## On the Analysis of a M/M/1 Queue with Bulk Services

(Article begins on next page)

25 April 2024

# On the Analysis of a M/M/1 Queue with Bulk Services

GIANFRANCO BALBO[1,2] AND MARIA GRAZIA VIGLIOTTI[3,*]

[1]*Dipartimento di Informatica,Università di Torino, C.so Svizzera 135, Torino, Italy*
[2]*Faculty of Information and Communication Technology at Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia*
[3]*Department of Computing, Imperial College London, South Kensington Campus, SW7 2AZ London, UK*
*Corresponding author: maria.vigliotti@imperial.ac.uk

**The analysis of a queue that serves batches of customers with a novel service policy is presented in this paper, showing that a closed steady-state distribution of the number of customers in the queue can be derived for a very general setting of its parameters. Two output processes are flowing out of this queue depending on the fact that single customers are removed from the queue upon completion of a service, or bulks of customers of a fixed size are departing from the queue at the completion of the service's. Based on the expression of the steady-state distribution which assumes a Poisson arrival process, specific parameter configurations are identified that make the queue quasi-reversible, depending on the output process of interest. Since quasi-reversible queues are very important in the context of product form queueing networks, these results have relevant impacts on their own as well as when considered as the basis for possible computationally efficient approximations. Comparisons among the results obtained for different parameter settings are provided using both stochastic order arguments and numerical experiments. Future research directions are proposed considering also the many practical applications of this model ranging from flexible manufacturing, to logistics, to transportation systems.**

## 1. INTRODUCTION

In this paper, we derive the closed-form solution for the steady-state distribution of a class of queues that serve batches of customers with different configurations of service rates. The basic queue consists of a system in which arrivals form a Poisson process with parameter $\lambda$, services are exponentially distributed with rates that may depend on the load of the server, and the queue is served with a non-standard service policy. A threshold $B$ is set on the queue so that when its length[1] exceeds (or is equal to) $B$ a bulk (or batch) of $B$ customers can be removed from the queue upon completion of a service. If less than $B$ customers are in the system, the completion of a service always removes one customer from the queue. The total rate of service in states $j \geq B$ is assumed to be constant and denoted by $\mu$. In any state

$j \geq B$, there is a positive probability $p$ that customers are served in bulks of size $B$. Alternatively, with probability $q = (1 - p)$ a single customer is served and removed from the queue upon service completion. From state 1 to $B - 1$ single customers are served with service rates $\eta$ which may be state dependent (in this last case, $\eta_i$, $i = 1, \ldots, B - 1$). The state space of the queue can be seen in Fig. 1. In the rest of the paper, we shall call this queue $M/M^{B^*}/1$.

In this work, we report on a comprehensive analysis on *five different variants* of the basic model depicted in Figs 1 and 2, showing that, not only we can fully characterize the steady-state distributions for all these queues, but we can also establish a clear relationship among the corresponding stationary solutions. The queues are compared via *stochastic order* [1], which essentially provides a ranking on the tails of the (cumulative) distribution functions and, consequently, on the expected values of the corresponding queue lengths. Two of the queues that we have studied enjoy the very important property of being

---

[1]In the rest of the paper, we will often refer to the number of customers present in the system as the number of clients in the queue, thus considering in the same manner those who are really queueing and those who are receiving service.
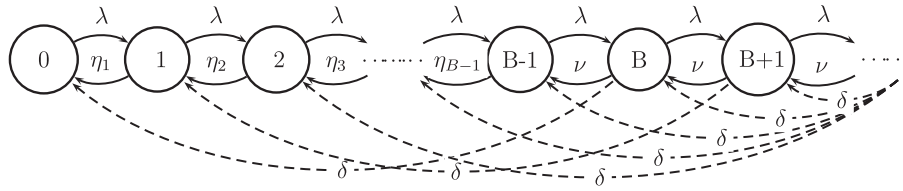
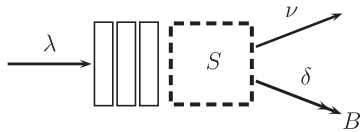**FIGURE 1.** State transition diagram of M/M/1 queue with bulk service $B$.



**FIGURE 2.** Diagram of M/M/1 queue with bulk service $B$.

*quasi-reversible* [2, 3]. Moreover, they turn out to be the upper and lower bounds for the others.

This result is rather important. Indeed, for the queues to be quasi-reversible, they must have very specific values of the service rates when less than $B$ customers are in the system. However, the queue length distributions, which have computationally simple closed-form expressions, do not actually depend on these values of the service rates when less than $B$ are present. This implies that these formulae can be used as approximations and bounds for a whole family of models whose service rates, when less than $B$ customers are in the queue, lay within the boundaries established by the parameters of the two quasi-reversible queues.

We believe that the $M/M^{B^*}/1$ is very relevant from a practical point of view as it is easy to envision many applications that can be modelled with any of its variants. In this paper, we will often refer to a specific example which comes from the *Manufacturing/Distribution* field and that we outline now.

Consider the case of a small production company whose sales department deals with both retail and wholesale orders. Finished parts are put into boxes and stored in the sales department. Orders handled by the department either refer to the purchase by clients of individual parts for personal use, or to the purchase of whole boxes used for subsequent processing. Orders are kept in two different stacks depending on the type of purchase they refer to. Assuming that there is a constant backlog of orders of the two types, when less than $B$ parts are in stock, orders for individual parts are considered only. When, instead, there are at least $B$ parts in stock, orders of the two types are considered with probabilities $p$ and $q$. The input of the sales department consists thus of a flow of single and indistinguishable parts, while the output is split into two separate streams consisting of different items (the individual parts and the boxes containing multiple parts) directed to completely different business 'circuits'. Let $\lambda$ (parts per second) represent the production rate of new parts, $B(> 1)$ the size of the boxes (i.e. the number of parts that fit into

a single box), $1/\mu$ (s) the average time required for handling an order, and $p = 1.0 - q$ the probability that an order refers to the purchase of a box. With a certain level of abstraction, the behaviour of this system can be approximately represented by *any of the five variants of the queue* $M/M^{B^*}/1$ (see Fig. 2), where customers correspond to the manufactured parts and in which services involve either a single customer or, whenever possible, a group of customers of a fixed size. The arrivals to the basic queue form a Poisson process with parameter $\lambda$, and services are exponentially distributed with rates $\eta$ or $\mu$ depending on the load of the server. The mathematical model that corresponds to this queue is a Continuous Time Markov Chain (CTMC) whose state transition diagram is depicted in Fig. 1, where we use the following notation:

(i) $\lambda$ is the arrival rate of new customers;
(ii) $\mu$ is the total service rate of the system when loaded with at least $B$ clients;
(iii) $\eta_i$ is the service rate of the system when loaded with $i$ clients ($i < B$);
(iv) $p$ is the probability that a service completion involves a bulk of $B$ customers;
(v) $q$ is the probability that a service completion involves a single customer when at least $B$ clients are at the server; $p + q = 1$.
(vi) $\nu = \mu * q$ is the rate at which a customer leaves the system alone when more than $B$ customers are present;
(vii) $\delta = \mu * p$ is the rate at which a bulk of $B$ customers leaves the system;
(viii) $\pi_n$ is the stationary probability of finding $n$ customers in the system ($n \geq 0$).

In a previous work (Balbo and Vigliotti, submitted for publication), we started the analysis of this kind of queue in the context of product-form queueing networks. For that study, the $M/M^{B^*}/1$ conditions were derived to ensure that the queue was *quasi-reversible* [2, 3]. Quasi-reversibility is a concept defined with respect to specific arrival and departure processes. We have shown in Balbo and Vigliotti (submitted for publication) that the $M/M^{B^*}/1$ is quasi-reversible with respect to both, the arrival process and the bulk's departure process. This implies that to obtain product-form solution, the $M/M^{B^*}/1$ is made to interact with other queues only via this very specific bulk departure flow.

In this paper, in contrast, we focus our attention on the queue in isolation and we study the behaviour of this queue in a
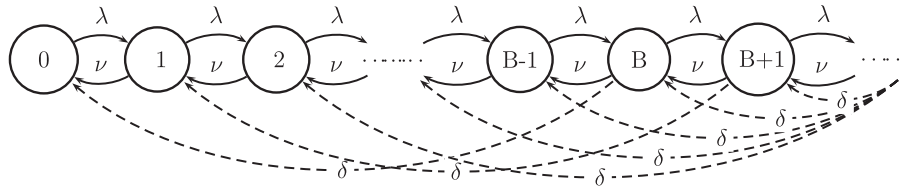
**FIGURE 3.** State transition diagram of M/M/1 queue with bulk service $B$.

more general setting (i.e. without restricting to product-form considerations), yet with the assumption of negative exponential service time distributions and Poisson arrival Process.

We begin our analysis by discussing the general results that can be derived for the case in which services are provided with a constant rate when more than $B$ customers are in the queue, while the service rate changes when less than B customers are requesting service.

Subsequently, we shall make our results more specific by considering four different versions of the $M/M^{B^*}/1$ characterized in the following manner:

(1) A queue with Poisson arrivals and a constant service rate equal to $\mu$ (i.e. the departure process of single customers when the length of the queue is smaller than $B$ is the same of the total service rate of the queue when more than $B$ customers are present: $\eta_1 = \cdots = \eta_{B-1} = \mu$); see Fig. 4.

(2) A queue with Poisson arrivals and service rate equals to $\nu$ for the departure process of single customers both when the length of the queue is smaller than $B$ and when bulk services are possible as well (i.e. $\eta_1 = \cdots = \eta_{B-1} = \nu = \mu * q$); see Fig. 3.

(3) A queue with Poisson arrivals that is quasi-reversible with respect to bulk departures.

(4) A queue with Poisson arrivals that is quasi-reversible with respect to single departures.

Finally, we prove a rather interesting and surprising result, showing that the quasi-reversible queue with respect to bulks of customers (number (3) in the list above) is the smallest of the four, while the quasi-reversible queue with respect to a single customer (number (4) in the list above) is the biggest according to our stochastic ordering. The other two queues in the list above are in between: the queue with service rates equal to $\mu$ (number (1) in the list above) being smaller than that with service rates equal to $\nu$.

The rest of the paper is organized as follows. In Section 2, we overview the state of the art emphasizing the differences of our model from those already published in the literature, and discussing Quasi-Reversibility at some length. Section 3 is the core of our paper and contains all the original results concerning the $M/M^{B^*}/1$. This section starts with some generic derivations which are independent of the properties that we are identifying in the specific configurations of our model. It

then continues with some results that can be derived when the initial service rates (those that we have previously denoted with $\eta_i$, $i = 1, \ldots, B - 1$) are assumed to be constant and, subsequently, we consider sets of variable service rates that make our output processes quasi-reversible. Section 4 uses results from the literature on *Stochastic Ordering* to compare the performances of $M/M^{B^*}/1$ in its different configurations. Section 5 builds on the results on stochastic ordering by quantifying the relations among the distributions derived for the different configurations of the $M/M^{B^*}/1$. Data coming from a specific case study allow one to draw some conclusions about the practical differences among the configurations of the $M/M^{B^*}/1$. Finally, Section 6 concludes the paper by proposing the use of the quasi-reversible configurations as computationally efficient approximations for realistic choices of the parameters of the model. Following this idea, this concluding session highlights some future research directions. An Appendix is also included to provide the details of the derivations of several expressions contained in the paper.

