Matrix Analytic Methods for Stochastic Fluid Flows

V. Ramaswami, AT&T Labs, 200 Laurel Avenue D5-3B22, Middletown, NJ 07748

We present an analysis of stochastic fluid flow models along the lines of matrix-analytic methods. The computation of the steady state distribution is reduced to the analysis of a *discrete time, discrete state space* quasi-birth-death model. This approach, which can be extended immediately to the infinite dimensional case, may enable the inclusion of heavy tails and self similarity in fluid models.

1. INTRODUCTION

The canonical model considered here is a stochastic fluid flow whose input rates are controlled by a continuous time, irreducible Markov chain on a finite state space $\{1,\dots,m,m+1,\dots,m+n\}$ and infinitesimal generator *T*. The net rate of input to the infinite fluid buffer is assumed to be $c_i > 0$ when the Markov chain is in state *i*, $1 \le i \le m$, and $-c_j < 0$ when the Markov chain is in state *j*, $m+1 \le j \le m+n$ and the fluid level is nonempty. Special cases have been considered in the literature by many authors [1],[3],[4],[9],[11],[15],[25],[26] for modeling data communication channels and have become a mainstay in the analysis of high speed communication networks.

The mathematical model is a continuous state space analogue of the discrete state space Quasi Birth-and-Death Process (QBD) [16],[21],[22] which can be analyzed efficiently using matrix analytic methods [12],[5]. We demonstrate that the matrix geometric approach of the QBD extends in a natural manner to the model at hand yielding a matrix-exponential form for the steady state distribution which can further be characterized as phase type. Furthermore, it is shown that computations can be accomplished via a related discrete time discrete state space QBD.

We have reason to believe that these methods will yield greater numerical accuracy and stability than spectral methods; for example, in the case of the Daigle-Lucantoni model [8], a discrete state space analogue whose spectral solution closely resembles the one for fluid flows and even requires only the solution of quadratic equations to determine the eigenvalues, comparisons in [12],[14] have shown that spectral methods lose accuracy for nearly saturated queues. Furthermore, our methods can provide transient results for fluid flows and characterization of upward level crossing times which are all important in evaluating queue control strategies.

The phase type characterization of the steady state fluid flow distribution is not new and has been obtained by Asmussen [3],[4] using the concept of Siegmund duality [24]. Indeed, a vast literature, unnoticed in the ongoing work on fluid flows in the communications context exists, and we refer to Asmussen [3] (end of Section 1) for a bibliography. While some of our results may bear a close resemblance to what is obtained by Asmussen and some of our equations are indeed identical, our conditioning arguments are quite different and simpler. Instead of conditioning on the maximum level of the fluid in a busy period, we are conditioning on the last epoch of increase in a busy period; this is a technique which we have found [18] valuable for matrix-geometric models. The duality invoked here is similar to the simpler dual process introduced by Ramaswami [19] for the analogous discrete state space problem; see also [2]. The connection to discrete state space, discrete time QBDs is quite new.

2. THE MATRIX-EXPONENTIAL FORM

Denote by $X(t)$ the fluid level and by $J(t)$ the state of the modulating Markov chain (hereafter to be called level and phase respectively) at time $t+$. Throughout, we assume this process to be ergodic. The familiar starting point [1],[26] is the partial differential equation

$$\frac{\partial}{\partial t} f_j(t,x) = \sum_{i=1}^{m+n} f_i(t,x) T(i,j) + \gamma_j \frac{\partial}{\partial x} f_j(t,x), \quad x>0,$$

where $\gamma_j = -c_j$ for $j \le m$ and $+c_j$ for $j \ge m+1$, for the density terms

$$f_j(t,x)\ dx = P[x<X(t) \le x+dx,\ \ J(t)=j],$$

which leads, upon letting $t \to \infty$, to the steady state equations

$$\gamma_j \frac{d}{dx} \pi(x,j) + \sum_{i=1}^{m+n} \pi(x,i) T(i,j) = 0, \quad x>0, \quad 1 \le j \le m+n.$$

The coefficient matrix of the above system for $\pi(x,j)$, the steady state probability of state (x,j), can have eigenvalues with nonnegative real part and that poses significant difficulties [1]. Indeed, this is not surprising given that this system is the analogue of the equations $\boldsymbol{\pi}_{n-1} A_0 + \boldsymbol{\pi}_n A_1 + \boldsymbol{\pi}_{n+1} A_2 = 0$ that govern (a doubly infinite) QBD corresponding to the $M/M/1$-queue in a random environment where some environment states can lead to only arrivals and others only to departures; its solutions are not well behaved. In the QBD case with a reflecting boundary, matrix-analytic methods help to avoid this problem by considering appropriate taboo processes avoiding the boundary, and that is the approach to be adopted below. The methods we mimic are those in [21] which rely on level crossings, Markov renewal theory and a certain duality introduced by Ramaswami [19] for the QBD case. A formal presentation will require us to use results on semi-Markov processes on arbitrary spaces, but details of those technicalities will not be presented here.

