \to \infty$. Thus, we can view our approximating queues as

also comprising of replacing the deterministic boundary β for the fluid by a sequence of random variables converging to β in the limit.
(b) Note that $\{a_k, k \geq 0\}$ emerging in the construction of $\mathcal{J}$ and $\mathcal{F}_\beta$ is the set of epochs of the Poisson process $\mathcal{N}_1$ and

$$\{a_k, k \geq 0\} \subset \{t_k^n : k \geq 0\} \text{ for all } n \geq 1.$$

Moreover, we can see also that for all $n = 1, 2, \cdots$,

$$\{t_k^n : k \geq 0\} \subset \{t_k^{n+1} : k \geq 0\}, \tag{13}$$

so that, these epochs form a set of nested sequences over n. This aspect of the construction is crucial in making all processes to be modulated by a common CTMC of phases and for obtaining the fluid process as an a.s. pathwise limit as $n \to \infty$ of the work-in-the-queue processes.

4 Limit Results

Note that the sequence of epochs $\{t_k^{(n)} : k \geq 0\}$ provide (over n) a sequence of nested spatial uniformizations for the fluid process on the common probability space $(\Omega, \mathcal{A}, P)$ and correspond (see Section 2) to $\theta_n = n\lambda$. Clearly, $\theta_n \to \infty$, as $n \to \infty$, and since we have shown that in the n-th queue $\mathcal{Q}^{(n)}$, the amount of work brought by each customer is exponentially distributed with mean $1/(n\lambda)$, we can now view the queues as providing a sequence of approximations to the fluid model through a set of queues whose work process corresponds to a stochastic quantization of the fluid and that these quanta converge to zero as $n \to \infty$. This discussion anticipates a limit theorem, namely, that the process $(W^{(n)}(t), J(t))$, where $W^{(n)}(t)$ is the amount of work at $t+$ in the queue $\mathcal{Q}^{(n)}$ will indeed converge (we denote this by $\Rightarrow$) as $n \to \infty$ to the process $(F_\beta(t), J(t))$. Here, convergence is in the sense of

$$\int_\Omega f(W^{(n)}, J)\, d\mathcal{P} \to \int_\Omega f(F_\beta, J)\, d\mathcal{P}, \quad \text{as } n \to \infty,$$

for all continuous functionals f, and it can be sharpened to almost sure pathwise convergence by assuming that the underlying probability space is a separable, metric space (see [15], [12]). Since for large n, $Y^{(n)}(t) \approx W^{(n)}(t)$, similar results are anticipated for the sequence $(Y^{(n)}, J)$. We only sketch the proof of that result; the details are similar to those employed in [2].

Theorem 2 *We have*

$$(Y^{(n)}(\cdot), J(\cdot)) \Rightarrow (F_\beta(\cdot), J(\cdot)) \text{ as } n \to \infty.$$

There exist versions of the processes (defined on a separable, metric space) such that the convergence is indeed an a.s. pathwise convergence. Finally, this a.s. convergence is uniform over bounded time intervals $[0, t]$.

Sketch of Proof: Consider the sequence of unrestricted queue processes and the unrestricted fluid process corresponding to the finite queues and finite fluid processes. (Here the boundaries are removed and the processes are allowed to range over $(-\infty, \infty)$.) Our approximation scheme through the nested spatial uniformization can be explained as follows: when the phase process stays in $j \in S_1$ during an interval $[a_{k-1}, a_k)$, where a_k's are defined in the construction of $\mathcal{J}$, the quantity $c_j(t - a_{k-1})$, which is the amount of incoming fluid during $[a_{k-1}, t)$, $a_{k-1} \leq t \leq a_k$, is approximated by

$$\sum_{i=1}^{N_n(t-a_{k-1})} X_{n,i}, \tag{14}$$

where $N_n(t)$ is Poisson distributed with mean $\theta_n c_j t$, and $X_{n,i}$'s are iid exponentially distributed with mean $1/n\lambda$ and independent of $N_n(t)$; when the phase process stays in $j \in S_2$ during $[a_{k-1}, a_k)$, the total depletion of fluid during $[a_{k-1}, t)$ equals the total depletion of work during the same period; when the phase process stays in $j \in S_3$ during $[a_{k-1}, a_k)$, there is no change either in the fluid or the work process. If we consider a process defined by (14), then it can be shown that this process converges to $\{c_j(t-a_{k-1}), t \in [a_{k-1}, a_k)\}$ for all $k \geq 1$. This implies that the unrestricted work processes do converge to the unrestricted fluid flow over any finite time interval. (The bounds obtained in [2], Section 9, also show that the convergence is uniform over bounded intervals — i.e., we have local uniform convergence.)

From the above, we can, by using the reflection map([15], [2]) prove that the sequence of work processes of the queues with infinite buffer sizes converges (in probability) locally uniformly to the fluid flow with an infinite buffer, where by "locally uniform," we mean that convergence is uniform in any finite interval [0, t]; see [2] for the details on the infinite buffer case. Similarly, if we use the two-sided reflection map ([15], Sec. 5.2.3) for the finite fluid flow and the associated finite queue process, it can be shown that the sequence of work processes of the finite queues $\mathcal{Q}^{(n)}$ converges locally uniformly to the finite fluid flow. The details are a bit arduous but similar to that in [2] . □

The above provides a means to obtain transient results for the finite buffer fluid flow model from those of the approximating queues in a manner similar to that developed in [2] for the infinite buffer case. In this paper, we will, however, restrict ourselves to steady state results concerning which we have the following theorem.

Theorem 3 *For all $-\infty < a < b < \infty$, and $j \in S$,*

$$\lim_{n\to\infty}\lim_{t\to\infty} \mathcal{P}[Y^{(n)}(t-) \in (a,b], J(t) = j] = \lim_{t\to\infty} \mathcal{P}[F_\beta(t) \in (a,b], J(t) = j].$$

Remark 2 *Usually, one justifies the interchange of the limits over n and t entailed in the above theorem by appealing to some tightness conditions; see e.g., [7]. Such proofs are known to be quite difficult. In our case, we can provide a simpler proof by direct computations, and we will take it up later in Section 6.*

Remark 3 *Later on, we will show that some key matrices needed to determine the steady state distribution are identical for each of the stochastic discretizations. That simplifies the steady state analysis drastically compared to the transient analysis and accounts for our limiting ourselves to the steady state analysis in the following sections.*

5 Steady State Analysis of the Queues

In this section, we present the steady state analysis of $(\mathcal{Y}^{(n)}, \mathcal{J})$. By Theorem 3, we can obtain the steady state results for $(\mathcal{F}_\beta, \mathcal{J})$ from the consideration of these processes. For the analysis, we need certain results on the n-th finite queue process $(\mathcal{Q}^{(n)}, \mathcal{J})$, especially, steady-state results for its embedded sequence $\{(Q_l^n,\, J_l^n) : l \geq 0\}$ defined in Section 3.2.

By the construction in the previous section, $\mathcal{Q}^{(n)}$ has a finite QBD structure. This lends to an analysis by the matrix-geometric method for finite queues, and the next theorem lists the relevant properties of the queues, which will be used in the analysis.

Theorem 4

(a) All the queues $\mathcal{Q}^{(n)}$ are modulated by the same continuous time phase process $\mathcal{J}$ 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, finite QBD with state space

$$\Omega_n = \cup_{i=1}^{q_n} (\{i\} \otimes S),$$

where $\otimes$ represents the Kronecker product of sets.

(c) *The embedded sequence* $\{(Q_k^n, J_k^n) : k \geq 0\}$ *is a discrete time, finite QBD ([8]) with the state space* Ω_n*, and has transition matrix*

$$\begin{pmatrix} M_0^n & B_{n0} & & & \\ B_{n2} & B_{n1} & B_{n0} & & \\ & \ddots & \ddots & \ddots & \\ & & B_{n2} & B_{n1} & B_{n0} \\ & & & B_{n2} & M_n^n \end{pmatrix}, \tag{15}$$

where $M_0^n = B_{n2} + B_{n1}$, $M_n^n = B_{n1} + B_{n0}$ *and*

$$B_{n2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2}I & 0 \\ 0 & 0 & 0 \end{pmatrix}, B_{n1} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2}P_{n21} & \frac{1}{2}P_{n22} & \frac{1}{2}P_{n23} \\ P_{n31} & P_{n32} & P_{n33} \end{pmatrix}, \tag{16}$$

$$B_{n0} = \begin{pmatrix} P_{n11} & P_{n12} & P_{n13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{17}$$

where the matrices P_{nij} *are defined by (8)-(11).*

(d) *The sojourn time* $t_k^n - t_{k-1}^n, k \geq 1$*, depends on the phase* $J(t_{k-1}^n)$*; if* $J(t_{k-1}^n) = j \in S_1 \cup S_3$*, then* $t_k^n - t_{k-1}^n$ *is exponentially distributed with mean* $1/(n\lambda)$*; if* $J(t_{k-1}^n) = j \in S_2$*, then* $t_k^n - t_{k-1}^n$ *exponentially distributed with mean* $1/(2n\lambda)$*.*

Proof: The results are easy to prove based on the the way the processes are constructed, and we omit the details. □

Before the analysis, we will define some notations to be used later. Denote the steady state probability vector related to the embedded sequence $\{(Q_l^n,\ J_l^n) : l \geq 0\}$ by $\mathbf{x}_{nk}$, $k = 0, \cdots, q_n$, such that its j-th element is $\mathcal{P}[Q_\infty^n = k,\ J_\infty^n = j]$, $j \in S$, and partition it in conformity with the partitioning of the state space as

$$\mathbf{x}_{nk} = (\mathbf{x}_{nk1} \ \ \mathbf{x}_{nk2} \ \ \mathbf{x}_{nk3}); \tag{18}$$

here $\mathbf{x}_{nki}$ is a vector of size $|S_i|$, and its components are indexed by phases in S_i.

5.1 Steady state analysis of $\{(Q_l^n, J_l^n)\}$

We shall analyze $\{(Q_l^n, J_l^n)\}$ using matrix-geometric methods. To that end, we need to define an associated "level reversed process" which is a Markov chain obtained from (15) by reversing the upward and downward movements in "levels" and has transition matrix

$$\begin{pmatrix} M_n^n & B_{n2} & & & & \\ B_{n0} & B_{n1} & B_{n2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & & B_{n0} & B_{n1} & B_{n2} \\ & & & & B_{n0} & M_0^n \end{pmatrix}. \tag{19}$$

We will call the queueing model defined by the above as "the level reversed queue."

