QUASI-BIRTH-AND-DEATH PROCESSES, LATTICE PATH COUNTING, AND HYPERGEOMETRIC FUNCTIONS

J. S. H. van Leeuwaarden [1]

M. S. Squillante [2]

E. M. M. Winands [3]

Abstract

This paper considers a class of quasi-birth-and-death processes for which explicit solutions can be obtained for the rate matrix $\mathbf{R}$ and the associated matrix $\mathbf{G}$. The probabilistic interpretations of these matrices allow us to describe their elements in terms of paths on the two-dimensional lattice. Then determining explicit expressions for the matrices becomes equivalent to solving a lattice path counting problem, the solution of which is derived using path decomposition, Bernoulli excursions, and hypergeometric functions. A few applications are provided, including classical models for which we obtain some new results.

Keywords: quasi-birth-and-death processes, matrix-analytic methods, rate matrix, lattice path counting, hypergeometric functions.

1 INTRODUCTION

Multidimensional Markov processes arise in many fields of science, engineering and business. The two-dimensional case is of particular theoretical and practical importance, often occurring directly or through decomposition of higher dimensional processes. In this paper our general interest is in a class of two-dimensional Markov processes, called quasi-birth-and-death (QBD) processes, whose transitions are skipfree to the left and to the right, with no restrictions upward or downward, in the two-dimensional lattice. A wide variety of stochastic models fall within this class, including random walks in the quarter-plane with unit-bounded jumps and many applications in queueing theory.

The invariant distributions of QBD processes, under appropriate conditions, are well known to have a matrix-geometric form. More precisely, the stationary probability vector has a geometric solution in terms of a so-called rate matrix $\mathbf{R}$, which is closely related to another matrix, typically denoted by $\mathbf{G}$. Together, $\mathbf{R}$ and $\mathbf{G}$

---

[1]Eindhoven University of Technology and EURANDOM, P.O. Box 513 - 5600 MB Eindhoven, The Netherlands. Email address: j.s.h.v.leeuwaarden@tue.nl

[2]IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA. Email address: mss@watson.ibm.com

[3]Eindhoven University of Technology, P.O. Box 513 - 5600 MB Eindhoven, The Netherlands. Email address: e.m.m.winands@tue.nl

play a fundamental role in the general theory of matrix-analytic methods. These two matrices are most often obtained by numerically computing the solutions of nonlinear matrix equations, and many algorithms have been developed for doing so.

On the other hand, the probabilistic significance of the $\mathbf{R}$ and $\mathbf{G}$ matrices have been well established. Hence, in theory, these probabilistic interpretations can be used together with the structural properties of the QBD process to derive explicit expressions for the elements of $\mathbf{R}$ and $\mathbf{G}$.

The research literature is quite limited with respect to explicit solutions for the $\mathbf{R}$ and $\mathbf{G}$ matrices. Ramaswami and Latouche [16] explicitly determine $\mathbf{R}$ and $\mathbf{G}$ in QBD processes for the special cases when the rows of the birth or death transition matrix are proportional to a common row vector, allowing the state space to be infinite in both dimensions. These results for the special birth transition case were much later extended by Liu and Zhao [12] to Markov processes of the GI/M/1-type and M/G/1-type. A few studies consider QBD processes with less restrictive transitions, but further restrict attention to the $\mathbf{R}$ matrix and to state spaces infinite in only one dimension. Squillante and Nelson [13, 18, 19, 20] exploit the probablistic interpretation of $\mathbf{R}$ to determine explicit solutions for its elements in various stochastic models based on path decomposition and lattice path counting. van Leeuwaarden and Winands [11] present an explicit description of $\mathbf{R}$ for a class of QBD processes, which includes those in [18, 19, 20], based on the probabilistic interpretation of $\mathbf{R}$ and lattice path counting.

In this paper we investigate a more general class of QBD processes for which we exploit the probabilistic interpretations of $\mathbf{R}$ and $\mathbf{G}$ and derive solutions for these matrices using path decomposition, Bernoulli excursions, and lattice path counting. Our results are then used to determine explicit expressions for the elements of $\mathbf{R}$ and $\mathbf{G}$ in terms of hypergeometric functions. These explicit expressions can be used in turn to obtain explicit solutions for the stationary distribution and other probability measures of interest. Our approach, which combines various areas of applied probability such as matrix-analytic methods and lattice path counting, is quite general and can be applied to a wide variety of QBD processes.

We make several contributions, the first being the introduction of a class of QBD processes for which the $\mathbf{R}$ and $\mathbf{G}$ matrices can be explicitly determined. This class includes QBD processes that are infinite in one or both dimensions and that allow less restrictive transitions than in previous studies, thus making the mathematical problem much more intricate and significantly increasing the applicability of the results. Stochastic models that turn out to be members of this class include certain random walks in the quarter-plane and various queueing systems such as the classical longest queue, priority and feedback models.

A second contribution is that our exact expressions for the fundamental matrix elements provide an explicit characterization of the probabilistic and dynamic behavior of the stochastic process itself. Moreover, the results derived within our framework for a general class of QBD processes lead to new results for some specific stochastic models, e.g., the longest queue model.

A final contribution is that our results provide an efficient alternative for computing the invariant distribution and other probability measures over those based on numerical algorithms. Moreover, these results may facilitate the implementation

and performance of more effective numerical methods. As an example, numerical algorithms for QBD processes that are infinite in both dimensions must be intrinsically based on truncation of one of the dimensions. The explicit results derived in this study can be used to reveal the potential errors induced by such truncations.

In Section 2 we consider a class of homogeneous QBD processes that serve as the initial basis for our analysis. In Section 3 we solve for each element of $\mathbf{R}$ and $\mathbf{G}$ in terms of counting paths on the two-dimensional lattice, and then in Section 4 we reformulate these solutions in terms of hypergeometric functions. In Section 5 we remove several initial assumptions made to simplify the presentation and discuss various generalizations of our analysis. Section 6 then presents some classical stochastic models that fall within the introduced class of QBD processes.

2 QBD PROCESSES

Consider a continuous-time Markov process $\{X(t),\, t \in \mathbb{R}_+\}$ on the two-dimensional state space $\{(n,j) : n \in \mathbb{Z}_+, j \in \{1,\ldots,M\}\}$ that is partitioned as $\bigcup_{n=0}^{\infty} l(n)$, where

$$l(n) = \{(n,1),(n,2),\ldots,(n,M)\} \tag{2.1}$$

and $\mathbb{Z}_+$, $\mathbb{R}_+$ denote the nonnegative integer and real numbers. The first coordinate n is called the *level* and the second coordinate j is called the *phase* of state (n,j), with the set $l(n)$ referred to as *level n*. Each level may have a finite or infinite number of states, M.

