

## Fluid queues and mountain processes

***Citation for published version (APA):***

Boxma, O. J., Perry, D., & Duyn Schouten, van der, F. A. (1998). *Fluid queues and mountain processes*. (Memorandum COSOR; Vol. 9828). Technische Universiteit Eindhoven.

***Document status and date:***

Published: 01/01/1998

***Document Version:***

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

***Please check the document version of this publication:***

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

***General rights***

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

***Take down policy***

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.



Eindhoven University  
of Technology

## Department of Mathematics and Computing Sciences

Memorandum COSOR 98-28

### **Fluid queues and mountain processes**

O.J. Boxma

D. Perry

F.A. van der Duyn Schouten

Eindhoven, November 1998  
The Netherlands

# Fluid Queues and Mountain Processes

O.J. Boxma<sup>1,2</sup>, D. Perry<sup>3</sup>, and F.A. van der Duyn Schouten<sup>4,2</sup>

<sup>1</sup> *Eindhoven University of Technology,*

*Department of Mathematics and Computing Science,*

*P.O. Box 513, 5600 MB Eindhoven, The Netherlands;*

<sup>2</sup> *CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands;*

<sup>3</sup> *Dept. of Statistics, University of Haifa, 31999 Haifa, Israel;*

<sup>4</sup> *CentER for Economic Research, Tilburg University,*

*P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

## Abstract

This paper is devoted to the analysis of a fluid queue with a buffer content that varies linearly during periods that are governed by a three-state semi-Markov process. Two cases are being distinguished: (i) two upward slopes and one downward slope, and (ii) one upward slope and two downward slopes. In both cases, at least one of the period distributions is allowed to be completely general. We obtain exact results for the buffer content distribution, the busy period distribution and the distribution of the maximal buffer content during a busy period. The results are obtained by establishing relations between the fluid queues and ordinary queues with instantaneous input, and by using level crossing theory.

*1991 Mathematics Subject Classification:* 60K25, 90B22.

*Keywords & Phrases:* Fluid queue, queue with work removal, workload, maximum content.

## 1 Introduction

In the classical queueing literature, the workload is almost invariably assumed to increase instantaneously upon the arrival of a customer, and to decrease linearly while a server is working. In the related literature of dam and storage processes, there are some studies where the buffer content alternately decreases and increases linearly; see, e.g., Gaver and Miller [12] and Miller [21]. In the beginning of the seventies, queueing models with gradual, non-instantaneous, input emerged, in the pioneering papers of Kosten

[16, 17, 19], Cohen [9], Rubinovitch [25] and Kaspi and Rubinovitch [14]. These concerned models of infinite buffers in which fluid was fed by a number of independent on/off sources, viz., sources that alternate between active (on) and silent (off) periods. The motivation for these ‘fluid queues’ was provided by data communication networks. Kosten [16, 17, 19] successively studied the buffer content for the cases of exponential, Erlang and hyperexponential on-period distributions. Cohen [9] considered the case of identical sources, with *general* on-period distribution. Rubinovitch [25] ( $N$  identical sources) and Kaspi and Rubinovitch [14] ( $N$  non-identical sources) were mainly concerned with the busy period distribution.

The interest in fluid queues was renewed by the fundamental paper of Anick, Mitra and Sondhi [2]. It considers  $N$  on/off sources with exponentially distributed on- and off-periods. Similar to Kosten’s work, the joint equilibrium distribution of the buffer content and the number of busy sources in [2] is described by a set of differential equations. All the eigenvalues are obtained explicitly, and simple expressions are presented for moments of the distribution and the asymptotic behaviour of the buffer content. This paper, along with a new paper of Kosten [18], has stimulated much further research, and has contributed much towards establishing the fluid queue fed by on/off sources as a key model for representing traffic behaviour in modern communication networks. In particular, bursts of cells in ATM-based communication networks naturally give rise to fluid queues. We refer the reader to the survey of Kulkarni [20]. In addition, we would like to mention the methodologically important contribution of Rogers [24], who applies Wiener-Hopf factorization of finite Markov chains to fluid queues.

Most fluid queues studied so far have a Markovian structure. In the last few years, the observation of long-range dependence and non-exponential tail behaviour of modern communication traffic (like Internet traffic, but also Ethernet traffic, VBR video and Wide Area Network traffic) has spurred an interest in non-Markovian fluid queues. The obtained results are mainly qualitatively oriented, proving a particular type of workload tail behaviour under assumptions on the tail behaviour of the on- and/or off-period distributions; see the survey [5].

The papers of Chen and Yao [7] and Kella and Whitt [15] study a broad class of non-Markovian fluid queues, and establish relations between such fluid queues and ordinary queues with instantaneous input. Our paper also aims to link fluid queues and queues with instantaneous input. We restrict

ourselves to the case of an infinite buffer, with a buffer content that varies linearly during periods that are governed by a three-state semi-Markov process (the survey of Kulkarni [20] already mentions the possibility of allowing a semi-Markov structure, but suggests that the analysis is going to be rather hard). We relate the buffer content to the workload of an ordinary single-server queue and to the workload of a single-server queue with additional negative customers that instantaneously remove a random amount of work. A key step in our establishment of such relations is the use of level crossing theory.

Our contribution is two-fold. Firstly, the combination of level crossing theory and the establishment of relations between fluid queues and queues with instantaneous input seems to be both novel and promising. Secondly, we obtain several new results that are quite explicit and that should be useful in various application areas: communication networks, but also, e.g., production systems. Indeed, fluid queues may also represent the content of a finite buffer in multi-stage production processes that are subject to interruptions of various kind. Buffers can prevent installations downwards in the production process from starvation when an upstream installation is not producing, but they also may prevent upstream machines from blocking when a downward installation is temporarily unavailable. There are various questions of interest here. In flexible manufacturing systems, where the same machines are used for a variety of products, a relevant question is how much of the intermediate products should be locally stored in order to guarantee a desired level of throughput. In the oil industry optimal physical sizes of intermediate buffers have to be determined. In this situation not only the throughput is of interest but also the probability of buffer overflow. Hence a key quantity of interest is the distribution of the maximum of the buffer content during a busy period ('top of the mountain'). This suggested the terminology 'mountain process' used in the title as an alternative to fluid queues.

The remainder of the paper is organized in the following way. The model under consideration is described in Section 2. Two cases are being distinguished. In Case (i) there are two upward slopes (the periods having an exponential and a general distribution, respectively) and one downward slope (the period having an exponential distribution). In Case (ii) there is one upward slope (the period having a general distribution) and two downward slopes (the periods having an exponential and a general distribution, respec-

tively). In Section 3 we present an exact analysis of the buffer content for Case (i), and in Section 4 for Case (ii). The busy period distribution, for a class of fluid queues containing Case (i), is determined in Section 5, while the distribution of the ‘top of the mountain’ is studied in Sections 6 and 7 for Case (i) and Case (ii) respectively.