Assume that $X(0)=0$ and $J(0)=i$. Since the fluid level process is skip-free upward, for $x,y>0$, the fluid level can reach $x+y$ only after it has crossed x. This leads, by conditioning on the last epoch of crossing the fluid level x, to the equation

$$f_j(t,x+y) = \int_0^t \sum_{i=1}^{m+n} f_i(t-\tau,x) [\phi(\tau,x,x+y)]_{ij}\ d\tau, \quad x,y>0,$$

where $[\phi(\tau,x,x+y)]_{ij}$ is the density of being at $(x+y,j)$ at time τ avoiding the set $L(x) = [0,\ x] \times \{1,...,m+n\}$ in the interval $(0,\ \tau)$, given that the process starts in (x,i). Upon letting $t \to \infty$, we then see that the steady state density $\boldsymbol{\pi}(x)$ must, in matrix notation, satisfy the equation

$$\boldsymbol{\pi}(x+y) = \boldsymbol{\pi}(x)\Phi(x,x+y), \quad x,y>0, \tag{2.1}$$

where $\Phi(x,x+y) = \int_0^\infty \phi(\tau,x,x+y)\,d\tau$ records the expected number of visits to level $x+y$ before returning to level x given that the process starts in level x. Note now that the spatial homogeneity of the model away from the boundary level 0 (the flow rates and transition rates do not depend on the level variable x for $x>0$) implies that

$$\Phi(x,x+y) = \Phi(0,y). \tag{2.2}$$

The skip-free upward property also implies that, for $x,y>0$, the taboo process avoiding level 0 cannot reach $x+y$ before reaching x, whence, again by conditioning on the last epoch of visit to level x, we can write in matrix form

$$\phi(t,0,x+y) = \int_0^t \phi(t-u,0,x)\ \phi(u,0,y)\ du,$$

which immediately yields

$$\Phi(0,x+y) = \Phi(0,x)\Phi(0,y), \quad x,y>0. \tag{2.3}$$

We now assert that for $y>0$, the last n rows of $\Phi(0,y)$ are zero; this is so, because the process enters a negative level immediately upon starting, if it starts off in phases $m+1,\ldots,m+n$. Taking note of this, we partition

$$\Phi(0,y) = \begin{bmatrix} \Phi_{11}(y) & \Phi_{12}(y) \\ 0 & 0 \end{bmatrix}, \quad y>0.$$

Now, (2.3) implies

$$\Phi_{11}(x+y) = \Phi_{11}(x)\Phi_{11}(y), \qquad \Phi_{12}(x+y) = \Phi_{11}(x)\Phi_{12}(y). \tag{2.4}$$

Thus,

$$\Phi_{11}(x) = e^{Kx}, \quad \Phi_{12}(x) = e^{Kx}\Psi, \tag{2.5}$$

where $\Psi = \Phi_{12}(0+)$. We have thus established the following theorem in which we partition the steady state probability density $\boldsymbol{\pi}(x)$ as $\boldsymbol{\pi}(x) = [\boldsymbol{\pi}_1(x),\ \boldsymbol{\pi}_2(x)]$, where $\boldsymbol{\pi}_1(\cdot)$ of order m gives the density of states of the form (x,i), $i \le m$ and $\boldsymbol{\pi}_2(\cdot)$ of order n gives the density of states of the form (x,i) with $i>m$.

Theorem 2.1: There exists a matrix K of order $m\times m$ and a matrix Ψ of order $m\times n$ such that the steady state density vector of the fluid flow model, when it exists, is of the form

$$[\boldsymbol{\pi}_1(x),\ \boldsymbol{\pi}_2(x)] = \boldsymbol{\alpha}[e^{Kx},\ \ e^{Kx}\Psi], \quad x>0. \tag{2.6}$$

The (i,j)-th element of the matrix e^{Kx} records the expected number of visits to (x,j), $1 \le j \le m$ before returning to level 0 given that the process starts in $(0,i)$, $1 \le i \le m$. The (i,j)-th element of $e^{Kx}\Psi$ gives the expected number of visits to (x,j) with $m+1 \le j \le m+n$ before a return to level 0 given that the process starts in $(0,i)$ with $1 \le i \le m$.