This Markov process is called a QBD process when its one-step transitions from each state are restricted to states in the same level or in the two adjacent levels, and a *homogeneous* QBD process when these transition rates are additionally level-independent. The infinitesimal generator $\mathbf{Q}$ of the Markov process then takes the block tridiagonal form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B} & \mathbf{A}_0 & & & \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{2.2}$$

where $\mathbf{A}_0, \mathbf{A}_2$ are nonnegative and $\mathbf{A}_1, \mathbf{B}$ have nonnegative off-diagonal elements and strictly negative diagonals. Each of these matrices has dimension $M \times M$.

Let $\boldsymbol{\pi}$ denote the stationary probability vector of this homogeneous QBD process. We shall assume throughout the paper that the QBD process is irreducible and ergodic. Hence, the stationary probability vector exists and is uniquely determined as the solution of $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{e} = 1$, where $\mathbf{e}$ denotes a column vector of appropriate dimension containing all ones. We partition the vector $\boldsymbol{\pi}$ by levels into subvectors $\boldsymbol{\pi}_n$, $n \in \mathbb{Z}_+$, where $\boldsymbol{\pi}_n$ has M components corresponding to the states of $l(n)$. The matrix-geometric solution of this partitioned stationary probability vector is given by the following theorem.

Theorem 1. *Consider a continuous-time QBD process with infinitesimal generator* $\mathbf{Q}$ *in the form of* (2.2). *Suppose that the QBD process is irreducible and ergodic.*

Then its stationary distribution $\boldsymbol{\pi}$ is given by

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \mathbf{R}^n, \quad n \in \mathbb{N}, \tag{2.3}$$

where $\mathbf{R}$ is the minimal nonnegative solution of the nonlinear matrix equation

$$\mathbf{A}_0 + \mathbf{R}\mathbf{A}_1 + \mathbf{R}^2\mathbf{A}_2 = \mathbf{0} \tag{2.4}$$

with $\mathrm{sp}(\mathbf{R}) < 1$. Furthermore, the stationary probability vector $\boldsymbol{\pi}_0$ exists and is uniquely determined by solving the boundary condition

$$\boldsymbol{\pi}_0\mathbf{B} + \boldsymbol{\pi}_1\mathbf{A}_2 \;=\; \boldsymbol{\pi}_0(\mathbf{B} + \mathbf{R}\mathbf{A}_2) \;=\; \mathbf{0} \tag{2.5}$$

and the normalization condition

$$\sum_{i=0}^{\infty} \boldsymbol{\pi}_i\mathbf{e} \;=\; \boldsymbol{\pi}_0(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = 1, \tag{2.6}$$

where $\mathbf{I}$ denotes the identity matrix of appropriate dimension.

This theorem is due to Neuts for the case $M < \infty$ (see [14]); the theorem for the case $M = \infty$ follows from the results of Tweedie [22] (see also [10, 16]) with the spectrum and inverse related to $\mathbf{R}$ appropriately defined (see [6, 7]). From Theorem 1 we know that the stationary distribution is determined once $\mathbf{R}$ is obtained. Several iterative algorithms exist for numerically solving (2.4); an overview of such algorithms is provided in [10].

A related matrix, typically denoted by $\mathbf{G}$, also plays an important role together with $\mathbf{R}$ in the general theory of matrix-analytic methods. This related $\mathbf{G}$ matrix is the minimal nonnegative solution of the nonlinear matrix equation

$$\mathbf{A}_0\mathbf{G}^2 + \mathbf{A}_1\mathbf{G} + \mathbf{A}_2 = \mathbf{0} \tag{2.7}$$

(see [10, 14, 15]). Some recent algorithms for numerically solving (2.4) involve first computing the matrix $\mathbf{G}$ and then computing the matrix $\mathbf{R}$ based on the relationship

$$\mathbf{R} = \mathbf{A}_0(-[\mathbf{A}_1 + \mathbf{A}_0\mathbf{G}])^{-1} \tag{2.8}$$

(see [10]).

2.1. Probabilistic interpretations

Our analysis of QBD processes is based on the probabilistic significance of $\mathbf{R}$ and $\mathbf{G}$. The matrix $\mathbf{R}$ records the expected rate of visits to the states of $l(n+1)$ per unit of local time in $l(n)$. More precisely, the element R_{ij} is the expected sojourn time in $(n+1, j)$ before the first return to $l(n)$ given that the process starts in (n, i), expressed in units of the mean sojourn time in (n, i) (see [10, 14]). Equivalently,

$$R_{ij} = V_{ij} \cdot \frac{\boldsymbol{\Delta}_{ii}}{\boldsymbol{\Delta}_{jj}}, \tag{2.9}$$

where V_{ij} is the expected number of visits to $(n+1, j)$ before the first return to $l(n)$ given that the process starts in (n, i), and $\mathbf{\Delta} = -\operatorname{diag}(\mathbf{A}_1)$ so that the fraction in (2.9) is the ratio of the ith and jth diagonal elements of $\mathbf{A}_1$.

The matrix $\mathbf{G}$ records the probability, starting from $l(n+1)$, of visiting level $l(n)$ in finite time. Suppose that the process is in $l(n+1)$ at time 0 and define τ as the first passage time from $l(n+1)$ to $l(n)$. The element G_{ij} then has the probabilistic description

$$G_{ij} = \mathbb{P}[\tau < \infty, \, X(\tau) = (n, j) \,|\, X(0) = (n+1, i)] \qquad (2.10)$$

(see [10, 14, 15]). Under our assumption that $\{X(t), \, t \in \mathbb{R}_+\}$ is irreducible and ergodic, it follows that $\tau < \infty$ with probability 1 and, thus,

$$G_{ij} = \mathbb{P}[X(\tau) = (n, j) \,|\, X(0) = (n+1, i), \, \tau < \infty]. \qquad (2.11)$$

Due to the homogeneity of $\mathbf{Q}$, the elements of $\mathbf{R}$ and $\mathbf{G}$ do not depend on $n \in \mathbb{Z}_+$. We shall exploit the probabilistic interpretations (2.9) and (2.11) in our analysis of the next two sections to obtain explicit solutions for $\mathbf{R}$ and $\mathbf{G}$.