## 2. RELATED WORK

In the last 40 years, many papers appeared in the literature on the analysis of queues with bulk services starting from the pioneering work reported in [4, 5] to the more comprehensive discussion reported in [6]. In our paper, we consider a very specific case of a queue with bulk service and Poisson input with a special attention to the conditions that must be satisfied to make it quasi-reversible. Since we believe that the results of our paper are original, we concentrate on reviewing the literature on this topic to show the novelty of our analysis.

The simplest extension of the M/M/1 with batch services consists in modifying the service policy in such a way that only service for a fixed batch of size $B$ is possible. Time of service is exponentially distributed, and when a customer arrives to an empty queue, it waits until a full batch size is reached in order to be served. Upon the completion of a service, if the number of customers is bigger than the batch size, then the server removes the first $B$ customers from the queue. In [7, 8], it is shown that M/M/1 queues with fixed batches enjoy geometric steady state distributions.

A variation of this model is obtained by allowing the service of partial batches. In this case, the server starts working as

soon as a customer joins the queue. Upon completion of the service, a batch of $B$ customers is removed from the queue, if more than $B$ customers are there; otherwise a partial batch of size $i \leq B$ is served and the queue becomes empty. The rate of service is constant in each state of the queue. Also in this case, the steady-state distribution of the model has a geometric shape [7, 8]. Although [7, 8] do not directly consider quasi-reversible queues, the geometric solutions guarantee that these queues are quasi-reversible.

More general is the model proposed by Chao *et al.* [3] and Harrison [9] and studied in the context of quasi-reversibility. Instead of considering only fixed batch sizes, as in the previous case, it is assumed that batch sizes are integer random variables with arbitrary distributions. Given a batch size $B$, in states $j < B$ partial batches are served and the queue is emptied. In states $j \geq B$ partial batches are served according to a discrete distribution. If, instead of considering an M/M/1 queue, we consider a $G$-queue [10], modifications of the service policy to deal with batches can be included in a similar fashion [3, 9].

All these models are different from that considered in this paper. Outside the context of quasi-reversibility, our three queue configurations are different from [3, 9] as we do not have batch arrivals. In the context of quasi-reversible processes with respect to bulk departures, we note that we derive a geometric distribution only for a very specific configuration of the service rates when less than $B$ customers are in the queue. Indeed, the peculiarity of our model is that we assume that in any state of the queue with at least $B$ customers present, either one customer is served or exactly $B$ clients are considered simultaneously and leave the queue at the end of their service. This model is different from both models introduced by Chao *et al.* [3], Kleinrock [7] and Gross and Harris [8]. With respect to the first one, we note that in our model the server has only two choices: either one customer is served or a batch of a fixed size. Batches of sizes in between 1 and $B$, for a fixed $B$, are not considered by the server. It is rather subtle to observe that, even in the case of $B = 2$, the model discussed in this paper and that presented in [7, 8] would still be different. We differ from the second one because our batch size is not arbitrarily distributed. Also we differ from the work by Henderson and Taylor [11] as we do not consider batch arrivals. In the context of quasi-reversible processes with respect to single departures, the observations above, still stand.

### 2.1. Quasi-reversibility and queues with batches

Identifying the conditions that make a queueing node suitable to be included in a product-form queueing network remains a challenging task with important practical consequences.

As two of the configurations of the queue analysed in this paper are quasi-reversible, we shall see why this is important. A network of quasi-reversible queues enjoys a product-form solution [2, 3]. In the literature, there are two equivalent definitions of quasi-reversibility: the original one that refers to the arrival and departure processes of a queue [2], and the

more recent one that enforces conditions on the rates of the CTMC underlying a queueing system (see [3, p. 63]). In the CTMC, which describes the temporal behaviour of this system, the transition rate $\mathbf{q}(s, s')$ for any two states $s, s'$ in the state space $\mathsf{S}$ can be decomposed as the sum of three component rates in the following manner:

$$\mathbf{q}(s, s') = \mathbf{q}_A(s, s') + \mathbf{q}_D(s, s') + \mathbf{q}_I(s, s'). \quad (1)$$

Intuitively, we can interpret $\mathbf{q}_A(s, s')$ as the rate of arrivals and $\mathbf{q}_D(s, s')$ as the rate of departures; $\mathbf{q}_I(s, s')$ accounts for some form of internal transitions. Within this context, the definition of quasi-reversibility follows.

DEFINITION 2.1 [3]. *A CTMC with transition rates defined as in Equation* (1) *is quasi-reversible if there exist two non-negative real numbers* $\alpha, \beta$ *(i.e.* $\alpha, \beta \in \mathbb{R}^+$*) such that the following holds*:

$$\sum_{s' \in \mathsf{S}} \mathbf{q}_A(s, s') = \alpha \quad s \in \mathsf{S}, \quad (2)$$

$$\frac{\sum_{s' \in \mathsf{S}} \pi(s') \mathbf{q}_D(s', s)}{\pi(s)} = \beta \quad s \in \mathsf{S}, \quad (3)$$

*where* $\pi(s)$ *represents the stationary (or steady-state) distribution of the CTMC. In* [3], *it is shown that this definition is equivalent to that provided by Kelly* [2].

Definition 2.1 says that a CTMC is quasi-reversible if the rate of arrivals ($\alpha$) is independent of the state of the chain, and if the rate at which the CTMC enters an arbitrary state due to the occurrence of some departure is proportional ($\beta$) to the steady-state probability of that same state. As far as our work is concerned, we will show that the queue with Poisson arrivals is quasi-reversible with respect to the bulk departure process if we can set the service rates in a proper manner. Such a queue enjoys a fully geometric distribution. Equally, a queue with Poisson arrivals is quasi-reversible with respect to the departure process of single customers if the service rates are set in a different, but specific manner too. Such a queue enjoys a generalized geometric distribution. We can categorically exclude that the queue with Poisson arrivals is quasi-reversible with respect to both departure processes, single and bulk, with the same parameter setting.

### 3. ANALYSIS OF BULK SERVICE IN M/M/1 QUEUE

In this paper, we shall consider various queues that serve batches of customers. The generic state space of such a model is depicted in Fig. 1.

First, we derive some general results that assume a constant service rate when more than $B$ customers are in the queue, while the service rate changes when less than B customers are requesting service. Subsequently, we specialize the results

focusing on different assignments of the service rates when less that $B$ customers are in the queue. The contents of this section can thus be divided into two parts which correspond to two different analysis perspectives. The first (Sections 3.1 and 3.2) derives a set of results under the assumption that the model is fully specified and, in particular, that the service rates with less than $B$ customers in the system ($\eta_1 \cdots = \eta_{B-1}$) are parameters of the model itself. Within this context we shall see that the distribution of the number of customers in the system has a form, that we call *quasi-geometric*, characterized by a geometric tail that depends on these previous service rates only through a multiplicative constant. The second (Section 3.3) exploits the results of the first part to identify the (very) specific values that must be assigned to $\eta_1 = \cdots = \eta_{B-1}$ to ensure the quasi-reversibility of the queue. In this case, the desired property is defined for the output process of interest, and the corresponding values of these rates are determined.

## 3.1. Bulk service in $M/M^{B^*}/1$ queue with variable service rates

We proceed to derive the stationary distribution for this first general case, which involves finding the solution of the Global Balance Equations (GBEs) in their most general setting. The initial equations are the same for all the queue configurations discussed in this paper. When expressed to explicitly give evidence to the parameter representing the arbitrary bulk's size $B$, these equations assume the following form:

$$
\begin{cases}
\pi_0 \lambda = \pi_1 \eta_1 + \pi_B \delta, \\
\pi_1 (\lambda + \eta_1) = \pi_0 \lambda + \pi_2 \eta_2 + \pi_{B+1} \delta, \\
\quad \vdots \quad \vdots \\
\pi_{B-2} (\lambda + \eta_{B-2}) = \pi_{B-3} \lambda + \pi_{B-1} \eta_{B-1} + \pi_{2B-2} \delta, \\
\pi_{B-1} (\lambda + \eta_{B-1}) = \pi_{B-2} \lambda + \pi_B \nu + \pi_{2B-1} \delta, \\
\pi_B (\lambda + \nu + \delta) = \pi_{B-1} \lambda + \pi_{B+1} \nu + \pi_{2B} \delta, \\
\quad \vdots \quad \vdots \\
\pi_{B+k} (\lambda + \nu + \delta) = \pi_{B+k-1} \lambda + \pi_{B+k+1} \nu + \pi_{2B+k} \delta.
\end{cases}
$$

A simple manipulation of this system of equations provides a new representation that will be useful for deriving an explicit expression of the desired distribution: [math1]

$$
\begin{cases}
\pi_1 \eta_1 = \pi_0 \lambda - \pi_B \delta, \\
\pi_2 \eta_2 = \pi_1 \lambda - \delta(\pi_B + \pi_{B+1}), \\
\quad \vdots \quad \vdots \\
\pi_{B-1} \eta_{B-1} = \pi_{B-2} \lambda - \delta \sum_{j=1}^{B-1} \pi_{(B-1)+j}, \\
\pi_B (\nu + \delta) = \pi_{B-1} \lambda - \delta \sum_{j=1}^{B} \pi_{B+j}, \\
\quad \vdots \quad \vdots \\
\pi_{B+k} (\nu + \delta) = \pi_{B+k-1} \lambda - \delta \sum_{j=1}^{B-1} \pi_{B+k+j}.
\end{cases}
\tag{4}
$$

Without any loss of generality, we can introduce the following notation:

$$
\begin{aligned}
\frac{\sum_{j=1}^{i+1} \pi_{(B-1)+j}}{\pi_i} \delta &= \phi_i \quad 0 \leq i \leq (B-2), \\
\frac{\sum_{j=1}^{B-1} \pi_{(n+1)+j}}{\pi_n} \delta &= \psi_n \quad n \geq (B-1).
\end{aligned}
\tag{5}
$$

Using these definitions, we can rewrite the GBEs in the following more compact manner:

$$
\begin{cases}
\pi_1 = \pi_0 \dfrac{(\lambda - \phi_0)}{\eta_1}, \\
\pi_2 = \pi_1 \dfrac{(\lambda - \phi_1)}{\eta_2}, \\
\pi_3 = \pi_2 \dfrac{(\lambda - \phi_2)}{\eta_3}, \\
\quad \vdots = \vdots \\
\pi_{(B-1)} = \pi_{(B-2)} \dfrac{(\lambda - \phi_{(B-2)})}{\eta_{(B-1)}}, \\
\pi_n = \pi_{(n-1)} \dfrac{(\lambda - \psi_{(n-1)})}{(\nu + \delta)} \quad n \geq B,
\end{cases}
\tag{6}
$$

which immediately yields the recursive solution of our model

$$
\pi_n =
\begin{cases}
\pi_0 \left[ \displaystyle\prod_{i=1}^{n} \dfrac{(\lambda - \phi_{i-1})}{\eta_i} \right] & n \leq (B-1), \\
\pi_0 \left[ \displaystyle\prod_{i=1}^{B-1} \dfrac{(\lambda - \phi_{i-1})}{\eta_i} \right] \left[ \displaystyle\prod_{j=B}^{n} \dfrac{(\lambda - \psi_{j-1})}{\nu + \delta} \right].
\end{cases}
\tag{7}
$$

This simple expression of the desired distribution, which is similar to that of the steady-state solution of a $M/M/k$ queue, is readable, and easy to understand on its own, but, contrary to the solution of a $M/M/k$, is not explicit.

The model is fully specified once the parameters $\eta_1, \eta_2, \ldots \eta_{B-1}, \mu, \delta, \lambda$ and $B$ are known and the solution represented by Equation (7) gives the steady-state probability for each state of the queue, provided that the system is stable, i.e. that there is a value $m \geq B$ such that, for all $i \geq m$, $(\lambda - \psi_i)/(\nu + \delta) < 1$.