Remarks: (a) The operators $\Phi(0,x)$, $x>0$ play the role of the matrix sequence $\{R^n,\ n\ge 1\}$ in the matrix geometric solution; $-K^{-1}$ is the analogue of the matrix $(I-R)^{-1}$ and is finite and has finite row sums; (b) the existence (and continuity/differentiability) of the densities above is established by appealing to standard constructions; (c) the proof of the finiteness of $\Phi(0,\cdot)$ is similar to that of R in [21] using the facts governing the total number of renewals in a terminating renewal process; (d) just as in [21], the conditioning arguments on the last epochs used above can be justified via Markov renewal theory (for semi-Markov processes on general state spaces).

Corollary 2.1: We have $\boldsymbol{\alpha} = -\boldsymbol{\pi}_1 K$, where $(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ is the steady state probability vector of the Markov process with infinitesimal generator T.

Proof: Clearly, for $1 \le j \le m$, the steady state probability of $(0,j)$ is zero since in j, the fluid level is increasing. Therefore, integrating the first m components of (2.6) should give us $\boldsymbol{\pi}_1$, the steady state probabilities for the phases. Thus, $-\boldsymbol{\alpha} K^{-1} = \boldsymbol{\pi}_1$.

Corollary 2.2: Let β_{m+i} denote the steady state probability that the fluid level is zero and the phase is $m+i$. The vector $\boldsymbol{\beta} = (\beta_{m+1}, \ldots, \beta_{m+n})$ is given by $\boldsymbol{\beta} = \boldsymbol{\pi}_2 - \boldsymbol{\pi}_1 \Psi$.

Proof: The result is obtained by evaluating the steady state probabilities $\boldsymbol{\pi}_2$ for the last n phases as the integral of the last n components of (2.6) plus $\boldsymbol{\beta}$.

Thus, the evaluation of the steady state distribution reduces to the determination of K and Ψ which we take up next.

Throughout the following, matrices A of order $m+n$ shall be partitioned into $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, where A_{11} is of order $m \times m$. When $A = diag(a_1, \ldots, a_{m+n})$, we shall denote $A_{11} = A_1$ and $A_{22} = A_2$; of course, A_{12} and A_{21} are zero matrices if A is diagonal. Not to confuse the partitioned matrices with individual elements, the (i,j)-th element of a matrix A will be denoted by $[A]_{ij}$. The acceptance of this uniform convention throughout avoids having to repeat many identical definitions of notations. Throughout, we shall denote by C the diagonal matrix $C = diag(c_1, \ldots, c_{m+n})$.

Lemma 2.1: For $x > 0$, we have

$$\frac{\partial}{\partial t}\phi_{11}(t,0,x) = \phi_{11}(t,0,x)T_{11} + \phi_{12}(t,0,x)T_{21} - \frac{\partial}{\partial x}\phi_{11}(t,0,x)C_1, \tag{2.7}$$

$$\frac{\partial}{\partial t}\phi_{12}(t,0,x) = \phi_{11}(t,0,x)T_{12} + \phi_{12}(t,0,x)T_{22} + \frac{\partial}{\partial x}\phi_{12}(t,0,x)C_2. \tag{2.8}$$

Proof: Equations (2.7), (2.8) are indeed the Kolmogorov differential equations.

Theorem 2.2: We have

$$K = (T_{11} + \Psi T_{21})C_1^{-1}. \tag{2.9}$$

and

$$\Psi = \int_0^\infty e^{Ky} T_{12} C_2^{-1} e^{T_{22} C_2^{-1} y} \, dy. \tag{2.10}$$

Proof: Integrating (2.7) over t in 0 to ∞, we get

$$\Phi_{11}(x) T_{11} - \frac{\partial}{\partial x} \Phi_{11}(x) C_1 + \Phi_{12}(x) T_{21} = 0, \quad x > 0,$$

since for fixed $x > 0$, $\phi_{11}(t,0,x) \to 0$ as $t \to 0$ and as $t \to \infty$. Using (2.5) in the above equation immediately gives (2.9).

Now, integrating (2.8) over t in 0 to ∞ and using (2.5), we get

$$T_{12} + \Psi T_{22} + K \Psi C_2 = 0. \tag{2.11}$$

Multiplying both sides of (2.10) on the left by K and then performing an integration by parts on the right will show immediatley that Ψ given by (2.10) indeed is a solution of the above equation.

Remark: We have used the fact that for all $x > 0$, $\phi_{1i}(\infty,0,x) = 0$, $\phi_{1i}(0+,0,x) = 0$ for i=1,2; the first is a consequence of the finiteness of $\Phi_{1i}(x)$ while the second is a consequence of the obvious fact that $[\phi_{1i}(t,0,x)]_{ij} = 0$ for $t < x/(\max_{1 \le i \le m} c_i)$. Finally, the system (2.11) can be shown to be a nonsingular system (using Kronecker products). We can also get (2.10) directly by an argument conditioning on the level y from which the process leaves the set of phases $\{1,...,m\}$ for the last time in its return to level 0.

