

Second order fluid models with general boundary behaviour

M. Gribaudo · D. Manini · B. Sericola · M. Telek

Published online: 5 December 2007
© Springer Science+Business Media, LLC 2007

Abstract A crucial property of second order fluid models is the behaviour of the fluid level at the boundaries. Two cases have been considered: the reflecting and the absorbing boundary. This paper presents an approach for the stationary analysis of second order fluid models with any combination of boundary behaviours. The proposed approach is based on the solution of a linear system whose coefficients are obtained from a matrix exponent. A practical example demonstrates the suitability of the technique in performance modeling.

Keywords Second order fluid models · Matrix exponent · Stationary distribution · Numerical analysis

1 Introduction

In the recent years fluid flow models have been widely used to study complex communication, computer and flexible manufacturing systems. In particular, fluid queues with Markov modulated input rates play a key role in queueing theory and performance evaluation of networks. The main feature of these models is to describe systems using discrete and continuous quantities. The use of continuous quantities can significantly reduce the well-known state space explosion problem. For example, a quantity that can take a very large number of

This work is partially supported by the Italian-Hungarian bilateral R&D programme, by OTKA grant n. T-34972, by Italian Ministry for University and Research (MIUR) through PRIN project Famous and by EEC project Crucial.

M. Gribaudo · D. Manini (✉)
Dip. di Informatica, Università di Torino, Torino, Italy
e-mail: manini@di.unito.it

B. Sericola
INRIA-IRISA, Rennes, France

M. Telek
Dept. of Telecommunications, Technical University of Budapest, Budapest, Hungary

possible values (like the queue size in a communication network) can be accurately approximated by a continuous quantity.

The first studies of such queueing systems can be dated back in the early 1980's. The works (Kosten 1984) and (Anick et al. 1982) analyze fluid models in connection with statistical multiplexing of several identical exponential on-off input sources in a buffer.

The above studies mainly focus on the analysis of the stationary regime and raise a series of theoretical developments. For instance, Mitra (1987) and Mitra (1988) generalize the analysis by considering multiple types of exponential on-off inputs and outputs. Stern and Elwalid (1991) consider separable Markov modulated processes and express the equilibrium equations as a sum of Kronecker product terms. Igel'nik et al. (1995) derive a new approach based on the use of interpolating polynomials for the computation of the buffer overflow probability. Using the Wiener-Hopf factorization of finite Markov chains, Rogers (1994) shows that the distribution of the buffer level has a matrix exponential form. Algorithmic issues of that factorization are explored in Rogers and Shi (1994).

The papers (Ramaswami 1996) and (da Silva Soares and Latouche 2002), respectively, exhibit and exploit the similarity between stationary fluid queues in a finite Markovian environment and quasi birth and death processes. In Ahn and Ramaswami (2003) a direct connection by stochastic coupling is established between fluid queues and quasi birth and death processes.

The transient analysis of fluid queues with exponential on-off sources is studied in Kobayashi and Ren (1992), Ren and Kobayashi (1995) by using Laplace transform. These studies are extended to the Markov modulated input rate model in Tanaka et al. (1995). Sericola (1998) obtains a transient solution based on simple recurrence relations that have interesting numerical properties. More recently, Ahn and Ramaswami (2004) approximate the fluid model by the amounts of work in a sequence of Markov modulated queues of the quasi birth and death type.

Second order models (also known as Markov modulated diffusion processes) are introduced in Karandikar and Kulkarni (1995) and Asmussen (1995). In these works the authors consider a "white noise" factor which represents the variability of the traffic during the transmission periods. The fluid level is described by a reflected Brownian motion modulated by a continuous time Markov chain (CTMC). When the CTMC is in state i , the fluid level is modeled by a reflected Brownian motion with drift r_i and variance parameter σ_i^2 .

In first order models the fluid level grows linearly with a deterministic rate r_i when the modulating Markov chain is in state i . If $x(t)$ denotes the fluid level at time t and t' is a time instant such that $t' > t$, then:

$$x(t') = x(t) + (t' - t)r_i, \quad (1)$$

provided that the Markov chain remains in state i and the continuous variable does not reach a boundary in the (t, t') interval. Markov modulated diffusion processes instead consider random fluid changes. In these models, we have that:

$$x(t') = x(t) + N((t' - t)r_i, (t' - t)\sigma_i^2), \quad (2)$$

that is, the fluid level change in the (t, t') interval is normally distributed with mean $(t' - t)r_i$ and variance $(t' - t)\sigma_i^2$. Note that (2) is valid if the Markov chain remains in state i and the continuous variable does not reach a boundary in the (t, t') interval.

Second order models have been addressed in several works. The authors of Rabehasaina and Sericola (2003) provide a stability analysis of such models when the modulating process is general stationary ergodic (not necessarily Markovian).

The differential equations that describe a fluid model are hard to solve. The symbolic solution of the equations can be obtained only for trivial cases. Different numerical solutions are available for transient and stationary analysis. In the case of transient analysis the system has an initial state which can be exploited as considered in Horton et al. (1998), Wolter (1997), Chen et al. (2002), Sericola (1998) to mention a few. In the case of stationary analysis, the equations that describe a fluid model are ordinary differential equations (ODEs), and there is no initial condition. Indeed this problem has been solved for first order models by the analysis of first passage time probabilities, see for instance Ramaswami (1996), da Silva Soares and Latouche (2002), Ahn and Ramaswami (2003), Bean et al. (2005a, 2005b), da Silva Soares and Latouche (2006) and the references therein. The key of these solutions lies in the matrix characterization of the distribution of the phase visited at the end of a busy period of the fluid queue.