## 2 Model formulation

Let  $\tilde{J} = \{\tilde{J}(t), t \geq 0\}$  be a semi-Markov process with state space  $\tilde{\mathcal{E}} = \{0, 1, 2\}$ , and let  $T_n, n = 0, 1, \dots$ , be the transition times (with  $T_0 = 0$ ). We define the fluid process  $\tilde{X} = \{\tilde{X}(t), t \geq 0\}$  by  $\tilde{X}(0) = 0$  and

$$\tilde{X}(t) = \tilde{X}(T_n) + (\tilde{a}\tilde{J}(t) - \tilde{b})(t - T_n), \quad T_n \leq t < T_{n+1}, \quad n = 0, 1, \dots,$$

where  $\tilde{a}$  and  $\tilde{b}$  are known constants. By applying the reflection map

$$X(t) := \tilde{X}(t) - \min_{0 \leq s \leq t} \tilde{X}(s),$$

and by defining  $J(t) := \tilde{a}\tilde{J}(t) - \tilde{b}$ , we obtain the so-called *mountain process*  $(X, J) = \{(X(t), J(t)), t \geq 0\}$ . We assume that the conditions for stability hold, and that  $(X, J)$  is a regenerative process.  $X$  can be interpreted as the content level of an infinite buffer that has net inflow rate  $J(t)$  when it is not empty, viz., that receives fluid input at rate  $\tilde{a}j$  when  $\tilde{J}(t) = j$  and which has outflow rate  $\tilde{b}$  as long as there is fluid in the buffer. The stability conditions in fact amount to the assumption that the total amount of fluid offered per time unit is less than  $\tilde{b}$ .

Our model is related to that of Anick et al. [2]. They do not restrict themselves to the case in which the state space of  $J$  has only three states; on the other hand, in their model the times between transitions are all exponential, so that  $(X, J)$  is a Markov process. When  $J$  is a Markov process, one can relate this model to that of two machines that are subject to breakdown and repair, each with production rate  $\tilde{a}$  when working, and with total demand rate  $\tilde{b}$ . A similar (but more general) interpretation holds when  $(X, J)$  is a Markov renewal process. Then, still,  $X$  is the content and  $J$  is the slope, respectively. However, the underlying model no longer is that of two machines that are subject to breakdown, since the lack-of-memory property does not hold.

We distinguish in the following between two cases:

Case (i):  $\tilde{a} > \tilde{b}$ . The mountain process now has one negative slope  $b = \tilde{b}$  and two positive slopes  $a_1 = \tilde{a} - \tilde{b}$  and  $a_2 = 2\tilde{a} - \tilde{b}$  (in fact, for the analysis in this paper it is sufficient to assume that there are two positive slopes  $a_2 > a_1$  and one negative slope  $b$ ). The time intervals in which  $J$  stays in state  $b$ ,  $a_1$ ,  $a_2$  are respectively called *b-intervals*, *a<sub>1</sub>-intervals* and *a<sub>2</sub>-intervals*. Case (i) is characterized by Figure 1.

FIGURE 1 HERE

The *b*-intervals are assumed to be exponentially distributed with rate  $\mu_b$ ; the *a<sub>1</sub>*-intervals are exponentially distributed with rate  $\lambda_1 + \mu_1$ ; but we allow the *a<sub>2</sub>*-intervals to be generally distributed, with distribution  $G_{a_2}(\cdot)$  and Laplace-Stieltjes transform (LST)  $G_{a_2}^*(\omega)$ .

Motivated by the concept of producing machines, it is assumed that transitions from states  $b$  and  $a_2$  can only be to state  $a_1$ ; see Figure 1.

Case (ii):  $\tilde{b}/2 < \tilde{a} < \tilde{b}$ . The mountain process now has two negative slopes  $b_1 = \tilde{b}$  and  $b_0 = \tilde{b} - \tilde{a}$ , and one positive slope  $a = 2\tilde{a} - \tilde{b}$  (again, for the analysis in this paper it is sufficient to assume that there are two negative slopes  $b_0$  and  $b_1$  and one positive slope  $a$ ). We now speak of *b<sub>1</sub>-intervals*, *b<sub>0</sub>-intervals*, and *a-intervals*. The *b<sub>0</sub>*-intervals are assumed to be exponentially distributed with rate  $\lambda_0 + \mu_0$ ; however, we allow the *b<sub>1</sub>*-intervals and the *a*-intervals to be generally distributed, with distribution  $G_{b_1}(\cdot)$  and  $G_a(\cdot)$ , respectively. Their respective LST's are denoted by  $G_{b_1}^*(\omega)$  and  $G_a^*(\omega)$ . As above, transitions only take place to neighbouring states; see Figure 2.

FIGURE 2 HERE

In Section 3 we discuss the buffer content process for Case (i), and in Section 4 for Case (ii).

### 3 The buffer content and its related jump processes – Case (i)

In this section we derive the steady-state distribution of the buffer content  $X$  for Case (i), i.e., the case with two positive slopes and one negative slope. Key elements of our approach are Level Crossing Theory (LCT) and the

establishment of a relation between buffer contents of fluid queues and of queues with instantaneous input and/or work removal. We observe that the mountain process  $(X, J)$  is a regenerative process. Each cycle is generated by an alternating renewal process of activity periods and silence periods. The first activity period starts at  $(X(0), J(0)) = (0, a_1)$  and terminates at  $P = \inf\{t > 0 : X(t) = 0\}$ . Since there is only one negative slope in Case (i),  $(X(P), J(P)) = (0, b)$ . Clearly, by the lack-of-memory property, the silence periods are  $\exp(\mu_b)$ . These are the periods during a cycle in which the buffer is empty.

As announced above, we now establish a relation between buffer contents of fluid queues and of queues with instantaneous input and/or work removal. Observe the unreflected fluid process  $(\tilde{X}, \tilde{J})$  with  $\tilde{X}(0) = 0$  and  $\tilde{J}(0) = 1$ , and construct two compound Poisson processes with an additional drift component, in the following way:

1. Delete the  $a_1$ -intervals and the  $a_2$ -intervals, and glue together the consecutive  $b$ -intervals. The resulting process is a compound Poisson process with a negative slope  $b$  and with upward jumps that have a phase-type distribution. Since  $(X(0), J(0)) = (0, a_1)$ , the first jump,  $S_0$ , occurs at 0. We call this process  $S_0 + L_F(t)$ , where  $S_0$  is independent of  $L_F(t)$ .

$$L_F(t) := S_1 + \dots + S_{N_1(t)} - bt,$$

where  $N_1(t)$  is a Poisson process with rate  $\mu_b$  and the  $S_i$  are i.i.d. random variables with distribution  $H(t)$  and LST  $H^*(\omega)$ . These  $S_i$ ,  $i = 0, 1, \dots$ , represent the total increments accumulated during a sequence of  $a_1$ -,  $a_2$ -,  $\dots$ ,  $a_1$ -intervals in between two consecutive  $b$ -intervals. It is easy to show that

$$H^*(\omega) = \frac{\frac{\lambda_{a_1}}{\lambda_{a_1} + \mu_{a_1} + \omega}}{1 - \frac{\mu_{a_1}}{\lambda_{a_1} + \mu_{a_1} + \omega} G_{a_2}^*(a_2\omega)}, \quad \text{Re } \omega \geq 0, \quad (3.1)$$

where  $\lambda_{a_1} := \lambda_1/a_1$ ,  $\mu_{a_1} := \mu_1/a_1$ .