Let $R_{O,n}$ and $R_{L,n}$ denote respectively the Neuts' R-matrices ([9], [8]) for the original QBD of (15) and the level-reversed QBD of (19). Then it is well-known that they are respectively the minimal non-negative solutions of the following equations:

$$R_{O,n} = B_{n0} + R_{O,n}B_{n1} + R_{O,n}^2 B_{n2} \tag{20}$$

$$R_{L,n} = B_{n2} + R_{L,n}B_{n1} + R_{L,n}^2 B_{n0}. \tag{21}$$

Hajek [6] has shown that the stationary probability vector of a finite QBD may be expressed as a linear combination of two matrix-geometric vectors determined by $R_{O,n}$ and $R_{L,n}$ respectively, and we summarize the relevant results in the following theorem.

Theorem 5 *Let ξ denote the stationary probability vector of the Markov process of phases defined by Q satisfying the equations*

$$\xi Q = 0, \quad \xi \mathbf{1} = 1,$$

and assume that it is partitioned according to the state space as $\xi = (\xi_1, \xi_2, \xi_3)$. Assume that $\xi_1 C_1 \mathbf{1} \neq \xi_2 C_2 \mathbf{1}$.

(a) (Hajek [6]) For $k = 0, \cdots, q_n$,

$$\mathbf{x}_{nk} = \eta_{O,n} R_{O,n}^k + \eta_{L,n} R_{L,n}^{q_n - k}, \tag{22}$$

where the vectors $\eta_{O,n}$ and $\eta_{L,n}$ satisfy the following equation,

$$0 = (\eta_{O,n} \quad \eta_{L,n}) \times \begin{pmatrix} M_0^n + R_{O,n}B_{n2} - I & R_{O,n}^{q_n}(M_n^n - I) + R_{O,n}^{q_n-1}B_{n0} \\ R_{L,n}^{q_n}(M_0^n - I) + R_{L,n}^{q_n-1}B_{n2} & M_n^n + R_{L,n}B_{n0} - I \end{pmatrix} \tag{23}$$

(b) If we define $\Delta = diag\{I_{|S_1|}, 2I_{|S_2|}, I_{|S_3|}\}$ and denote the steady state probability vector of the infinitesimal generators $C^{-1}\Delta^{-1}Q$ by ζ, then

$$\zeta(B_{n2} + B_{n1} + B_{n0}) = \zeta \quad \text{and} \quad \sum_{k=0}^{q_n} \mathbf{x}_{nk} = \zeta \ \forall n \geq 1, \tag{24}$$

and

$$\zeta = \frac{1}{(\xi \Delta C \mathbf{1})} \xi \Delta C \tag{25}$$

Proof: The proof of (a) can be seen in [6] and [8], and (b) comes from the structure of the transition matrix in (15). □

For later use, we consider ζ, $\eta_{O,n}$ and $\eta_{L,n}$ also to be partitioned (in conformity with the partitioning of the state space) as

$$\zeta = (\zeta_1 \quad \zeta_2 \quad \zeta_3),$$
$$\eta_{O,n} = (\eta_{O,n1} \quad \eta_{O,n2} \quad \eta_{O,n3}), \quad \eta_{L,n} = (\eta_{L,n1} \quad \eta_{L,n2} \quad \eta_{L,n3}).$$

Before concluding this brief section, we note that from (20) and the structure of the matrices B_{n0} and B_{n2}, the matrices $R_{O,n}$ and $R_{L,n}$ have the following partitioned form:

$$R_{O,n} = \begin{pmatrix} R_{O,n11} & R_{O,n12} & R_{O,n13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{26}$$

and

$$R_{L,n} = \begin{pmatrix} 0 & 0 & 0 \\ R_{L,n21} & R_{L,n22} & R_{L,n23} \\ 0 & 0 & 0 \end{pmatrix}. \tag{27}$$

These reducible structures are known to reduce computational complexity in evaluating these matrices numerically.

5.2 The Rate Reversed Fluid Flow

We introduce now one more intermediate object we need to complete our analysis, namely, the rate reversed fluid flow.

Analogous to the level reversed queue in the finite state space case, we can construct a "rate reversed fluid model" in the fluid case. This is the fluid flow model modulated by $\mathcal{J}$ such that the fluid level increases at rate

c_j when the phase process is in $j \in S_2$, decreases at rate c_j when the phase process is in $j \in S_1$, and remains constant when the phase process is in $j \in S_3$. (Note that the roles of S_1 and S_2 have been reversed.) Then, in the same way as we constructed the approximating queue process $\mathcal{Q}^{(n)}$ for the given fluid process, we can construct the approximating queues, say $\mathcal{Q}_*^{(n)} = \{(Q_{*,l}^{(n)}, J_{*,l}^{(n)} : l \geq 0\}$, for the level reversed fluid flow process as well. (Note that the embedding points for these are different and hence this process is not the same as the level reversed queue process introduced earlier.) The n-th queue thus obtained is also a discrete time QBD whose transition matrix has a similar structure as in (15) with $B_{*,n2}$, $B_{*,n1}$ and $B_{*,n0}$ in place of B_{n2}, B_{n1} and B_{n0} respectively, where

$$\begin{aligned} B_{*,n2} &= \begin{pmatrix} \frac{1}{2}I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; B_{*,n1} = \begin{pmatrix} \frac{1}{2}P_{n11} & \frac{1}{2}P_{n12} & \frac{1}{2}P_{n13} \\ 0 & 0 & 0 \\ P_{n31} & P_{n32} & P_{n33} \end{pmatrix}, \\ B_{*,n0} &= \begin{pmatrix} 0 & 0 & 0 \\ P_{n21} & P_{n22} & P_{n23} \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{28}$$

If we let $R_{*,n}$ denote the Neuts' matrix of this queue, then it is the minimal non-negative solution of the equation

$$R_{*,n} = B_{*,n0} + R_{*,n}B_{*,n1} + R_{*,n}^2 B_{*,n2}, \tag{29}$$

and due to the structure of the block matrices has the following partitioned form:

$$R_{*,n} = \begin{pmatrix} 0 & 0 & 0 \\ R_{*,n21} & R_{*,n22} & R_{*,n23} \\ 0 & 0 & 0 \end{pmatrix}. \tag{30}$$

We will need the matrices $R_{*,n}$ for defining some key quantities needed in the analysis of the bounded fluid process. They are given by:

$$\tilde{K}_O = \theta_n C_1 (R_{O,n11} - I) C_1^{-1}, \; \tilde{\Psi}_O = \frac{1}{2} C_1 R_{O,n12} C_2^{-1}, \tilde{\Theta}_O = C_1 R_{O,n13}, \tag{31}$$

and

$$\tilde{K}_* = \theta_n C_2 (R_{*,n22} - I) C_2^{-1}, \; \tilde{\Psi}_* = \frac{1}{2} C_2 R_{*,n21} C_1^{-1}, \tilde{\Theta}_* = C_2 R_{*,n23}. \tag{32}$$

For the original and the rate reversed fluid models, these sets of matrices are none other than the matrices $\tilde{K}(0+)$, $\tilde{\Psi}(0+)$ and $\tilde{\Theta}(0+)$ introduced in

[11] and [2]. For their probabilistic interpretations and use in the analysis of the infinite buffer case, we refer the reader to [2]. These matrices do not depend on n; see Theorem 6 below. This fact is what leads to the simplicity of evaluating the steady state distribution of the finite fluid model; we need to evaluate only one set of these matrices, and not a set for each queue separately. (Unfortunately, the kernels $\tilde{K}_{On}(s)$ etc., derived from the n-th approximations (see [2]) needed in the transient analysis do not have this nice property.)

Theorem 6 *The matrices* $(\tilde{K}_O, \tilde{\Psi}_O, \tilde{\Theta}_O)$ *defined in (31) and the matrices* $(\tilde{K}_*, \tilde{\Psi}_*, \tilde{\Theta}_*)$ *defined in (32) are independent of* n*.*

Proof:
For a $|S| \times |S|$ dimensional matrix A, we use the notation $\{A\}_{ij},\ i,j = 1,2,3$ to denote the (i,j)-th submatrix when A is partitioned in conformity with the partitioning of the state space S.

Using (20) and (26), we see that for all $n \geq 1$, the matrices $R_{O,n11}$, $R_{O,n12}$ and $R_{O,n13}$ satisfy the following equations:

$$\begin{aligned} R_{O,n11} &= I + \frac{1}{n\lambda} C_1^{-1} Q_{11} + \frac{1}{2n\lambda} R_{O,n12} C_2^{-1} Q_{21} + \frac{1}{n\lambda} R_{O,n13} Q_{31}, \\ R_{O,n12} &= \frac{1}{n\lambda} C_1^{-1} Q_{12} + \frac{1}{2} R_{O,n12} + \frac{1}{2n\lambda} R_{O,n12} C_2^{-1} Q_{22} \\ &\qquad + \frac{1}{n\lambda} R_{O,n13} Q_{32} + \frac{1}{2} R_{O,n11} R_{O,n12}, \\ R_{O,n13} &= \frac{1}{n\lambda} C_1^{-1} Q_{13} + \frac{1}{2n\lambda} R_{O,n12} C_2^{-1} Q_{23} + R_{O,n13} + \frac{1}{n\lambda} R_{O,n13} Q_{33}, \end{aligned} \tag{33}$$

and it follows

$$\begin{aligned} \frac{1}{2} R_{O,n12} &= \frac{1}{n\lambda} C_1^{-1} Q_{12} + \frac{1}{2n\lambda} R_{O,n12} C_2^{-1} Q_{22} + \frac{1}{n\lambda} R_{O,n13} Q_{32} \\ &\qquad + \frac{1}{2} R_{O,n11} R_{O,n12} \\ 0 &= \frac{1}{n\lambda} C_1^{-1} Q_{13} + \frac{1}{2n\lambda} R_{O,n12} C_2^{-1} Q_{23} + \frac{1}{n\lambda} R_{O,n13} Q_{33}. \end{aligned} \tag{34}$$

Now, let

$$H = \begin{pmatrix} I + \frac{1}{n}(R_{O,111} - I) & R_{O,112} & R_{O,113} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{35}$$

Since $I + \frac{1}{n}(R_{O,111} - I) = \frac{n-1}{n} I + \frac{1}{n} R_{O,111}$ is nonnegative, H is a non-negative matrix.