For the purpose of deriving an explicit expression of the solution of the model, we make the following *conjecture*:

$$
\boxed{\frac{\sum_{j=1}^{B-1} \pi_{(n+1)+j}}{\pi_n} \delta = \psi_n = \psi \quad n \geq (B-1)}
$$

which is inspired by the structure of the last equations of (4) (those for $n \geq B - 1$), and which implies a substantial 'regularity' of the stationary distribution.

Indeed, using this conjecture, the system of equations (6) simplifies drastically

$$
\begin{cases}
\pi_n = \pi_{n-1} \dfrac{(\lambda - \phi_{n-1})}{\eta_n} & 1 \le n \le (B-1), \\
\vdots & \vdots \\
\pi_n = \pi_{n-1} \dfrac{(\lambda - \psi)}{\nu + \delta} & n \ge B,
\end{cases}
\tag{8}
$$

so that the following (generalized) geometric distribution can be easily derived:

$$
\pi_n =
\begin{cases}
\pi_0 \left[ \displaystyle\prod_{i=1}^{n} \sigma_i \right], & n \le (B-1), \\
\pi_0 \left[ \displaystyle\prod_{i=1}^{B-1} \sigma_i \right] \rho^{n-B+1}, & n \ge B,
\end{cases}
\tag{9}
$$

where

$$
\sigma_i = \frac{(\lambda - \phi_{i-1})}{\eta_i} \quad 1 \le i \le (B-1),
$$
$$
\rho = \frac{(\lambda - \psi)}{\nu + \delta}.
\tag{10}
$$

It is easy to observe that the system becomes certainly unstable when the arrival rate $\lambda$ exceeds the maximum exit speed represented by $\mu(q + pB) = \mu + (B-1)\delta$, i.e. the number of customers arriving at the queue per unit of time exceeds the maximum number of clients that can leave the system per unit of time. More specifically, the stability condition for our system requires $\rho < 1.0$. In these conditions, $\pi_0$ assumes the following form:

$$
\pi_0 = \left[ 1 + \sum_{n=1}^{B-1} \left[ \prod_{i=1}^{n} \sigma_i \right] + \left[ \prod_{i=1}^{B-1} \sigma_i \right] \frac{\rho}{(1-\rho)} \right]^{-1}, \tag{11}
$$

which cannot be simplified further since, contrary to a full geometric distribution, we cannot find a closed-form expression in this case.

Even in this simple and elegant form our solution is not yet directly computable since all the factors that appear in (9) depend on the distribution itself.

We can nevertheless find the desired result proving a theorem that relies on the following lemma.

LEMMA 3.1. *Consider the polynomial function of degree $B+1$, $f_B(x) = \lambda - x(\nu + \delta + \lambda) + x^2\nu + x^{B+1}\delta$, where $\delta, \lambda, \nu$ are positive real constants. For all $B > 1$, there exists a unique root, $r$, such that $f_B(r) = 0$ with $r \in (0, 1)$.*

*Proof.* We compute the value of $f_B$ at the boundaries of the open interval $(0, 1)$: $f_B(0) = \lambda$ and $f_B(1) = 0$. The first derivative of $f_B$, evaluated again at the boundaries of the open interval, shows that $f'_B(x) < 0$ at $x = 0$, while it is larger than 0 at

$x = 1$ as long as $\lambda < \mu + (B-1)\delta$ (which is obviously true in our system as we observed before), thus implying that at least one root exists for $r \in (0, 1)$. The second derivative of $f_B$ is $f''_B(x) = 2\nu + B(B+1)x^{B-1}\delta$. For all $x \in (0, 1)$ it holds that $f''_B(x) > 0$ showing that the function is concave. This implies that the root $r$ is unique in the open interval $(0, 1)$. □

An explicit value for $\rho$ can be obtained only in the case of $B = 2$. In all the other cases, the solution can be derived using numerical methods such as Newton–Raphson's algorithm.

With this result, we are now in the position of stating our main theorem where $(a - b)^+ = \max(a - b, 0)$.

THEOREM 3.1. *The steady-state distribution for the queue with state space as depicted in Fig. 1 can be written as*

$$
\pi_n = \pi_0 \left[ \prod_{i=1}^{\min(n, B-1)} \sigma_i \right] \rho^{(n-(B-1))^+},
$$

*where $\rho$ is the unique root of the equation*

$$
f_B(x) = \lambda - x(\nu + \delta + \lambda) + x^2\nu + x^{B+1}\delta
$$
$$
\text{with } \rho \in (0, 1),
$$
$$
\sigma_{B-1} = \frac{\lambda\rho}{\lambda - \rho\delta + \rho(\eta_{B-1} - \nu)}
$$
$$
\text{and} \quad \sigma_i = \frac{\lambda}{\eta_i + \delta[\prod_{k=1}^{i} \sigma_k][\rho((1 - \rho^i)/(1 - \rho))]}
$$
$$
1 \le i \le B - 2
$$

*Proof.* Let us first observe that, when $n \ge (B-1)$, the expression contained in Equation (9) allows one to write

$$
\sum_{j=1}^{B-1} \pi_{n+j} = \pi_0 \left[ \prod_{i=1}^{B-1} \sigma_i \right] \sum_{j=1}^{B-1} \rho^{n-B+1+j}
$$
$$
= \pi_0 \left[ \prod_{i=1}^{B-1} \sigma_i \right] \rho^{n-B+1} \sum_{j=1}^{B-1} \rho^{j}
$$
$$
= \pi_n \frac{\rho(1 - \rho^{B-1})}{1 - \rho}. \tag{12}
$$

Noting also that, always in the case of $n \ge B$, we have $\pi_{n-1} = \pi_n/\rho$, we can make the appropriate substitutions and observe that the last equation of (4) can be written in the following way:

$$
\pi_{B+k}(\nu + \delta) = \frac{\pi_{B+k}}{\rho}\lambda - \delta\pi_{B+k}\frac{\rho(1 - \rho^{B-1})}{1 - \rho} \quad k \ge 0,
$$

which immediately yields

$$
(\nu + \delta) = \frac{\lambda}{\rho} - \delta\frac{\rho(1 - \rho^{B-1})}{1 - \rho}, \tag{13}
$$

**TABLE 1.** Algorithm to calculate the components, $\sigma_i$, of the steady-state distribution with $B - 2 \geq i \geq 0$ for a queue with arbitrary departures size $B$.

---

1: **procedure** FindSigma(Rate of the arrivals is $\lambda$, rate of the single departure from state $n \geq B$ to state $n - 1$ is $\nu$, rate of bulk departures is $\delta$)

2:      Calculate the root $\rho$ such that $0 < \rho < 1$ of the function $f_B(x) = \lambda - x(\nu + \delta - \lambda) - x^2\nu + x^{B+1}\delta$

3:      $\sigma_{B-1} = \dfrac{\lambda\rho}{\lambda + \rho(\eta_{B-1} - \mu)}$;

4:      $i = B - 2$;

5:      **while** $i >= 1$ **do**

6:        $\sigma_i = \dfrac{\lambda}{\eta_i + \delta(\prod_{k=i+1}^{B-1}\sigma_k)(\rho\frac{1-\rho^{i+1}}{1-\rho})}$;

7:        $i = i - 1$;

8:      **end while**

9: **end procedure**

---

where some simple manipulations produce the following equation of degree $B + 1$:

$$\lambda - \rho(\nu + \delta + \lambda) + \rho^2\nu + \rho^{B+1}\delta = 0. \tag{14}$$

For the first part of the theorem to be proved, we need to show that there exists a unique positive solution for $0 < \rho < 1$, but this is certainly true as Lemma 3.1 guarantees that such $\rho$ can be uniquely identified.

We can now turn our attention to prove the form of the $\sigma_i$ ($i = 1, \ldots, B - 1$). Indeed, the system of equations (4), which defines $\pi_n$ for $2 \leq n \leq (B - 1)$ (but which is valid also for $n = 1$), allows one to obtain an explicit expression for $\sigma_i$. For this purpose, let us first derive the expression of $\sigma_{(B-1)}$.

Consider the equation that defines $\pi_{(B-1)}$ among those of the system of Equations (4)

$$\pi_{(B-1)}\eta_{(B-1)} = \pi_{(B-2)}\lambda - \delta\sum_{j=1}^{B-1}\pi_{(B-1)+j}. \tag{15}$$

Recalling that $\pi_n = \pi_{(n-1)}\sigma_n$, $1 \leq n \leq (B - 1)$, and using the result expressed by Equation (12), we can substitute in the previous equation to write

$$\pi_{(B-1)}\eta_{(B-1)} = \frac{\pi_{(B-1)}}{\sigma_{(B-1)}}\lambda - \delta\pi_{(B-1)}\frac{\rho(1 - \rho^{B-1})}{1 - \rho} \tag{16}$$

from which we obtain

$$\sigma_{(B-1)} = \frac{\lambda}{\eta_{B-1} + \delta\rho[(1 - \rho^{B-1})/(1 - \rho)]} \tag{17}$$

which, using Equation (13), can be rewritten as

$$\sigma_{(B-1)} = \frac{\lambda\rho}{\lambda - \rho\delta + \rho(\eta_{B-1} - \nu)}. \tag{18}$$

The expression of Equation (18) proves the second item of the theorem, but cannot be generalized to obtain the values of $\sigma_i$, $i = 1, 2, \ldots, (B - 2)$. However, using the same arguments of the previous derivation (i.e. starting from the equations of

system (4)), we can obtain a general expression valid for any $i = 1, 2, \ldots, (B - 1)$:

$$\sigma_i = \frac{\lambda}{\eta_i + \delta[\prod_{k=i+1}^{B-1}\sigma_k][\rho((1 - \rho^i)/(1 - \rho))]}, \tag{19}$$

which completes the proof of the theorem. □

Note that in Equation (19), the computation of $\sigma_i$ requires knowing the values of $\rho$ and $\sigma_k$, $k = i + 1, \ldots, B - 1$. Thus, once we have computed $\sigma_{(B-1)}$, we can compute $\sigma_{(B-2)}$, and from there on all the other values of $\sigma_i$, $i = B - 3, \ldots, 1$. Equation (19) thus gives rise to an algorithm that allows one to derive all values of $\sigma_i$ as shown in Table 1.

Moreover, we can observe that, for the derivation of the generalized geometric solution for the steady-state distribution, namely $\pi_n = \pi_0[\prod_{i=1}^{\min(n,B-1)}\sigma_i]\rho^{(n-(B-1))^+}$, we *conjectured* that, for all $n$, there exists (and it is unique) a constant $\psi$ such that $\psi_n = \psi$.

The reader should note that, in the proof of Theorem (3.1), we have implicitly shown by substitution in the GBEs that such generalized geometric solution is indeed the correct one. For clarity, we state the following corollary.

COROLLARY 3.1. *For CTMC with state space as depicted in Fig. 1 the steady-state probability function can be written as*

$$\pi_n = \pi_0\left[\prod_{i=1}^{\min(n,B-1)}\sigma_i\right]\rho^{(n-(B-1))^+}$$

*if and only if, for the recursive solution derived in Equation (6), it holds that, for all $n$, there exists a unique constant $\psi$ such that $\psi_n = \psi$.*

*Proof.* (*If*).