By applying the reflection map

$$V_F(t) := L_F(t) - \min_{0 \leq s \leq t} L_F(s),$$

we obtain that  $V_F(t)$  is also a regenerative process and by scaling time such that  $\mu := \mu_b/b$ ,  $V_F(t)$  can be interpreted as the work process of the  $M/G/1$  queue with arrival rate  $\mu$  and service times having LST  $H^*(\omega)$  as given in



(3.1). The subscript  $F$  for  $L_F(t)$  and  $V_F(t)$  designates that the work process  $V_F(t)$  proceeds "forward" in time.

2. Delete the  $a_2$ -intervals and the  $b$ -intervals, and glue together the consecutive  $a_1$ -intervals. The resulting process  $L_B(t)$  is a compound Poisson process with positive slope  $a_1$  and with upward and downward jumps. The positive jumps occur according to a Poisson process with rate  $\mu_1$ , and the jump size distribution has LST  $G^*(a_2\omega)$ ; the negative jumps occur according to a Poisson process with rate  $\lambda_1$ , and the jump sizes are exponentially distributed of rate  $\mu = \mu_b/b$ . As before we apply the reflection map

$$V_B(t) := L_B(t) - \min_{0 \leq s \leq t} L_B(s).$$

The subscript  $B$  for  $L_B(t)$  and  $V_B(t)$  designates that  $V_B(t)$  is a process that proceeds "backward" in time. Typical sample paths of  $(X, J)$  and their constructed reflected processes  $V_F(t)$  and  $V_B(t)$  are given in Figure 3.

FIGURE 3 HERE

Note that  $V_F(t)$  as defined above should not start with a jump at the origin; but we have drawn it with such a jump to enable a comparison between the laws of  $V_F$  and  $V_B$  in the next Lemma. Here  $V_F$  and  $V_B$  denote random variables with distribution the limiting distribution of  $V_F(t)$  and  $V_B(t)$ , respectively.

**Lemma 3.1** *Denote by  $(V_F|V_F > 0)$  the random variable generated by conditioning the law of  $V_F$  on the event  $(V_F > 0)$ . Let  $S$  and  $S_e$  be random variables having LST  $H^*(\omega)$  and  $(1 - H^*(\omega))/\omega \mathbb{E}S$ , respectively; i.e.,  $S_e$  is the equilibrium random variable of  $S$ . Then*

$$S + (V_F|V_F > 0) \simeq S_e + V_B, \tag{3.2}$$

where  $\simeq$  denotes equality in distribution, and where the various terms are independent.

**Proof:**

$V_F(t)$  and  $V_B(t)$  are Lévy processes starting at the origin. Thus, we can identify the law of  $S_0 + \sup_{0 \leq t < \infty} L_F(t)$  with  $S + V_F$ , and the law of  $\sup_{0 \leq t < \infty} L_B(t)$

with that of  $V_B$ . By construction of  $L_F(t)$  and  $L_B(t)$ , it follows that  $S_0 + \sup_{0 \leq t < \infty} L_F(t) = \sup_{0 \leq t < \infty} L_B(t)$  with probability 1, so that  $S + V_F \simeq V_B$ . This implies that

$$S + V_F + S_e \simeq V_B + S_e, \quad (3.3)$$

where  $S_e$  is independent of  $S + V_F$  and of  $V_B$ . Now remark that the sum of  $V_F$  and  $S_e$  in the  $GI/G/1$  queue has the conditional steady-state distribution of  $(V_F | V_F > 0)$ , cf. [11] p. 296. The lemma is proven.  $\spadesuit$

In terms of LST's, Lemma 3.1 says simply that if  $\psi_F(\omega)$  and  $\psi_B(\omega)$  are the LST's of  $V_F$  and  $V_B$ , respectively, then:

### Corollary 3.2

$$\psi_B(\omega) = \frac{\omega \text{E}SH^*(\omega) \psi_F(\omega) - F_F(0)}{1 - H^*(\omega) \frac{\psi_F(\omega) - F_F(0)}{1 - F_F(0)}}, \quad \text{Re } \omega \geq 0, \quad (3.4)$$

where  $1 - F_F(0)$  is the probability of the event  $(V_F > 0)$ .

The next theorem relates the steady-state law of  $X$  to that of  $V_F$  (or  $V_B$ ). First some definitions. Let  $F_X(\cdot)$  denote the steady-state distribution of  $X$ , with LST  $\psi_X(\omega)$ ,  $P_b := \lim_{t \rightarrow \infty} \mathbf{P}(J(t) = b)$ ,  $P_{a_1} := \lim_{t \rightarrow \infty} \mathbf{P}(J(t) = a_1)$ , and  $P_{a_2} := \lim_{t \rightarrow \infty} \mathbf{P}(J(t) = a_2)$ . In addition,  $F_F(x) = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) \leq x | J(t) = b)$ ,  $F_B(x) = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) \leq x | J(t) = a_1)$ ,  $F^*(x) = \lim_{t \rightarrow \infty} \mathbf{P}(X(t) \leq x | J(t) = a_2)$ .

### Theorem 3.3

$$\psi_X(\omega) - F_X(0) = K^*(\omega)[\psi_F(\omega) - F_F(0)], \quad \text{Re } \omega \geq 0, \quad (3.5)$$

where

$$K^*(\omega) := \left(1 + \frac{b}{a_2}\right)P_b + \left(1 - \frac{a_1}{a_2}\right)P_{a_1} \frac{\omega \text{E}SH^*(\omega)}{1 - H^*(\omega)} \frac{1}{1 - F_F(0)}, \quad \text{Re } \omega \geq 0.$$

**Proof:**

The state space of  $J(t)$  is  $\mathcal{E} = \{a_1, a_2, b\}$  and according to the construction of  $L_F(t)$  and  $L_B(t)$  we have

$$\begin{aligned} F_X(x) &= \lim_{t \rightarrow \infty} \sum_{j \in \mathcal{E}} \mathbf{P}(X(t) \leq x | J(t) = j) \mathbf{P}(J(t) = j) \\ &= F_F(x)P_b + F_B(x)P_{a_1} + F^*(x)P_{a_2}. \end{aligned}$$

By LCT,  $F_B(\cdot)$  and  $F^*(\cdot)$  are absolutely continuous distributions for all  $x \geq 0$  and  $F_F(\cdot)$  is an absolutely continuous distribution for all  $x > 0$ , so that  $F_X(x) = F_X(0) + \int_0^x \pi(y)dy$ , where  $\pi(\cdot)$  is the steady-state density of  $X$ . Clearly,  $F_X(0) = F_F(0)P_b > 0$ . In terms of densities we have

$$\pi(x) = f_F(x)P_b + f_B(x)P_{a_1} + f^*(x)P_{a_2}, \quad x > 0,$$

where  $f_F(\cdot)$ ,  $f_B(\cdot)$  and  $f^*(\cdot)$  are the densities associated with  $F_F(\cdot)$ ,  $F_B(\cdot)$  and  $F^*(\cdot)$  for all  $x > 0$ . However, by LCT,  $b f_F(x)P_b$  is the long-run average number of downcrossings of level  $x > 0$  by  $(X, J)$  and  $a_1 f_B(x)P_{a_1} + a_2 f^*(x)P_{a_2}$  is the long-run average number of upcrossings of that level (cf. [10]). Therefore, since  $(X, J)$  is a regenerative process, we obtain:

$$\pi(x) = \left(1 + \frac{b}{a_2}\right) f_F(x)P_b + \left(1 - \frac{a_1}{a_2}\right) f_B(x)P_{a_1}, \quad x > 0. \quad (3.6)$$

Taking Laplace transforms and invoking Corollary 3.2, the theorem is proven.