3 Lattice path counting

Let us denote by $\langle e_1, e_2 \rangle$ a one-step transition of the QBD process from state (n, j) to state $(n+e_1, j+e_2)$, $n \in \mathbb{N}$, where the probability that this step occurs is denoted by $\varphi \langle e_1, e_2 \rangle$ and obtained straightforwardly from uniformization using $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{\Delta}$. We shall restrict our attention to the class of QBD processes for which the only transitions from state (n, j), $n \in \mathbb{N}$, $j \in \{1, \dots, M-1\}$, are horizontal steps $\langle -1, 0 \rangle$ and $\langle 1, 0 \rangle$, diagonal steps $\langle -1, 1 \rangle$ and $\langle 1, 1 \rangle$, and vertical step $\langle 0, 1 \rangle$; and the only transitions from state (n, M), $n \in \mathbb{N}$, are horizontal steps $\langle -1, 0 \rangle$ and $\langle 1, 0 \rangle$.

The above restrictions imply that, from the states in $l(n)$, $n \in \mathbb{N}$, the QBD process cannot move in the downward direction, and thus the matrices $\mathbf{R}$ and $\mathbf{G}$ are of uppertriangular form. To elucidate the exposition, we further restrict our attention to the case where M is infinite and the diagonal elements of $\mathbf{A}_1$ are all equal. The former restriction is not prohibitive and its relaxation is discussed in Section 5, whereas the latter simply implies $R_{ij} = V_{ij}$ such that we would instead use (2.9) whenever the restriction does not hold. Then, with $R_h \equiv R_{i,i+h}$ and $G_h \equiv G_{i,i+h}$, we obtain

$$\mathbf{R} = \begin{pmatrix} R_0 & R_1 & R_2 & \dots \\ & R_0 & R_1 & \dots \\ & & R_0 & \dots \\ & & & \ddots \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} G_0 & G_1 & G_2 & \dots \\ & G_0 & G_1 & \dots \\ & & G_0 & \dots \\ & & & \ddots \end{pmatrix}. \qquad (3.1)$$

3.1. G matrix

We now derive expressions for each element of $\mathbf{G}$ in (3.1) using its probabilistic interpretation. Assume that at time τ the QBD process undergoes its νth transition. In order to calculate the probability in (2.11), we then need to count all of the paths on the two-dimensional lattice that start in $(n+1, i)$ and end in $(n, i+h)$ after ν

steps, without visiting any of the states in $l(n)$ prior to the νth transition. Each step of the path will be one of the steps $\langle -1, 0 \rangle$, $\langle 1, 0 \rangle$, $\langle -1, 1 \rangle$, $\langle 1, 1 \rangle$, and $\langle 0, 1 \rangle$.

The main idea in deriving a tractable lattice path counting problem is the decoupling of the paths into horizontal and vertical directions.

(i) Consider a path from state $(n+1, i)$ to state $(n, i+h)$, $h \in \mathbb{Z}_+$, that consists of ν steps and that goes from $l(n+1)$ to $l(n)$ only at the last (νth) step. Assume this path includes s $\langle -1, 1 \rangle$ steps, u $\langle 1, 1 \rangle$ steps and, hence, $t = h - s - u$ $\langle 0, 1 \rangle$ steps.

(ii) We first would like to consider the path in the horizontal direction only. The diagonal steps $\langle -1, 1 \rangle$ and $\langle 1, 1 \rangle$ influence both the horizontal and vertical directions. Therefore, we decompose these diagonal steps into

$$\langle -1, 1 \rangle \;=\; \langle 0, 1 \rangle \;+\; \langle -1, 0 \rangle, \tag{3.2}$$
$$\langle 1, 1 \rangle \;=\; \langle 0, 1 \rangle \;+\; \langle 1, 0 \rangle. \tag{3.3}$$

(iii) The decomposition of the diagonal steps ensures that the path includes at least s $\langle -1, 0 \rangle$ steps and u $\langle 1, 0 \rangle$ steps. Now denote the total number of $\langle 1, 0 \rangle$ steps by m. We then know that the total number of $\langle -1, 0 \rangle$ steps is $m+1$ (including the νth step). Furthermore, it should hold that $m \geq \max(u, s-1)$.

(iv) The path then consists of a total number of $\nu = 2m + t + 1$ steps, $2m + 1$ of which are in the horizontal direction. When we omit the νth step, the $2m$ horizontal steps form a Bernoulli excursion (see [21]). The excursion starts in $(n+1, i)$ and consists of m $\langle 1, 0 \rangle$ steps and m $\langle -1, 0 \rangle$ steps. Any sequence of steps may occur provided that there are, at each point during the excursion, at least as many $\langle 1, 0 \rangle$ steps as $\langle -1, 0 \rangle$ steps; for otherwise, $l(n)$ is visited, and the condition in (i) is violated. The number of possible Bernoulli excursions is given by the mth Catalan number:

$$\frac{1}{m+1} \binom{2m}{m}. \tag{3.4}$$

To further elucidate the main idea, let us illustrate the above procedure with an example. In Figure 1(a) we depict a path from state $(n+1, i)$ to state $(n, i+5)$ that consists of $\nu = 21$ steps and that goes from $l(n+1)$ to $l(n)$ only at the νth step. We have indicated in Figure 1(a) the number of times each state is visited. The path includes $s = 2$ $\langle -1, 1 \rangle$ steps, $u = 1$ $\langle 1, 1 \rangle$ step, and $t = 2$ $\langle 0, 1 \rangle$ steps. The diagonal steps are decomposed, leading to the path in Figure 1(b). We then remove the vertical steps to obtain the Bernoulli excursion shown in Figure 1(c). This path is just one out of $\frac{1}{10}\binom{18}{9} = 4862$ possible Bernoulli excursions that consist of $2m = 18$ steps. In the above procedure we consider just one path from state (n, i) to state $(n, i+h)$, and reduce it to a Bernoulli excursion.