A simple minded scheme for computing K then would be to iterate on the equations (2.9), (2.10) starting with, say, $K = T_{11} C_1^{-1}$. This would be analogous to the scheme of Sengupta [23] for the *GI/PH/*1 queue and can be shown to converge to the matrix K of interest to us. But there are better techniques that avoid having to deal with possibly negative quantities, numerical integrations, etc., as we shall show later. Our discussion will also shed further probabilistic insights into the nature of K as well as of the algorithms to be proposed.

3. A DUALITY RESULT

Recall that $\boldsymbol{\pi}$ is the steady state probability vector of the Markov chain governed by T. Associated with the fluid flow defined by the Markov chain with generator T and flows of rate $+c_i$ in states $i \le m$ and $-c_i$ in states $i > m$, we can associate another flow as follows:

Let the diagonal matrix $\Delta = diag(\pi_1, ..., \pi_{m+n})$. We define $\tilde{T} = \Delta^{-1} T' \Delta$, where T' is the transpose of T, and recognize $\tilde{T}$ to be the infinitesimal generator of a Markov chain which is indeed the time reversed version of the chain governed by T. Associated with $\tilde{T}$, we now consider a fluid flow such that the net rate in state i is equal to $-c_i$ for $1 \le i \le m$, and the net rate in state j for $j > m$ is $+c_j$. This fluid flow can be called a dual of the flow model considered in the previous sections (which we shall henceforth call the primal) and is the analogue of the dual defined by Ramaswami [19] for matrix-geometric models in that (without reflection at 0) they can be viewed as the time reversed versions of each other; see [2] for details of the analogous discrete state space case.

With the conventions established for partitioning, we note the relationships

$$\tilde{T}_{ij} = \Delta_i^{-1} T'_{ji} \Delta_j \tag{3.1}$$

The dual model can be analyzed from scratch along parallel lines, but it is much cheaper to appeal to the duality. To save space, we take a direct algebraic approach, although the steps are well rooted in probabilistic methods involving time reversal. We start by defining for $x > 0$, the matrix $\begin{bmatrix} \Omega_{11}(x) & 0 \\ \Omega_{21}(x) & 0 \end{bmatrix}$ of order $(m+n)\times(m+n)$ by the equation

$$\begin{bmatrix} \Omega_{11}(x) & 0 \\ \Omega_{21}(x) & 0 \end{bmatrix} = \Delta^{-1} \begin{bmatrix} \Phi_{11}(x) & \Phi_{12}(x) \\ 0 & 0 \end{bmatrix}' \Delta \tag{3.2}$$

Then we have the following theorem.

Theorem 3.1: We have

$$\Omega_{11}(x) = e^{Ux}, \tag{3.3}$$

where $U = \Delta_1^{-1} K' \Delta_1$ satisfies the equation

$$U = C_1^{-1} \tilde{T}_{11} + C_1^{-1} \tilde{T}_{12} \int_0^{\infty} e^{C_2^{-1} \tilde{T}_{22} y} C_2^{-1} \tilde{T}_{21} e^{Uy} \, dy \tag{3.4}$$

Furthermore,

$$\Omega_{21}(x) = W \Omega_{11}(x), \tag{3.5}$$

where $W = \Delta_2^{-1} \Psi' \Delta_1$ satisfies the equation

$$W = \int_0^{\infty} e^{C_2^{-1} \tilde{T}_{22} y} C_2^{-1} \tilde{T}_{21} e^{Uy} \, dy. \tag{3.6}$$

Proof: Equation (3.3) with $U = \Delta_1^{-1} K' \Delta_1$ is obtained from the first equation in (2.5). Equation (3.4) is obtained from (2.9),(2.10); take transposes in them and multiply both sides on the left by Δ_1^{-1} and on the right by Δ_1. With the matrix W defined as $\Delta_2^{-1} \Psi' \Delta_1$, we get (3.6) by transposing both sides of the second equation in (2.5) and both sides of (2.10) and supplying the premultipliers and postmultipliers as needed.

The following theorem is important for the interpretation of the quantities Ω_{ij} introduced above. We first define the matrices $G_{11}(x)$ of order $m\times m$ and $G_{21}(x)$ of order $n\times m$ such that $[G_{11}(x)]_{ij}$ is the probability that the dual's fluid level ultimately becomes zero with the phase of the dual Markov chain at j, given that it starts off with fluid level x and phase i, and $G_{21}(x)$ records the emptiness probabilities when the process starts with fluid level x and phase $i > m$.

In the following, we set $\tilde{S} = C^{-1} \tilde{T}$, where $C = diag(c_1, \ldots, c_{m+n})$. We note that $\tilde{S}$ is the infinitesimal generator of a continuous time Markov chain.