♠

**Remark 3.1.** (i) The probabilities  $P_b$ ,  $P_{a_1}$  and  $P_{a_2}$  are the limiting probabilities of a (three-state) semi-Markov process. Therefore, they can be easily computed; for more details see [8], Ch. 10. Using  $F_X(0) = F_F(0)P_b$ , one can subsequently eliminate  $F_F(0)$ ; finally,  $F_X(0)$  is obtained via the normalization condition  $\psi_X(0) = 1$ .

(ii) Naturally,  $a_2 > a_1$ , but Theorem 3.3 is valid for any positive constants  $a_1$  and  $a_2$ . Even if the second term of  $K^*(\omega)$  is negative for  $\omega \geq 0$ , it follows by LCT that  $K^*(\omega) \geq 0$ .

(iii) The special case  $a_1 = a_2$  implies that  $K^*(\omega)$  is a constant (independent of  $\omega$ ). Therefore, by substituting  $\omega = 0$  in Theorem 3.3, we obtain that this

constant must be 1. For this special case it has already been shown by Kella and Whitt [15] that

$$\frac{F_X(x) - F_X(0)}{1 - F_X(0)} = \frac{F_F(x) - F_F(0)}{1 - F_F(0)}.$$

## 4 The buffer content and its related compound Poisson processes – Case (ii)

In this section we derive the steady-state distribution of the buffer content  $X$  for Case (ii), i.e., the case with one positive slope and two negative slopes. The state space of  $J(t)$  now is  $\mathcal{E} = \{a, b_0, b_1\}$ . We assume that the  $a$ -intervals and the  $b_1$ -intervals are generally distributed, while the  $b_0$ -intervals are exponentially distributed. By deleting the  $a$ -intervals and the  $b_1$ -intervals and glueing together the  $b_0$ -intervals we obtain a compound Poisson process with a negative slope and with positive and negative jumps. By subsequently applying the reflection map we obtain the workload process of a certain queueing model which we call the  $M/G_U/G_D/1$  queue.  $G_U$  and  $G_D$ , respectively, refer to upward and downward jumps. This queueing model has been analyzed in [4]. It is an  $M/G/1$ -type queue with the extra feature that, according to a Poisson process, stochastic amounts of work are removed from the system.

Let  $V_{NF}(t)$  be the workload of this  $M/G_U/G_D/1$  queue, and let  $V_{NF} := \lim_{t \rightarrow \infty} V_{NF}(t)$  in the sense of weak convergence. More specifically, we let

$$L_{NF}(t) := \hat{S}_1 + \dots + \hat{S}_{\hat{N}_1(t)} - (\hat{Z}_1 + \dots + \hat{Z}_{\hat{N}_2(t)}) - b_0 t,$$

where  $\hat{N}_1(t)$  and  $\hat{N}_2(t)$  are independent Poisson processes with rates  $\mu_0$  and  $\lambda_0$ , respectively. The  $\hat{S}_i$  are i.i.d. random variables having distribution  $G_a(x/a)$  and the  $\hat{Z}_j$  are i.i.d. random variables, also independent of the  $\hat{S}_i$ , having distribution  $G_{b_1}(x/b_1)$ . Also assume with no loss of generality that  $b_0 = 1$ .  $V_{NF}(t)$  is obtained by applying the reflection

$$V_{NF}(t) = L_{NF}(t) - \min_{0 \leq s \leq t} L_{NF}(s).$$

The subscript  $NF$  designates "negative-forward".

Next, by deleting the (exponentially distributed)  $b_0$ -intervals and the (generally distributed)  $b_1$ -intervals, we obtain a process with positive slope. Formally, this process can be written in the following way:

$$L_E(t) := at - (\bar{S}_1 + \dots + \bar{S}_{\bar{N}(t)}).$$

Here the  $\bar{S}_i$  are i.i.d. positive random variables;  $\bar{N}(t) := \inf\{n \geq 1 : \bar{Z}_1 + \dots + \bar{Z}_n \geq t\} - 1$ , with the  $\bar{Z}_j$  i.i.d. positive random variables, which are also independent of the  $\bar{S}_i$ . The  $\bar{S}_i$  represent jumps composed of decrements during a succession of  $b_0$ - and  $b_1$ -periods in between two successive  $a$ -periods. The  $\bar{Z}_j$  represent the lengths of successive  $a$ -periods. By applying the reflection map

$$V_E(t) = L_E(t) - \min_{0 \leq s \leq t} L_E(s),$$

we obtain that  $V_E(t)$  is the so-called *elapsed waiting time* process (EWT) of a certain  $GI/G/1$  queue (cf. [22]). The subscript  $E$  designates "elapsed". The EWT is the time elapsed since the arrival of the customer in service. Note that under this interpretation the system is with probability 1 never empty, since the idle periods are deleted. The  $\bar{S}_i$  are the interarrival times and the  $\bar{Z}_j$  are the interdeparture times of customers within the busy period; the time to departure of the first customer in the busy period is truncated to its service requirement. Now reverse time. Departures and arrivals are reversed so that interarrival times become interdeparture times and vice versa. This technique is used to construct the EWT from the work process. Generally, the duality existing between the  $M/G/1$ -type and the  $G/M/1$ -type dam models is shown by time reversal (see [22] and [13]). Typical realisations of  $X(t)$ ,  $V_{NF}(t)$  and  $V_E(t)$  are shown in Figure 4.

FIGURE 4 HERE

Again,  $V_{NF}(t)$  as defined above should not start with a jump at the origin, but we have drawn it with such a jump to enable a comparison of the laws of  $V_{NF}$  and  $V_E$  (cf. Remark 4.1 below). Here  $V_{NF}$  and  $V_E$  are random variables with distribution the limiting distribution of  $V_{NF}(t)$  and  $V_E(t)$ , respectively. Since  $L_{NF}(t)$  is a Lévy process, one can identify the law of  $\sup_{0 \leq t < \infty} L_{NF}(t)$  with that of  $V_{NF}$  [3].

In [4] it is shown that the limiting distribution of  $V_{NF}(t)$  is equal to the steady-state distribution of the actual waiting time of a certain  $GI/G/1$

queue. This  $GI/G/1$  queue is obtained from the  $M/G_U/G_D/1$  queue by replacing the downward jumps by elongations of the interarrival times, in such a way that the workload at arrival epochs of ordinary customers (upward jumps) is the same in the  $M/G_U/G_D/1$  queue and the  $GI/G/1$  queue. In that  $GI/G/1$  queue, the interarrival times are exactly the  $\bar{S}_i$  and the service times the  $a\bar{Z}_j$ . This is the  $GI/G/1$  queue whose EWT is  $V_E(t)$ . In the next lemma we introduce the formal relations between the laws of  $V_{NF}$  and  $V_E$ .