Let us now change perspective to reflect our original lattice counting problem. Namely, starting from a Bernoulli excursion of length $2m$, we consider how many paths from state (n, i) to state $(n, i+h)$ can be constructed with s $\langle -1, 1 \rangle$ steps, u $\langle 1, 1 \rangle$ steps and, thus, $t = h - s - u$ $\langle 0, 1 \rangle$ steps. First, some definitions are needed.
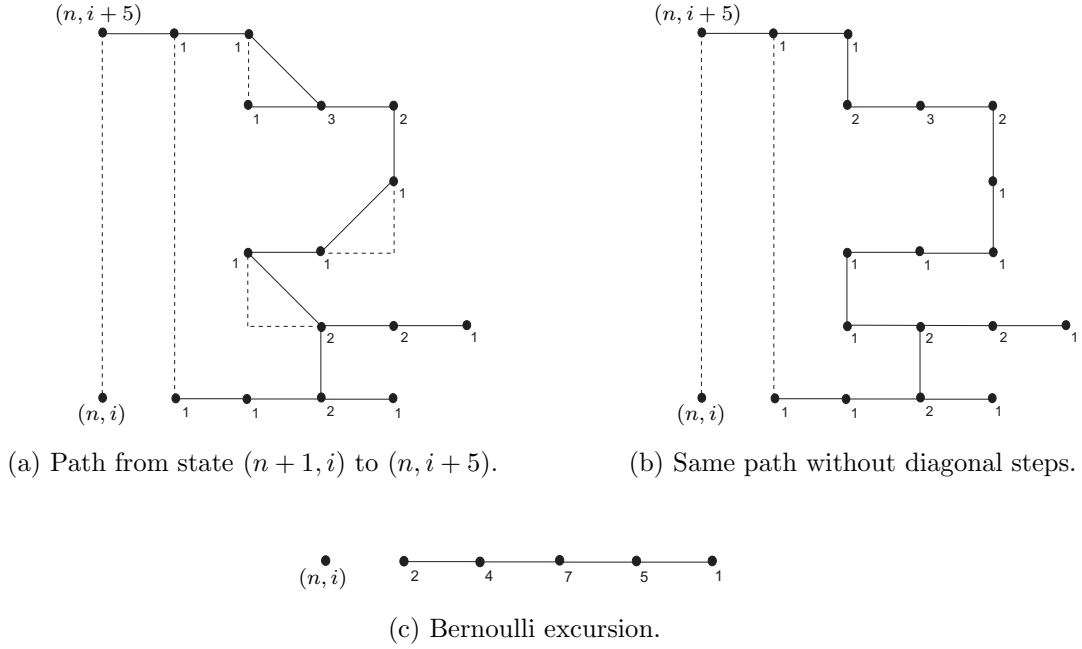
6

(a) Path from state $(n+1, i)$ to $(n, i+5)$.

(b) Same path without diagonal steps.

(c) Bernoulli excursion.

Figure 1: Example of path decomposition and corresponding Bernoulli excursion.

Definition 1. *Define $L_h(s, u, m)$ to be the number of paths from $(n+1, i)$ to $(n, i+h)$, $h \in \mathbb{Z}_+$, that consist of $\nu = 2m+1+t$ steps and that go from $l(n+1)$ to $l(n)$ only at the νth step. Assume each path includes $s \langle -1, 1 \rangle$ steps, $u \langle 1, 1 \rangle$ steps and, hence, $t = h - s - u \langle 0, 1 \rangle$ steps. Define $P_h(s, u, m)$ to be the probability of each such path.*

Obviously, we have

$$P_h(s, u, m) = \varphi \langle -1, 1 \rangle^s \varphi \langle 0, 1 \rangle^t \varphi \langle 1, 1 \rangle^u \varphi \langle 1, 0 \rangle^{m-u} \varphi \langle -1, 0 \rangle^{m+1-s}. \qquad (3.5)$$

For $L_h(s, u, m)$ we establish the following result.

Lemma 1. *Given values of s and u such that $s + u + t = h$, we have*

$$L_h(s, u, m) = \frac{1}{m+1} \binom{2m}{m} \binom{m+1}{s} \binom{m}{u} \binom{2m+t}{t}. \qquad (3.6)$$

Proof.

(v) Consider a Bernoulli excursion of length $2m$. We will extend the Bernoulli excursion, which describes the path in the horizontal direction, by the vertical steps to reconstruct a path from state $(n+1, i)$ to state $(n, i+h)$. The vertical steps consist of $s+t+u \langle 0, 1 \rangle$ steps. However, because of the decomposition in (ii), s steps should be matched to $\langle -1, 0 \rangle$ steps and u steps should be matched to $\langle 1, 0 \rangle$ steps. The number of ways to do this is

$$\binom{m+1}{s} \binom{m}{u}. \qquad (3.7)$$

7

(vi) This leaves the placement of the t original $\langle 0, 1 \rangle$ steps, which can be placed in every state visited by the Bernoulli excursion. This means that we have $2m + 1$ possible states in which the t $\langle 0, 1 \rangle$ steps can be placed. Note that it is possible for multiple steps to be placed in one state. Therefore, placing the t $\langle 0, 1 \rangle$ steps is equivalent to distributing t balls over $2m + 1$ bins, and the number of ways to do so is

$$\binom{2m + t}{t}. \tag{3.8}$$

Combining (3.4), (3.7) and (3.8) completes the proof. $\qquad\square$

Using (2.11) and $G_h = G_{i,i+h}$, we arrive at our main result for the elements of $\mathbf{G}$.

Theorem 2. *For all $h = 0, 1, \ldots$, we have*

$$G_h = \sum_{s=0}^{h} \sum_{u=0}^{h-s} \sum_{m=\max(u,s-1)}^{\infty} L_h(s, u, m) \cdot P_h(s, u, m), \tag{3.9}$$

with $P_h(s, u, m)$ as in (3.5) and $L_h(s, u, m)$ as in (3.6).

To illustrate this result, we return to the example path in Figure 1(a)-(c). Starting from the Bernoulli excursion in Figure 1(c), the path in Figure 1(a) is just one of the $\binom{10}{2}\binom{9}{1}\binom{20}{2} = 76950$ paths that can be constructed with $s = 2$ $\langle -1, 1 \rangle$ steps, $u = 1$ $\langle 1, 1 \rangle$ step, and $t = 2$ $\langle 0, 1 \rangle$ steps.

3.2. $\mathbf{R}$ MATRIX

We shall now derive expressions for each element of $\mathbf{R}$ in (3.1) using the probabilistic interpretation in (2.9). This requires us to determine $V_{i,i+h}$, the expected number of visits to state $(n + 1, i + h)$ in the two-dimensional lattice before a return to $l(n)$ given that the process starts in (n, i). The main idea is to decompose the lattice path counting problem into first passage and return visits.