Recall the definition of $\psi_n$ for any $n \geq (B - 1)$:

$$\psi_n = \frac{\sum_{j=1}^{B-1}\pi_{(n+1)+j}}{\pi_n}\delta.$$

Note that, in Equation (9), the two expressions on the right-hand side are identical for $n = B - 1$. We can thus rewrite

Equation (3.1) in the following way:

$$\psi_n = \frac{\sum_{j=1}^{B-1}\pi_0[\prod_{j=1}^{B-1}\sigma_i]\rho^{n+1+j-B+1}}{\pi_0[\prod_{j=1}^{B-1}\sigma_i]\rho^{n-B+1}}\delta$$

$$= \frac{\pi_0[\prod_{j=1}^{B-1}\sigma_i]\rho^{n-B+2}\sum_{j=1}^{B-1}\rho^j}{\pi_0[\prod_{j=1}^{B-1}\sigma_i]\rho^{n-B+1}}\delta$$

$$= \rho^2\left[\sum_{j=0}^{B-2}\rho^j\right]\delta$$

$$= \rho^2\delta\frac{1-\rho^{B-1}}{1-\rho},$$

which shows that $\psi_n$ is independent of $n$, thus proving this part of our conjecture.

(*Only if*).

If, for all $n$, there exists a unique constant $\psi$ such that $\psi_n = \psi$, then clearly Equation (9) is derived. □

Theorem 3.1 shows that the tail of the distribution depends on $\rho$ up to a multiplicative constant that accounts for the normalization ($\pi_0$ and for the product of the $\sigma_i$, $i = 1, \dots, B-1$). Since $\rho$ depends on $\lambda$, $\mu$ and $p$, but not on the values of $\eta_i$ ($i = 1, \dots, B-1$), different configurations of these service rates will only have the effect of changing the multiplicative constant and of scaling the tail of the distribution $\pi_n$ for $n \geq B$. In the next two subsections, we will now consider two interesting configurations of the $\eta_i$ ($i = 1, \dots, B-1$) that will be useful for comparison purposes, which will be discussed in Sections 4 and 5.

### 3.2. Bulk service in $M/M^{B^*}/1$ queue with constant service rates

As already observed, the derivation conducted up to Equation (14) does not depend on the rates of single service from state 1 to $B-1$. If we suppose that the speed of the server remains constant when the number of customers in the queue does not reach the threshold value of $B$, two possibilities exist that are worthy of being analysed.

*3.2.1. Case $\eta_i = \nu$*
The first refers to the case in which

$$\eta_i = \eta = \nu \quad 1 \leq i \leq (B-1) \tag{20}$$

corresponding to the fact that single customers are served with the same rate, independently of the fact that the number of customers at the server is larger or smaller than the threshold value $B$, which identifies the bulk's size. Changes in the rate of service of a single customer will not affect the derivation of the recursive solution of the steady-state distribution, but it has an impact on the definition of $\sigma_i$ that we denote in this case in the

following manner:

$$\hat{\sigma}_i = \frac{(\lambda - \phi_{i-1})}{\nu} \quad 1 \leq i \leq (B-1). \tag{21}$$

We can thus state now the following result.

LEMMA 3.2. *Consider the queue with state space as depicted in Fig. 3 with $\eta_i = \nu$ for $i = 1, \dots, B-1$; the recursive solution for the steady-state distribution is*

$$\pi_n = \pi_0\left[\prod_{i=1}^{\min(n,B-1)}\hat{\sigma}_i\right]\rho^{(n-(B-1))^+},$$

*where $\rho$ is the unique root of the equation*

$$f_B(x) = \lambda - x(\nu + \delta + \lambda) + x^2\nu + x^{B+1}\delta \quad with \; \rho \in ]0,1[,$$

$$\hat{\sigma}_{B-1} = \frac{\lambda\rho}{\lambda - \delta\rho}$$

*and* $\quad \hat{\sigma}_i = \frac{\lambda}{\nu + \delta[\prod_{k=i+1}^{B-1}\hat{\sigma}_k][\rho((1-\rho^i)/(1-\rho))]}$
$$1 \leq i \leq B-2.$$

*Proof.* The proof that $\pi_n = [\prod_{i=1}^{B-1}\hat{\sigma}_i]\rho^{(n-(B-1))^+}$ is the steady-state probability for this queue is identical to that of Theorem 3.1. It remains thus to show the analytical formulation for $\hat{\sigma}_i$ with $i \leq B-1$.

Using the simple expression represented by Equation (18) and introducing the assumption that is specific of this case, it is straightforward to show that

$$\hat{\sigma}_{B-1} = \frac{\lambda\rho}{\lambda - \delta\rho} \tag{22}$$

and, in general, that

$$\hat{\sigma}_i = \frac{\lambda}{\nu + \delta[\prod_{k=i+1}^{B-1}\hat{\sigma}_k][\rho((1-\rho^i)/(1-\rho))]}. \tag{23}$$

□

For this case the algorithm in Table 1 can still be used, changing the equation of Step 3 with Equation (22) and that in the while loop (Step 6) with Equation (23).

*3.2.2. Case $\eta_i = \mu$*
The second possibility refers to the case in which

$$\eta_n = \eta = \mu \quad 1 \leq i \leq (B-1). \tag{24}$$

With this other definition, we assume that the (overall) server rate is constant, independently of the number of customers leaving the server upon the completion of a service.

Again, this assumption impacts on the definition of $\tilde{\sigma}_i$ that now is expressed in the following way:

$$\tilde{\sigma}_i = \frac{(\lambda - \phi_{i-1})}{\mu} \quad 1 \leq i \leq (B-1). \tag{25}$$

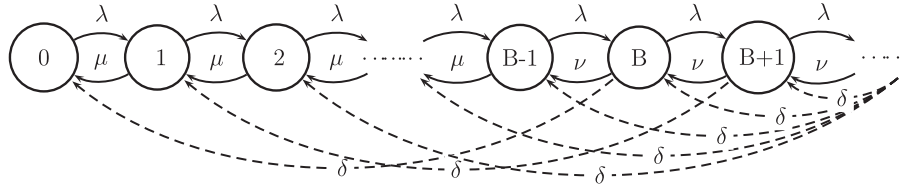For this case, we have the following result.

**FIGURE 4.** State transition diagram of M/M/1 queue with bulk service $B$ where $\mu = \delta + \nu$.

LEMMA 3.3. *Consider the queue with state space as depicted in Fig. 4 with $\eta_i = \mu$ for $i = 1, \ldots, B - 1$; the recursive solution for the steady-state distribution is*

$$\pi_n = \pi_0 \left[ \prod_{i=1}^{\min(n, B-1)} \tilde{\sigma}_i \right] \rho^{(n-(B-1))^+},$$

*where $\rho$ is the unique root of the equation*

$$f_B(x) = \lambda - x(\nu + \delta + \lambda) + x^2\nu + x^{B+1}\delta \quad \text{with } \rho \in ]0, 1[,$$
$$\tilde{\sigma}_{B-1} = \rho$$

*and* $\quad \tilde{\sigma}_i = \dfrac{\lambda}{\mu + \delta[\prod_{k=i+1}^{B-1} \tilde{\sigma}_k][\rho((1-\rho^i)/(1-\rho))]}$
$$1 \le i \le B - 2.$$

*Proof.* As for the previous lemma, we only have to show the analytical formulation for $\tilde{\sigma}_i$ with $i \le B - 1$.

Using again the simple expression represented by Equation (18) and introducing the assumption of the lemma, it is straightforward to show that

$$\tilde{\sigma}_{B-1} = \frac{\lambda\rho}{\lambda + (\mu - \nu - \delta)\rho} = \rho. \tag{26}$$

The general expression for $\tilde{\sigma}_n$, $n = 1, \ldots, B - 2$ does not have a close form and we can only say that

$$\tilde{\sigma}_i = \frac{\lambda}{\mu + \delta[\prod_{k=i+1}^{B-1} \tilde{\sigma}_k][\rho((1-\rho^i)/(1-\rho))]}. \tag{27}$$

$\square$

Also in this case the algorithm in Table 1 can be used by setting $\sigma_{B-1} = \rho$ in Step 3 and changing the equation in the while loop (Step 6) with Equation (27).

### 3.3. Variable service rates for the quasi-reversible process

We can now analyse further the $M/M^{B^*}/1$ building upon the results derived in Section 3, which showed that variable service rates, when the population of the system is smaller than the bulk size, are compatible with the generalized geometric distribution. As we pointed out at the beginning of Section 3, the analysis that will follow starts from the different perspective of identifying the values that need to be assigned to the rates $\eta_i$ ($i = 1, \ldots, B - 1$) in order to make the queue quasi-reversible. We will show that these rates are very specific and that queues with these service rates will be very difficult to encounter in practical applications. The results are nevertheless relevant because they will allow one to bound the performances of more realistic models with closed-form expressions that enjoy nice computational properties.

For this purpose, we recall that because of the definition of quasi-reversibility stated in Section 2.1, we need to fix the rate for the arrival $\mathbf{q}_A(s, s')$, as well as that for the departure $\mathbf{q}_D(s, s')$. In what follow, we assume that

$$\mathbf{q}(n, n + 1) = \mathbf{q}_A(n, n + 1).$$

Since our model exhibits two different output processes, we have to choose which of the departures corresponds to the rate $\mathbf{q}_D(s, s')$ that we are considering. Given the batch size $B$, we analyse a first case in which we assume that the departure rate is associated with the bulk-service:

$$\mathbf{q}(n, n - B) = \mathbf{q}_D(n, n - B),$$

while in the second case we assume that the departure rate is associated with the single-service:

$$\mathbf{q}(n, n - 1) = \mathbf{q}_D(n, n - 1),$$

#### 3.3.1. The bulk output process

Let us refer again for a moment to our motivating example of the sales department. When considering the modelling of this department within the organization of the whole production company, it is conceivable to assume that of main importance is the process of boxes of parts that are flowing out of the sales department to other enterprises which will process these items further. If the interaction with the 'outside world' depends on this process, it can be interesting to know whether there are conditions under which this process is quasi-reversible [2, 3]. Given that $\delta$ is the rate at which whole boxes are coming out of the system (when enough parts are in stock), we have in this case $\mathbf{q}_D(n, n - B) = \delta$ and the requirement for the process to be quasi-reversible corresponds to ensuring that the following relationship holds:

$$\frac{\pi_{n+B}\delta}{\pi_n} = \chi, \quad n \ge 0, \tag{28}$$

where $\chi$ is an appropriate constant.

Considering the generalized geometric distributions derived in the previous sections (Equations (11) and (9)), it is easy to see that this condition holds naturally as long as $n \geq B$. Indeed,

$$\frac{\pi_{n+B}\delta}{\pi_n} = \frac{\pi_0[\prod_{i=1}^{B-1}\sigma_i]\rho^{n+1}\,\delta}{\pi_0[\prod_{i=1}^{B-1}\sigma_i]\rho^{n-B+1}}$$
$$= \delta\rho^B \quad n \geq B, \tag{29}$$

so that we can conclude that $\chi = \delta\rho^B$.

In order for the relation to hold also for $n = B - 1$, we must have $\sigma_{B-1} = \rho$. Referring back to Equation (18), it is easy to see that this can be achieved as long as $\eta_{B-1} = \mu$.