From equations (33) and (34), we can now easily verify the following equations:

$$\begin{aligned}
\{H\}_{11} &= \left\{B_{n0} + HB_{n1} + H^2 B_{n2}\right\}_{11}, \\
\{H\}_{12} &= \left\{B_{n0} + HB_{n1} + H^2 B_{n2}\right\}_{12}, \\
\{H\}_{13} &= \left\{B_{n0} + HB_{n1} + H^2 B_{n2}\right\}_{13}.
\end{aligned}$$

which shows that H is a non-negative matrix solution of the equation (20) and since $R_{O,n}$ is the non-negative minimal solution of the equation (20), it follows that

$$I + \frac{1}{n}(R_{O,111} - I) \geq R_{O,n11} \ \Leftrightarrow\ \lambda(R_{O,111} - I) \geq n\lambda(R_{O,n11} - I), \qquad (36)$$
$$R_{O,112} \geq R_{O,n12} \text{ and } R_{O,113} \geq R_{O,n13}.$$

To show inequalities in the reverse direction, we fix an $n \geq 2$ and define

$$\tilde{H} = \begin{pmatrix} I + n(R_{O,n11} - I) & R_{O,n12} & R_{O,n13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad (37)$$

Using the equations (33) and (34), we can then verify that the following equations hold:

$$\begin{aligned}
\{\tilde{H}\}_{11} &= \left\{B_{10} + \tilde{H}B_{11} + \tilde{H}^2 B_{12}\right\}_{11}, \\
\{\tilde{H}\}_{12} &= \left\{B_{10} + \tilde{H}B_{11} + \tilde{H}^2 B_{12}\right\}_{12}, \\
\{\tilde{H}\}_{13} &= \left\{B_{10} + \tilde{H}B_{11} + \tilde{H}^2 B_{12}\right\}_{13}.
\end{aligned}$$

Moreover, since $\lambda = \max_{i \in S}\{-[C^{-1}Q]_{ii}\}$, the matrix

$$I + n(R_{O,n11} - I) = I + \frac{1}{\lambda}C_1^{-1}Q_{11} + \frac{1}{2\lambda}R_{O,n12}C_2^{-1}Q_{21} + \frac{1}{\lambda}R_{O,n13}Q_{31},$$

is a non-negative matrix. Thus, it follows that $\tilde{H}$ is a non-negative matrix solution of the equation (20) with $n = 1$. Since R_1 is the non-negative minimal solution of the equation (20) with $n = 1$, we then have that

$$I + n(R_{O,n11} - I) \geq R_{O,111} \ \Leftrightarrow\ n\lambda(R_{O,n11} - I) \geq \lambda(R_{O,111} - I), \qquad (38)$$
$$R_{O,n12} \geq R_{O,112} \text{ and } R_{O,n13} \geq R_{O,113}.$$

From the equations (36) and (38), we can see that for all $n \geq 1$,

$$n\lambda(R_{O,n11} - I) = \lambda(R_{O,111} - I), \;\; R_{O,n12} = R_{O,112} \;\; \text{and} \;\; R_{O,n13} = R_{O,113}.$$

This shows that the matrices $\tilde{K}_O$, $\tilde{\Psi}_O$ and $\tilde{\Theta}_O$ do not depend on n. A similar proof holds for the matrices $\tilde{K}_*$, $\tilde{\Psi}_*$ and $\tilde{\Theta}_*$. □

For later use, we summarize, in the following two theorems, some properties related to the various R-matrices and others derived from them.

Theorem 7

(a) Under the condition $\xi_1 C_1 \mathbf{1} - \xi_2 C_2 \mathbf{1} < 0$, which is the stability condition for the infinite buffer fluid model corresponding to the given finite buffer model,

$$sp(R_{O,n}) < 1 = sp(R_{L,n}) = sp(R_{*,n}),$$

where $sp(A)$ denotes the spectral radius of a matrix A. On the contrary, if $\xi_1 C_1 \mathbf{1} - \xi_2 C_2 \mathbf{1} > 0$,

$$sp(R_{O,n}) = 1, \;\; sp(R_{L,n}) < 1, \;\; sp(R_{*,n}) < 1.$$

(b) The matrix $R_{L,n22}$ is invertible and $sp(R_{L,n22}) = sp(R_{L,n})$.

Proof: Part (a) follows from the well-known properties of the R-matrices for stable and unstable QBDs; see [8]. The invertibility of $R_{L,n22}$ is a consequence of the invertibility of $B_{n2} = \frac{1}{2}I$, since from (20), we have

$$R_{L,n} = B_{n2}(I - B_{n1} - R_{L,n}B_{n0})^{-1},$$

and the second assertion of (b) follows easily from (27). □

Theorem 8

(a) Under the condition $\xi_1 C_1 \mathbf{1} - \xi_2 C_2 \mathbf{1} < 0$, the matrix $\tilde{K}_O$ is invertible, and the eigenvalues of $\tilde{K}_O$ have negative real parts. Furthermore, $\tilde{K}_*$ is then a singular matrix whose eigenvalues have non-positive real parts.

(b) Under the condition $\xi_1 C_1 \mathbf{1} - \xi_2 C_2 \mathbf{1} > 0$, $\tilde{K}_*$ is invertible, and the eigenvalues of $\tilde{K}_*$ have negative real parts. But, $\tilde{K}_O$ is then a singular matrix whose eigenvalues have non-positive real parts.

(c) *The matrices $\tilde{K}_O$, $\tilde{\Psi}_O$ and $\tilde{\Theta}_O$ do not depend on the uniformization parameter θ_n and satisfy the following equations:*

$$\begin{aligned} \tilde{K}_O &= [Q_{11} + \tilde{\Psi}_O Q_{21} + \tilde{\Theta}_O Q_{31}]C_1^{-1}, \\ 0 &= Q_{12} + \tilde{\Psi}_O Q_{22} + \tilde{\Theta}_O Q_{32} + \tilde{K}_O \tilde{\Psi}_O C_2 \\ 0 &= Q_{13} + \tilde{\Psi}_O Q_{23} + \tilde{\Theta}_O Q_{33}. \end{aligned} \tag{39}$$

Furthermore, the matrices $\tilde{K}_O$ and $\tilde{\Theta}_O$ are determined by $\tilde{\Psi}_O$. Finally, among the solutions of (39), the set we have defined has the property that its $\tilde{\Psi}_O$ is the minimal nonnegative matrix for which (39) are satisfied.

(d) *The matrices $\tilde{K}_*$, $\tilde{\Psi}_*$ and $\tilde{\Theta}_*$ do not depend on the uniformization parameter θ_n and satisfy the following equations,*

$$\begin{aligned} \tilde{K}_* &= [Q_{22} + \tilde{\Psi}_* Q_{12} + \tilde{\Theta}_* Q_{32}]C_2^{-1}, \\ 0 &= Q_{21} + \tilde{\Psi}_* Q_{11} + \tilde{\Theta}_* Q_{31} + \tilde{K}_* \tilde{\Psi}_* C_1 \\ 0 &= Q_{23} + \tilde{\Psi}_* Q_{13} + \tilde{\Theta}_* Q_{33}. \end{aligned} \tag{40}$$

Furthermore, the matrices $\tilde{K}_*$ and $\tilde{\Theta}_*$ are determined by $\tilde{\Psi}_*$. Finally, among the solutions of (40), the set we have defined has the property that its $\tilde{\Psi}_*$ is the minimal nonnegative matrix for which (40) are satisfied.

(e) *For all θ_n,*

$$\begin{aligned} \tilde{K}_* &= \theta_n C_2 (I - R_{L,n22}^{-1}) C_2^{-1}, \\ \tilde{\Psi}_* &= 2C_2 R_{L,n21} C_1^{-1}, \\ \tilde{\Theta}_* &= 2C_2 R_{L,n22}^{-1} R_{L,n23}. \end{aligned} \tag{41}$$

Proof: The asserted results in Parts (a) and (b) follow from Theorem 10 and (28) of [2] by noting that the appropriate $\tilde{K}$-matrices here are indeed $\tilde{K}(0+)$ of that paper and are related to the expected total time spent in a phase during a busy period of the unrestricted fluid flow model. The equations in Parts (c) and (d) were also derived in [2], Theorem 11, and the minimality of the Ψ matrices follows from the minimality of the corresponding R-matrices defining them. We thus need to prove only (e). From the equation of $R_{L,n}$ in (20) and Theorem 6(b), we can show as in Theorem 12 in [2] that

$$\begin{aligned} \theta_n (I - R_{L,n22}^{-1}) &= C_2^{-1} Q_{22} + 2R_{L,n21} C_1^{-1} Q_{12} + 2R_{L,n22}^{-1} R_{23} Q_{32}, \\ 0 &= C_2^{-1} Q_{21} + 2R_{L,n21} C_1^{-1} Q_{11} + 2R_{L,n22}^{-1} R_{L,n23} Q_{31} \\ &\quad + 2\theta_n (I - R_{L,n22}^{-1}) R_{L,n21}, \\ 0 &= C_2^{-1} Q_{23} + 2R_{L,n22}^{-1} R_{L,n23} Q_{33} + 2R_{L,n21} C_1^{-1} Q_{13}. \end{aligned}$$

If we pre-multiply all these equations by the matrix C_2 and post-multiply by C_2^{-1} in the first equation, then we get the equations in (40) for the matrices defined by the right hand sides in (41). Now, the minimality of $R_{L,n21}$ in both implies that these matrices are indeed $\tilde{K}_*$, $\tilde{\Psi}_*$ and $\tilde{\Theta}_*$. □

5.3 Steady State Analysis of $Y^{(n)}$

By Theorem 3, we can obtain the steady state distribution of the finite fluid flow from the steady state distributions of the process $Y^{(n)}(\cdot)$. We first characterize the latter steady state distributions. The key tool is Markov renewal theory and the Key Renewal Theorem [4].

To this end, for any Borel measurable set G, let $\mathbf{V}_n(G)$ be a vector such that its j-th element is

$$[V_n(G)]_j = \lim_{t\to\infty} \mathcal{P}[Y^{(n)}(t-) \in G, J(t) = j],\ j \in S.$$

Its limit as $n \to \infty$ yields, by Theorem 3, the joint steady state distribution of the fluid level and phase for the fluid process (F_β, J); for this reason, we shall denote that limit by $V_F(\cdot)$.