Definition 2. *Define κ_h to be the probability that the QBD process reaches state $(n + 1, i + h)$, $h \in \mathbb{Z}_+$, before the first visit to $l(n)$, given that the process started in state $(n + 1, i)$.*

Definition 3. *Define γ to be the expected number of visits to state $(n + 1, i + h)$ before the first visit to $l(n)$ given that state $(n + 1, i + h)$ is reached at least once.*

It is then clear that

$$R_h = V_{i,i+h} = (\varphi\langle 1, 0 \rangle \kappa_h + \varphi\langle 1, 1 \rangle \kappa_{h-1}) \cdot \gamma. \tag{3.10}$$

We first consider γ and suppose that the QBD process enters state $(n + 1, i + h)$ for the first time. Then, the path will go through state $(n + 1, i + h)$ for a second time (before visiting $l(n)$) when the process moves to $(n + 2, i + h)$, starts a Bernoulli excursion of $2m$ steps $(m = 0, 1, \ldots)$, and returns from state $(n + 2, i + h)$ to state

8

$(n + 1, i + h)$. With $\omega = \varphi\langle 1, 0\rangle\varphi\langle -1, 0\rangle \leq 1/4$, the probability that this event happens is given by

$$\eta = \omega \sum_{m=0}^{\infty} \frac{1}{m + 1} \binom{2m}{m} \omega^m = \frac{1 - \sqrt{1 - 4\omega}}{2} \tag{3.11}$$

(see, e.g., [21], p. 561). Thus, since $\gamma = 1 + \eta\gamma$, we eventually obtain

$$\gamma = \frac{2}{1 + \sqrt{1 - 4\varphi\langle 1, 0\rangle\varphi\langle -1, 0\rangle}}. \tag{3.12}$$

Now turning to the probabilities κ_h, it obviously holds that $\kappa_0 = 1$ and some reasoning yields

$$\kappa_h = \varphi\langle 0, 1\rangle\kappa_{h-1} + \varphi\langle 1, 0\rangle \sum_{j=0}^{h} G_{h-j}\kappa_j + \varphi\langle 1, 1\rangle \sum_{j=0}^{h-1} G_{h-j-1}\kappa_j, \quad h = 1, 2, \ldots. \tag{3.13}$$

Rewriting this expression renders, for $h = 1, 2, \ldots$,

$$\kappa_h = \frac{\varphi\langle 0, 1\rangle\kappa_{h-1} + \varphi\langle 1, 0\rangle \sum_{j=0}^{h-1} G_{h-j}\kappa_j + \varphi\langle 1, 1\rangle \sum_{j=0}^{h-1} G_{h-j-1}\kappa_j}{1 - \varphi\langle 1, 0\rangle G_0}. \tag{3.14}$$

A combination of (3.10), (3.12) and (3.14) leads to our main result for a description of the elements of $\mathbf{R}$.

Theorem 3. *For all $h = 0, 1, \ldots$, we have*

$$R_h = \frac{2(\varphi\langle 1, 0\rangle\kappa_h + \varphi\langle 1, 1\rangle\kappa_{h-1})}{1 + \sqrt{1 - 4\varphi\langle 1, 0\rangle\varphi\langle -1, 0\rangle}}, \tag{3.15}$$

with κ_h as in (3.14), $\kappa_0 = 1$, and κ_{-1} taken to be zero.

4 HYPERGEOMETRIC FUNCTIONS

We next express our results for the elements of $\mathbf{G}$ and $\mathbf{R}$ in terms of hypergeometric functions, which are defined as

$$F(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a + n)\Gamma(b + n)}{\Gamma(c + n)} \frac{z^n}{n!}, \tag{4.1}$$

with $\Gamma(\cdot)$ the Gamma function. We shall need the following result.

Lemma 2. *For $n = 1, 2, \ldots$ and $t = 1, 2, \ldots$,*

$$\Gamma(n + t/2) = \frac{\sqrt{\pi}}{2^{2(n-1)+t}} \frac{\Gamma(2n + t - 1)}{\Gamma(n + t/2 - 1/2)}. \tag{4.2}$$

Proof. The case $t = 1$ can be found in [1], pp. 255-6, and is given by

$$\Gamma(p + 1/2) = \frac{\sqrt{\pi}}{2^{2p-1}} \frac{\Gamma(2p)}{\Gamma(p)}, \quad p = 1, 2, \ldots . \tag{4.3}$$

For $t = 2m$, $m = 1, 2, \ldots$, identity (4.2) reduces to

$$\Gamma(n + m) = \frac{\sqrt{\pi}}{2^{2(n+m-1)}} \frac{\Gamma(2(n+m) - 1)}{\Gamma(n + m - 1/2)}. \tag{4.4}$$

This yields

$$\begin{aligned}
\Gamma(n + m - 1/2) &= \frac{\sqrt{\pi}}{2^{2(n+m-1)}} \frac{\Gamma(2(n+m) - 1)}{\Gamma(n + m)} \\
&= \frac{\sqrt{\pi}}{2^{2(n+m-1)-1}} \frac{\Gamma(2(n+m-1))}{\Gamma(n+m-1)},
\end{aligned} \tag{4.5}$$

which is equivalent to (4.3) for $p = m + n - 1$. In the case $t = 2m + 1$, $m = 0, 1, \ldots$, identity (4.2) can be written as

$$\Gamma(m + n + 1/2) = \frac{\sqrt{\pi}}{2^{2(n+m)-1}} \frac{\Gamma(2(n+m))}{\Gamma(n + m)}, \tag{4.6}$$

which is equivalent to (4.3) for $p = m + n$. This completes the proof. $\square$

We now consider the infinite series in (3.9):

$$\sum_{m=\max(u,s-1)}^{\infty} L_h(s, u, m) \cdot P_h(s, u, m), \tag{4.7}$$

with $P_h(s, u, m)$, $L_h(s, u, m)$ as in (3.5), (3.6). For the case $s - 1 \geq u$, we can express (4.7) in terms of the Gamma function (see [1], pp. 255ff) as

$$\frac{\varphi\langle -1, 1\rangle^s \varphi\langle 0, 1\rangle^t \varphi\langle 1, 1\rangle^u \varphi\langle 1, 0\rangle^{s-1-u}}{s! \, t! \, u!} \sum_{m=0}^{\infty} \frac{\Gamma(2m + 2s + t - 1)}{\Gamma(m + s - u)} \frac{\omega^m}{m!}, \tag{4.8}$$

where $\omega = \varphi\langle 1, 0\rangle \varphi\langle -1, 0\rangle$ as previously defined. From Lemma 2 we obtain

$$\Gamma(2m + 2s + t - 1) = \frac{2^{2(m+s-1)+t}}{\sqrt{\pi}} \Gamma(m + s + \tfrac{t}{2}) \Gamma(m + s + \tfrac{t-1}{2}) \tag{4.9}$$

and

$$\frac{2^{2s-2+t}}{\sqrt{\pi}} = \frac{\Gamma(2s + t - 1)}{\Gamma(s + \tfrac{t}{2}) \Gamma(s + \tfrac{t-1}{2})}. \tag{4.10}$$