The problem remains open for modulated diffusion processes, where the solution is obtained from a set of boundary equations, ODEs and a normalizing condition. For example, in case of fluid level independent transition and fluid drift, the solution of the ODE is obtained by the computation of eigenvalues and eigenvectors of a matrix (Karandikar and Kulkarni 1995). Usually those approaches are very sensitive to the computation of the eigenvalues and may lead to severe numerical errors. An alternative approach using modal decomposition is proposed in Agapie and Sohraby (2001).

In this work, we address the problem of performing steady state solution of modulating diffusion processes using neither discretization nor singular value decomposition. Our approach is similar to the one used in Gribaudo and German (2001), German et al. (2003) for first order models. In particular, we focus on the boundary behaviours. Second order models can have two different types of boundaries: absorbing or reflecting (Cox and Miller 1972). We are able to consider models where the upper and lower boundary of each state can be either absorbing or reflecting. To the best of our knowledge second order fluid models were applied in a wide range of application fields, but always with reflecting boundaries. A special approximation method is proposed in Ang and Barria (2000) for approximating the absorbing boundaries based on the solution of the system with reflecting boundaries.

The paper is structured as follows. First it presents a motivating example in Sect. 2. Section 3 introduces the second order fluid models we consider, and it presents their transient and steady state equations. The analysis technique is considered in Sect. 4. Section 5 provides an applicative example and the paper is concluded in Sect. 6.

2 A motivating example

We introduce the simple queueing system, represented in Fig. 1, to show the impact of the boundary behaviour in second order fluid models.

The considered queueing system has a single server and a finite buffer of size B . Customers arrive to the queue according to an on-off arrival process. Customers arrive only during the on period according to a Poisson process of rate λ . The length of the on and the

Fig. 1 The discrete Markov model of the queueing system

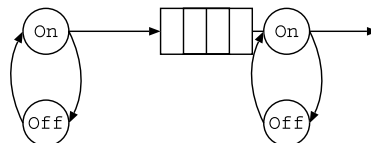


Fig. 2 The discrete Markov model of the queueing system

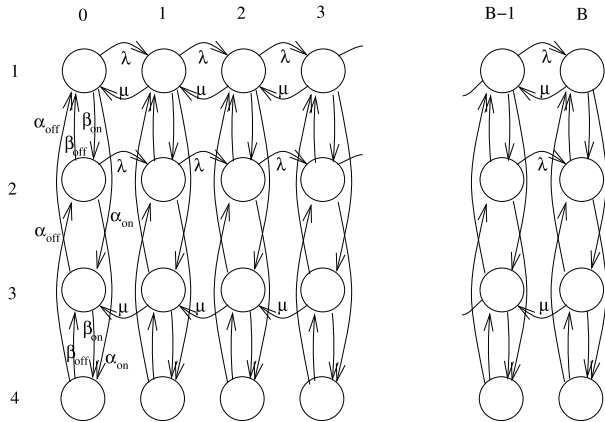


Table 1 Fluid behaviour versus system state

$J(t)$	1	2	3	4
Arrival-server	On-on	On-off	Off-on	Off-off
$\lim_{h \rightarrow 0} [E(X(t+h) - X(t))/h]$	$\lambda - \mu$	λ	$-\mu$	0
$\lim_{h \rightarrow 0} [Var(X(t+h) - X(t))/h]$	$\lambda + \mu$	λ	μ	0

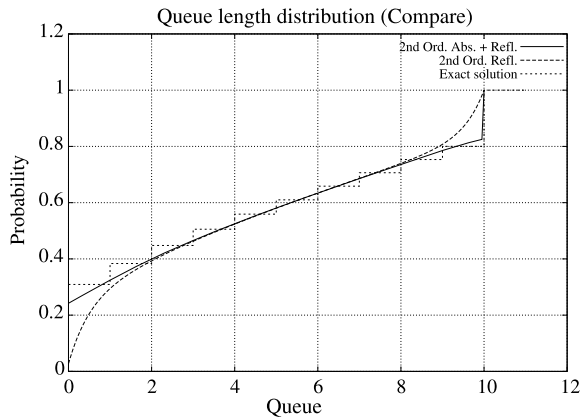
off periods are exponentially distributed with parameters α_{on} and α_{off} . The server follows an on-off behaviour as well. During its on period it serves customers at rate μ and during its off period it goes on vacation. Interrupted services are resumed as soon as the server comes back from vacation. The on and the off period of the server are exponentially distributed with parameters β_{on} and β_{off} . Let $N(t)$ ($0 \leq N(t) \leq B$) be the number of customers in the buffer and $J(t)$ be the combined state of both the arrival process and the server ($1 \leq J(t) \leq 4$). We define $J(t) = 1$ when both the arrival and the service are on, $J(t) = 2$ when the arrival is on and the service is off, $J(t) = 3$ when the arrival is off and the service is on, and $J(t) = 4$ when both the arrival and the service are off.

The Markov chain of the queueing system is depicted in Fig. 2. When B is large the number of customers in the system ($N(t)$) can be represented with a continuous variable $X(t)$ commonly referred to as fluid level. The behaviour of the fluid level depends on the state of both the arrival process and the server, i.e., on $J(t)$. Between the boundaries, we have the cases reported in Table 1.

Conventional second order fluid model approximations consider only states with reflecting boundaries. The comparison of the fluid model approximation with the exact solution, as depicted in Fig. 3, shows that the former fails to describe the behaviour at the boundaries. In particular it does not consider the jumps created by probability masses. This can be a very critical issue, since many important performance indices rely on the probability at the boundaries (i.e., utilization and blocking probability).