In the following, we shall denote by $\hat{\mathbf{V}}_n(s)$ the Laplace Stieltjes transform (LST) of the matrix of distribution functions $V_n((-\infty, x])$. We shall demonstrate that these distributions have a limit $V_F((-\infty, x])$ as $n \to \infty$. (Our notation anticipates a later identification of the limit as the stationary distribution of the fluid model.) The distribution function $V_F(\cdot)$ will be shown to have a mass at 0 and β and a density in the open interval $(0, \beta)$; the resulting joint density function in $(0, \beta)$ will be denoted in the following by $\mathbf{v}_F(x)$, and its elements are indexed by states in S. Also, we consider a partitioning of all key quantities according to the partitioning of the state space. Thus, for example,

$$\begin{aligned}
\mathbf{V}_n(\cdot) &= (\mathbf{V}_{n,1}(\cdot)\ \ \mathbf{V}_{n,2}(\cdot),\ \ \mathbf{V}_{n,3}(\cdot)) \\
\mathbf{V}_F(\cdot) &= (\mathbf{V}_{F,1}(\cdot)\ \ \mathbf{V}_{F,2}(\cdot)\ \ \mathbf{V}_{F,3}(\cdot)),\ \text{and} \\
\mathbf{v}_\mathbf{F}(\cdot) &= (\mathbf{v}_{F,1}(\cdot)\ \ \mathbf{v}_{F,2}(\cdot)\ \ \mathbf{v}_{F,3}(\cdot)).
\end{aligned}$$

Similar partitionings hold for the LSTs also.

We also define the vectors

$$\alpha_{O,n} = (\alpha_{O,n1}\ \ \alpha_{O,n2}\ \ \alpha_{O,n3}) = (\theta_n \eta_{O,n1}\ \ \eta_{O,n2}\ \ \eta_{O,n3}) \tag{42}$$

$$\alpha_{L,n} = (\alpha_{L,n1}\ \ \alpha_{L,n2}\ \ \alpha_{L,n3}) = (\eta_{L,n1}\ \ \theta_n \eta_{L,n2}\ \ \eta_{L,n3}), \tag{43}$$

and

$$c^* = \xi \Delta C \mathbf{1}. \tag{44}$$

The following result characterizes the joint steady state distribution as $t \to \infty$ of $(Y^{(n)}(t), J(t))$.

Theorem 9 *Let $Re(s) > 0$ be sufficiently large so that $(sI + \tilde{K}_*)^{-1}$ exists. Then, the LST $\hat{\mathbf{V}}_n(s)$ is given by its partitioned components defined below.*
(a) For S_1,

$$\begin{aligned}
&\hat{\mathbf{V}}_{n,1}(s) \qquad (45)\\
= \;& c^* \alpha_{O,n1} C_1^{-1} (sI - \tilde{K}_O)^{-1} C_1 \\
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} (I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1) \right)^{q_n} \right] C_1^{-1} \\
+ \;& (c^*/2) \alpha_{L,n2} C_2^{-1} (sI + \tilde{K}_*)^{-1} C_2 \\
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} (I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2) \right)^{q_n} \right] \\
& \times \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right)^{-q_n+1} C_2^{-1} \tilde{\Psi}_* \\
+ \;& c^* \theta_n^{-1} \alpha_{O,n1} \left(\frac{\theta_n}{s+\theta_n} (I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1) \right)^{q_n} C_1^{-1} \\
+ \;& c^* \alpha_{L,n1} \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n} C_1^{-1}.
\end{aligned}$$

(b) For S_2,

$$\begin{aligned}
&\hat{\mathbf{V}}_{n,2}(s) \qquad (46)\\
= \;& c^* \alpha_{O,n1} C_1^{-1} (sI - \tilde{K}_O)^{-1} C_1 \\
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \left(I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1 \right) \right)^{q_n} \right] C_1^{-1} \tilde{\Psi}_O \\
+ \;& c^* \alpha_{L,n2} \frac{s+\theta_n}{2\theta_n} C_2^{-1} (sI + \tilde{K}_*)^{-1} C_2 \\
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right) \right)^{q_n+1} \right] \\
& \times \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right)^{-q_n} C_2^{-1} \\
+ \;& \frac{c^*}{2} \alpha_{O,n2} C_2^{-1}.
\end{aligned}$$

(c) For S_3,

$$\begin{aligned}
&\hat{\mathbf{V}}_{n,3}(s) \qquad (47)\\
= \;& c^* \alpha_{O,n1} C_1^{-1} (sI - \tilde{K}_O)^{-1} C_1
\end{aligned}$$

$$
\begin{aligned}
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \left(I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1 \right) \right)^{q_n} \right] C_1^{-1} \tilde{\Theta}_O \\
+ & \frac{c^*}{2} \alpha_{L,n2} C_2^{-1} (sI + \tilde{K}_*)^{-1} C_2 \\
& \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right) \right)^{q_n+1} \right] \\
& \times \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right)^{-q_n+1} C_2^{-1} \tilde{\Theta}_* \\
+ & c^* \alpha_{O,n3} + c^* \alpha_{L,n3} \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n} .
\end{aligned}
$$

Proof: Consider the SMP $\{(Q_k^{(n)}, J_k^{(n)}, t_k^n) : k \geq 0\}$. The fundamental mean (see [4]) of this SMP can be computed from the stationary distribution of the embedded Markov chain and the mean sojourn times. A direct calculation would show that fundamental mean to be given by $c^* \theta_n$, where c^* was defined by (44) and $\theta_n = n\lambda$. For the following, let us define $\rho_{(0,i);(k,j)}^{(n)}(du)$ to be the element in the Markov renewal kernel (see [4]) associated with this SMP that can be interpreted as the elementary probability of a transition by the SMP into the state (k, j) at u, given that it starts in $(0, i)$ at time 0.

Conditioning on the last epoch of jump of the SMP u before time t, we evaluate joint transform

$$
E_{(0,i)}[e^{-sY^{(n)}(t)} I(J(t) = j)], \tag{48}
$$

for $j \in S_1$ as being given by

$$
\begin{aligned}
& \sum_{k=0}^{q_n-1} \int_0^t \rho_{(0,i);(k,j)}^{(n)}(du) \left(\frac{\theta_n}{s+\theta_n} \right)^k e^{-sc_j(t-u)} \; e^{-\theta_n c_j (t-u)} \\
& \quad + \int_0^t \rho_{(0,i);(q_n,j)}^{(n)}(du) \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n} e^{-\theta_n c_j (t-u)} .
\end{aligned}
$$

Now, taking the limit as $t \to \infty$ in the above using the Key Renewal Theorem (see [4]), we get the limit of that transform as

$$
c^* \theta_n \left[\sum_{k=0}^{q_n-1} [\mathbf{x}_{nk}]_j \left(\frac{\theta_n}{s+\theta_n} \right)^k [(s+\theta_n) c_j]^{-1} + [\mathbf{x}_{nq_n}]_j \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n} (c_j \theta_n)^{-1} \right] .
$$

We can write the above in vector form as

$$
\hat{\mathbf{V}}_{n,1}(s) \quad = \quad c^* \theta_n \sum_{k=0}^{q_n-1} \mathbf{x}_{nk1} \left(\frac{\theta_n}{s+\theta_n} \right)^k \frac{1}{s+\theta_n} C_1^{-1}
$$

$$+ c^* \mathbf{x}_{nq_n1} \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n} C_1^{-1}.$$

Equation (45) is now obtained by simplifying the above using Theorem 5 and the following relations which are easy to derive using the definitions of $\tilde{K}_O$ and $\tilde{K}_*$:

$$\sum_{k=0}^{q_n} R_{O,n11}^k \left(\frac{\theta_n}{s+\theta_n} \right)^k = (s+\theta_n) C_1^{-1} (sI - \tilde{K}_O)^{-1} C_1 \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n+1} \left(I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1 \right)^{q_n+1} \right],$$

$$\sum_{k=0}^{q_n} R_{L,n22}^{-k} \left(\frac{\theta_n}{s+\theta_n} \right)^k = (s+\theta_n) C_2^{-1} (sI + \tilde{K}_*)^{-1} C_2 \times \left[I - \left(\frac{\theta_n}{s+\theta_n} \right)^{q_n+1} \left(I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2 \right)^{q_n+1} \right].$$

Coming to the set S_2, we get for $j \in S_2$,

$$\begin{aligned} & E_{(0,i)}[e^{-sY^{(n)}(t)} I(J(t)=j)] \\ = & \sum_{k=0}^{q_n} \int_0^t \rho_{(0,i);(k,j)}^{(n)}(du) e^{-\theta_n c_j (t-u)} \left(\frac{\theta_n}{s+\theta_n} \right)^k \\ \to & \; c^* \theta_n \sum_{k=0}^{q_n} [\mathbf{x}_{nk}]_j \left(\frac{\theta_n}{s+\theta_n} \right)^k (2c_j\theta_n)^{-1} \quad \text{as} \quad t \to \infty, \end{aligned}$$

This, after some minor simplifications, gives us equation (46).

Finally, for $j \in S_3$, we have

$$\lim_{t\to\infty} E_{(0,i)}[e^{-sY^{(n)}(t)} I(J(t)=j)] = c^* \theta_n \sum_{k=0}^{q_n} [\mathbf{x}_{nk}]_j \left(\frac{\theta_n}{s+\theta_n} \right)^k \theta_n^{-1},$$

and this yields equation (47) after some minor simplifications. $\square$.

Remark 4 *Except possibly for a finite number of singularities (which should also be removable since the LST is analytic), the above formulae determine the LST* $\hat{\mathbf{V}}_n(\cdot)$ *in the entire right half plane. These singularities occur at eigenvalues of the matrix* $-\tilde{K}_*$ *and are the ones one would encounter if one were to take a spectral approach to the analysis.*

6 Steady State Fluid Flow

The final steps in our analysis consist of taking the limit as $n \to \infty$ in the expressions given in Theorem 9 and demonstrating that the limits indeed yield the stationary distribution of the finite buffer fluid flow. We need several preliminary results.

We will see later that in the expressions for the steady state distribution of the given finite fluid flow, we have the appearance of a certain inverse. Its existence is established in the following lemma.

Lemma 1 *Assume that $\xi_1 C_1 \mathbf{1} \neq \xi_2 C_2 \mathbf{1}$ so that at least one of the two infinite buffer fluid models corresponding to the given finite flow model or its rate reversal is stable. Then the matrix $e^{\tilde{K}_O \beta} \tilde{\Psi}_O \, e^{\tilde{K}_* \beta} \tilde{\Psi}_*$ which is a square matrix of order $|S_1|$ has spectral radius less than unity. Furthermore, the matrix*

$$[I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O \, e^{\tilde{K}_* \beta} \tilde{\Psi}_*]^{-1}$$

exists and is nonnegative.