Substituting (4.9) and (4.10) into the series in (4.8), and using (4.1), yields

$$\begin{aligned}
\sum_{m=0}^{\infty} \frac{\Gamma(2m + 2s + t - 1)}{\Gamma(m + s - u)} \frac{\omega^m}{m!} &= \frac{2^{2s-2+t}}{\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{\Gamma(m + s + \tfrac{t}{2}) \Gamma(m + s + \tfrac{t-1}{2})}{\Gamma(m + s - u)} \frac{(4\omega)^m}{m!} \\
&= \frac{(2s + t - 2)!}{(s - u - 1)!} \, F(s + \tfrac{t}{2}, s + \tfrac{t-1}{2}; s - u; 4\omega), \quad (4.11)
\end{aligned}$$

10

and hence (suppressing the indices of L_h and P_h)

$$
\sum_{m=s-1}^{\infty} L_h \cdot P_h = \varphi\langle -1, 1\rangle^s \varphi\langle 0, 1\rangle^t \varphi\langle 1, 1\rangle^u \varphi\langle 1, 0\rangle^{s-1-u}
$$
$$
\times \frac{(2s+t-2)!}{s!t!u!(s-u-1)!} \, F(s+\tfrac{t}{2}, s+\tfrac{t-1}{2}; s-u; 4\omega).
$$
(4.12)

In a similar manner, we obtain for the case $u > s - 1$ (again suppressing indices)

$$
\sum_{m=u}^{\infty} L_h \cdot P_h = \varphi\langle -1, 1\rangle^s \varphi\langle 0, 1\rangle^t \varphi\langle 1, 1\rangle^u \varphi\langle -1, 0\rangle^{u+1-s}
$$
$$
\times \frac{(2u+t)!}{s!t!u!(u-s+1)!} \, F(u+1+\tfrac{t}{2}, u+1+\tfrac{t-1}{2}; u-s+2; 4\omega).
$$
(4.13)

The hypergeometric function $F(a + \tfrac{1}{2}, a; c; z)$ that emerges in both cases of the infinite series can be rewritten as

$$
F(a+\tfrac{1}{2}, a; c; z) = F(a, a+\tfrac{1}{2}; c; z)
$$
$$
= 2^{c-1}\Gamma(c) z^{\frac{1}{2}-\frac{1}{2}c}(1-z)^{\frac{1}{2}c-a-\frac{1}{2}} P_{2a-c}^{1-c}((1-z)^{-\frac{1}{2}}), \quad (4.14)
$$

with $P_\nu^\mu(\cdot)$ the Legendre function of the first kind (see [1], p. 562). Using the relation

$$
P_\nu^{\mu-1}(z) = \frac{P_{\nu+1}^\mu(z) - P_{\nu-1}^\mu(z)}{(2\nu+1)\sqrt{z^2-1}}
$$
(4.15)

(see [1], p. 334), we can recursively express the Legendre function P_ν^μ in terms of Legendre polynomials P_ν^0. In turn, these Legendre polynomials have the following explicit form

$$
P_\nu^0(x) = \frac{1}{2^\nu} \sum_{m=0}^{\lfloor \frac{\nu}{2} \rfloor} (-1)^m \binom{\nu}{m}\binom{2(\nu-m)}{m} x^{\nu-2m}
$$
(4.16)

(see [1], p. 775), which implies that all of the above hypergeometric functions can be computed in a *finite* number of steps.

5 Generalizations

Many assumptions in the foregoing analysis were made to simplify the presentation. Indeed, our results hold for the class of nonhomogeneous QBD processes in which the levels are divided into a nonhomogeneous boundary, comprised of $l(0), \dots, l(N)$, and a homogeneous portion, comprised of $l(N+1), l(N+2), \dots, N < \infty$. The generator $\mathbf{Q}$ in this case takes the form (2.2) but with the four matrices in the upper leftmost corner replaced by $\mathbf{B}_{00}, \mathbf{B}_{01}, \mathbf{B}_{10}, \mathbf{B}_{11}$ having dimensions $NM \times NM$, $NM \times M$, $M \times NM$, $M \times M$. Equations (2.3) and (2.5), (2.6) in Theorem 1 are then replaced by

$$
\boldsymbol{\pi}_{N+n} = \boldsymbol{\pi}_N \mathbf{R}^n, \quad n \in \mathbb{N}
$$
(5.1)

and

$$(\boldsymbol{\pi}_0, \, \boldsymbol{\pi}_1, \, \ldots, \, \boldsymbol{\pi}_N) \begin{pmatrix} \mathbf{B}_{00} & \mathbf{B}_{01} \\ \mathbf{B}_{10} & \mathbf{B}_{11} + \mathbf{R}\mathbf{A}_2 \end{pmatrix} = \mathbf{0}, \tag{5.2}$$

$$(\boldsymbol{\pi}_0, \, \boldsymbol{\pi}_1, \, \ldots, \, \boldsymbol{\pi}_{N-1}) \, \mathbf{e} + \boldsymbol{\pi}_N(\mathbf{I} - \mathbf{R})^{-1}\mathbf{e} = 1. \tag{5.3}$$

The number of phases for $l(n)$, $n \in \{0, \ldots, N\}$, can be further generalized to an arbitrary M_n. Due to the homogeneous portion of $\mathbf{Q}$, the elements of $\mathbf{R}$ and $\mathbf{G}$ do not depend on $l(N + n)$, $n \in \mathbb{N}$, and are given by the results in Sections 3 and 4.

Furthermore, our analysis directly applies to discrete-time Markov processes since the probabilistic significance of $\mathbf{R}$ and $\mathbf{G}$ in this case are $R_{ij} = V_{ij}$ and the discrete-time version of (2.11) (see [10, 14, 15]). The restriction of $M = \infty$ does not cause any major technical difficulties and our analysis can be extended to handle the $M < \infty$ case; the technical details are somewhat tedious due to the additional boundary and the different transitions for phase M in each level of the homogeneous portion (since no vertical or diagonal transitions are allowed), and therefore are omitted in the present paper. As previously noted, when $\boldsymbol{\Delta}_{ii} \neq \boldsymbol{\Delta}_{jj}$ for one or more $i, j \in \{1, \ldots, M\}$, then our results are straightforwardly adjusted according to (2.9).