Previous results prove that the reflecting behaviour is not appropriate in describing the actual model (in some of its states). To investigate this problem, we analyze the Markov chain of the queueing system represented in Fig. 2. The structure of the Markov chain indicates that in state $J(t) = 1$ the fluid level can vary between the barriers independent of the fact that the buffer gets empty or full. In state $J(t) = 2$ the fluid level can leave the lower

Fig. 3 Various solution of the queueing system



barrier, but if it reaches the upper barrier it remains there as long as $J(t) = 2$. Similar observations are valid for $J(t) = 3$ and $J(t) = 4$. It follows that in state $J(t) = 1$ both the lower and the upper boundaries are reflecting, in state $J(t) = 2$ the lower boundary is reflecting and the upper boundary is absorbing, in state $J(t) = 3$ the lower boundary is absorbing and the upper boundary is reflecting, and finally in state $J(t) = 4$ both the lower and the upper boundaries are absorbing. If we exploit the correct barrier behaviour at both boundaries, we obtain more accurate results as shown in Fig. 3.

This example is only to justify the importance of models with mixed (reflecting and absorbing) boundary behaviour. In the following sections we show how such continuous models can be solved, but this solution is independent of the discrete model presented here.

3 Notation and preliminaries

3.1 Notation

The transient analysis of Markov modulated diffusion processes with finite buffer is based on the transition matrix \mathbf{Q} , the drift matrix \mathbf{R} , the variance matrix \mathbf{S} , the buffer size B and the description of the boundary behaviour. The corresponding stochastic process is defined by a couple $\{Z(t), X(t)\}$. $Z(t)$ is a CTMC with generator matrix \mathbf{Q} , and $X(t)$ represents the fluid level process. During a sojourn in state i , i.e. $Z(t) = i$, the fluid level $X(t)$ follows a Brownian motion with rate r_i and variance s_i^2 . \mathbf{R} and \mathbf{S} are diagonal matrices containing the state dependent drift and variance parameters ($\mathbf{R} = \text{diag}(r_i)$, $\mathbf{S} = \text{diag}(s_i^2/2)$). For each state j we define

- The fluid distribution: $F_j(t, x) = \Pr(Z(t) = j, X(t) \leq x)$,
- The fluid density: $f_j(t, x) = \lim_{h \rightarrow 0} [F_{i,j}(t, x + h) - F_{i,j}(t, x)]/h$,
- The mass at the lower boundary: $\ell_j(t) = \Pr(Z(t) = j, X(t) = 0)$,
- The mass at the upper boundary: $u_j(t) = \Pr(Z(t) = j, X(t) = B)$.

3.2 State space partitioning and boundary behaviour

We partition the state space into six disjoint subsets $\mathcal{S} = \mathcal{S}^+ \cup \mathcal{S}^0 \cup \mathcal{S}^- \cup \mathcal{S}^\oplus \cup \mathcal{S}^\ominus \cup \mathcal{S}^\circ$, where:

$$\begin{aligned} \mathcal{S}^+ &= \{i : r_i > 0, s_i = 0\}, & \mathcal{S}^0 &= \{i : r_i = 0, s_i = 0\}, & \mathcal{S}^- &= \{i : r_i < 0, s_i = 0\}, \\ \mathcal{S}^\oplus &= \{i : r_i > 0, s_i > 0\}, & \mathcal{S}^\ominus &= \{i : r_i = 0, s_i > 0\}, & \mathcal{S}^\ominus &= \{i : r_i < 0, s_i > 0\}. \end{aligned}$$

We refer to the states belonging to $\mathcal{S}^+ \cup \mathcal{S}^0 \cup \mathcal{S}^-$ as *first order states*, and to the ones belonging to $\mathcal{S}^\oplus \cup \mathcal{S}^\ominus \cup \mathcal{S}^\circ$ as *second order states*. Furthermore, we introduce

$$\mathcal{S}^\circ = \mathcal{S}^\oplus \cup \mathcal{S}^\ominus \cup \mathcal{S}^\circ.$$

The cardinality of a subset U of the state space is denoted by $\#U$. The background CTMC is supposed to be irreducible, hence it stays for a finite amount of time in all non-empty and valid subsets of states. The reflecting and absorbing behaviours determine the fluctuation of the fluid level when it reaches a boundary. Let us focus on the lower boundary. The absorbing boundary retains the fluid at the border till the background CTMC moves to a new state where the boundary behaviour is reflecting. This implies that for an absorbing state j we have $\ell_j(t) > 0$. Instead, the reflecting boundary immediately reflects the fluid. In this case for a reflecting state i we have $\ell_i(t) = 0$. Similar considerations are valid for the upper boundary.

It is important to note that in a system where all the states have an absorbing behaviour the fluid is confined at the boundaries. States with reflecting behaviour allow the fluid to leave the barriers.

3.3 Transient behaviour

We denote respectively by $f(t, x)$, $F(t, x)$, $\ell(t)$ and $u(t)$ the row vectors containing the $f_j(t, x)$, $F_j(t, x)$, $\ell_j(t)$ and $u_j(t)$. They satisfy the equations (Karandikar and Kulkarni 1995):

$$\begin{aligned} \frac{\partial F(t, x)}{\partial t} + \frac{\partial F(t, x)}{\partial x} \mathbf{R} - \frac{\partial^2 F(t, x)}{\partial x^2} \mathbf{S} &= F(t, x) \mathbf{Q}, \quad \text{for } 0 < x < B, \\ \frac{\partial f(t, x)}{\partial t} + \frac{\partial f(t, x)}{\partial x} \mathbf{R} - \frac{\partial^2 f(t, x)}{\partial x^2} \mathbf{S} &= f(t, x) \mathbf{Q}, \quad \text{for } 0 < x < B, \\ \frac{\partial \ell(t)}{\partial t} + f(t, 0) \mathbf{R} - f'(t, 0) \mathbf{S} &= \ell(t) \mathbf{Q}, \quad \text{for } x = 0, \\ \frac{\partial u(t)}{\partial t} - f(t, B) \mathbf{R} + f'(t, B) \mathbf{S} &= u(t) \mathbf{Q}, \quad \text{for } x = B, \end{aligned} \tag{3}$$

where $f'(t, 0)$ (resp. $f'(t, B)$) is the derivative, with respect to x , of the density density $f(t, x)$ taken at point $x = 0^+$ (resp. $x = B^-$). In order to simplify the notation we write $f(t, 0)$ and $f(t, B)$ instead of $f(t, 0^+)$ and $f(t, B^-)$. Boundary equations are defined as follows:

- Lower boundary:
 - $\ell_j(t) = 0$ for $j \in \mathcal{S}^+$,
 - $\ell_j(t) = 0$ for $j \in \mathcal{S}^\circ$ and reflecting lower boundary in j ,
 - $f_j(t, 0) = 0$ for $j \in \mathcal{S}^\circ$ and absorbing lower boundary in j .

- Upper boundary:
 - $u_j(t) = 0$ for $j \in \mathcal{S}^-$,
 - $u_j(t) = 0$ for $j \in \mathcal{S}^\circ$ and reflecting upper boundary in j ,
 - $f_j(t, B) = 0$ for $j \in \mathcal{S}^\circ$ and absorbing upper boundary in j .

A rigorous proof of these boundary equations is given in Cox and Miller (1972) (absorbing in Sect. 5.7[i], reflecting in Sect. 5.7[ii]).

3.4 Stationary behaviour

Since the buffer capacity is finite, the stationary distribution exists and is given by the following limits:

- $f_j(x) = \lim_{t \rightarrow \infty} f_j(t, x)$,
- $F_j(x) = \lim_{t \rightarrow \infty} F_j(t, x)$,
- $\ell_j = \lim_{t \rightarrow \infty} \ell_j(t)$,
- $u_j = \lim_{t \rightarrow \infty} u_j(t)$,

which results in the following equations

- Differential equations:

$$F'(x)\mathbf{R} - F''(x)\mathbf{S} = F(x)\mathbf{Q}, \tag{4}$$

$$f'(x)\mathbf{R} - f''(x)\mathbf{S} = f(x)\mathbf{Q}, \tag{5}$$

where $f(x)$, $f'(x)$, $f''(x)$, $F(x)$, $F'(x)$, $F''(x)$, ℓ and u are the row vectors containing the $f_j(x)$, $f'_j(x)$, $f''_j(x)$, $F_j(x)$, $F'_j(x)$, $F''_j(x)$, ℓ_j and u_j . Note also that we have $F'(x) = f(x)$, $F(0) = \ell$, $F(B^-) = \pi - u$ and $F(B) = \pi$, where π is the stationary distribution of the driving CTMC ($Z(t)$), i.e., the probability distribution which satisfies $\pi Q = 0$.

- Boundary conditions:

$$f(0)\mathbf{R} - f'(0)\mathbf{S} = \ell\mathbf{Q}, \tag{6}$$

$$-f(B)\mathbf{R} + f'(B)\mathbf{S} = u\mathbf{Q}. \tag{7}$$

- Constraints on lower boundary:
 - $\ell_j = 0$ for $j \in \mathcal{S}^+$,
 - $\ell_j = 0$ for $j \in \mathcal{S}^\circ$ and reflecting lower boundary in j ,
 - $f_j(0) = 0$ for $j \in \mathcal{S}^\circ$ and absorbing lower boundary in j ,
- Constraints on upper boundary:
 - $u_j = 0$ for $j \in \mathcal{S}^-$,
 - $u_j = 0$ for $j \in \mathcal{S}^\circ$ and reflecting upper boundary in j ,
 - $f_j(B) = 0$ for $j \in \mathcal{S}^\circ$ and absorbing upper boundary in j .

4 Solution

This section presents the main result of this paper, that is the stationary analysis of second order fluid models with finite buffer. First we consider the known case in which the system has only reflecting boundaries. Then we generalize the approach to study state dependent absorbing or reflecting boundaries.

4.1 Fully second order fluid models with reflecting boundaries

We define fully second order model cases where $S = S^\circ$ ($s_i > 0, \forall i \in S$), $S^+ = S^- = S^0 = \emptyset$, and \mathbf{S} is non-singular. Due to the reflecting behaviour at both boundaries, we have $\ell = u = 0$. We introduce the row vector $G(x) = (F(x), F'(x))$ and we insert it in (4):

$$G'(x) \begin{pmatrix} \mathbf{R} & \mathbf{I} \\ -\mathbf{S} & \mathbf{0} \end{pmatrix} = G(x) \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{8}$$

where the row vector $G'(x)$ is the derivative of $G(x)$ with respect to x , that is $G'(x) = (f(x), f'(x))$. Multiplying with the inverse of the matrix on the left hand side, we get

$$G'(x) = G(x) \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{0} & -\mathbf{S}^{-1} \\ \mathbf{I} & \mathbf{RS}^{-1} \end{pmatrix} = G(x) \begin{pmatrix} \mathbf{0} & -\mathbf{QS}^{-1} \\ \mathbf{I} & \mathbf{RS}^{-1} \end{pmatrix}. \tag{9}$$

The solution to (9) is given, for $x \in [0, B]$, by

$$G(x) = G(0)e^{\mathbf{M}x}, \quad \text{where } \mathbf{M} = \begin{pmatrix} \mathbf{0} & -\mathbf{QS}^{-1} \\ \mathbf{I} & \mathbf{RS}^{-1} \end{pmatrix}. \tag{10}$$