Theorem 3.2: $G_{11}(x)$ and $G_{21}(x)$ are respectively the minimal nonnegative solutions of the equations

$$G_{11}(x) = e^{\tilde{S}_{11}x} + \int_0^x e^{\tilde{S}_{11}(x-t)} \tilde{S}_{12} \int_0^\infty e^{\tilde{S}_{22}y} \tilde{S}_{21} G_{11}(t+y)\ dy\ dt \tag{3.7}$$

$$G_{21}(x) = \int_0^\infty e^{\tilde{S}_{22}t} \tilde{S}_{21} G_{11}(x+t)\ dt. \tag{3.8}$$

Also, $G_{11}(x) = \Omega_{11}(x)$ and $G_{21}(x) = \Omega_{21}(x)$.

Proof: An elementary level crossing argument using the skip-free downward property of the fluid level will show that $G_{11}(x)$ is a matrix exponential. Assuming a LIFO draining of the fluid, Equation (3.7) is obtained routinely by conditioning on the fluid level t at the epoch of the first increase and the amount y of increase that results in the fluid level before it empties out. Multiplying both sides of (3.7) by $e^{-\tilde{S}_{11}x}$ and then differenting with respect to x, we get the integro differential equation

$$\frac{d}{dx} G_{11}(x) = \tilde{S}_{11} G_{11}(x) + \tilde{S}_{12} \int_0^\infty e^{\tilde{S}_{22}t} \tilde{S}_{21} G_{11}(x+t)\ dt.$$

This admits a solution of the form $G_{11}(x) = e^{Ux}$ iff U satisfies the equation (3.4), as it is seen by letting $x \to 0$ in the above equation. Therefore, $G_{11}(x) = \Omega_{11}(x)$.

Equation (3.8) is obtained by conditioning on the fluid level t that has been added up to the first epoch when the phase process enters the set $\{1,...,m\}$. Now, consider a solution of (3.8). It satisfies

$$\frac{d}{dx} G_{21}(x) = \int_0^\infty e^{\tilde{S}_{22}t} \tilde{S}_{21} \frac{\partial}{\partial x} G_{11}(x+t)\ dt = We^{Ux}U$$

whence $G_{21}(x) = We^{Ux} = \Omega_{21}(x)$.

The interesting probabilistic interpretation of (3.7) based on LIFO ideas is once again similar to those leading to the matrix-exponential forms of the G-matrix that have been explored by several authors [13],[17],[20] in the matrix analytic literature. Under this, one views the situation as that of draining a heavy fluid from the top where the oldest inputs are at the bottom! Now the first term in the right of (3.7) represents a case where no overload occurs while the original backlog is being drained out while the second term represents the case where an overload starts occurring after we have cleared some $x-t$ units of the original backlog. Turning to (3.4), U and the Markov process defined by U model the phase evolution *only* during those times when the initial backlog is being cleared! Indeed, the second term in (3.4) corresponds to excision of the portions of the paths where the fluid being drained is not from the initial content of the buffer. After excisions, we can treat the phase as a function of the fluid level which plays the role of time in reverse; see also Asmussen [3].

The above probabilistic interpretations are not mere mathematical curiosities because they immediately lead to interesting constructions yielding powerful algorithms for the fluid flow model. They are further testimony to the reach of the (simple) ideas underlying matrix analytic methods and the importance of maintaining probabilistic interpretations in the analysis. We demonstrate that in the next section after providing the following phase type characterization for the steady state fluid level.

Theorem 3.3: Assume that the stationary distribution of the fluid flow (of the primal model) exists. Then it is phase type with representation $PH(\boldsymbol{\alpha}, U)$ of order m, where $\boldsymbol{\alpha} = \boldsymbol{\pi}_1 + \boldsymbol{\pi}_2 W$.

Proof: If the primal is ergodic, then clearly the dual is not and therefore the generator U is defective. From (2.6) and Corollary 2.1, we have that the steady state fluid density is given by

$$
\begin{aligned}
f(x) &= -\boldsymbol{\pi}_1 K e^{Kx}[\mathbf{1} + \Psi\mathbf{1}] \\
&= -[\mathbf{1}' + \mathbf{1}'\Psi']\Delta_1\Delta_1^{-1} e^{K'x}\Delta_1\Delta_1^{-1} K'\Delta_1\Delta_1^{-1}\boldsymbol{\pi}_1' \\
&= -\boldsymbol{\alpha}\, e^{Ux} U\mathbf{1}
\end{aligned}
$$

where $\boldsymbol{\alpha} = [\mathbf{1}' + \mathbf{1}'\Psi']\Delta_1$. The first term in the right hand side for the expression for $\boldsymbol{\alpha}$ is clearly $\boldsymbol{\pi}_1$; the second term simplifies further to yield the given form for $\boldsymbol{\alpha}$ upon noting that $\Psi'\Delta_1 = \Delta_2 W$ (see the line after (3.5)). Other representations of the distribution are possible, but not considered here.