More fundamental extensions of the introduced class of QBD processes are also possible. A critical factor in our analysis is that these QBD processes cannot move downward due to directional restrictions on vertical and diagonal transitions. Similar results may be obtained for classes of QBD processes that cannot move upward (due to the opposite directional restrictions), leftward or rightward (both due to directional restrictions on horizontal and diagonal transitions); see [11] for analogous rotations of a strict subset of these transitions. Although the basic approach may not be fundamentally different from our analysis here, many subtleties need to be addressed and specific probabilistic arguments are required in the analysis of this extended class of QBD processes; these technical details remain for future research. Stochastic models falling in the extended class include a make-to-order/make-to-stock system [2], a two-machine re-entrant line [3], and many more.

6   APPLICATIONS

We end this paper by briefly discussing three classical stochastic models that are members of the introduced class of QBD processes. To simplify the presentation, we primarily focus on *symmetric* systems in which two queues are served by a single server. Namely, at each queue customers arrive according to a Poisson process with rate λ and both customer types require exponentially distributed service times with mean μ^{-1}. The models differ, however, in the way the server capacity is shared among the two queues as explained below. Once again, we stress that many more stochastic models fall within the introduced class as touched upon in Section 5.

6.1.   LONGEST QUEUE MODEL

In the first model, the server always serves the longest queue, where ties are resolved randomly and the service policy is applied preemptively. This model has been investigated in several studies; see, e.g., [5, 8, 23]. If (n, j) denotes the state where, in

equilibrium, the difference $D = |L_1 - L_2|$ between the two queue lengths is n and the shortest queue length $L^* = \min(L_1, L_2)$ is j, then the QBD queue length process under the single-server longest queue policy is described by the matrices

$$
\mathbf{A}_0 = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & \ddots \end{pmatrix}, \ \mathbf{B}_{01} = \begin{pmatrix} 2\lambda & & \\ \mu & 2\lambda & \\ & \mu & 2\lambda \\ & & \ddots & \ddots \end{pmatrix}, \ \mathbf{B}_{00} = \begin{pmatrix} -2\lambda & & \\ & \delta & \\ & & \ddots \end{pmatrix}, \quad (6.1)
$$

and

$$
\mathbf{A}_1 = \mathbf{B}_{11} = \begin{pmatrix} \delta & & \\ & \delta & \\ & & \ddots \end{pmatrix}, \quad \mathbf{A}_2 = \mathbf{B}_{10} = \begin{pmatrix} \mu & \lambda & \\ & \mu & \lambda \\ & & \ddots & \ddots \end{pmatrix}, \quad (6.2)
$$

where $N = 1$ and $\delta = -(\mu + 2\lambda)$. The multi-server version of this model also falls within the introduced class of QBD processes, where N reflects the number of servers and the nonhomogeneous boundary is appropriately adjusted.

For the single-server and multi-server longest queue policy, where both $t = 0$ and $u = 0$, (4.12) significantly simplifies as shown below.

Proposition 1. *If $s = 0$ we have*

$$
\sum_{m=0}^{\infty} L_h(0, 0, m) \cdot P_h(0, 0, m) = \frac{1 - \sqrt{1 - 4\omega}}{2\varphi\langle 1, 0\rangle}, \quad (6.3)
$$

and if $s \geq 1$ we have

$$
\sum_{m=s-1}^{\infty} L_h(s, 0, m) \cdot P_h(s, 0, m) = \varphi\langle -1, 1\rangle^s \varphi\langle 1, 0\rangle^{s-1} \frac{(2s - 2)!}{s!(s - 1)!}(1 - 4\omega)^{\frac{1}{2} - s}. \quad (6.4)
$$

Proof. Equation (6.3) follows directly from (3.11). From (4.12) it follows that

$$
\sum_{m=s-1}^{\infty} L_h(s, 0, m) \cdot P_h(s, 0, m) = \varphi\langle -1, 1\rangle^s \varphi\langle 1, 0\rangle^{s-1} \frac{(2s - 2)!}{s!(s - 1)!} F(s, s - \tfrac{1}{2}; s; 4\omega).
$$
$$(6.5)$$

The fact that $F(s, s - \frac{1}{2}; s; 4\omega) = F(s - \frac{1}{2}, s; s; 4\omega)$ together with the identity $F(a, b; b; z) = (1 - z)^{-a}$ (see [1], p. 556) then completes the proof. $\square$

6.2. PRIORITY MODEL

In the second application, we investigate the single-server case where queue 1 has (fixed) priority over queue 2. More specifically, we consider the preemptive priority rule which implies that type-1 customers are allowed to interrupt type-2 customers. This model has a long history; see, e.g., [4, 9]. Let (n, j) denote the state where,

13

in equilibrium, the lengths of queues 1 and 2 are n and j, respectively. The QBD queue length process is described by the matrices

$$\mathbf{A}_0 = \mathbf{B}_{01} = \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & \ddots \end{pmatrix}, \quad \mathbf{A}_2 = \mathbf{B}_{10} = \begin{pmatrix} \mu & & \\ & \mu & \\ & & \ddots \end{pmatrix}, \quad (6.6)$$

and

$$\mathbf{A}_1 = \mathbf{B}_{11} = \begin{pmatrix} \delta & \lambda & \\ & \delta & \lambda \\ & & \ddots & \ddots \end{pmatrix}, \quad \mathbf{B}_{00} = \begin{pmatrix} -2\lambda & \lambda & \\ \mu & \delta & \lambda \\ & \ddots & \ddots & \ddots \end{pmatrix}, \quad (6.7)$$

where $N = 1$ and $\delta = -(\mu + 2\lambda)$.

Since both $s = 0$ and $u = 0$ in the priority model, (4.13) can be simplified as follows.