Differentiating relation (10), we get

$$G'(x) = (f(x), f'(x)) = G(0)(e^{\mathbf{M}x})' = (\ell, f(0))(e^{\mathbf{M}x})'. \tag{11}$$

Using the sub-matrix decomposition

$$e^{\mathbf{M}x} = \begin{pmatrix} \Phi_1(x) & \Phi_2(x) \\ \Phi_3(x) & \Phi_4(x) \end{pmatrix},$$

we can compute $G'(B)$ from (11) as

$$f(B) = \ell \Phi_1'(B) + f(0) \Phi_3'(B), \tag{12}$$

$$f'(B) = \ell \Phi_2'(B) + f(0) \Phi_4'(B). \tag{13}$$

Note that the relation $(e^{\mathbf{M}x})' = \mathbf{M}e^{\mathbf{M}x} = e^{\mathbf{M}x}\mathbf{M}$ leads to

$$\begin{aligned} \Phi_1'(x) &= -\mathbf{QS}^{-1}\Phi_3(x) = \Phi_2(x) \\ \Phi_2'(x) &= -\mathbf{QS}^{-1}\Phi_4(x) = -\Phi_1(x)\mathbf{QS}^{-1} + \Phi_2(x)\mathbf{RS}^{-1} \\ \Phi_3'(x) &= \Phi_1(x) + \mathbf{RS}^{-1}\Phi_3(x) = \Phi_4(x) \\ \Phi_4'(x) &= \Phi_2(x) + \mathbf{RS}^{-1}\Phi_4(x) = -\Phi_3(x)\mathbf{QS}^{-1} + \Phi_4(x)\mathbf{RS}^{-1}. \end{aligned} \tag{14}$$

Since $F(0) = \ell = 0$ and $u = 0$, we have

$$f(B) = f(0)\Phi_3'(B) = f(0)\Phi_4(B), \tag{15}$$

$$f'(B) = f(0)\Phi_4'(B) = f(0)(\Phi_2(B) + \mathbf{RS}^{-1}\Phi_4(B)), \tag{16}$$

and the boundary condition (7) can be expressed as

$$f'(B) = f(B)\mathbf{RS}^{-1}. \tag{17}$$

Combining (15), (16) and (17) we have

$$(f(0), f(B)) \begin{pmatrix} \Phi_4(B) & \Phi_2(B) + \mathbf{RS}^{-1}\Phi_4(B) \\ -\mathbf{I} & -\mathbf{RS}^{-1} \end{pmatrix} = 0. \tag{18}$$

The normalizing condition of (18) can be obtained from (10) since for $x = B$ we have

$$\pi = F(B) = f(0)\Phi_3(B),$$

and multiplying it with $\mathbb{1}$, the column vector of ones, we get

$$f(0)\Phi_3(B)\mathbb{1} = 1.$$

4.2 Fully second order fluid models with state dependent absorbing and reflecting boundaries

In this section we consider the fully second order fluid models, where in each state both the upper and the lower boundaries may have either an absorbing or reflecting barrier. We partition the state space S° in four subsets

$$S^\circ = S_{aa} \cup S_{ar} \cup S_{ra} \cup S_{rr}, \tag{19}$$

where the first subscript refers to the lower boundary and the second one to the upper boundary. In both cases a stands for absorbing and r for reflecting.

The boundary conditions (6) and (7) can be expressed as

$$f'(0) = f(0)\mathbf{RS}^{-1} - \ell\mathbf{QS}^{-1}, \tag{20}$$

$$f'(B) = f(B)\mathbf{RS}^{-1} + u\mathbf{QS}^{-1}. \tag{21}$$

Using (14), relations (12) and (13) can be rewritten as

$$f(B) = \ell\Phi_2(B) + f(0)\Phi_4(B), \tag{22}$$

$$f'(B) = -\ell\mathbf{QS}^{-1}\Phi_4(B) + f(0)(\Phi_2(B) + \mathbf{RS}^{-1}\Phi_4(B)). \tag{23}$$

The constraints on lower and upper boundaries given at the end of Sect. 3.4 can be written using the decomposition of the state space in (19):

$$\begin{aligned} f(0) &= (0, 0, f_{ra}(0), f_{rr}(0)), \\ \ell &= (\ell_{aa}, \ell_{ar}, 0, 0), \\ f(B) &= (0, f_{ar}(B), 0, f_{rr}(B)), \\ u &= (u_{aa}, 0, u_{ra}, 0). \end{aligned} \tag{24}$$

Note that each element of the vectors on the r.h.s. of (24) represents a boundary combination. Let us define the vectors v and w as

$$v = f(0) + \ell = (\ell_{aa}, \ell_{ar}, f_{ra}(0), f_{rr}(0)), \tag{25}$$

$$w = f(B) + u = (u_{aa}, f_{ar}(B), u_{ra}, f_{rr}(B)). \tag{26}$$

Using vector v relations (22) and (23) can be rewritten as

$$f(B) = v(\Phi_2(B)^{a\bullet} + \Phi_4(B)^{r\bullet}), \tag{27}$$

$$f'(B) = v([-QS^{-1}\Phi_4(B)]^{a\bullet} + [\Phi_2(B) + RS^{-1}\Phi_4(B)]^{r\bullet}), \tag{28}$$

where, for any matrix H of dimension $\#S$, we denote by $H^{a\bullet}$ (resp. $H^{r\bullet}$) the matrix obtained from H by setting to zero the rows not corresponding to states of $S_{aa} \cup S_{ar}$ (resp. $S_{ra} \cup S_{rr}$). In the same way, we denote by $H^{a\bullet}$ (resp. $H^{r\bullet}$) the matrix obtained from H by setting to zero the rows not corresponding to states of $S_{aa} \cup S_{ra}$ (resp. $S_{ar} \cup S_{rr}$).