Proof: Let $\Delta_1 = diag(\frac{1}{(\xi_1 + \xi_2)\mathbf{1}} \xi_1)$ and $\Delta_2 = diag(\frac{1}{(\xi_1 + \xi_2)\mathbf{1}} \xi_2)$ be diagonal matrices of orders $|S_1|$ and $|S_2|$ respectively. The matrices

$$U_1 = \Delta_1^{-1} (\tilde{K}_O)' \Delta_1 \quad \text{and} \quad W_1 = \Delta_2^{-1} (\tilde{\Psi}_O)' \Delta_1,$$

where $'$ denotes transpose, are precisely the U-matrix and W-matrix introduced in [11] through a time reversal argument on the infinite buffer fluid flow model associated with the given fluid model. From the interpretations there, U_1 is the generator of a Markov chain and $[e^{U_1 \beta}]_{ij}$ gives the probability that a busy period of that time reversed version ends in phase j given that it starts in state (β, i), and in addition we have shown that W_1 is row sub-stochastic. Furthermore, $e^{U_1 \beta}$ is strictly substochastic if the fluid model is stable (since that makes the reversed version unstable), and stochastic otherwise. Now, similar comments hold for the matrices

$$U_2 = \Delta_2^{-1} (\tilde{K}_*)' \Delta_2 \quad \text{and} \quad W_2 = \Delta_1^{-1} (\tilde{\Psi}_*)' \Delta_2$$

which relate to the rate reversed fluid flow. Now, if the infinite buffer fluid flow corresponding to the given model is stable, then its rate reversed version is unstable and vice versa. Thus, one of the matrices $e^{U_1 \beta}$ or $e^{U_2 \beta}$ is strictly substochastic. These entail that the matrix

$$W_2 e^{U_2 \beta} W_1 e^{U_1 \beta} W_1 = \Delta_1^{-1} [e^{\tilde{K}_O \beta} \tilde{\Psi}_O \, e^{\tilde{K}_* \beta} \tilde{\Psi}_*]' \Delta_1$$

is strictly substochastic. Hence the proof is complete. □

Remark 5 *From now on, we will assume the technical condition*

$$\xi_1 C_1 \mathbf{1} \neq \xi_2 C_2 \mathbf{1}; \tag{49}$$

see Remark 6.

The next result we need concerns the quantities defined below:

$$\begin{aligned}
\alpha_O &= (\alpha_{O1} \;\; \alpha_{O2} \;\; \alpha_{O3}) = \lim_{n\to\infty} (\alpha_{O,n1} \;\; \alpha_{O,n2} \;\; \alpha_{O,n3}) \\
\alpha_L &= (\alpha_{L1} \;\; \alpha_{L2} \;\; \alpha_{L3}) = \lim_{n\to\infty} (\alpha_{L,n1} \;\; \alpha_{L,n2} \;\; \alpha_{L,n3}).
\end{aligned}$$

Concerning these, we prove the following:

Lemma 2 *Assume $\xi_1 C_1 \mathbf{1} \neq \xi_2 C_2 \mathbf{1}$. Then,*
(a) the limit vectors α_O and α_L exist,
(b) and the vectors α_O and α_L are given by

$$\begin{aligned}
\alpha_{O1} &= c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} (-\tilde{K}_O)C_1 \\
\alpha_{O2} &= 2c^{*-1}\xi_2 C_2 - 2c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} \\
&\qquad \times \left(\tilde{\Psi}_O - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\right) C_2 \\
\alpha_{O3} &= c^{*-1}\xi_3 - c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} \\
&\qquad \times \left(\tilde{\Theta}_O - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Theta}_*\right) \\
\alpha_{L1} &= c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}(I - \tilde{\Psi}_O\tilde{\Psi}_*)C_1 \\
\alpha_{L2} &= 2c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O(-\tilde{K}_*)C_2 \\
\alpha_{L3} &= c^{*-1}\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}(\tilde{\Theta}_O - \tilde{\Psi}_O\tilde{\Theta}_*)
\end{aligned} \tag{50}$$

Proof: The involved proof of this result is given in Section 8. □

Armed with these preliminaries, we are now ready to evaluate the limit as $n \to \infty$ of the steady state distributions of $(Y^{(n)}, J)$. Immediately after that, we will prove Theorem 3 and identify the resulting limit as the stationary distribution of the finite buffer fluid flow of interest. Although the path to these formulae has been pretty long and grim, the final results are quite appealing in that we only need the K, Ψ, Θ matrices (for exactly one uniformized version), and these in turn are obtained from the R-matrices through which they were defined and for which we have in the literature extremely efficient algorithms.

Theorem 10 *Assume* $\xi_1 C_1 \mathbf{1} \neq \xi_2 C_2 \mathbf{1}$*. Then*
(a) For S_1 *and* $0 < x < \beta$*, the joint density*

$$\begin{aligned} \mathbf{v}_{F,1}(x) &= \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} (-\tilde{K}_O) e^{\tilde{K}_O x} \qquad (51) \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O \beta} \tilde{\Psi}_O (-\tilde{K}_*) e^{\tilde{K}_* (\beta - x)} \tilde{\Psi}_*. \end{aligned}$$

Furthermore, the masses at 0 *and* β *for the set* S_1 *are respectively given by*

$$\mathbf{V}_{F,1}(\{0\}) = 0, \qquad (52)$$

$$\mathbf{V}_{F,1}(\{\beta\}) = \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O \beta} (I - \tilde{\Psi}_O \tilde{\Psi}_*). \qquad (53)$$

(b) For the set S_2 *and* $0 < x < \beta$*, we have,*

$$\begin{aligned} \mathbf{v}_{F,2}(x) &= \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} (-\tilde{K}_O) e^{\tilde{K}_O x} \tilde{\Psi}_O \qquad (54) \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O \beta} \tilde{\Psi}_O (-\tilde{K}_*) e^{\tilde{K}_* (\beta - x)}. \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathbf{V}_{F,2}(\{0\}) &= \xi_2 - \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} \\ &\qquad \times (\tilde{\Psi}_O - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta}), \qquad (55) \\ \mathbf{V}_{F,2}(\{\beta\}) &= 0. \qquad (56) \end{aligned}$$

(c) For S_3 *and* $0 < x < \beta$*,*

$$\begin{aligned} \mathbf{v}_{F,3}(x) &= \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} (-\tilde{K}_O) e^{\tilde{K}_O x} \tilde{\Theta}_O \qquad (57) \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O \beta} \tilde{\Psi}_O (-\tilde{K}_*) e^{\tilde{K}_* (\beta - x)} \tilde{\Theta}_*, \end{aligned}$$

and

$$\begin{aligned} \mathbf{V}_{F,3}(\{0\}) &= \xi_3 - \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} (\tilde{\Theta}_O - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Theta}_*), \\ \mathbf{V}_{F,3}(\{\beta\}) &= \xi_1 \left(I - e^{\tilde{K}_O \beta} \tilde{\Psi}_O e^{\tilde{K}_* \beta} \tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O \beta} (\tilde{\Theta}_O - \tilde{\Psi}_O \tilde{\Theta}_*). \qquad (58) \end{aligned}$$

Proof: (a). Taking the limit as $n \to \infty$ in the equation (45) and using Lemma 3 in Section 8, we get

$$\begin{aligned} &\hat{\mathbf{V}}_{F,1}(s) \equiv \lim_{n \to \infty} \hat{\mathbf{V}}_{n,1}(s) \\ &= c^* \alpha_{O1} C_1^{-1} (sI - \tilde{K}_O)^{-1} C_1 \left[I - e^{-\beta s} e^{\beta C_1^{-1} \tilde{K}_O C_1}\right] C_1^{-1} \end{aligned}$$

$$
\begin{aligned}
+ & \quad \frac{c^*}{2}\alpha_{L2}C_2^{-1}(sI+\tilde{K}_*)^{-1}C_2\left[I-e^{-\beta s}e^{-\beta C_2^{-1}\tilde{K}_* C_2}\right]e^{\beta C_2^{-1}\tilde{K}_* C_2}C_2^{-1}\tilde{\Psi}_* \\
+ & \quad c^* e^{-\beta s}\alpha_{L1}C_1^{-1}.
\end{aligned}
$$

If we denote a cdf concentrated at y by $\delta_y(\cdot)$, then we can see that $\hat{\mathbf{V}}_{F,1}$ is the Laplace-Stieltjes function of

$$
c^*\alpha_{L1}C_1^{-1}\delta_\beta(x)+\int_0^{\min(x,\beta)}\left[c^*\alpha_{O1}C_1^{-1}e^{\tilde{K}_O y}+\frac{c^*}{2}\alpha_{L2}C_2^{-1}e^{\tilde{K}_*(\beta-y)}\tilde{\Psi}_*\right]dy.
$$

The result (a) now follows by using the expressions in Part (b) of Lemma 2.