When $B > 2$, Equation (29) applied for any $n \leq B - 1$ requires that $\sigma_k = \rho$ for any $k = 1, \ldots, B-2$. This implies that, for the queue to be quasi-reversible with respect to the bulk's output process, the steady-state distribution must be geometric and expressed in the following simple form:

$$\pi_n = (1 - \rho)\rho^n \quad n \geq 0. \tag{30}$$

Indeed, it is possible to show by induction that this solution is achievable as long as the service rates $\eta_i$, $i = 1, \ldots, B - 2$, assume specific values. If we assume that $\sigma_i = \rho$, $i = 1, \ldots, B - 2$, we identify the value of $\eta_i$ by rewriting $\sigma_i$ in the following manner:

$$\sigma_i = \frac{\lambda}{\eta_i + \delta\rho^{B-i-1}[\rho((1-\rho^i)/(1-\rho))]}$$
$$= \frac{\lambda}{\eta_i + \delta\rho^{B-i}((1-\rho^i)/(1-\rho))} \quad 1 \leq i \leq (B-2), \tag{31}$$

so that, for $\sigma_i$ to be equal to $\rho$, $\eta_i$ must assume the following value:

$$\eta_i = \frac{\lambda}{\rho} - \delta\,\rho^{B-i}\frac{1-\rho^i}{1-\rho} \quad 1 \leq i \leq (B-2). \tag{32}$$

Note that Equation (9) is also valid for $n = (B-1)$, but it is not needed since we have already proved that $\eta_{B-1}$ must be equal to $\mu$. However, this observation allows one to easily show that the following result holds:

$$\eta_{B-2} = \mu + \delta\rho. \tag{33}$$

This derivation can be carried over to show that, in general,

$$\eta_i = \eta_{i+1} + \delta\rho^{B-(i+1)} \quad 1 \leq i \leq (B-2). \tag{34}$$

This last result allows one also to obtain a direct expression for $\eta_i$ that must be satisfied in order for our $M/M^{B^*}/1$ queue to be quasi-reversible:

$$\eta_i = \mu + \rho\delta\frac{1-\rho^{(B-1)-i}}{1-\rho} \quad 1 \leq i \leq (B-1). \tag{35}$$

All these results can be summarized in the following lemma.

LEMMA 3.4. *Consider the queue with state space as depicted in Fig. 1. The steady-state distribution of the number of customers in the queue has the following geometric form*:

$$\pi_n = (1 - \rho)\rho^n \quad n \geq 0,$$

*if and only if there exists a constant $\chi$ such that, for all $n \geq 0$,*

$$\frac{\pi_{n+B}\delta}{\pi_n} = \chi$$

*and $\rho$ is the unique root of the equation $f_B(x) = \lambda - x(\nu + \delta + \lambda) + x^2\nu + x^{B+1}\delta$ with $\rho \in ]0, 1[$. Furthermore, it holds that $\eta_{B-1} = \mu$ and*

$$\eta_i = \mu + \rho\delta\frac{1-\rho^{(B-1)-i}}{1-\rho} \quad 1 \leq i \leq B - 2.$$

In this case, the solution can be computed directly without the help of the algorithm of Table 1 which reduces to the computation of $\rho$.

### 3.3.2. The single customer output process

We can now turn our attention to the single-customer output process. Referring again to our motivating example of the sales department, we could be interested in the flow of single items. If the interaction with the 'outside world' depends on this process, it can be interesting to know whether there are conditions under which this process is quasi-reversible [2, 3]. In this case, we observe $\mathbf{q}_D(n, n-1)$ cannot be written as a constant rate. Thus, the requirement for the process to be quasi-reversible corresponds to ensuring that the following relationship holds:

$$\frac{\pi_{n+1}\mathbf{q}_D(n, n-1)}{\pi_n} = C, \quad n \geq 0, \tag{36}$$

where $C$ is an appropriate constant.

Considering the different rates at which individual customers are served by the system, we can make the previous condition more explicit:

$$C = \begin{cases} \dfrac{\pi_{n+1}\,\eta_{n+1}}{\pi_n}, & n \leq (B-2), \\ \dfrac{\pi_{n+1}\,\nu}{\pi_n}, & n \geq (B-1). \end{cases} \tag{37}$$

Given the general expression for the solution of this model (see Equation (9)), and considering the second of the previous conditions, it is easy to see that $C$ must be equal to $\rho\nu$ so that we must also have

$$C = \rho\nu = \frac{\pi_{n+1}\eta_{n+1}}{\pi_n} \quad n \leq (B-2). \tag{38}$$

Using again the explicit expression of the solution (Equation (9)), we can write

$$\rho\nu = \sigma_i\eta_i \quad 1 \leq i \leq (B-1) \tag{39}$$

so that, assuming that we are able to get an explicit expression for $\sigma_i$, we can identify the rates that make this process reversible

in the following manner:

$$\eta_i = \frac{\rho v}{\sigma_i} \quad 1 \le i \le (B-1). \tag{40}$$

Consider first the case of $i = B-1$ for which we have a direct and explicit expression for $\sigma_{(B-1)}$ (Equation (18)). Substituting in the previous equation, we obtain

$$\rho v = \sigma_{B-1} \eta_{B-1}$$
$$= \frac{\lambda \rho}{\lambda - \rho \delta + \rho(\eta_{B-1} - v)} \eta_{B-1}, \tag{41}$$

which we can easily rewrite to obtain

$$\eta_{B-1} = \frac{\rho v(\lambda - \rho \delta - \rho v)}{\rho(\lambda - \rho v)}$$
$$= v \frac{(\lambda - \rho \mu)}{(\lambda - \rho v)}. \tag{42}$$

The expression of $\eta_{B-1}$ allows one to derive two alternative forms for $\sigma_{B-1}$. Indeed, recalling that $\sigma_{B-1} \eta_{B-1} = \rho v$, we easily obtain

$$\sigma_{B-1} = \rho \frac{(\lambda - \rho v)}{(\lambda - \rho \mu)}. \tag{43}$$

However, a simple manipulation of Equation (13) allows one to also obtain

$$\eta_{B-1} \frac{\lambda - \rho v}{\rho \delta} \frac{1-\rho}{(1-\rho^{B-1})} = \rho v \tag{44}$$

but recalling again that $\sigma_{B-1} \eta_{B-1} = \rho v$, we can observe that

$$\sigma_{B-1} = \frac{(\lambda - \rho v)(1-\rho)}{\rho \delta(1-\rho^{B-1})}. \tag{45}$$

Using now the general result that we derived before for $\sigma_i$, $i = 1, \ldots, B-2$, we can obtain the explicit expression of $\sigma_{B-2}$. Indeed,

$$\sigma_{B-2} = \frac{\lambda}{\eta_{B-2} + \delta \sigma_{B-1} \rho((1-\rho^{B-2})/(1-\rho))} \tag{46}$$

yields

$$\sigma_{B-2} \eta_{B-2} + \sigma_{B-2} \delta \sigma_{B-1} \rho \frac{1-\rho^{B-2}}{1-\rho} = \lambda \tag{47}$$

and recalling that in this case $\sigma_{B-2} \eta_{B-2} = \rho v$, we obtain

$$\rho v + \sigma_{B-2} \delta \sigma_{B-1} \rho \frac{1-\rho^{B-2}}{1-\rho} = \lambda \tag{48}$$

from which, using the explicit expression of $\sigma_{B-1}$ and the result expressed by Equation (13), we obtain

$$\sigma_{B-2} = \frac{1-\rho^{B-1}}{1-\rho^{B-2}} \tag{49}$$

so that

$$\eta_{B-2} = \rho v \frac{1-\rho^{B-2}}{1-\rho^{B-1}}. \tag{50}$$

Following a similar approach and using an inductive argument, it is easy to prove that

$$\sigma_i = \frac{1-\rho^{i+1}}{1-\rho^i} \quad 1 \le i \le B-2 \tag{51}$$

and

$$\eta_i = \rho v \frac{1-\rho^i}{1-\rho^{i+1}} \quad 1 \le i \le B-2. \tag{52}$$

As in the previous case, all these results can be summarized with the following lemma.

LEMMA 3.5. *Consider the queue with state space as depicted in Fig. 1. The steady-state distribution of the number of customers in the queue is*

$$\pi_n = \pi_0 \left[ \prod_{i=1}^{\min(n, B-1)} \sigma_i \right] \rho^{(n-(B-1))^+},$$

*if and only if*

$$C = \begin{cases} \dfrac{\pi_{n+1}\, \eta_{n+1}}{\pi_n}, & n \le (B-2), \\[2mm] \dfrac{\pi_{n+1}\, v}{\pi_n}, & n \ge (B-1), \end{cases}$$

*where $\rho$ is the unique root of the equation $f_B(x) = \lambda - x(v + \delta + \lambda) + x^2 v + x^{B+1} \delta$ with $\rho \in ]0, 1[$.*

*Furthermore, it holds that*

$$\sigma_{B-1} = \rho \frac{(\lambda - \rho v)}{(\lambda - \rho \mu)} \quad \eta_{B-1} = v \frac{(\lambda - \rho v)}{(\lambda - \rho \mu)}$$

*and, for all $i \le B-2$,*

$$\sigma_i = \frac{1-\rho^{i+1}}{1-\rho^i} \quad \eta_i = \rho \mu \frac{1-\rho^i}{1-\rho^{i+1}}.$$

Again, the algorithm in Table 1 can be used also in this case by setting $\sigma_{B-1} = \rho((\lambda - \rho v)/(\lambda - \rho \mu))$ in Step 3 and changing the equation in the while loop (Step 6) with Equation (52).

Note that substituting in the formula for the queue length distribution the explicit expressions of the $\sigma_i$, we obtain a new version of this result which allows the derivation of a closed-form expression of $\pi_0$. Indeed we can state the following corollary.

COROLLARY 3.2. *Consider the queue with state space as depicted in Fig. 1. The steady-state distribution of the number*

*of customers in the queue is*

$$
\pi_n = \begin{cases} \pi_0 \dfrac{1 - \rho^{n+1}}{1 - \rho}, & n \leq (B - 2), \\[2ex] \pi_0 \dfrac{1 - \rho^{B-1}}{1 - \rho} \dfrac{\lambda - \rho\nu}{\lambda - \rho\mu} \rho^{n-B+1}, & n \geq (B - 1), \end{cases} \tag{53}
$$

*where $\pi_0$ has the following closed form*:

$$
\pi_0 = \frac{(1 - \rho)}{(B - 1) - \rho((\lambda - \rho\nu)/(\lambda - \rho\mu))((1 - \rho^{B-1})/(1 - \rho))}.
$$

## 4. BOUNDS

In the previous section, we studied four variants of the same queue with bulk service: two versions where the service of single customers are state dependent, and two queues with different constant service rates for a single customer when the service of bulks is not possible. A natural question that arises after having derived all these results concerns the relationships among the steady-state distributions of these different versions of the queue. In this section, we provide an answer to this question by comparing their corresponding stochastic processes [1] from the point of view of their cumulative distribution functions. For this purpose, we use the notion of *stochastic order* over totally ordered sets in CTMCs [1].

The basic idea in what follows is to derive the ranking over the probability distributions of two (or more) chains, assuming a pre-defined ordering among the corresponding states of the chain. The order on the states of the chain is simply a relation that is reflexive, transitive, but not symmetric. In this work, as we are dealing with queues in which the states represent the number of customers in the system and are indexed with that same number, the order of the states is assumed to be the natural order $\leq$ on the natural numbers, $\mathbb{N}$. Moreover, in what follows, it is assumed that the order on the states of the chain is total.

We use the symbols $X, Y, V, U$ for random variables and $\{X(t) : t \geq 0\}, \{Y(t) : t \geq 0\}$ for stochastic processes. Two random variables may be compared with respect to their expectations.