Using vector w , relation (21) can be written as

$$f'(B) = w([\mathbf{RS}^{-1}]^{r\bullet} + [\mathbf{QS}^{-1}]^{a\bullet}). \tag{29}$$

Combining (24) and (26), we have

$$f(B) = w\mathbf{I}^{r\bullet}. \tag{30}$$

Putting together relations (27), (28), (29) and (30), we obtain the following system of dimension $2\#S$

$$(v, w)\mathbf{U} = 0 \quad \text{where } \mathbf{U} = \begin{pmatrix} \mathbf{A} & \mathbf{\Theta} \\ -\mathbf{I}^{r\bullet} & -\mathbf{C} \end{pmatrix}, \tag{31}$$

and

$$\begin{aligned} \mathbf{A} &= \Phi_2(B)^{a\bullet} + \Phi_4(B)^{r\bullet}, \\ \mathbf{\Theta} &= [-QS^{-1}\Phi_4(B)]^{a\bullet} + [\Phi_2(B) + RS^{-1}\Phi_4(B)]^{r\bullet}, \\ \mathbf{C} &= [\mathbf{RS}^{-1}]^{r\bullet} + [\mathbf{QS}^{-1}]^{a\bullet}. \end{aligned}$$

The normalizing condition is given by writing $G(B^-) = G(0)e^{MB}$. Since $F(B^-) = \pi - u$, it corresponds to

$$\begin{aligned} \pi - u &= \ell\Phi_1(B) + f(0)\Phi_3(B), \\ F'(B^-) &= f(B^-) = f(B) = \ell\Phi_2(B) + f(0)\Phi_4(B). \end{aligned}$$

This leads, using (25) and (26), to

$$\begin{aligned} u &= \pi - v(\Phi_1(B)^{a\bullet} + \Phi_3(B)^{r\bullet}), \\ f(B) &= w - u = v(\Phi_2(B)^{a\bullet} + \Phi_4(B)^{r\bullet}). \end{aligned}$$

If we compute the sum of these two relations, we get

$$w = \pi + v(\Phi_2(B)^{a\bullet} - \Phi_1(B)^{a\bullet} + \Phi_4(B)^{r\bullet} - \Phi_3(B)^{r\bullet}),$$

that is

$$w\mathbf{1} + v(\Phi_1(B)^{a\bullet} - \Phi_2(B)^{a\bullet} + \Phi_3(B)^{r\bullet} - \Phi_4(B)^{r\bullet})\mathbf{1} = 1.$$

5 Application example

In this section we apply the proposed technique to model the receiving buffer of the video streaming application depicted in Fig. 4. The model takes into account videos encoded with variable bit rate (VBR), and it is divided into two components: the network and the video player application. Fluid is used to model the receiving buffer. The streaming application cycles among three states: *buffering*, *playing*, and *finishing*. In state *buffering*, the application receives data from the network without playing it. In state *playing*, the application decodes the data stored in the buffer. In this state the application continues to receive data from the network. State *finishing* models the completion of the transmission while the player decodes the last part of the stream. The network can be in two states: *normal* and *congested*. In state *congested*, data arrive at a lower speed. The discrete part of the model is composed by five states as depicted in Fig. 5. Each state represents a combination of the status of the two components. Note that when the application does not receive data (in state *finishing*) the status of the network can be neglected. Parameters $1/\beta_B$, $1/\beta_P$ and $1/\beta_F$ represent the mean sojourn time spent by the application in state *buffering*, *playing*, and *finishing*. Parameters α_{HL} and α_{LH} describe the rate at which the network switches from the normal (H) to congested (L) states and vice versa. The infinitesimal generator Q of the underlying Markov chain is given by:

$$Q = \begin{pmatrix} -\beta_B - \alpha_{LH} & \beta_B & \alpha_{LH} & 0 & 0 \\ 0 & -\beta_P - \alpha_{LH} & 0 & \alpha_{LH} & \beta_P \\ \alpha_{HL} & 0 & -\beta_B - \alpha_{HL} & \beta_B & 0 \\ 0 & \alpha_{HL} & 0 & -\beta_P - \alpha_{HL} & \beta_P \\ \beta_F \frac{\alpha_{HL}}{\alpha_{HL} + \alpha_{LH}} & 0 & \beta_F \frac{\alpha_{LH}}{\alpha_{HL} + \alpha_{LH}} & 0 & -\beta_F \end{pmatrix}$$