For S_2 and S_3, we can also derive the following equations by taking the limits as $n\to\infty$ in the expressions in Theorem 9:

$$
\begin{aligned}
\hat{\mathbf{V}}_{F,2}(s) & \equiv \lim_{n\to\infty}\hat{\mathbf{V}}_{n,2}(s) \\
= & \quad \frac{c^*}{2}\alpha_{O2}C_2^{-1}+c^*\alpha_{O1}C_1^{-1}(sI-\tilde{K}_O)^{-1}C_1\left[I-e^{-\beta s}e^{\beta C_1^{-1}\tilde{K}_O C_1}\right]C_1^{-1}\tilde{\Psi}_O \\
+ & \quad \frac{c^*}{2}\alpha_{L2}C_2^{-1}(sI+\tilde{K}_*)^{-1}C_2\left[I-e^{-\beta s}e^{-\beta C_2^{-1}\tilde{K}_* C_2}\right]e^{\beta C_2^{-1}\tilde{K}_* C_2}C_2^{-1}.
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\mathbf{V}}_{F,3}(s) & \equiv \lim_{n\to\infty}\hat{\mathbf{V}}_{n,3}(s) \\
= & \quad c^*\alpha_{O3}+c^*\alpha_{L3}e^{-\beta s} \\
+ & \quad c^*\alpha_{O1}C_1^{-1}(sI-\tilde{K}_O)^{-1}C_1\left[I-e^{-\beta s}e^{\beta C_1^{-1}\tilde{K}_O C_1}\right]C_1^{-1}\tilde{\Theta}_O \\
+ & \quad \frac{c^*}{2}\alpha_{L2}C_2^{-1}(sI+\tilde{K}_*)^{-1}C_2\left[I-e^{-\beta s}e^{-\beta C_2^{-1}\tilde{K}_* C_2}\right]e^{\beta C_2^{-1}\tilde{K}_* C_2}C_2^{-1}\tilde{\Theta}_*.
\end{aligned}
$$

The above are respectively the Laplace-Stieltjes transforms of

$$
\frac{c^*}{2}\alpha_{O2}C_2^{-1}\delta_0(x)+\int_0^{\min(x,\beta)}\left[c^*\alpha_{O1}C_1^{-1}e^{\tilde{K}_O y}\tilde{\Psi}_O+\frac{c^*}{2}\alpha_{L2}C_2^{-1}e^{\tilde{K}_*(\beta-y)}\right]dy,
$$

and

$$
\begin{aligned}
& c^*\alpha_{O3}\delta_0(x)+c^*\alpha_{L3}\delta_\beta(x) \\
& \int_0^{\min(x,\beta)}\left[c^*\alpha_{O1}C_1^{-1}e^{\tilde{K}_O y}\tilde{\Theta}_O+\frac{c^*}{2}\alpha_{L2}C_2^{-1}e^{\tilde{K}_*(\beta-y)}\tilde{\Theta}_*\right]dy.
\end{aligned}
$$

From these, (b) and (c) of the theorem follow. □

Remark 6 *The case* $\xi_1 C_1 \mathbf{1} = \xi_2 C_2 \mathbf{1}$*, which is not covered by the above can be handled through the consideration of a generalized inverse of the matrix* $I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*$*; we omit the details.*

We can now prove the following result which is essentially a restatement of Theorem 3.

Theorem 11 *The expressions obtained in Theorem 10 yield the joint steady state distribution of the process* $(F(t), J(t))$ *as* $t \to \infty$*.*

Proof: The distribution functions corresponding to the density in Theorem 10 are given by

$$\begin{aligned} \mathbf{V}_{F,1}(x) &= \xi_1 - \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x} \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}\tilde{\Psi}_*. \end{aligned} \tag{59}$$

$$\begin{aligned} \mathbf{V}_{F,2}(x) &= \xi_2 - \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}\tilde{\Psi}_O \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}. \end{aligned} \tag{60}$$

$$\begin{aligned} \mathbf{V}_{F,3}(x) &= \xi_3 - \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}\tilde{\Theta}_O \\ &\quad + \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}\tilde{\Theta}_*. \end{aligned} \tag{61}$$

Let $D = diag(C_1, -C_2, 0)$, and let

$$\mathbf{V}_F(x)Q = (\mathbf{U}_1(x) \;\; \mathbf{U}_2(x) \;\; \mathbf{U}_3(x))$$

Then,

$$\begin{aligned} \mathbf{U}_1(x) &= \mathbf{V}_{F,1}(x)Q_{11} + \mathbf{V}_{F,2}(x)Q_{21}\mathbf{V}_{F,3}(x)Q_{31} \\ &= \xi_1 Q_{11} + \xi_2 Q_{21} + \xi_3 Q_{31} \\ &\quad -\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}(Q_{11} + \tilde{\Psi}_O Q_{21} + \tilde{\Theta}_O Q_{31}) \\ &\quad +\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}(\tilde{\Psi}_* Q_{11} + Q_{21} + \tilde{\Theta}_* Q_{31}). \end{aligned}$$

Since $\xi Q = 0$, it follows from Theorem 8(a, b) and Theorem 10 that

$$
\begin{aligned}
\mathbf{U}_1(x) &= -\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}\tilde{K}_O C_1 \\
&\quad +\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}(-\tilde{K}_*\tilde{\Psi}_* C_1) \\
&= \mathbf{v}_{F,1}(x)C_1.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbf{U}_2(x) &= \mathbf{V}_{F,1}(x)Q_{12} + \mathbf{V}_{F,2}(x)Q_{22}\mathbf{V}_{F,3}(x)Q_{32} \\
&= \xi_1 Q_{12} + \xi_2 Q_{22} + \xi_3 Q_{32} \\
&\quad -\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}(Q_{12} + \tilde{\Psi}_O Q_{22} + \tilde{\Theta}_O Q_{32}) \\
&\quad +\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}(\tilde{\Psi}_* Q_{12} + Q_{22} + \tilde{\Theta}_* Q_{32}) \\
&= \xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}\tilde{K}_O\tilde{\Psi}_O C_2 \\
&\quad +\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}\tilde{K}_* C_2 \\
&= -\mathbf{v}_{F,2}(x)C_2
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{U}_3(x) &= \mathbf{V}_{F,1}(x)Q_{13} + \mathbf{V}_{F,2}(x)Q_{23}\mathbf{V}_{F,3}(x)Q_{33} \\
&= \xi_1 Q_{12} + \xi_2 Q_{23} + \xi_3 Q_{32} \\
&\quad -\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O x}(Q_{13} + \tilde{\Psi}_O Q_{23} + \tilde{\Theta}_O Q_{33}) \\
&\quad +\xi_1 \left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*(\beta-x)}(\tilde{\Psi}_* Q_{13} + Q_{23} + \tilde{\Theta}_* Q_{33}) \\
&= 0.
\end{aligned}
$$

Therefore, we have

$$
\mathbf{v}_F(x)D = \frac{d}{dx}\mathbf{V}_F(x)D = \mathbf{V}_F(x)Q, \quad \text{for } x \in (0, \beta). \tag{62}
$$

By considering the state probabilities at times t and $t + \Delta t$ and letting $\Delta t \to 0$, we can get a partial differential equation for the state probabilities of the fluid model, from which after letting $t \to \infty$, a set of linear differential equations for the stationary probabilities can be obtained. These routine calculations will reveal that the equations given in (62) are indeed the differential equations governing the stationary distribution of the fluid

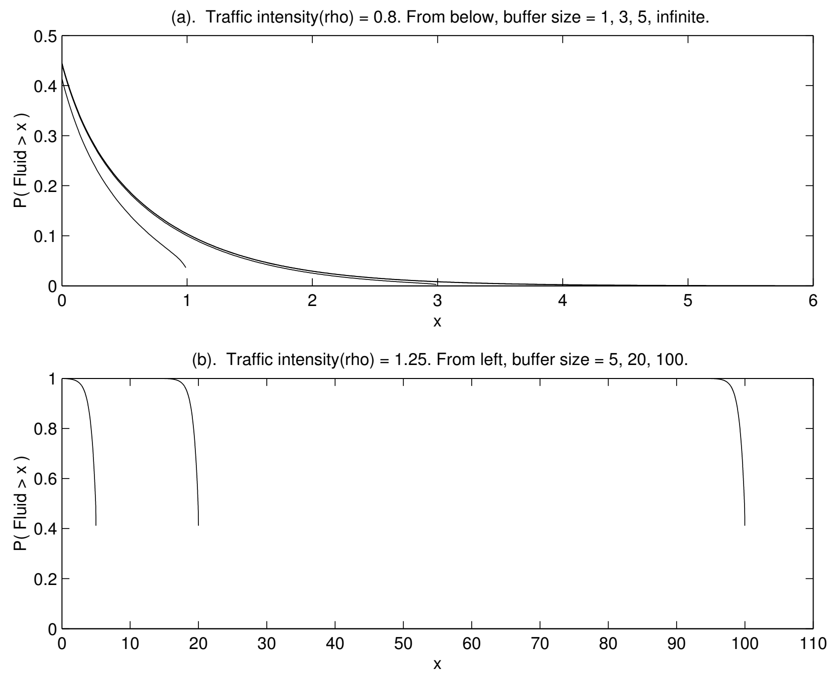

Figure 3: Complementary steady state probabilities of finite fluid flows for $m = 5$ and $\rho = 0.8,\ 1.25$.

flow model under consideration. In the course of the proof Lemma 2, we have shown that the boundary conditions at 0 and at β are also satisfied by the solution given in Theorem 10. Since by the general results on linear differential equations [5], the solution satisfying the boundary condition(s) is also unique, the proof of Theorem 11 is complete. $\square$

7 Numerical Examples

We consider a fluid flow system with m statistically independent and identical on-off input sources, a channel with constant output rate δ, and finite buffer of size of β. For each source, we assume that the on periods and the off periods form an alternating renewal process, and that their durations are exponentially distributed with mean μ^{-1} and ν^{-1}, respectively. When a source is in the on-state, it is assumed to generate fluid at rate ϑ. In addition, we shall assume that whenever all sources become idle, the system stops serving until at least one source becomes busy at which point of time it resumes with the accumulated work, if any, intact. Note that if the server stops, then there is no change of fluid levels until the time a source comes on again.

x	$\beta = 5$	$\beta = 10$	$\beta = 20$	$\beta = \infty$
0	0.4450797021	0.4452689649	0.4452693145	0.4452693145
0.50	0.1986696531	0.1989429573	0.1989434621	0.1989434621
1.00	0.1035983645	0.1039040940	0.1039046588	0.1039046588
1.50	0.0549253342	0.0552476646	0.0552482600	0.0552482600
2.00	0.0290996771	0.0294308169	0.0294314286	0.0294314286
2.50	0.0153450321	0.0156808702	0.0156814905	0.0156814905
3.00	0.0080164868	0.0083548645	0.0083554895	0.0083554895
3.50	0.0041114574	0.0044513949	0.0044520223	0.0044520223
4.00	0.0020295850	0.0023715251	0.0023721539	0.0023721539
4.50	0.0009136638	0.0012633162	0.0012639456	0.0012639456
5.00		0.0006728335	0.0006734633	0.0006734633
$\mathcal{P}[F_\beta = \beta]$	0.0001958552	0.0000003611	0.0000000000	N/A

Table 1: Tables of $\mathcal{P}[F_\beta > x]$ for $m = 5$ and $\rho = 0.8$.

x	$\beta = 5$	$\beta = 20$	$\beta = 100$
0	0.9997778325	1.0000000000	1.0000000000
1	0.9981418688	1.0000000000	1.0000000000
2	0.9918008864	1.0000000000	1.0000000000
3	0.9663936583	1.0000000000	1.0000000000
4	0.8645337214	0.9999999999	1.0000000000
5		0.9999999995	1.0000000000
10		0.9999994930	1.0000000000
18		0.9661584781	1.0000000000
19		0.8643234002	1.0000000000
20			1.0000000000
90			0.9999994930
98			0.9661584781
99			0.8643234002
$\mathcal{P}[F_\beta = \beta]$	0.3432839851	0.3432004768	0.3432004768

Table 2: Tables of $\mathcal{P}[F_\beta > x]$ for $m = 5$ and $\rho = 1.25$.