DEFINITION 4.1. *Let $X, Y$ be two random variables. We write $X \preceq_{\mathrm{st}} Y$ if and only if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$, for any increasing function $f$ for which both expectations exist.*

DEFINITION 4.2. *Let $\{X(t) : t \geq 0\}$ and $\{Y(t) : t \geq 0\}$ be two stochastic processes. We write $\{X(t)\} \preceq_{\mathrm{st}} \{Y(t)\}$ if, for all $t$, it holds that*:

$$
X(t) \preceq_{\mathrm{st}} Y(t).
$$

PROPOSITION 4.1 [12, 13]. *Let $\{X(t) : t \geq 0\}$ and $\{Y(t) : t \geq 0\}$ be two CTMCs with state space $S = \{0, 1, 2, 3, \ldots\}$ with*

*generator matrices $\mathbf{A}, \mathbf{B}$ such that $X(0) \preceq_{\mathrm{st}} Y(0)$. It holds that $\{X(t)\} \preceq_{\mathrm{st}} \{Y(t)\}$, if*

$$
\sum_{k \geq m} \mathbf{A}(i, k) \leq \sum_{k \geq m} \mathbf{B}(j, k)
$$

*for all $i \leq j$ and for all $m$ such that $m \leq i$ or $m > j$.*

This last result is very useful as it allows one to compare CTMCs by looking at their transition matrices only.

Assume that $\{X(t) : t \geq 0\}$ and $\{Y(t) : t \geq 0\}$ are the stochastic processes corresponding to two M/M/1 queues, with arrival rates $\lambda_1, \lambda_2$ and service rates $\mu_1, \mu_2$, respectively. Here $\{X(t)\}$ and $\{Y(t)\}$ represent the number of customers in queue (and in service as well) in the two systems at time $t$. Proposition 4.1 says that $\{X(t)\} \preceq_{\mathrm{st}} \{Y(t)\}$ if $\lambda_1 \leq \lambda_2$ and $\mu_1 \geq \mu_2$.

We can compare now the queues we have analysed in this paper. In each queue we have changed only the variable rates of the single departures. We write $\{X(t) : t \geq 0\}$ for our generic queue with variable rates of the single departures, $\{X^\nu(t) : t \geq 0\}$ for the queue with fixed rate $\nu$ of the single departures and $\{X^\mu(t) : t \geq 0\}$ for the queue with fixed rate $\mu$ of the single departures. Using the arguments, we have just developed, it is easy to see that the following holds.

LEMMA 4.1.

(i) $\{X^\mu(t) : t \geq 0\} \preceq_{\mathrm{st}} \{X^\nu(t) : t \geq 0\}$.
(ii) *Assume that, for all $i$, $\mu \geq \eta_i \geq \nu$; then $\{X(t) : t \geq 0\} \preceq_{\mathrm{st}} \{X^\nu(t) : t \geq 0\}$ and $\{X^\mu(t) : t \geq 0\} \preceq_{\mathrm{st}} \{X(t) : t \geq 0\}$.*

For what concerns the other two cases corresponding to queues that exhibit peculiar properties, in the sense that they are quasi-reversible with respect to different transitions, we can make the following observations. The quasi-reversible queue with respect to batch departures has the relevant property that, for all $i$ such that $1 \leq i \leq B - 2$, it holds that $\eta_i > \eta_{i+1}$ and $\eta_{B-1} = \mu$ as can be easily seen from the expression of Equation (35). The quasi-reversible queue with respect to single departures has the property instead that $\eta_i < \nu$ for any $i$ such that $1 \leq i \leq B - 1$, as we can see from Equations (42) and (52). We write $\{X^{[R,\delta]}(t) : t \geq 0\}$ for the queue that is quasi-reversible with respect to batch departures, and $\{X^{[R,\nu]}(t) : t \geq 0\}$ for the queue that is quasi-reversible with respect to single departures. We thus obtain the following lemma.

LEMMA 4.2.

(1) $\{X^\nu(t) : t \geq 0\} \preceq_{\mathrm{st}} \{X^{[R,\nu]}(t) : t \geq 0\}$.
(2) $\{X^{[R,\delta]}(t) : t \geq 0\} \preceq_{\mathrm{st}} \{X^\mu(t) : t \geq 0\}$.

*Proof.* The first statement simply follows from Proposition 4.1; the second statement follows from Equation (35). □

The previous lemma is rather important because it gives a total order among the various queue configurations studied in this paper, namely:

$$\{X^{[R,\delta]}(t) : t \geq 0\} \preceq_{st} \{X^{\mu}(t) : t \geq 0\},$$
$$\{X^{\mu}(t) : t \geq 0\} \preceq_{st} \{X^{\nu}(t) : t \geq 0\},$$
$$\{X^{\nu}(t) : t \geq 0\} \preceq_{st} \{X^{[R,\nu]}(t) : t \geq 0\}.$$

By transitivity of $\preceq_{st}$ we conclude that $\{X^{[R,\delta]}(t) : t \geq 0\} \preceq_{st} \{X^{[R,\nu]}(t) : t \geq 0\}$.

This apparently obvious observation allows one to draw an important conclusion referring to the fact that the performance figures computed for the choices of parameters that yield the quasi-reversible behaviours bracket similar performance results computed for $M/M^{B^*}/1$ queues with parameters belonging to the parameter space defined by these two extreme cases.

Specifically, since the queue length distributions for the two quasi-reversible models have closed-form expressions that depend only on the input parameters ($\lambda$, $\mu$, $p$, $B$) and on the loading factor $\rho$, which is the root of the polynomial of degree $B$ of Theorem (3.1), they easily provide bounds for the solutions of variants of the models that do not enjoy these properties, as long as their service rates belong to the space identified are bound by the state dependent service rates of these two reference models.

## 5. DISCUSSION

The results derived in the previous section ensure that a ranking exists among the different versions of the $M/M^{B^*}/1$ analysed in this paper. No quantitative information is however provided by these results for what concerns the actual distance among the performance figures computed in the different cases. We can say that the extreme situations corresponding to the quasi-reversible parametrizations provide approximations for the behaviour of any other $M/M^{B^*}/1$ with intermediate parameters. However, in order to get a feeling for the quality and the effectiveness of these approximations, we must refer the reader to a specific case as we will do in the sequel of this section.

Consider an $M/M^{B^*}/1$ with bulk's size $B = 20$ (clients), total service rate $\mu = 2$ (clients/s) and probability $p = 0.6$ of removing a bulk of $B$ clients from the queue upon completion of a service when more that $B$ customers are at the station. Depending on the value of $\lambda$ which represents the arrival rate of clients at the $M/M^{B^*}/1$, we can easily identify the loading factor $\rho$ of the system and compute the distribution of the number of customers in our queue for the case of the bulk's reversible process using Equation (30). Most important for our purposes is however the computation of the variable service rates $\eta_i$ that our $M/M^{B^*}/1$ must exhibit with less than $B$ clients. Figure 5a depicts the sets of values that must be assumed for different values of $\lambda$ ranging from the case of a very lightly loaded queue ($\lambda = 0.4$ clients/s) to an almost saturated one ($\lambda = 22$ clients/s). Note that all these sets have the common
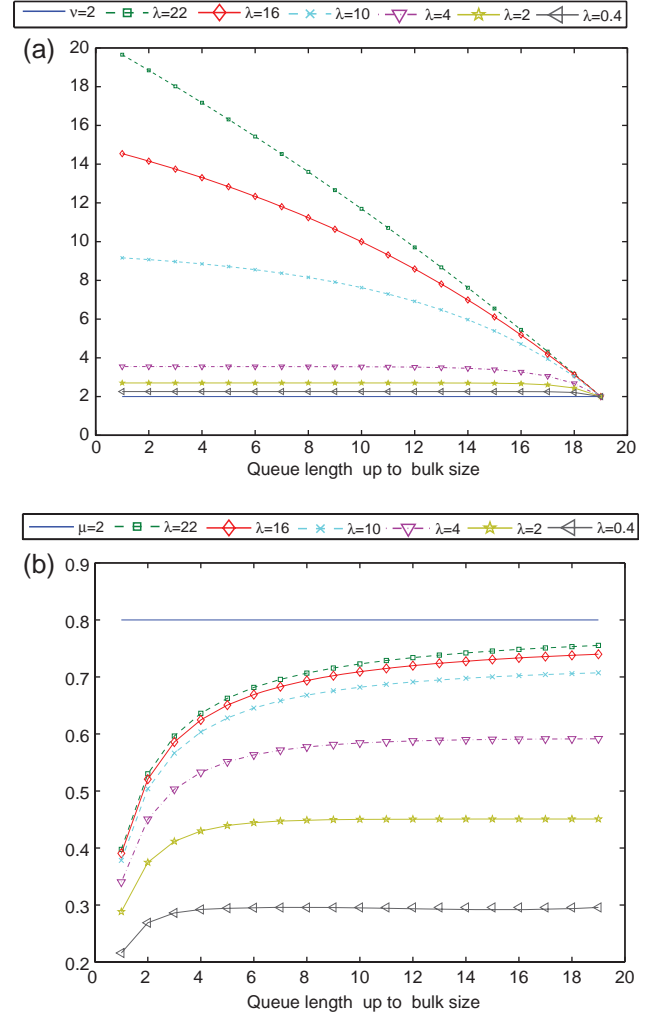


**FIGURE 5.** Plots of single departures rates $\eta$. (**a**) Plot of the value of the $\eta$s versus different arrival rates for a quasi-reversible queue with respect to bulk departures with bulk size $B = 20$. (**b**) Plot of the value of the $\eta$s versus different arrival rates for a quasi-reversible queue with respect to single departures with bulk size $B = 20$.

value of $\eta_{19} = 2.0$ as discussed in the section devoted to the analysis of this case (Section 3.3.1). Similar sets of values can also be computed for the single customer reversible process as depicted in Fig. 5b. Note that in this case all the values are smaller than $\nu = 0.8$ with a substantial discontinuity between the rates required for $n = 19$ in the case of different load configurations and that for $B = 20$ with the difference being substantial in the case of a lightly loaded queue. Figure 6 combines the information contained in the two previous figures, selecting a specific arrival rate ($\lambda = 10$ clients/s) and pointing out the parameter space for the possible service rates for less than $B$ customers in the queue (portion (a) of the figure). Any $M/M^{B^*}/1$ with an $\eta_i$ 'curve' positioned between the two extremal curves of the diagram will have a cumulative queue length distribution bounded by those of the two reversible cases,

and more specifically will be such that its mean queue length will be larger than that computed for the bulk's reversible process case and smaller than that of the single customer reversible process.

Turning now our attention to the comparison among the distribution functions and the cumulative distributions for the different cases, we can start with the discussion of the results depicted in Fig. 6 and computed for $\lambda = 10$, where the left one (part b) corresponds to the distribution function, while the right one (part c) reports the cumulative distribution. The shapes of the distribution functions are quite different for values of $n$ smaller than $B$, while they become comparable for larger values of $n$ (Fig. 7). Moreover, we can note the similarity and the closeness among three of the four cases studied in this paper singling out the peculiar behaviour of the $M/M^{B^*}/1$ when the bulk output process is reversible. A similar remark can be made also with respect to the comparison among the cumulative distributions where it becomes evident the lower bound role played by the $M/M^{B^*}/1$ with reversible bulk's output process. The differences in performance of the several versions of the $M/M^{B^*}/1$ become smaller with the increasing load of the queue, while some similarities among the distribution functions of the different models start to appear for extremely lightly loaded models (see Fig. 8a and b). Moreover, we can observe that, for $n$ larger than $B$, the ratios among different distribution functions remain constant for any value of $n$. The behaviours of the different versions of the $M/M^{B^*}/1$ for increasing values of the arrival rate $\lambda$ are summarized in Fig. 8, which provide a comprehensive view of the differences and similarities of the various cases.