Proposition 2. *If $s = 0$ and $u = 0$ we have*

$$\sum_{m=0}^{\infty} L_h(0,0,m) \cdot P_h(0,0,m) = \varphi\langle -1, 0\rangle \varphi\langle 0, 1\rangle^t F(1 + \tfrac{t}{2}, 1 + \tfrac{t-1}{2}; 2; 4\omega). \quad (6.8)$$

Proof Follows immediately from (4.13). □

6.3. Feedback model

In our final example we consider a single-server, two-queue system with feedback, where queue 1 has preemptive priority over queue 2; see, e.g., [17]. Upon arrival, customers enter queue 1 and after receiving service, they either depart the system or join queue 2 (with equal probability); in the latter case, customers depart the system after being served at queue 2. Let (n, j) denote the state where, in equilibrium, the length of queue 1 is n and the number of customers that eventually will ask for service in queue 2 (including those still in queue 1) is j. The associated QBD process is described by the matrices

$$\mathbf{A}_0 = \mathbf{B}_{01} = \begin{pmatrix} \lambda & \lambda & \\ & \lambda & \lambda \\ & & \ddots & \ddots \end{pmatrix}, \quad \mathbf{A}_2 = \mathbf{B}_{10} = \begin{pmatrix} \mu & & \\ & \mu & \\ & & \ddots \end{pmatrix}, \quad (6.9)$$

and

$$\mathbf{A}_1 = \mathbf{B}_{11} = \begin{pmatrix} \delta & & \\ & \delta & \\ & & \ddots \end{pmatrix}, \quad \mathbf{B}_{00} = \begin{pmatrix} \delta & & \\ \mu & \delta & \\ & \ddots & \ddots \end{pmatrix}, \quad (6.10)$$

where $N = 1$ and $\delta = -(\mu + 2\lambda)$. We again obtain a simplified version of (4.13).

Proposition 3. *If $s = 0$ and $t = 0$ we have*

$$\sum_{m=u}^{\infty} L_h(s, u, m) \cdot P_h(s, u, m) = \varphi(1,1)^u \varphi(-1,0)^{u+1} \frac{(2u)!}{u!u!} (4\omega)^{-u-1} B_{4\omega}(1 + u, \tfrac{1}{2} - u),$$

(6.11)

where $B_x(\cdot, \cdot)$ is the incomplete Beta function defined by (see [1], p. 263)

$$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt.$$

(6.12)

Proof In the case $s = 0$ and $t = 0$, (4.13) reduces to

$$\sum_{m=u}^{\infty} L_h(s, u, m) \cdot P_h(s, u, m) = \varphi(1,1)^u \varphi(-1,0)^{u+1} \frac{(2u)!}{(u+1)!u!} F(u+1, u+\tfrac{1}{2}; u+2; 4\omega).$$

(6.13)

The identity

$$B_x(a, b) = \frac{x^a}{a} F(a, 1 - b; a + 1; x)$$

(6.14)

(see [1], p. 263) then completes the proof. □

REFERENCES

[1] Abramowitz, M., Stegun, I.A. (1964). *Handbook of Mathematical Functions* (US Government Printing Office, Washington D.C.).

[2] Adan, I.J.B.F., Wal, J. van der (1998). *Combining make to order and make to stock* (OR Spektrum, vol. 20, pp. 73-81).

[3] Adan, I.J.B.F., Weiss, G. (2006). *Analysis of a simple Markovian re-entrant line with infinite supply of work under the LBFS policy* (Queueing Systems, vol. 54, pp. 169-183).

[4] Cobham, A. (1954). *Priority assignment in waiting line problems* (Operations Research, vol. 2, pp. 70-76).

[5] Cohen, J.W. (1987). *A two-queue, one-server model with priority for the longer queue* (Queueing Systems, vol. 2, no. 3, pp. 261-283).

[6] Dunford, N., Schwartz, J.T. (1963). *Linear Operators Part II: Spectral Theory* (John Wiley and Sons, New York).

[7] Dunford, N., Schwartz, J.T. (1971). *Linear Operators Part III: Spectral Operators* (John Wiley and Sons, New York).

[8] Flatto, L. (1989). *The longer queue model* (Probability in the Engineering and Informational Sciences, vol. 3, pp. 537-559).

[9] Jaiswal, N.K. (1968). *Priority Queues* (Academic Press, London).

[10] Latouche, G., Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling* (SIAM, Philadelphia).

[11] van Leeuwaarden, J.S.H., Winands, E.M.M. (2006). *Quasi-birth-and-death processes with an explicit rate matrix* (Stochastic Models, vol. 22, no. 1, pp. 77-98).

[12] Liu, D., Zhao, Y.Q. (1997). *Determination of explicit solutions for a general class of Markov processes* (In Matrix-analytic methods in stochastic models, Dekker, New York, Lecture Notes in Pure and Appl. Math., vol. 183, pp. 343-357).

[13] Nelson, R.D., Squillante, M.S. (1996). *Stochastic analysis of affinity scheduling and load balancing in parallel processing systems* (In Interfaces in Computer Science and Operations Research, Proceedings of Fifth INFORMS Computer Science Technical Section Conference).

[14] Neuts, M.F. (1981). *Matrix-geometric Solutions in Stochastic Models, An Algorithmic Approach* (The Johns Hopkins Press, Baltimore).

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

[16] Ramaswami, V., Latouche, G. (1986). *A general class of Markoc processes with explicit matrix-geometric solutions* (OR Spektrum, vol. 8, pp. 209-218).

[17] Schrage, L.E. (1967). *The queue M/G/1 with feedback to lower priority queues* (Management Science, vol. 13, pp. 466-474).

[18] Squillante, M.S. (1998). *Matrix-analytic methods in stochastic parallel-server scheduling models* (In Advances in matrix analytic methods for stochastic models, Notable, New Jersey, pp. 311-340).

[19] Squillante, M.S. (2005). *Stochastic analysis of resource allocation in parallel processing systems* (In Computer System Performance Modeling in Perspective: A Tribute to the Work of Prof. K.C. Sevcik, Imperial College Press, London, pp. 227-256).

[20] Squillante, M.S., Nelson, R.D. (1991). *Analysis of task migration in shared-memory multiprocessors* (In Proceedings of ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems, pp. 143-155).

[21] Takács, L. (1991). *A Bernoulli excursion and its various applications* (Journal of Applied Probability, vol. 23, pp. 557-585).

[22] Tweedie, R.L. (1982). *Operator-geometric stationary distributions for Markov chains, with applications to queueing models* (Advances in Applied Probability, vol. 14, pp. 368-391).

[23] Zheng, Y., Zipkin, P.H. (1990). *A queueing model to analyze the value of centralized inventory information* (Operations Research, vol. 38, no. 2, pp. 296-307).