4. CONNECTION TO DISCRETE TIME QUEUES

The connections to be discussed here are quite intuitive, but their formal probabilistic proofs require major technicalities involving the construction of the primal and dual processes on a common probability space, stochastic coupling, etc. So, here we will once again use an algebraic approach that, though not fully enlightening, nevertheless gives us the results we want [the probabilistic details will be given in the extended version of this paper].

First of all, note from (3.4) that $U_{ii} > [\tilde{S}_{11}]_{ii}$ for all i. Thus, we can choose a number $\theta \geq \max_{1 \leq i \leq m+n} (-\tilde{S}_{ii})$ and use it to uniformize all the Markov processes considered thus far. Applying the routine calculations of uniformization to (3.4), we can rewrite that equation in terms of the nonnegative substochastic matrices $M_1 = \theta^{-1}U + I$, $\tilde{P}_{ii} = \theta^{-1}\tilde{S}_{ii} + I$, and $\tilde{P}_{ij} = \theta^{-1}\tilde{S}_{ij}$, for $i \neq j$ as follows:

$$
M_1 = \tilde{P}_{11} + \tilde{P}_{12} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \gamma_{nk}\, \tilde{P}_{22}^{n} \tilde{P}_{21} M_1^k \tag{4.1}
$$

where

$$
\gamma_{nk} = \int_0^{\infty} e^{-2\theta y} \frac{(\theta y)^{n+k}}{n!\, k!} \theta\, dy = \frac{(n+k)!}{n!\, k!} \left(\frac{1}{2}\right)^{n+k+1}. \tag{4.2}
$$

We recognize γ_{nk} to be the negative binomial probability of obtaining n failures before the $(k+1)$-st success in a sequence of Bernoulli trials with probability 1/2 of success. That enables us to simplify (4.1) to

$$
M_1 = B_0 + \sum_{k=1}^{\infty} B_k M_1^k \tag{4.3}
$$

where

$$
B_0 = \tilde{P}_{11} + \frac{1}{2}\tilde{P}_{12}(I - \frac{1}{2}\tilde{P}_{22})^{-1}\tilde{P}_{21},
$$

and

$$B_k = 2^{-(k+1)} \tilde{P}_{12} (I - \frac{1}{2} \tilde{P}_{22})^{-(k+1)} \tilde{P}_{21}, \quad k \ge 1.$$

It is trivial to verify that $\sum_{k=0}^{\infty} B_k = \tilde{P}_{11} + \tilde{P}_{12}(I - \tilde{P}_{22})^{-1} \tilde{P}_{21}$ and hence that $\sum_{k=0}^{\infty} B_k \mathbf{1} = \mathbf{1}$. Thus, M_1 is the G-matrix of the $M/G/1$-type model defined by the sequence $\{B_k\}$. [17],[21]; see also the Appendix. Incidentally, this means that we can compute M_1 by the powerful algorithms of [5]-[7],[11] and get all the required quantities therefrom, although we are in even greater luck than this entails.

Corollary 4.1: We have

$$W = \sum_{k=0}^{\infty} 2^{-(k+1)} (I - \frac{1}{2} \tilde{P}_{22})^{-(k+1)} \tilde{P}_{21} M_1^k \tag{4.4}$$

Proof: This is obtained by uniformization in (3.6).

It is elementary to verify that the discrete time, discrete state space $M/G/1$-type model is the embedded process at visits to phases $1,...,m$ in the discrete time, discrete state space QBD (see Appendix) defined by the matrices

$$A_2 = \begin{bmatrix} \tilde{P}_{11,} & 0 \\ \frac{1}{2} \tilde{P}_{21} & 0 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & \tilde{P}_{12} \\ 0 & \frac{1}{2} \tilde{P}_{22} \end{bmatrix}, \quad A_0 = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} I \end{bmatrix} \tag{4.5}$$

What is important is the following result which shows that M_1 and W can be determined from the blocks of the G-matrix of the QBD determined by the matrices A_i, $i=0,1,2$ defined in (4.5). We know [12] that that G-matrix itself can be computed efficiently by quadratically convergent iterative techniques which are numerically very reliable.