Despite the fact that, in order to satisfy the GBEs, the queues exhibiting quasi-reversible behaviours with respect to the two different output processes must have very specific load-dependent service rates when the queue length is smaller than $B$ (which make them very unlikely to occur in practice), the queue length distributions expressed by Equations (30) and (53) provide computationally efficient bounds for the solution of similar models, but such that $\eta_i^{[R,\nu]} \leq \nu_i \leq \eta_i^{[R,\nu]}$ (where we borrowed the notation from the previous Section 4 to identify the service rates of the two quasi-reversible models). Moreover, the small distance that exists for many loading factors among the cumulative distributions of the cases with constant service rates equal to $\mu$ and $\nu = p\mu$ and that corresponding to the quasi-reversible single customer output process allows one to say that such a model can be often used as a good approximation of the other ones.

## 6. CONCLUSION

In this paper, we presented the results of the analysis of the $M/M^{B^*}/1$, where bulks of customers of size $B$ are possibly removed from the queue upon completion of a service when more than $B$ clients are present in the system. When the
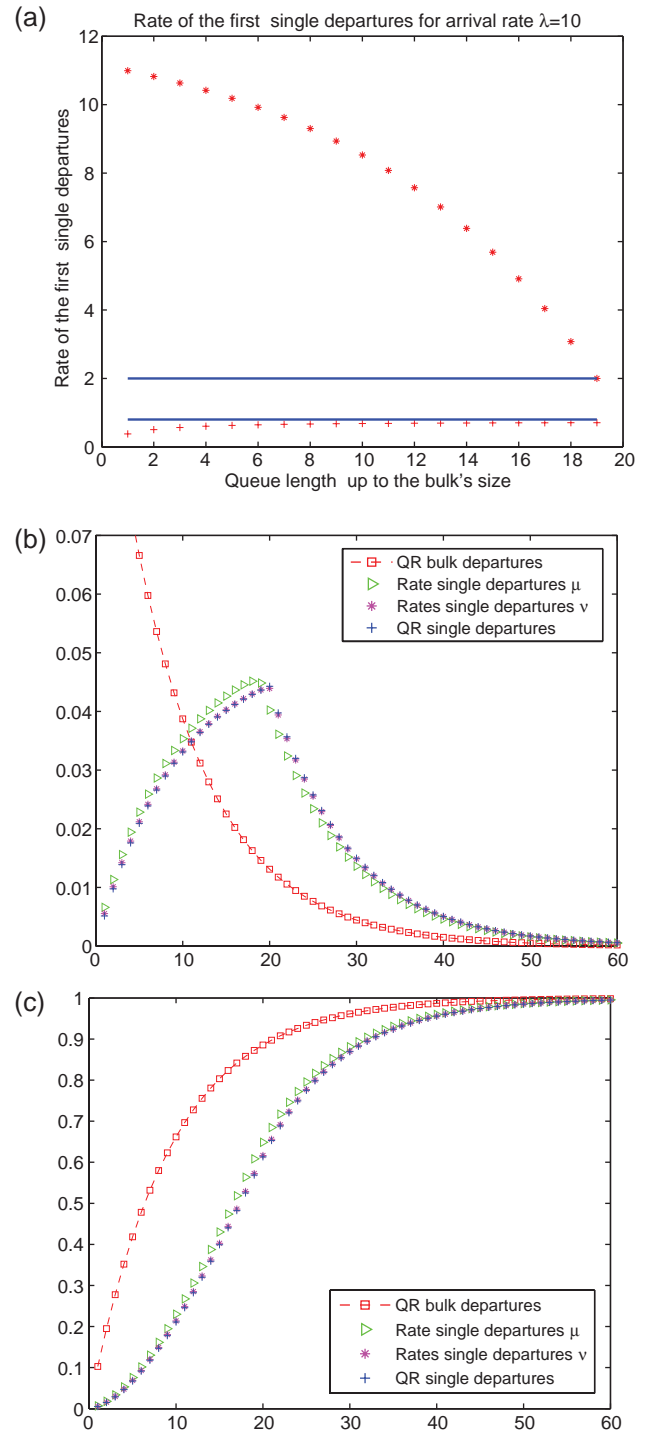


**FIGURE 6.** Plots of the density function, distribution function and the rates of single departures $\eta$ for a queue with bulk size 20 and arrival rate 10. (**a**) Plot of the value of the $\eta$s for the quasi-reversible queues with arrival rate $\lambda = 10$ and with bulk size $B = 20$. (**b**) Plot of the density function for the queues in steady state with arrival rate $\lambda = 10$ and with bulk size $B = 20$. (**c**) Plot of the cumulative distribution function for the queues in steady state with arrival rate $\lambda = 10$ and with bulk size $B = 20$.
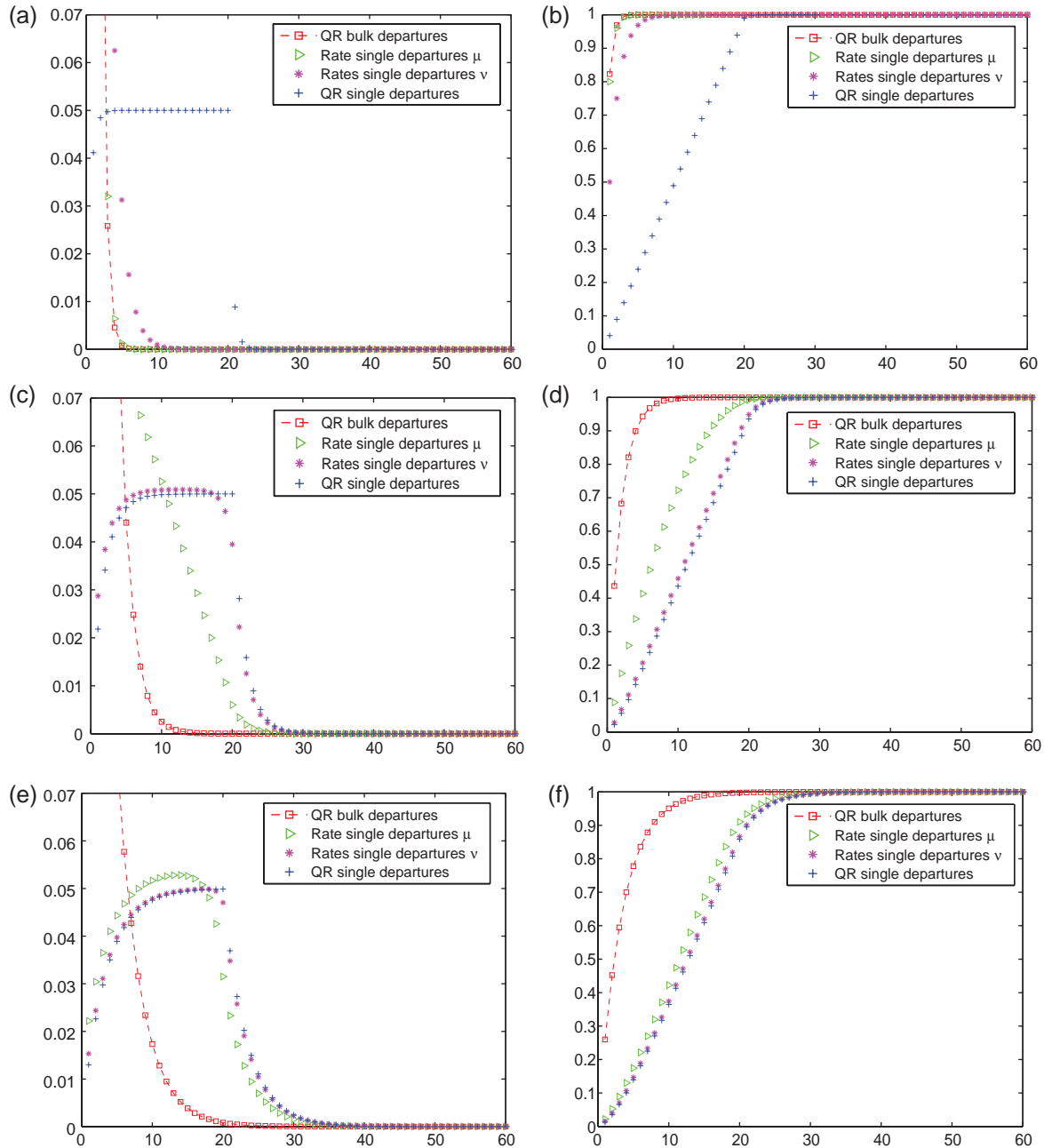
**FIGURE 7.** Plots of the various density function and distribution functions of the queues with bulk size 20 and different arrival rates. (**a**) Graph of the pdf with arrival rate $\lambda = 0.4$. (**b**) Graph of the cdf with arrival rate $\lambda = 0.4$. (**c**) Graph of the pdf with arrival rate $\lambda = 2$. (**d**) Graph of the cdf with arrival rate $\lambda = 2$. (**e**) Graph of the pdf with arrival rate $\lambda = 4$. (**f**) Graph of the cdf with arrival rate $\lambda = 4$.

global service rate $\mu$ is kept constant for any value of $n$ larger than $B$, the distribution function of this queue shows a geometric tail independently of the service rates presented by the system when less than $B$ clients are there. The 'regularity' of this result relies on the computation of the load factor that derives from the solution of an equation of degree $B$. We have shown that conditions may be put on the rates of the server when less than $B$ customers are there, in