**Lemma 4.1**

$$V_{NF} \simeq S_e + V_E. \quad (4.1)$$

**Proof:**

Let  $W_n^{GI/G/1}$  and  $V^{GI/G/1}(t)$  be, respectively, the actual waiting time of the  $n$ -th customer and the workload at time  $t$  of a certain  $GI/G/1$  queue in which the service requirements are the  $\bar{S}_i$  and the interarrival times are the  $\bar{Z}_j$ . Also let  $W^{GI/G/1} = \lim_{n \rightarrow \infty} W_n^{GI/G/1}$  and  $V^{GI/G/1} = \lim_{t \rightarrow \infty} V^{GI/G/1}(t)$ , both in the sense of weak convergence. Let  $S_e$  be a random variable that is independent of  $V_{NF}$ , with as distribution the equilibrium distribution of  $S$ . Then

$$\begin{aligned} V_{NF} + S_e &\simeq W^{GI/G/1} + S_e \\ &\simeq (V^{GI/G/1} | V^{GI/G/1} > 0) \\ &\simeq V_E. \end{aligned}$$

Here the first equality follows from [4], the second from p. 296 of [11], and the third by time reversal. ♠

**Remark 4.1** Below we sketch a different proof of Lemma 4.1. It follows from renewal theory that  $V_E \simeq R + S_e$ , where  $R$  is a stochastic variable with distribution the steady-state distribution of the  $\{V_E(t), t \geq 0\}$  process just after downward jumps. But note that there is a 1 – 1 correspondence between the values of  $V_E(t)$  just after downward jumps and the values of  $V_{NF}(t)$  just before a upward jumps. Finally, PASTA implies that  $V_{NF}$  has the same distribution as  $R$ .

**Remark 4.2** Having obtained a relation between the distributions of  $V_{NF}$  and  $V_E$ , the density  $\pi(x)$  of the buffer content for Case (ii) can be obtained similarly as in Theorem 3.3. Introducing the densities  $f_{NF}(x)$  and  $f_E(x)$  of  $V_{NF}$  and  $V_E$  for  $x > 0$ , LCT leads as in (3.6) to:

$$\pi(x) = \left(1 - \frac{b_0}{b_1}\right)f_{NF}(x)P_{b_0} + \left(1 + \frac{a}{b_1}\right)f_E(x)P_a, \quad x > 0. \quad (4.2)$$

Finally,  $\pi(x)$  follows from known results for  $f_{NF}(x)$  in [4] and Lemma 4.1.

## 5 Busy period analysis

In this section we present a method for determining the busy period distribution for a class of fluid queues *with one negative slope*. We then apply that method to the model of Case (i). The problem of obtaining the busy period distribution for fluid queues in which more than one negative slope can occur (like the fluid queue of Case (ii)) is still open.

Consider a fluid queue with i.i.d. activity periods  $A_1, A_2, \dots$  that are alternated by i.i.d. silence periods  $I_1, I_2, \dots$ . All activity and silence periods are independent. The activity period distribution remains unspecified; the silence periods are exponentially distributed with rate  $\mu_b$ . During an activity period  $A_i$ , the buffer content monotonously increases; the total increment is  $B_i(A_i)$ . During a silence period, the buffer content decreases at constant rate  $b$ . So in the framework of Case (i), the  $b$ -periods are the silence periods, while the activity periods are compositions of  $a_1$ - and  $a_2$ -periods.

A busy period is a period during which the buffer is never empty. Consider a busy period  $P$  that starts with activity period  $A_1$ . Its length will be at least equal to  $A_1 + B_1(A_1)/b$ . Adapting Takács' argument for the busy period of the ordinary  $M/G/1$  queue, we can write:

$$P \simeq A_1 + \frac{B_1(A_1)}{b} + P_1 + \dots + P_{K(\frac{B_1(A_1)}{b})}, \quad (5.1)$$

where  $K(t)$  is the counting function of a Poisson process of intensity  $\mu_b$ , that is independent of the sequence  $P_1, P_2, \dots$ . To understand (5.1), first observe that a busy period does not change when the order of handling the fluid is changed (as long as there is an outflow at constant rate  $b$  whenever the buffer

is not empty). Now assume that the fluid  $B_1(A_1)$ , that remains in the buffer after  $A_1$ , flows out of the buffer until the start of the second activity period,  $A_2$ . The outflow of what is left of  $B_1(A_1)$  is then assumed to be delayed until all the fluid generated in  $A_2$ , plus all the fluid that was generated during its handling, has left the buffer. Note that this delay can be viewed as a busy period that was started by  $A_2$ ; because of the independence assumption and the memoryless property of the exponential silence distribution, this busy period has the same distribution as  $P$ . The outflow of what is left of  $B_1(A_1)$  is now resumed; there may be other interruptions, according to a Poisson( $\mu_b$ ) process. This reasoning results in (5.1).

Taking LST in (5.1), we readily obtain the following functional equation for the LST of the busy period distribution. For  $\text{Re } \omega \geq 0$ :

$$\begin{aligned} \mathbb{E}[e^{-\omega P}] &= \mathbb{E}\left[e^{-\omega\left(A_1 + \frac{B_1(A_1)}{b} + P_1 + \dots + P_{K\left(\frac{B_1(A_1)}{b}\right)}\right)}\right] \\ &= \mathbb{E}\left[e^{-\omega\left(A_1 + \frac{B_1(A_1)}{b}\right) - \mu_b(1 - \mathbb{E}[e^{-\omega P}])\frac{B_1(A_1)}{b}}\right]. \end{aligned} \quad (5.2)$$

Note that  $B_i(A_i) = B_i$  and  $A_i = 0$  would yield Takács' functional equation for the busy period LST in an ordinary  $M/G/1$  queue with service times  $B_i$  and service speed (slope)  $b$ . Similarly as in the ordinary  $M/G/1$  queue, one can prove that  $\mathbb{E}[e^{-\omega P}]$  is, for  $\text{Re } \omega \geq 0$ , the unique solution of the functional equation

$$x = \mathbb{E}\left[e^{-\omega\left(A_1 + \frac{B_1(A_1)}{b}\right) - \mu_b(1-x)\frac{B_1(A_1)}{b}}\right],$$

with an absolute value of at most one.

Let us now turn to Case (i), cf. Sections 2 and 3. In this case, the activity period  $A_1$  consists of a succession of  $a_1$ -intervals  $R_{11}, R_{12}, \dots$  and  $a_2$ -intervals  $R_{21}, R_{22}, \dots$ , that are all independent:

$$A_1 = R_{11} + R_{21} + R_{12} + \dots + R_{2,M-1} + R_{1M}. \quad (5.3)$$

The  $R_{1i}$  are exponentially distributed with rate  $\lambda_1 + \mu_1$ , and the  $R_{2i}$  have distribution  $G_{a_2}(\cdot)$ . The random variable  $M$  is geometrically distributed:

$$\mathbf{P}(M = k) = \frac{\mu_1}{\lambda_1 + \mu_1} \left(\frac{\lambda_1}{\lambda_1 + \mu_1}\right)^{k-1}, \quad k = 1, 2, \dots \quad (5.4)$$

Observing that

$$B_1(A_1) = a_1(R_{11} + \dots + R_{1M}) + a_2(R_{21} + \dots + R_{2,M-1}), \quad (5.5)$$



we have:

$$\begin{aligned} \mathbb{E}[e^{-\omega A_1 - \zeta B_1(A_1)}] &= \sum_{k=1}^{\infty} \mathbf{P}(M = k) \mathbb{E}^k[e^{-(\omega + \zeta a_1)R_{11}}] \mathbb{E}^{k-1}[e^{-(\omega + \zeta a_2)R_{21}}] \\ &= \frac{\mu_1}{\lambda_1 + \mu_1 + \omega + \zeta a_1 - \lambda_1 G_{a_2}^*(\omega + \zeta a_2)}. \end{aligned} \quad (5.6)$$

Combination of (5.2) and (5.6) yields a functional equation for the busy period LST in the model of Case (i).  $\mathbb{E}[e^{-\omega P}]$  satisfies:

$$x = \frac{\mu_1}{\lambda_1 + \mu_1 + \omega(1 + a_1/b) + a_1\mu_b(1-x)/b - \lambda_1 G_{a_2}^*(\omega(1 + a_2/b) + a_2\mu_b(1-x)/b)}. \quad (5.7)$$

**Remark 5.1.** It is possible to generalize (5.6) to the case of generally distributed  $a_1$ -intervals, and even to the case in which an activity period consists of a succession of intervals that is determined by an underlying semi-Markov process. However, the expressions become very cumbersome, and these generalizations do not fit well into the present paper.