Theorem 4.1: The G-matrix of the QBD defined by (4.5) is given by

$$G = \begin{bmatrix} M_1 & 0 \\ W & 0 \end{bmatrix} \tag{4.6}$$

Proof: The embedded process in the QBD on phases $1,...,m$ is the $M/G/1$-problem defined by $\{B_i\}$, and therefore M_1 in (4.6) is the same M_1 in (4.3). As for the block below M_1 in the right of (4.6), note that if the QBD starts in level 1 in the set of phases $m+1,...,m+n$ and goes up a level before entering level 0, then it must first visit the set of phases $1,...,m$ before it can make a downward journey to level 0. Keeping track of how many upward level transitions occur during the first entry into the set 1,...,m leads to (4.4) as one may verify directly. Hence the result.

Remarks: It is trivial to verify that the stationary vector of the matrix $A = A_2 + A_1 + A_0$ is proporational to the vector $\boldsymbol{\xi} = (\boldsymbol{\pi}_1 C_1, \ 2\boldsymbol{\pi}_2 C_2)$. The familiar results [16] on QBDs, show that M_1 is substochastic (the QBD is transient) iff $\boldsymbol{\xi}_1 A_2 \mathbf{1} < \boldsymbol{\xi}_2 A_0 \mathbf{1}$, a condition which is equivalent to the assertion $\boldsymbol{\pi}_1 C_1 \mathbf{1} < \boldsymbol{\pi}_2 C_2 \mathbf{1}$; we recognize the latter to be the condition for ergodicity of the primal.

It is worth noting that so far we have made no assumptions about the Markov chain of phases except that it is irreducible. Under further assumptions, it is possible to simplify computations much further. As an example, when T is tri-diagonal (the birth-death case which includes on-off sources), we note that except for the last row, all other rows of the matrix $\tilde{P}_{12}$ are zero. Now, from (4.3), it follows that all rows of M_1 except for the last row are explicitly determined. The fact that only one row of M_1 needs to be determined is one that will lead to major simplifications in the algorithmic procedures similar to those for reducible chains which have been discussed in the matrix-analytic literature. Further consideration of this and other similar special cases is deferred to the detailed version of this paper.

5. CONCLUDING REMARKS

We have demonstrated a strong connection of the fluid flow model with a discrete time, discrete state space QBD, and that immediately provides powerful algorithms to compute the steady state distribution of the stochastic fluid flow model. We consider this to be an important advance in the analysis of fluid flow models for many reasons: (a) the matrix-analytic algorithms are demonstrably [12],[14] more stable and accurate than spectral methods; (b) the connection to QBDs can provide transient results in a natural manner; (c) the results on excursion to higher levels available for QBDs can be carried over to fluid models, and these should be useful for evaluating congestion control techniques and buffer sizing; (d) a connection of the (elusive) bounded fluid buffer problem to a bounded QBD problem now appears possible, and the latter can be solved by powerful matrix techniques; (e) just as the theory of matrix analytic methods is evolving now into techniques for level dependent queues, so also a potential exists to develop algorithms where the flow rates are dependent on the fluid level (as would be when flow controls are exercised.) From a theoretical perspective, our work demonstrates the greater simplicity of the conditioning arguments underlying the matrix-analytic approach and the duality of [19] as compared with Wiener-Hopf methods and Siegmund duality. These will be explored in full detail in an extended version of this short note which is but a preliminary announcement of our main results.

APPENDIX

Following is a quick overview of the basic elements of QBDs and *M/G/*1-type chains referred to in the Section 4.

An *M/G/*1 type Markov chain is a Markov chain on a two-dimensional state space $\bigcup_{i=-\infty}^{+\infty} \{(i,1),\, \ldots,\, (i,n)\}$ which has a transition matrix (after arranging states in a lexicographically increasing order) of the block partitioned form

$$
\begin{bmatrix}
\cdots & & & & & & \\
\cdots & 0 & 0 & B_0 & B_1 & B_2 & \cdots \\
\cdots & 0 & 0 & 0 & B_0 & B_1 & \cdots \\
\cdots & 0 & 0 & 0 & 0 & B_0 & \cdots \\
\cdots & 0 & 0 & 0 & 0 & 0 & \cdots \\
 & & & & & & \cdots
\end{bmatrix}
$$

where B_k are of order $n \times n$. Here B_0 records the one step probabilites of going one level down, B_1 the probabilities of remaining in the same level, and B_{k+1} the probabilities of going k levels up. A key quantity in analyzing it is the matrix G, whose (i,j)-th element is the probability that starting in (m,i), the process eventually visits a state of the form $(m-1,\cdot)$ by visiting precisely the state $(m-1,j)$. It is well known [17],[21] that G is the minimal nonnegative solution of the matrix equation $G = \sum_{k=0}^{\infty} B_k G^k$ and can be computed efficiently using matrix iterative techniques [5]-[7],[10].

The special case when $B_k = 0$ for $k \geq 3$, and $B_k = A_{2-k}$ for $k = 0,1,2$ is called a Quasi birth and death process (QBD). For a QBD, the equation for G becomes a quadratic equation $G = A_2 + A_1 G + A_0 G^2$, and it can be solved by quadratically convergent iterative techniques [12].