Let $J(t)$ be the number of sources in the state on at time t. Then, the phase process $\mathcal{J} = \{J(t), t \geq 0\}$ becomes a Markov process with the state space $S = \{0, \cdots, m\}$ and the infinitesimal generator Q given by

$$\begin{aligned} [Q]_{i,i+1} &= (m-i)\nu, \quad \text{for } 0 \leq i \leq m-1, \\ [Q]_{i,i-1} &= i\mu, \quad \text{for } 1 \leq i \leq m, \end{aligned}$$

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

If we think of a fluid process $\mathcal{F}_\beta = \{F_\beta(t), t \geq 0\}$ modulated by $\mathcal{J}$, then the net rate of change in the fluid level in phase $i \geq 1$ is $i\vartheta - \delta$, and for $i \geq 1$, $c_i = |i\vartheta - \delta|$. The net rate is zero when $i = 0$. We can also see that the traffic intensity ρ of the system can be represented by

$$\rho = \frac{m\vartheta\nu}{\delta(\mu+\nu)\left[1 - \left(1 + \frac{\nu}{\mu}\right)^{-m}\right]}. \tag{63}$$

For the numerical examples, we fixed the parameters as $m = 5$, $\vartheta = \mu = \nu = 1$ and consider several values of the buffer size β. Each figure in Figure 3 shows the complementary steady state probabilities of the finite flow, i.e. $\mathbf{P}[F_\beta > x]$, for $0 < x < \beta$, for $\rho = 0.8$ and 1.25 respectively. Also, we provide tables of some selected probability values.

8 Proof of Lemma 2

We begin with some quick lemmas.

Lemma 3

$$\lim_{n\to\infty} \left(\frac{\theta_n}{s+\theta_n}\right)^{q_n} = e^{-\beta s}, \tag{64}$$

$$\lim_{n\to\infty} R_{O,n11}^{q_n} = C_1^{-1} e^{\tilde{K}_O \beta} C_1, \quad \textit{and} \quad \lim_{n\to\infty} R_{L,n22}^{-q_n} = C_2^{-1} e^{-\tilde{K}_* \beta} C_2. \tag{65}$$

Proof: Note that $\theta_n = q_n/\beta$ and $q_n \to \infty$ as $n \to \infty$. Thus, the proofs can be directly obtained by the relations $R_{O,n11} = I + \theta_n^{-1} C_1^{-1} \tilde{K}_O C_1$ and $R_{L,n22}^{-1} = I - \theta_n^{-1} C_2^{-1} \tilde{K}_* C_2$. □

Lemma 4 *Let H be a $|S_1| \times |S_2|$ matrix such that for $i \in S_1$ and $j \in S_2$, its (i,j)-th element $H(i,j)$ denotes the probability that a busy period of the finite buffer fluid flow model ends in phase j given that it starts in phase i. Then* *under the assumption that the phase process defined by Q is irreducible, H is a strictly positive matrix, and furthermore the matrix*

$$M = \begin{pmatrix} P_{22} + P_{21}H & P_{23} \\ P_{32} + P_{31}H & P_{33} \end{pmatrix}$$

is irreducible.

Proof: Fix $j \in S_2$ and assume that the busy period starts in phase i for some $i \in S_1$. Since the CTMC of phases is irreducible, there is a time t at which the phase makes a first entrance into S_2, and that epoch of first entrance is in the busy period. Consider the case where the phase at this first entrance epoch is some $k \in S_2$, but $k \neq j$. Assume further that the fluid level at that epoch is $x > 0$. Since the phase process is irreducible, there is a positive probability of entering j within a time interval x/d (note $e^{Qt} > 0$ for all $t > 0$), where $d = max_{r \in S_2} c_r$, and the definition of d entails that an entrance into j is made during the busy period. Thus, j is visited in the busy period with positive probability. Now, there is a positive probability that the phase remains at j continuously to the end of the busy period. Thus, there is a positive probability that the busy period will end in j. That is, $H(i,j) > 0$ for all $i \in S_1,\ j \in S_2$. That proves the first statement.

Since $H >> 0$, there exists an $\epsilon > 0$ such that

$$H > \epsilon(I - P_{11})^{-1}P_{12} = \epsilon N,$$

where $N = (I - P_{11})^{-1}P_{12}$; the inverse here exists due to the irreducibility of P which is a consequence of the irreducibility of Q. Now, we have

$$M > \begin{pmatrix} P_{22} + \epsilon P_{21}N & P_{23} \\ P_{32} + \epsilon P_{31}N & P_{33} \end{pmatrix}.$$

The nonnegative matrix on the right of this inequality is irreducible since the matrix

$$\begin{pmatrix} P_{22} + P_{21}N & P_{23} \\ P_{32} + P_{31}N & P_{33} \end{pmatrix},$$

being the embedded Markov chain on $S_2 \cup S_3$ in the uniformized phase process, is irreducible. That completes the proof of the lemma. □

Corollary 1 *The CTMC with generator*

$$A = \begin{pmatrix} Q_{22} + Q_{21}H & Q_{23} \\ Q_{32} + Q_{31}H & Q_{33} \end{pmatrix}$$

is irreducible.

Proof The result follows from the fact that uniformization by λ of this CTMC yields a discrete time Markov chain with transition matrix M. $\square$

Proof of Lemma 2

(a): Consider the CTMCs of phases on the set $S_2 \cup S_3$ obtained by restricting $(Y^{(n)}(t), J(t))$ and $(F(t), J(t))$ on the set where their first coordinates are zero. (These processes are obtained by excising out from the respective processes the intervals of time during which their first coordinates are non-zero and "gluing" the remaining pieces together to get a right continuous process.) Let the infinitesimal generators of these processes be given respectively by A_n and A. By Theorem 2, we have

$$lim_{n\to\infty}\mathcal{P}[Y^{(n)}(t)=0,\ J(t)=j]=\mathcal{P}[F(t)=0,\ J(t)=j],\ \text{ for all } t\in[0,\infty).$$

This entails that $e^{A_n t} \to e^{At}$ as $n \to \infty$ for all t. Therefore, we must have $A_n \to A$ as $n \to \infty$. Now, A is the matrix in Corollary 1 and since A is irreducible by that Corollary, we can assert the convergence of the stationary distributions of these Markov chains, and claim that for $j \in S_2 \cup S_3$,

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

We also note that the above equality holds also for $j \in S_1$ since (as can be easily verified by direct computations) all the limits there as $t \to \infty$ are zero. Now,

$$\mathcal{P}[Y^{(n)}(\infty)=0,J(\infty)=j]=\lim_{s\to\infty}[\hat{\mathbf{V}}_{n,i}(s)]_j.$$

Evaluating the right side of the above from the expressions in Theorem 9, we get

$$\lim_{s\to\infty}[\hat{\mathbf{V}}_{n,2}(s)]=\frac{c^*}{2}\alpha_{O,n2}C_2^{-1}\qquad\text{and}$$
$$\lim_{s\to\infty}[\hat{\mathbf{V}}_{n,3}(s)]=c^*\alpha_{O,n3}.$$

Therefore, the vectors α_{O2} and α_{O3} exist, and

$$\begin{aligned}[(c^*/2)\alpha_{O2}C_2^{-1}]_j &= \mathcal{P}[F(\infty)=0,J(\infty)=j],\quad j\in S_2\\ [c^*\alpha_{O3}]_j &= \mathcal{P}[F(\infty)=0,J(\infty)=j],\quad j\in S_3.\end{aligned}\tag{66}$$

Now, we will show that the vectors α_{O1} and α_{L2} exist. If we define

$$\Psi_O=C_1^{-1}\tilde{\Psi}_O C_2,\Theta_O=C_1^{-1}\tilde{\Theta}_O,\Psi_*=C_2^{-1}\tilde{\Psi}_* C_1,\Theta_*=C_2^{-1}\tilde{\Theta}_*,\tag{67}$$

then, these sets of matrices are none other than the matrices $\Psi(0+)$ and $\Theta(0+)$ introduced in [11] and [2]. Moreover, under the assumption $\xi_1 C_1 \mathbf{1} \neq$

$\xi_2 C_2 \mathbf{1}$, one of the matrices Ψ_O and Ψ_* becomes substochastic and the other becomes a stochastic matrix and this yields that the matrices $I_{|S_2|} - \Psi_* \Psi_O$ and $I_{|S_1|} - \Psi_O \Psi_*$ are both invertible. Now if we multiply θ_n on the equation (23), then we can get

$$\begin{aligned} \alpha_{O,n1} &= \frac{1}{2}\alpha_{O,n2} C_2^{-1} Q_{21} + \alpha_{O,n3} Q_{31} \\ &\quad + \alpha_{L,n2} R_{L,n22}^{q_n - 1} \left[-\frac{1}{2}\Psi_* + \frac{1}{2\theta_n} R_{L,n22} C_2^{-1} Q_{21} + \frac{1}{\theta_n} R_{L,n22} \Theta_* Q_{31} \right] \end{aligned} \tag{68}$$

and

$$\begin{aligned} & -\frac{1}{2}\alpha_{O,n2} C_2^{-1}(Q_{21}\Psi_O + Q_{22}) - \alpha_{O,n3}(Q_{31}\Psi_O + Q_{32}) \\ = {} & \alpha_{L,n2} R_{L,n22}^{q_n - 1} \left[\frac{1}{2} - \frac{1}{2}\Psi_*\Psi_O + \frac{1}{2\theta_n} R_{L,n22} C_2^{-1}(Q_{21}\Psi_O + Q_{22}) \right. \\ & \left. + \frac{1}{\theta_n} R_{L,n22} \Theta_* (Q_{31}\Psi_O + Q_{32}) \right]. \end{aligned} \tag{69}$$

If we let $n \to \infty$, then we can obtain from Lemma 3 and the equations (31), (32) and Theorem 6 that

$$\begin{aligned} & \frac{1}{2}(\lim_{n\to\infty} \alpha_{L,n2}) C_2^{-1} e^{\tilde{K}_* \beta} C_2 [I - \Psi_* \Psi_O] \\ & = -\frac{1}{2}\alpha_{O2} C_2^{-1}(Q_{21}\Psi_O + Q_{22}) - \alpha_{O3}(Q_{31}\Psi_O + Q_{32}). \end{aligned}$$

In the above, the existence of the inverse matrix $(I - \Psi_*\Psi_O)^{-1}$ implies the existence of α_{L2} which is the limit of $\alpha_{L,n2}$. Now, the existence of α_{O1} can be directly derived from the equation (68).