Fig. 4 Video streaming application model

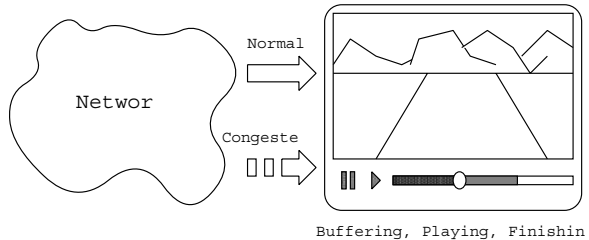


Fig. 5 Discrete part of the fluid model

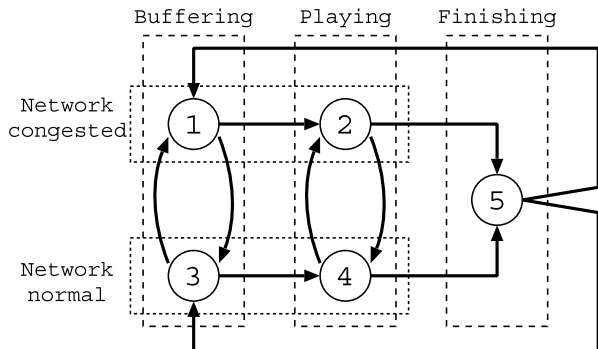
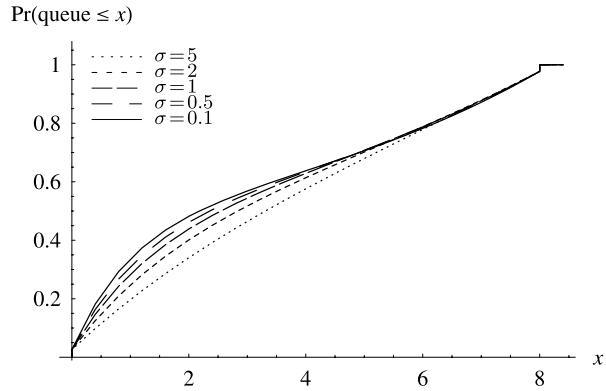


Fig. 6 Buffer occupancy (measured in 128 kb chunks) distribution versus the compression rate variance



Video streaming packets are decoded at rate μ with variance σ . The variance is due to the variable bit rate encoding. Data streaming arrives at rate λ_L with variance γ_L in state *congested* and at rate λ_H with variance γ_H in state *normal*. This allows us to compute matrix **R** and matrix **S** as follows:

$$\mathbf{R} = \begin{pmatrix} \lambda_L & & & & \\ & \lambda_L - \mu & & & \\ & & \lambda_H & & \\ & & & \lambda_H - \mu & \\ & & & & -\mu \end{pmatrix},$$

$$\mathbf{S} = \begin{pmatrix} \frac{1}{2}\gamma_L & & & & \\ & \frac{1}{2}(\gamma_L + \sigma) & & & \\ & & \frac{1}{2}\gamma_H & & \\ & & & \frac{1}{2}(\gamma_H + \sigma) & \\ & & & & \frac{1}{2}\sigma \end{pmatrix}.$$

We suppose that the streaming application exchanges chunks of 128 kb. The buffer size is 1 Mb (i.e. 8 chunks/s). The mean arrival rate is 256 kbps (i.e. 2 chunks/s) in state *congested* and 640 kbps (i.e. 5 chunks/s) in state *normal*. The video is encoded at a mean rate of 512 kbps (i.e. 4 chunks/s). We set the variance equal to the corresponding rate. The mean time spent by the application in state *buffering* and *finishing* is approximately 10 sec. Parameters can be summarized as follows: $\lambda_L = \gamma_L = 0.25$, $\lambda_H = \gamma_H = 0.625$, $\mu = 0.5$, $\beta_F = 0.1$, $\beta_B = 0.1$.

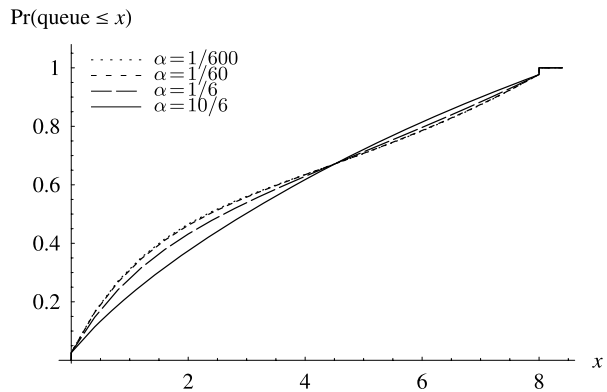
According to the discussion of Sect. 2, we have an absorbing behaviour at the lower boundary in state *finishing* and at the upper boundary in state *buffering*. All the other boundaries are reflecting.

First, we study the behaviour of the buffer occupancy distribution as function of the compression rate variance. We vary σ from 0.1 to 5. Other parameters are configured as follows: $B = 1\text{Mb}$, $\beta_P = 0.03$, $\alpha_{LH} = \alpha_{HL} = 1/60$. Results are reported in Fig. 6. The buffer occupancy distribution tends to become uniform as the variance increases. Table 2 shows that the empty and full buffer probabilities are only slightly affected by the variance.

Then we observe the buffer occupancy distribution as function of the network behaviour. In particular, we vary the mean duration of state *normal* and *congested*. We fix $\alpha_{LH} = \alpha_{HL}$ such that the mean sojourn time in both states ranges from 0.6 to 600 sec. Results are

Table 2 Empty and full buffer probability

σ	0.1	0.5	1.0	2	5
Empty buffer probability	0.02628	0.02624	0.02620	0.02614	0.02608
Full buffer probability	0.02257	0.02258	0.02259	0.0226	0.02263

Fig. 7 Buffer occupancy (measured in 128 kb chunks) distribution versus the network behaviour

reported in Fig. 7. When the network status changes more rapidly the distribution is more uniform, otherwise when changes are less frequent the distribution is more skewed.

6 Conclusion

This paper considers the stationary solution of second order bounded fluid models. In particular it focuses on the boundary equations for the absorbing and reflecting cases. The set of second order equations is reduced to a set of first order equations by adding extra variables. The resulting equations are expressed in matrix form in a way that can be easily implemented into an algorithm.

The equations for the general case where the model can have states belonging to all the possible subsets are not presented for space constraint. However, they can be derived from the equations provided in Sect. 4.

We implemented the numerical analysis in Mathematica, and used it to compute the solutions for the example presented in Sect. 5. We are currently working on a more efficient implementation of the algorithm in C. Further improvements can be achieved by developing a matrix exponentiation algorithm that exploits the particular block structure of the matrix defining the second order fluid model.