order to obtain quasi-reversible behaviours. Quasi-reversibility provides a general set of conditions to guarantee a product-form solution. It has been shown that quasi-reversibility is a necessary condition for this result to be true [2, 3, 14, 15], and that nearly all networks that enjoy product-form solutions are formed of queues that are quasi-reversible, or can be seen as such, including the problematic $G$-networks with triggers [16].
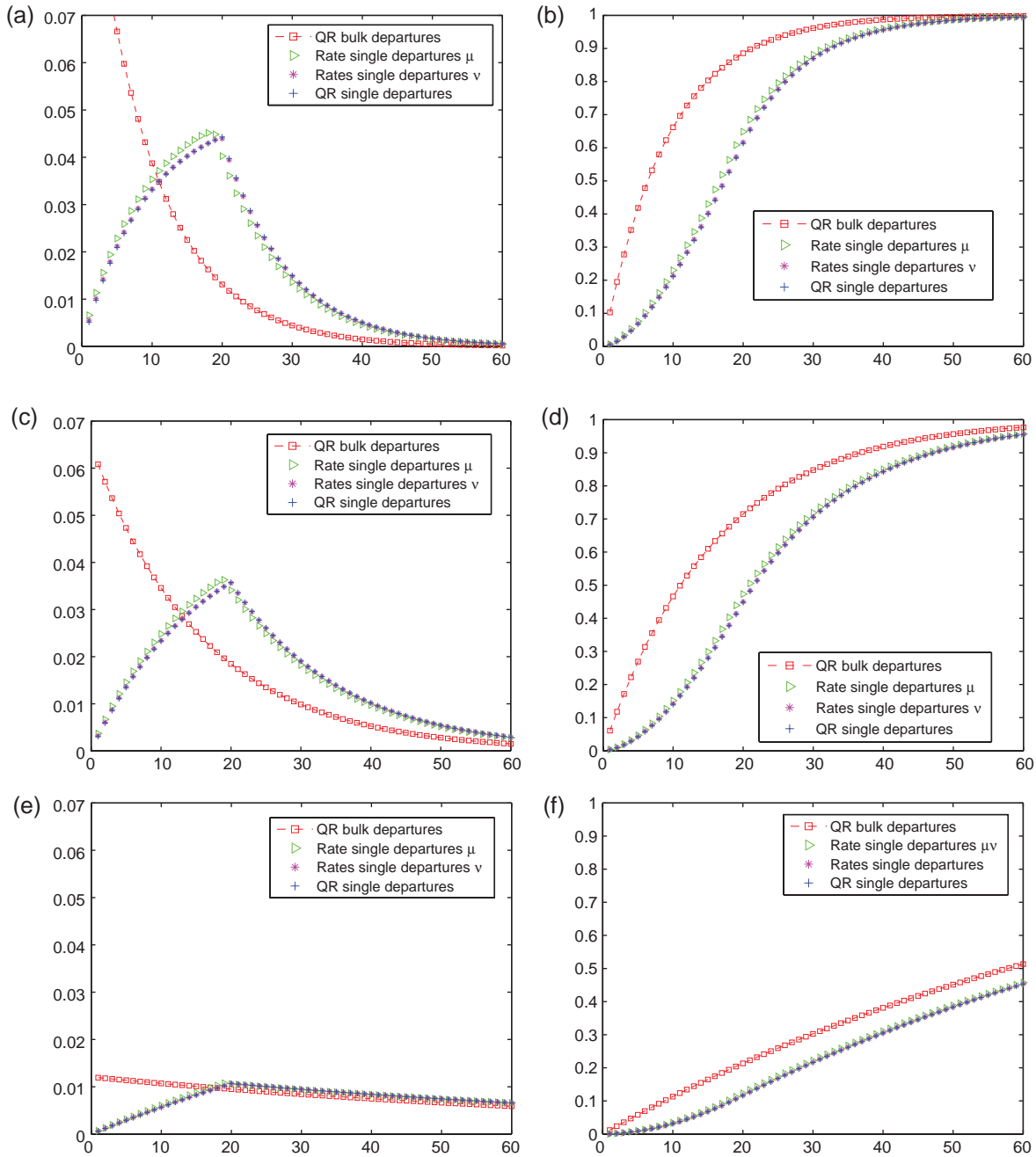
**FIGURE 8.** Plots of the various density function and distribution functions of the queues with bulk size 20 and different arrival rates. (**a**) Graph of the pdf with arrival rate $\lambda = 10$, (**b**) Graph of the cdf with arrival rate $\lambda = 10$, (**c**) Graph of the pdf with arrival rate $\lambda = 14$, (**d**) Graph of the cdf with arrival rate $\lambda = 14$, (**e**) Graph of the pdf with arrival rate $\lambda = 22$ and (**f**) Graph of the cdf with arrival rate $\lambda = 22$.

The model discussed in this paper identifies two output processes corresponding to the fact that a single customer or a bulk of customers leave the queue upon a service completion. Assuming that the process of arriving customers is Poisson, then the service rates of the $M/M^{B^*}/1$ can be adjusted to make the bulk output process or the single customer output process Poisson too. These two processes cannot be Poisson at the same time however (i.e. for the same set of service rates $\eta_i$), showing that they cannot be considered as the result of a simple Bernoulli decomposition of Poisson processes. This last feature is quite important when the $M/M^{B^*}/1$ has to be connected with other servers within a queueing network. Indeed, we can envision the two types of departure processes as two different 'interfaces' that can be used to connect the queue

within a network. Considering the $M/M^{B^*}/1$ as a module (component) of a queueing network, if the interface used for the connection corresponds to a reversible process, then the module becomes a candidate for ensuring that the stationary distribution of the whole network has a product form with the important computational consequences that make it suitable for practical applications. The bounds and the approximations that we propose in this paper thus become relevant in this context where an approximate model can be acceptable if the computational cost of its solution is limited.

Since the results in this paper refer to the computation of the queue length distributions for different model's configurations, performance figures such as the expected value of the queue length (as well as higher moments of the queue length distribution) and expected waiting times are easy to derive. Obviously, the order relationships among the queue length distributions that have been obtained for these various configurations can be easily extended to the comparisons among derived performance indices. The derivation and the analysis of the distribution of the waiting time is instead a complex topic on its own and must be considered to be outside the scope of this paper.

A full understanding of the relevance of this results for product form queueing network applications will be the topic of future research where queues with bulk services and bulk arrivals will be considered as components of complex queueing networks generalizing the work that we have already discussed in Balbo and Vigliotti (submitted for publication) and showing the practical relevance of these new results in many application scenarios.

## REFERENCES

[1] Muller, A. and Stoyan, D. (2002) *Comparison Methods for Stochastic Models and Risks*. Wiley.
[2] Kelly, F. (1979) *Reversibility and Stochastic Networks*. Wiley.
[3] Chao, X., Miyazawa, M. and Pinedo, M. (1999) *Queueing Networks*. Wiley.
[4] Neuts, M. (1967) A general class of bulk queues with Poisson input. *Ann. Math. Stat.*, **38**, 759–770.
[5] Bagchi, T. and Templeton, J. (1973) Finite waiting space bulk queuing systems. *J. Eng. Math.*, **7**, 313–317.
[6] Chaudhry, M. and Templeton, J. (1983) *A First Course in Bulk Queues*. Wiley.
[7] Kleinrock, L. (1975) *Queueing Systems*, Volume I (Theory). Wiley.
[8] Gross, D. and Harris, C.M. (1998) *Fundamentals of Queueing Theory*. Wiley.
[9] Harrison, P.G. (2004) Compositional reversed Markov processes, with applications to G-networks. *Perform. Eval.*, **57**, 379–408.
[10] Gelenbe, E. (1989) Random neural networks with positive and negative signals and product form solution. *Neural Comput.*, **1**, 502–510.
[11] Henderson, W. and Taylor, P. (1990) Product form in networks of queues with batch arrivals and batch services. *Queueing Syst.*, **6**, 71–88.
[12] Massey, W.A. (1987) Stochastic orderings for Markov processes on partially ordered spaces. *Math. Oper. Res.*, **17**, 350–367.
[13] Castel-Taleb, H. and Pekergin, N. (2009) Stochastic Monotonicity in Queueing Networks. *EPEW 2009: Proc. European Performance Evaluation Workshop*, pp. 116–130. Springer, London, LNCS.
[14] Serfozo, R. (1999) *Introduction to Stochastic Networks*. Springer.
[15] Chao, X., Serfozo, R.F., Miyazawa, M. and Takada, H. (1998) Markov network processes with product form stationary distribution. *Queuing Syst.*, **28**, 377–401.
[16] Gelembe, E. (1991) Product-form queueing networks with negative and positive customers. *J. Appl. Probab.*, **28**, 656–663.

## APPENDIX. DERIVATION DETAILS

Derivation of Equation (19):

$$\pi_i \eta_i = \frac{\pi_i}{\sigma_i}\lambda - \delta\pi_i \left[\prod_{k=i+1}^{B-1}\sigma_k\right]\left[\sum_{j=1}^{i}\rho^j\right],$$

$$\pi_i \sigma_i \eta_i + \pi_i \sigma_i \delta \left[\prod_{k=i+1}^{B-1}\sigma_k\right]\left[\sum_{j=1}^{i}\rho^j\right] = \pi_i\lambda,$$

$$\sigma_i = \frac{\lambda}{\eta_i + \delta[\prod_{k=i+1}^{B-1}\sigma_k][\sum_{j=1}^{i}\rho^j]}$$

$$= \frac{\lambda}{\eta_i + \delta[\prod_{k=i+1}^{B-1}\sigma_k][\rho((1-\rho^i)/(1-\rho))]}. \tag{A.1}$$

Derivation of Equation (33):

$$\eta_{B-2} = \frac{\lambda}{\rho} - \delta\,\rho^2 \frac{1-\rho^{B-2}}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\,\rho\,\frac{\rho - \rho^{B-1}}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\,\rho\,\frac{(\rho-1) + 1 - \rho^{B-1}}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\,\rho\,\frac{(1-\rho^{B-1})}{1-\rho} + \delta\,\rho$$

$$= \eta_{B-1} + \delta\,\rho$$

$$= \mu + \delta\,\rho. \tag{A.2}$$

Derivation of Equation (34):

$$\eta_i = \frac{\lambda}{\rho} - \delta\rho^{B-i}\frac{1-\rho^i}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\rho^{B-i-1}\frac{\rho - \rho^{i-1}}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\rho^{B-i-1}\frac{(\rho-1) + 1 - \rho^{i-1}}{1-\rho}$$

$$= \frac{\lambda}{\rho} - \delta\rho^{B-(i+1)}\frac{(1-\rho^{i+1})}{1-\rho} + \delta\,\rho^{B-(i-1)}$$

$$= \eta_{i+1} + \delta\,\rho^{B-(i+1)} \quad i = 1, 2, \ldots, (B-2). \tag{A.3}$$

Derivation of Equation (35):

$$\eta_i = \eta_{i+1} + \delta\rho^{B-(i+1)}$$

$$= \eta_{i+2} + \delta\rho^{B-(i+2)} + \delta\rho^{B-(i+1)}$$

$$\vdots$$

$$= \eta_{B-1} + \sum_{h=1}^{(B-1)-i} \delta\,\rho^h$$

$$= \mu + \rho\delta\frac{1-\rho^{(B-1)-i}}{1-\rho} \quad 1 \le i \le (B-1). \tag{A.4}$$

Derivation of Equation (44):

$$\frac{\lambda - \rho(\nu + \delta)}{\rho\delta} = \frac{\rho(1-\rho^{B-1})}{1-\rho},$$

$$\nu\,\frac{\lambda - \rho\mu}{\rho\delta}\,\frac{\lambda - \rho\nu}{\lambda - \rho\nu} = \rho\nu\,\frac{(1-\rho^{B-1})}{1-\rho},$$

$$\nu\,\frac{\lambda - \rho\mu}{\lambda - \rho\nu}\frac{\lambda - \rho\nu}{\rho\delta}\frac{1-\rho}{(1-\rho^{B-1})} = \rho\nu, \tag{A.5}$$

$$\eta_{B-1}\frac{\lambda - \rho\nu}{\rho\delta}\frac{1-\rho}{(1-\rho^{B-1})} = \rho\nu.$$

Derivation of Equation (49):

$$\sigma_{B-2} = \frac{\lambda}{\eta_{B-2} + \delta\sigma_{B-1}\rho((1-\rho^{B-2})/(1-\rho))},$$

$$\sigma_{B-2}\eta_{B-2} + \sigma_{B-2}\delta\sigma_{B-1}\rho\,\frac{1-\rho^{B-2}}{1-\rho} = \lambda,$$

$$\rho\nu + \sigma_{B-2}\delta\sigma_{B-1}\rho\,\frac{1-\rho^{B-2}}{1-\rho} = \lambda,$$

$$\sigma_{B-2} = \frac{\lambda - \rho\nu}{\sigma_{B-1}\delta\rho\,((1-\rho^{B-2})/(1-\rho))}$$

$$= \frac{\lambda - \rho\nu}{\frac{((\lambda-\rho\nu)(1-\rho)/\delta\rho(1-\rho^{B-1}))}{\delta\rho((1-\rho^{B-2})/(1-\rho))}}$$

$$= \frac{1-\rho^{B-1}}{1-\rho^{B-2}}. \tag{A.6}$$

Derivation of Equation (49):

$$\sigma_{B-2} = \frac{(\lambda - \rho\nu)}{\rho\delta((1-\rho^{B-2})/(1-\rho))\sigma_{B-1}}$$

$$= \frac{(\lambda - \rho\nu)}{\rho\delta((1-\rho^{B-2})/(1-\rho))((\lambda-\rho\nu)(1-\rho)/}{\delta\rho(1-\rho^{B-1}))}$$

$$= \frac{1-\rho^{B-1}}{1-\rho^{B-2}}. \tag{A.7}$$