**Remark 5.2.** In the special case  $a_2 = a_1$  one has  $B_1(A_1) = a_1 A_1$ . Then the busy period LST satisfies the functional equation

$$x = \mathbb{E}[e^{-\{(1+a_1/b)\omega + \mu_b a_1(1-x)/b\}A_1}]. \quad (5.8)$$

Comparison with the functional equation for the LST of the busy period  $P_{M/G/1}$  of an ordinary  $M/G/1$  queue with arrival rate  $\mu_b$ , service speed one, and service times  $A_1, A_2, \dots$ , reveals that, for this special case,  $P \simeq (1 + a_1/b)P_{M/G/1}$ . This result could also have been obtained via a more direct reasoning.

**Remark 5.3.** Boxma and Dumas [6], see also Aalto [1], have derived the busy period LST in a fluid queue fed by  $N$  independent on/off sources, with general on-period distributions and exponential off-period distributions. Note that, in the case of  $N = 2$  sources, this model is a special case of the one considered in the present section.

## 6 The top of the mountain - Case (i)

In this section and the next one, we study the distribution of  $X^{max} := \max\{X(t), 0 < t \leq P\}$ , the ‘top of the mountain’, viz., the maximum of the workload level during one busy period  $P$ . The present section is devoted to Case (i); Case (ii) is discussed in the next section.

In Section 3 we have related the workload process for Case (i) to the workload process  $\{V_F(t), t \geq 0\}$  of an  $M/G/1$  queue with arrival rate  $\mu_b$  and service times  $S_1, S_2, \dots$  with LST given by (3.1). In the sequel,  $S$  denotes a generic service time with this LST. This  $M/G/1$  queue was obtained from the fluid queue by deleting the  $a_1$ - and  $a_2$ -intervals and glueing together the  $b$ -intervals. For simplicity we assume that the slope in the  $b$ -intervals equals  $b = 1$ ; this is just a matter of scaling.

Obviously,

$$\mathbf{P}(X^{max} \leq x) = \mathbf{P}(V_F^{max} \leq x), \quad x \geq 0, \quad (6.1)$$

where  $V_F^{max}$  is the maximum of the workload during one busy period of the process  $\{V_F(t), t \geq 0\}$ . The distribution of  $V_F^{max}$  for the  $M/G/1$  queue has been extensively discussed in [3] and [11]. We are thus able to prove the following

**Theorem 6.1** *In Case (i), the steady-state distribution of the top of the mountain is given by:*

$$\mathbf{P}(X^{max} \leq x) = \frac{\mathbf{P}(V_F + S \leq x)}{\mathbf{P}(V_F \leq x)} = 1 - \frac{1}{\mu} \frac{d}{dx} \log \mathbf{P}(V_F \leq x), \quad x \geq 0. \quad (6.2)$$

**Proof:**

The first equality in (6.2) follows from (6.1) and [3]. Agreement of the first and last term follows from (6.1) and [11], p. 618. Agreement of the two representations is also easily shown by using Formula (II.5.110) on p. 297 of [11]. ♠

**Remark 6.1.** The above results remain valid when Case (i) is generalized to the fluid queue introduced in the beginning of the previous section. That is a queue in which activity periods  $A_1, A_2, \dots$  are alternated by exponentially distributed silence periods. During the silence periods, the buffer content

decreases at constant rate  $b$  (when non-empty), and during an activity period, the buffer content increases monotonously (but not necessarily linearly). Glue the silence periods together, with upward jumps in between that are equal to the  $B_i(A_i)$ . Then the maximum buffer content  $X^{max}$  during a busy period coincides with the maximum level  $V_F^{max}$  of the corresponding  $M/G/1$  queue with arrival rate  $\mu$ , slope  $b$  (for simplicity assumed to equal one), and service times  $B_1(A_1), B_2(A_2), \dots$ . In the special Case (i), it is easy to obtain the distribution of those service times, and hence that of  $V_F^{max}$ ; but it is also possible to obtain those distributions in, e.g., the semi-Markov case that was mentioned in Remark 5.1.

In Section 3 we have related the workload process for Case (i) not only to the process  $\{V_F(t), t \geq 0\}$  but also to the process  $\{V_B(t), t \geq 0\}$ . The latter process was obtained by glueing together successive  $a_1$ -intervals. We now show how this leads to yet another expression for the distribution of the top of the mountain.

The reflected process  $V_B(t)$  of  $L_B(t)$  is the EWT of the queueing system  $M/G_U/M_D/1$ . This is an  $M/G/1$ -type queue with the additional feature that at exponentially distributed intervals, an exponentially distributed amount of work is removed from the system (such an  $M/G/1$ -type queue with work removal has already been discussed in the beginning of Section 4). We identify the law of  $V_B$  with that of  $\sup_{0 \leq t < \infty} L_B(t)$ . We then have:

**Theorem 6.2**

$$\mathbf{P}(X^{max} \leq x) = \frac{\mathbf{P}(V_B \leq x)}{\mathbf{P}(V_B \leq x + U)}, \quad x \geq 0, \quad (6.3)$$

where  $U$  is exponentially distributed with rate  $\mu = \mu_b/b$  and independent of  $V_B$ .

**Proof:**

Let  $T$  be the busy period of the above  $M/G_U/M_D/1$  queue. Then

$$\begin{aligned} \mathbf{P}(V_B \leq x) &= \mathbf{P}(\sup_{0 \leq t < \infty} L_B(t) \leq x) \\ &= \mathbf{P}(\sup_{0 \leq t < T} L_B(t) \leq x) \mathbf{P}(\sup_{T \leq t < \infty} L_B(t) \leq x) \\ &= \mathbf{P}(V_B^{max} \leq x) \mathbf{P}(\sup_{0 \leq t < \infty} L_B(t) \leq x + U). \end{aligned}$$

Here  $V_B^{max}$  denotes the maximum of the workload process in the  $M/G_U/M_D/1$  queue during a busy period  $T$ . The second equality follows from the strong Markov property and the third from the lack-of-memory property of  $U$ . The theorem is proved since  $V_B^{max}$  and  $X^{max}$  have the same law. ♠

**Remark 6.2.** In the proof of Lemma 3.1 it was observed that  $S + V_F \simeq V_B$ . In combination with the first equality in (6.2), this relation easily leads to Theorem 6.2. Indeed, it then suffices to prove that

$$\mathbf{P}(V_F \leq x) = \mathbf{P}(S + V_F \leq x + U), \quad x \geq 0. \quad (6.4)$$