REFERENCES

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

[2] S. Asmussen & V. Ramaswami (1990): Probabilistic interpretations of some duality results for the matrix paradigms in queueing theory, *Stochastic Models,* 6, 715-734.

[3] S. Asmussen (1995): Stationary distributions for fluid flow models with or without Brownian noise, *Stochastic Models,* 11, 1-20.

[4] S. Asmussen (1995): Stationary distributions via first passage times, in *Advances in Queueing: Theory, Methods and Open Problems,* J.H. Dshalalow (Ed.), CRC Press, NY.

[5] D. Bini & B. Meini (1996): On the solution of a nonlinear matrix equation arising in queueing problems, *SIAM J. Matrix Anal. Appl.,* 17, 906-926.

[6] D. Bini & B. Meini (1997): Improved cyclic reduction for solving queueing problems, *Numerical Algorithms,* 15: 57-74.

[7] D. Bini & B. Meini (1997): On cyclic reduction applied to a class of Teoplitz-like matrices arising in queueing problems, in *Computations with Markov Chains,* W.J. Stewart (Ed.), pp. 21-38, Kluwer Academic Publisher.

[8] J.N. Daigle & D.M. Lucantoni (1991): Queueing systems having phase dependent arrival and service rates, in *Numerical Solution of Markov Chains,* W.J. Stewart (Ed.), 161-202, Marcel Dekker, NY.

[9] A.I. Elwalid & D. Mitra (1991): Analysis and design of rate-based congestion control of high speed networks, I: stochastic fluid models, access regulation, *QUESTA,* 9, 19-64.

[10] P. Favati & B. Meini (1998): On functional iteration methods for solving M/G/1 type Markov chains, in *Advances in Matrix Analytic Methods for Stochastic Models,* A.S. Alfa & S. Chakravarthy (Eds.), Notable Publications Inc., NJ., pp.39-54.

[11] B. Igelnik, Y. Kogan, V. Kriman & D. Mitra (1995): A new computational approach for stochastic fluid models of multiplexers with heterogeneous sources, *QUESTA,* 20, 85-116.

[12] G. Latouche & V. Ramaswami (1993): A logarithmic reduction algorithm for Quasi-Birth-Death processes, *J. Appl. Prob.,* 30, 650-674.

[13] F. Machihara (1990): A new approach to the fundamental period of a queue with phase type Markov renewal arrivals, *Stochastic Models,* 6, 551-559.

[14] B. Meini (1998): Solving QBD problems: the cyclic reduction algorithm versus the invariant subspace method, *Adv. in Perf. Anal.,* 1(2),215-225.

[15] D. Mitra (1988): Stochastic theory of a fluid model of producers and customer coupled by a buffer, *Adv. Appl. Prob.,* 20, 646-676

[16] M.F. Neuts (1981): *Matrix-geometric Solutions in Stochastic Models,* Johns Hopkins University Press, Baltimore.

[17] M.F. Neuts (1989): *Structured Stochastic Matrices of M/G/1 Type and Their Applications,* Marcel Dekker, NY.

[18] V. Ramaswami (1982): The busy period of queues which have a matrix-geometric steady state probability vector, *Opsearch,* 19, 238-261.

[19] V. Ramaswami (1990): A duality theorem for the matrix paradigms in queueing theory, *Stochastic Models,* 6, 151-161.

[20] V. Ramaswami (1990): From the matrix-geometric to the matrix-exponential, *QUESTA,* 6, 229-260.

[21] V. Ramaswami (1997): Matrix analytic methods: a tutorial overview with some extensions and new results, in *Matrix Analytic Methods in Stochastic Models,* S.R. Chakravarthy & A.S. Alfa (Eds.), Marcel Dekker, NY, pp. 261-296.

[22] V. Ramaswami (1998): The generality of Quasi Birth and Death Processes, in *Advances in Matrix Analytic Methods for Stochastic Models,* A.S. Alfa & S. Chakravarthy (Eds.), Notable Publications Inc., NJ., pp.93-114.

[23] B. Sengupta (1989): Markov processes whose steady state distribution is matrix-exponential with an application to the GI/PH/1 queue, *Adv. Appl. Prob.,* 21, 159-180.

[24] D. Siegmund (1976): The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes, *Ann. Prob.,* 4, 914-924.

[25] A. Simonian & J. Virtamo (1991): Transient and stationary distribution for fluid queues and input processes with a density, *SIAM J. Appl. Math,* 51, 1731-1739.

[26] T.E. Stern & A.I. Elwalid (1991): Analysis of separable Markov modulated rate models for information handling systems, *Adv. Appl. Prob.,* 23, 105-191