References

- Agapie, M., & Sohraby, K. (2001). Algorithmic solution to second order fluid flow. In *Proc. of IEEE infocom*, Anchorage, Alaska, Usa, Apr. 2001.
- Ahn, S., & Ramaswami, V. (2003). Fluid flow models and queues—a connection by stochastic coupling. *Communications in Statistics. Stochastic Models*, 19(3), 325–348.
- Ahn, S., & Ramaswami, V. (2004). Transient analysis of fluid flow models via stochastic coupling. *Communications in Statistics. Stochastic Models*, 20(1), 71–101.

- Ang, E.-J., & Barria, J. (2000). The Markov modulated regulated Brownian motion: a second-order fluid flow model of a finite buffer. *Queueing Systems*, 35, 263–287.
- Anick, D., Mitra, D., & Sondhi, M. M. (1982). Stochastic theory of a data-handling system with multiple sources. *Bell System Technical Journal*, 61(8), 1871–1894.
- Asmussen, S. (1995). Stationary distributions for fluid flow models with or without Brownian noise. *Stochastic Models*, 11, 1–20.
- Bean, N. G., O'Reilly, M. M., & Taylor, P. G. (2005a). Algorithms for the first return probabilities for stochastic fluid flows. *Stochastic Models*, 21(1).
- Bean, N. G., O'Reilly, M. M., & Taylor, P. G. (2005b). Hitting probabilities and hitting times for stochastic fluid flows. *Stochastic Processes and their Applications*, 115, 1530–1556.
- Chen, D.-Y., Hong, Y., & Trivedi, K. S. (2002). Second order stochastic fluid flow models with fluid dependent flow rates. *Performance Evaluation*, 49(1–4), 341–358.
- Cox, D. R., & Miller, H. D. (1972). *The theory of stochastic processes*. London: Chapman & Hall.
- da Silva Soares, A., & Latouche, G. (2002). Further results on the similarity between fluid queues and QBDs. In G. Latouche & P. Taylor (Eds.), *Proc. of the 4th int. conf. on matrix-analytic methods* (pp. 89–106). Adelaide, 2002. Singapore: World Scientific.
- da Silva Soares, A., & Latouche, G. (2006). Matrix-analytic methods for fluid queues with finite buffers. *Performance Evaluation*, 63(4), 295–314.
- German, R., Gribaudo, M., Horváth, G., & Telek, M. (2003). Stationary analysis of FSPNs with mutually dependent discrete and continuous parts. In *International conference on petri net performance models—PNPM 2003* (pp. 30–39). Urbana, IL, USA, Sept. 2003. New York: IEEE CS Press.
- Gribaudo, M., & German, R. (2001). Numerical solution of bounded fluid models using matrix exponentiation. In *Proc. 11th GI/ITG conference on measuring, modelling and evaluation of computer and communication systems (MMB)*. Aachen, Germany, Sep. 2001. VDE Verlag.
- Horton, G., Kulkarni, V. G., Nicol, D. M., & Trivedi, K. S. (1998). Fluid stochastic Petri nets: theory, application, and solution techniques. *European Journal of Operations Research*, 105(1), 184–201.
- Igel'nik, B., Kogan, Y., Kriman, V., & Mitra, D. (1995). A new computational approach for stochastic fluid models of multiplexers with heterogeneous sources. *Queueing Systems—Theory and Applications*, 20, 85–116.
- Karandikar, R. L., & Kulkarni, V. G. (1995). Second-order fluid flow models: reflected Brownian motion in a random environment. *Operations Research*, 43, 77–88.
- Kobayashi, H., & Ren, Q. (1992). A mathematical theory for transient analysis of communication networks. *IEICE Transactions Communications*, E75-B(12), 1266–1276.
- Kosten, L. (1984). Stochastic theory of data-handling systems with groups of multiple sources. In *Proceedings of the IFIP WG 7.3/TC 6 second international symposium on the performance of computer-communication systems* (pp. 321–331). Zurich, Switzerland.
- Mitra, D. (1987). Stochastic fluid models. In *Proceedings of performance '87* (pp. 39–51). Brussels, Belgium.
- Mitra, D. (1988). Stochastic theory of a fluid model of producers and consumers coupled by a buffer. *Advances in Applied Probability*, 20, 646–676.
- Rabehasaina, L., & Sericola, B. (2003). Stability analysis of second order fluid flow models in a stationary ergodic environment. *Annals of Applied Probability*, 13(4).
- Ramaswami, V. (1996). Matrix analytic methods for stochastic fluid flows. In D. Smith & P. Hey (Eds.), *Proc. ITC 16* (pp. 1019–1030). Edinburgh. Amsterdam: Elsevier.
- Ren, Q., & Kobayashi, H. (1995). Transient solutions for the buffer behavior in statistical multiplexing. *Performance Evaluation*, 23, 65–87.
- Rogers, L. C. G. (1994). Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Annals of Applied Probability*, 4(2), 390–413.
- Rogers, L. C. G., & Shi, Z. (1994). Computing the invariant law of a fluid model. *Journal of Applied Probability*, 31(4), 885–896.
- Sericola, B. (1998). Transient analysis of stochastic fluid models. *Performance Evaluation*, 32(4).
- Stern, T. E., & Elwalid, A. I. (1991). Analysis of separable Markov-modulated rate models for information-handling systems. *Advances in Applied Probability*, 23, 105–139.
- Tanaka, T., Hashida, O., & Takahashi, Y. (1995). Transient analysis of fluid models for ATM statistical multiplexer. *Performance Evaluation*, 23, 145–162.
- Wolter, K. (1997). Second order fluid stochastic petri nets: an extension of GSPNs for approximate and continuous modelling. In *Proc. of world congress on system simulation* (pp. 328–332). Singapore, Sep. 1997.