But since  $V_F$  is the workload in an  $M/G/1$  queue with service time  $S$  and interarrival time  $U$ , it has the same distribution as the waiting time in that  $M/G/1$  queue; and the steady-state waiting time  $W$  in that  $M/G/1$  queue satisfies the relation

$$W \simeq \max[0, W + S - U].$$

## 7 The top of the mountain - Case (ii)

In this section we study the distribution of  $X^{max}$ , the top of the mountain, for the model of Case (ii). In Section 4 we have related the workload process of Case (ii) to the workload process  $\{V_{NF}(t), t \geq 0\}$  of an  $M/G_U/G_D/1$  queue. Recall that  $L_{NF}(t)$  is the Lévy process generated by deleting the  $a$ -intervals and the  $b_1$ -intervals, and assume without loss of generality that the slope  $b_0 = 1$ .  $V_{NF}(t)$  is the reflected process. In the sequel, we assume that the  $b_1$ -intervals are *exponentially* distributed; this leads to more explicit results.  $V_{NF}(t)$  is the workload process in an  $M/G_U/M_D/1$  queue; i.e., an  $M/G/1$ -type queue with arrival rate  $\mu_0$  and service time distribution  $G_a(x/a)$ , in which downward jumps occur according to a Poisson process with rate  $\lambda_0$ , the downward jumps being exponentially distributed with mean  $1/\nu$ . As before, the upward jumps (service times) are indicated by  $\hat{S}$ , and the downward jumps are denoted by  $\hat{Z}$ . The busy period  $T = \inf\{t > 0 : \hat{S} + L_{NF}(t) \leq 0\}$  is terminated in one of two ways. With probability  $p$  one has  $\hat{S} + L_{NF}(T) = 0$ , and with probability  $q = 1 - p$  one has  $\hat{S} + L_{NF}(T) < 0$  (the busy period ends with a downward jump). The probabilities  $p$  and  $q$  are easily determined.

Furthermore, the distribution of  $V_{NF}$  is known, cf. [4] where the case of generally distributed downward jumps is treated. Obviously,

$$\mathbf{P}(X^{max} \leq x) = \mathbf{P}(V_{NF}^{max} \leq x), \quad x \geq 0, \quad (7.1)$$

where  $V_{NF}^{max}$  is the maximum of the workload during one busy period of the process  $\{V_{NF}(t), t \geq 0\}$ .

**Theorem 7.1** *In Case (ii), the steady-state distribution of the top of the mountain is given by:*

$$\mathbf{P}(X^{max} \leq x) = \frac{\mathbf{P}(V_{NF} + \hat{S} \leq x)}{p\mathbf{P}(V_{NF} \leq x) + q\mathbf{P}(V_{NF} \leq x + \hat{Z})}, \quad x \geq 0. \quad (7.2)$$

**Proof:**

We identify the law of  $V_{NF}$  with that of  $\sup_{0 \leq t < \infty} L_{NF}(t)$ . Then

$$\begin{aligned} \mathbf{P}(\hat{S} + V_{NF} \leq x) &= \mathbf{P}(\sup_{0 \leq t < \infty} (\hat{S} + L_{NF}(t)) \leq x) \\ &= \mathbf{P}(\sup_{0 \leq t \leq T} (\hat{S} + L_{NF}(t)) \leq x) \mathbf{P}(\sup_{T \leq t < \infty} (\hat{S} + L_{NF}(t)) \leq x) \\ &= \mathbf{P}(V_{NF}^{max} \leq x)[p\mathbf{P}(V_{NF} \leq x) + q\mathbf{P}(V_{NF} \leq x + \hat{Z})], \end{aligned}$$

where the second equality follows by the strong Markov property and the third equality follows since the cycle maximum of the ‘sojourn’ time is the maximum achieved during the busy period. ♠

Let us now also assume that the  $a$ -intervals are exponentially distributed, with rate, say,  $\mu$ . For the moment, the  $b_1$ -intervals are assumed to be generally distributed. Recall that  $L_E(t)$  is the Lévy process generated by deleting the  $b_0$ -intervals and the  $b_1$ -intervals. Assume without loss of generality that the slope  $a = 1$ . The reflected process  $V_E(t)$  can then be interpreted as the EWT in the  $GI/M/1$  queue with  $\exp(\mu)$  service times and with interarrival times that are generated as a geometric sum of buffer increments during successive phases composed of  $b_0$ -intervals and  $b_1$ -intervals. More precisely, the LST of these interarrival times is given by

$$\mathbf{E}[e^{-\omega Y}] = \frac{\frac{\mu}{\lambda_0 + \mu_0 + \omega}}{1 - \frac{\lambda_0}{\lambda_0 + \mu_0 + \omega} G_{b_1}^*(b_1 \omega)}, \quad \text{Re } \omega \geq 0.$$

Let  $I$  denote a random variable with distribution the steady-state distribution of the idle period in this  $G/M/1$  queue; its distribution can be found in [23], p. 36.

**Theorem 7.2**

$$\mathbf{P}(X^{max} \leq x) = \frac{\mathbf{P}(V_E \leq x)}{\mathbf{P}(V_E \leq x + I)}, \quad x \geq 0. \quad (7.3)$$

**Proof:**

Let  $V_E^{max}$  be the cycle maximum of the EWT process in the  $GI/M/1$  queue. We can identify the law of  $\sup_{0 \leq t < \infty} L_E(t)$  with that of  $V_E$ , the steady-state law of  $V_E(t)$ . Therefore,

$$\begin{aligned} \mathbf{P}(V_E \leq x) &= \mathbf{P}(\sup_{0 \leq t < \infty} L_E(t) \leq x) \\ &= \mathbf{P}(\sup_{0 \leq t < T} L_E(t) \leq x) \mathbf{P}(\sup_{T \leq t < \infty} L_E(t) \leq x) \\ &= \mathbf{P}(V_E^{max} \leq x) \mathbf{P}(V_E \leq x + I). \end{aligned}$$

The theorem is proved since  $V_E^{max}$  and  $X^{max}$  have the same law. ♠

**Remark 7.1.** Theorem 7.2 remains valid when the  $b_0$ -intervals have a general distribution.

In case all intervals are exponentially distributed, Theorems 7.1 and 7.2 result in relatively simple expressions. We end the paper by giving another derivation of the distribution of the top of the mountain in this exponential case. Apart from presenting the simple expression, our goal also is to demonstrate the elegance and power of the martingale approach.