The proof of the existence of α_{L1} and α_{L3} is similar and is obtained by considering the CTMC of phases obtained by restricting the queueing process to the set $Q^{(n)}(t) = q_n$ which converge to the CTMC of phases obtained by restricting to $F(t) = \beta$, and we omit the details. In addition, it can be seen that

$$\begin{aligned} [c^* \alpha_{L1} C_1^{-1}]_j &= \mathcal{P}[F(\infty) = \beta, J(\infty) = j], \quad j \in S_1 \\ [c^* \alpha_{L3}]_j &= \mathcal{P}[F(\infty) = \beta, J(\infty) = j], \quad j \in S_3. \end{aligned} \tag{70}$$

(b): If we multiply by θ_n both sides of the equation (23) and take limits with respect to n, then we can get

$$0 = -\alpha_{O1} + 2^{-1}\alpha_{O2} C_2^{-1} Q_{21} + \alpha_{O3} Q_{31} - 2^{-1}\alpha_{L2} C_2^{-1} e^{\tilde{K}_* \beta} \tilde{\Psi}_* C_1$$

$$
\begin{aligned}
0 &= \alpha_{O1}C_1^{-1}\tilde{\Psi}_O C_2 + 2^{-1}\alpha_{O2}C_2^{-1}Q_{22} + \alpha_{O3}Q_{32} + 2^{-1}\alpha_{L2}C_2^{-1}e^{\tilde{K}_*\beta}C_2 \\
0 &= 2^{-1}\alpha_{O2}C_2^{-1}Q_{23} + \alpha_{O3}Q_{33} \qquad (71) \\
0 &= \alpha_{O1}C_1^{-1}e^{\tilde{K}_O\beta}C_1 + \alpha_{L1}C_1^{-1}Q_{11} + 2^{-1}\alpha_{L2}C_2^{-1}\tilde{\Psi}_* C_1 + \alpha_{L3}Q_{31} \\
0 &= -\alpha_{O1}C_1^{-1}e^{\tilde{K}_O\beta}\tilde{\Psi}_O C_2 + \alpha_{L1}C_1^{-1}Q_{12} - 2^{-1}\alpha_{L2} + \alpha_{L3}Q_{32} \\
0 &= \alpha_{L1}C_1^{-1}Q_{13} + \alpha_{L3}Q_{33}.
\end{aligned}
$$

Besides, from the equation $\zeta = \sum_{k=0}^{q_n} \mathbf{x}_{nk}$ in Theorem 5 and (22), we can also get

$$
\begin{aligned}
\zeta_1 &= \eta_{O,n1}\sum_{k=0}^{q_n} R_{O,n11}^k + \eta_{L,n2}\sum_{k=0}^{q_n-1} R_{L,n22}^{q_n-i-1}R_{L,n21} + \eta_{L,n1} \\
\zeta_2 &= \eta_{O,n1}\sum_{k=0}^{q_n-1} R_{O,n11}^k R_{O,n12} + \eta_{O,n2} + \eta_{L,n2}\sum_{k=0}^{q_n} R_{L,n22}^{q_n-k} \\
\zeta_3 &= \eta_{O,n1}\sum_{k=0}^{q_n-1} R_{O,n11}^k R_{O,n13} + \eta_{O,n3} + \eta_{L,n2}\sum_{k=0}^{q_n} R_{L,n22}^{q_n-k-1}R_{L,n23} + \eta_{L,n3}.
\end{aligned}
$$

At this time, assume that $\xi_1 C_1 \mathbf{1} < \xi_2 C_2 \mathbf{1}$ so that $-\tilde{K}_O^{-1}$ exists. Now, if we define $L_\beta = \lim_{n\to\infty}\theta_n^{-1}\sum_{k=0}^{q_n} R_{L,n22}^{-k}$ and take the limit of the above equation w.r.t n, then we can get

$$
\begin{aligned}
\zeta_1 &= \alpha_{O1}C_1^{-1}(-\tilde{K}_O)^{-1}(I - e^{\tilde{K}_O\beta})C_1 + 2^{-1}\alpha_{L2}L_\beta C_2^{-1}e^{\tilde{K}_*\beta}\tilde{\Psi}_* C_1 + \alpha_{L1} \\
\zeta_2 &= \alpha_{O2} + 2C_1^{-1}(-\tilde{K}_O)^{-1}(I - e^{\tilde{K}_O\beta})\tilde{\Psi}_O C_2 + \alpha_{L2}L_\beta C_2^{-1}e^{\tilde{K}_*\beta}C_2 \qquad (72) \\
\zeta_3 &= \alpha_{O3} + \alpha_{O1}C_1^{-1}(-\tilde{K}_O)^{-1}(I - e^{\tilde{K}_O\beta})\tilde{\Theta}_O + 2^{-1}\alpha_{L2}L_\beta C_2^{-1}e^{\tilde{K}_*\beta}\tilde{\Theta}_* + \alpha_{L3},
\end{aligned}
$$

and L_β satisfies $I - C_2^{-1}\tilde{K}_* C_2 L_\beta = C_2^{-1}e^{-\tilde{K}_*\beta}C_2$.

Piecing together the equations (71), (72), (39), (40) and the fact that ζ is the stationary probability vector of $C^{-1}\Delta^{-1}Q$ (see (b) in Theorem 5), the following equations can be derived by some laborious elimination techniques.

$$
\begin{aligned}
\alpha_{O1} &= \zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1}(-\tilde{K}_O)C_1 \\
\alpha_{O2} &= \zeta_2 - 2\zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} \\
&\qquad \times\left(\tilde{\Psi}_O - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\right)C_2 \\
\alpha_{O3} &= \zeta_3 - \zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} \\
&\qquad \times\left(\tilde{\Theta}_O - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Theta}_*\right)
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{L1} &= \zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}(I - \tilde{\Psi}_O\tilde{\Psi}_*)C_1 \qquad (73)\\
\alpha_{L2} &= 2\zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}\tilde{\Psi}_O(-\tilde{K}_*)C_2\\
\alpha_{L3} &= \zeta_1 C_1^{-1}\left(I - e^{\tilde{K}_O\beta}\tilde{\Psi}_O e^{\tilde{K}_*\beta}\tilde{\Psi}_*\right)^{-1} e^{\tilde{K}_O\beta}(\tilde{\Theta}_O - \tilde{\Psi}_O\tilde{\Theta}_*)
\end{aligned}
$$

The equations in (50) can be easily derived using $\zeta = \xi\Delta C/(\xi\Delta C\mathbf{1}) = c^{*-1}\xi\Delta C$. (See Theorem 5.)

In a similar fashion, in the case $\xi_1 C_1\mathbf{1} > \xi_2 C_2\mathbf{1}$, we exploit the fact that $\tilde{K}_*$ is invertible, define the matrix $L_\beta = \lim_{n\to\infty}\theta_n^{-1}\sum_{k=0}^{q_n} R_{O,n11}^k$, and proceed exactly in an analogous manner to get the equations in (64), and the rest of the arguments follow. □

Acknowledgment: Soohan Ahn and Jongwoo Jeon were supported (partially) by KOSEF (Korea Science & Engineering Foundation) through the Statistical Research Center for Complex Systems at Seoul National University. Soohan Ahn thanks AT&T for its hospitality and post-doctoral support during the period when this work was done. The authors thank Ward Whitt for many helpful discussions, and the referees and editors for many incisive comments which improved the rigor of some proofs.

References

[1] Ahn, S & Ramaswami, V (2003), Fluid flow models & queues - A connection by stochastic coupling, *Stochastic Models*, 19.3, 325-348.

[2] Ahn, S. & Ramaswami, V. (2004) : Transient analysis of fluid flow models via stochastic coupling to a queue, 2003, *Stochastic Models*, 20.1, 71-101.

[3] Anick, D, Mitra, D & Sondhi, M.M (1982), Stochastic theory of a data handling system with multiple sources, *Bell Syst. Tech. J.*, 61, 1871-1894.

[4] Çinlar (1975): *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, New Jersey.

[5] Coddington, E.A. & Levinson, N (1955), *Theory of ordinary differential equations*, Mc-Graw Hill, New York.

[6] B. Hajek (1982), Birth-and-Death processes on the integers with phases and general boundaries. *J. Appl. Prob.* V 19, 488-499.

[7] Halfin, S & Whitt, W (1981): Heavy-traffic limits for queues with many exponential servers, *Operations Research*, **29**, pp. 567-588.

[8] Latouche, G & Ramaswami, V (1999): *Introduction to matrix analytic methods in Stochastic Modeling,* SIAM & ASA, Philadelphia.

[9] Neuts, M.F (1981): *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore, MD.

[10] Nielsen, B.F. & Ramaswami, V (1997): A computational framework for quasi birth and death processes with a continuous phase variable, in *Teletraffic Contributions for the Information Age: Proc. of the 16th International Teletraffic Congress,* V. Ramaswami & P.E. Wirth (Eds.), Elsevier, pp. 477–486.

[11] Ramaswami, V (1999): Matrix analytic methods for stochastic fluid flows, in *Teletraffic Engineering in a Competitive World - Proc. of the 16th International Teletraffic Congress,* D. Smith and P. Key (eds.), pp. 1019–1030, Elsevier.

[12] Skorohod, A.V (1956): Limit theorems for stochastic processes, *Theor. of Probability Appl.*, 1, pp. 261-290. (Republished also in *Skorohod's Ideas in Probability Theory*, V. Korolyuk, N. Portenko & H. Syta (eds.), Institute of Mathematics of the National Academy of Sciences of the Ukraine, Kyiv, Ukraine, 2000, pp. 23-52.)

[13] da Silva Soares, A & Latouche, G (2003): Matrix-analytic methods for fluid queues with finite buffers, Technical Report TR-509, Universite Libre de Bruxelles, Belgium.

[14] Tweedie, R.L (1982): Operator-geometric stationary distributions for Markov chains with application to queueing models, *Adv. Appl. Prob.,* 14, 369-391.

[15] Whitt, W (2002): *Stochastic Process Limits - An Introduction to Stochastic-Process Limits and their Application to Queues*, Springer.