We assume that the upward jump process is a Poisson process with rate  $\mu_0$ , the downward jump process is a Poisson process with rate  $\lambda_0$ , the upward jumps are exponentially distributed with rate  $\mu$ , and the downward jumps are exponentially distributed with rate  $\nu$ . Fix  $x > 0$  and assume that the first upward jump occurs at 0; it is denoted by  $\hat{S}$ . We define four stopping times.

$$T_x(s) := \inf\{t : s + L_F(t) \geq x\},$$

$$\begin{aligned} T_0(s) &:= \inf\{t : s + L_F(t) = 0\}, \\ T_-(s) &:= \inf\{t : s + L_F(t) < 0\}, \end{aligned}$$

and

$$T := \min(T_x(s), T_0(s), T_-(s)). \quad (7.4)$$

Also, define the conditional probabilities  $\phi_x(s) := \mathbf{P}(T = T_x(s) | \hat{S} = s)$ ,  $\phi_0(s) := \mathbf{P}(T = T_0(s) | \hat{S} = s)$ , and  $\phi_-(s) := \mathbf{P}(T = T_-(s) | \hat{S} = s)$ . Notice that  $T = \min(T_0(s), T_-(s))$  is the busy period generated by  $V_{NF}$ , and write  $V_{NF}^{max} = \max_{0 \leq t < T} V_{NF}(t)$ . Since  $\hat{S} \simeq \exp(\mu)$  we have:

$$\mathbf{P}(V_{NF}^{max} \leq x) = \int_0^x \mu e^{-\mu s} (\phi_0(s) + \phi_-(s)) ds = \int_0^x \mu e^{-\mu s} (1 - \phi_x(s)) ds. \quad (7.5)$$

Equation (7.5) follows since the event  $\{V_{NF}^{max} \leq x\}$  occurs if and only if level 0 is reached or downcrossed before level  $x$  is reached or upcrossed. We are now left with the problem of determining  $\phi_x(s)$ . This is accomplished by applying Wald's martingale to the Lévy process  $L_{NF}(t)$ . Note that, for this Lévy process,

$$\mathbf{E}(e^{-\omega L_{NF}(t)}) = e^{\phi(\omega)t}, \quad (7.6)$$

where

$$\phi(\omega) := \omega - \mu_0 \frac{\omega}{\mu + \omega} + \lambda_0 \frac{\omega}{\nu - \omega}. \quad (7.7)$$

Wald's martingale of the Lévy process  $L_{NF}(t)$  is

$$M(t) := e^{-\omega(s+L_{NF}(t))} / \mathbf{E}(e^{-\omega(s+L_{NF}(T))}). \quad (7.8)$$

Application of the martingale stopping theorem w.r.t.  $T$  yields:

$$e^{-\omega s} = \mathbf{E}(e^{-\omega(s+L_{NF}(T)) - \phi(\omega)T}). \quad (7.9)$$

Now consider  $s + L_{NF}(T)$  for the three disjoint events  $T = T_x(s)$ ,  $T = T_0(s)$  and  $T = T_-(s)$ . In the first event,  $s + L_{NF}(T) - x$  is exponentially distributed with rate  $\mu$ ; in the second event, it is zero; and in the third event,  $-(s + L_{NF}(T))$  is exponentially distributed with rate  $\nu$ . Therefore, it follows from (7.9) that (with  $I$  denoting an indicator function)

$$\begin{aligned} e^{-\omega s} &= \frac{\mu}{\mu + \omega} e^{-\omega x} \mathbf{E}(e^{-\phi(\omega)T} I_{s+L_{NF}(T) > x}) \\ &+ \mathbf{E}(e^{-\phi(\omega)T} I_{s+L_{NF}(T)=0}) + \frac{\nu}{\nu - \omega} \mathbf{E}(e^{-\phi(\omega)T} I_{s+L_{NF}(T) < 0}). \quad (7.10) \end{aligned}$$

Let us now choose  $\omega$  such that  $\phi(\omega) = 0$ . This happens for three, real, values: For  $\omega_0 = 0$  and for

$$\omega_{1,2} := \frac{1}{2}[\mu_0 + \nu + \lambda_0 - \mu \pm \sqrt{(\mu_0 + \nu + \lambda_0 - \mu)^2 + 4(\mu\nu + \lambda_0\mu - \mu_0\nu)}].$$

The stability condition  $\mu_0/\mu < 1 + \lambda_0/\nu$  (cf. [4]) guarantees that  $\omega_1$  and  $\omega_2$  are real. Substitution of  $\omega = \omega_0, \omega_1, \omega_2$  in (7.10) yields three equations which just suffice to determine the three unknowns  $\phi_x(s)$ ,  $\phi_0(s)$  and  $\phi_-(s)$ . The root  $\omega_0 = 0$  gives the obvious relation  $\phi_x(s) + \phi_0(s) + \phi_-(s) = 1$ . After some calculations we get:

$$\phi_x(s) = \frac{\omega_2(\nu - \omega_1)(1 - e^{-\omega_1 s}) - \omega_1(\nu - \omega_2)(1 - e^{-\omega_2 s})}{\omega_2(\nu - \omega_1)(1 - \frac{\mu}{\mu + \omega_1}e^{-\omega_1 x}) - \omega_1(\nu - \omega_2)(1 - \frac{\mu}{\mu + \omega_2}e^{-\omega_2 x})}. \quad (7.11)$$

Finally, substituting (7.11) in (7.5), we obtain:

$$\begin{aligned} \mathbf{P}(V_{NF}^{max} \leq x) &= \left(1 - \frac{k_1 - k_2}{k_3}\right)(1 - e^{-\mu x}) - \frac{\mu k_1}{k_3(\mu + \omega_1)}(1 - e^{-(\omega_1 + \mu)x}) \\ &+ \frac{\mu k_2}{k_3(\mu + \omega_2)}(1 - e^{-(\omega_2 + \mu)x}), \end{aligned} \quad (7.12)$$

where

$$k_1 = \omega_2(\nu - \omega_1), \quad k_2 = \omega_1(\nu - \omega_2),$$

and

$$k_3 = k_1\left(1 - \frac{\mu}{\mu + \omega_1}e^{-\omega_1 x}\right) - k_2\left(1 - \frac{\mu}{\mu + \omega_2}e^{-\omega_2 x}\right).$$

**Acknowledgement.** The authors would like to thank Offer Kella for his insightful remarks. Financial support of the Dutch research organization NWO, that has enabled the visit of David Perry to The Netherlands, is gratefully acknowledged.

## References

- [1] S. Aalto (1998). *Characterization of the output rate process for a Markovian storage model*, J. Appl. Probab. **35**, 184-199.



- [2] D. Anick, D. Mitra and M.M. Sondhi (1982). *Stochastic theory of a data-handling system with multiple sources*, The Bell System Techn. J. **61**, 1871-1894.
- [3] S. Asmussen and D. Perry (1992). *On cycle maxima, first passage problems and extreme value theory for queues*, Stochastic Models **8**, 421-458.
- [4] R.J. Boucherie and O.J. Boxma (1996). *The workload in the M/G/1 queue with work removal*, Probab. Eng. Inf. Sci. **10**, 261-277.
- [5] O.J. Boxma and V. Dumas (1998). *Fluid queues with long-tailed activity period distributions*. Computer-Communications, to appear.
- [6] O.J. Boxma and V. Dumas (1998). *The busy period in the fluid queue*, Perf. Eval. Review **26**, 100-110 (Special issue, with the Proceedings of ACM Sigmetrics/Performance '98).
- [7] H. Chen and D.D. Yao (1992). *A fluid model for systems with random disruptions*, Operations Research **S40**, S239-S247.
- [8] E. Cinlar (1975). *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs (NJ)).
- [9] J.W. Cohen (1974). *Superimposed renewal processes and storage with gradual input*. Stochastic Processes and their Applications **2**, 31-58.
- [10] J.W. Cohen (1977). *On up- and downcrossings*. J. Appl. Probab. **14**, 405-410.
- [11] J.W. Cohen (1982). *The Single Server Queue* (North-Holland Publ. Cy., Amsterdam; revised edition).
- [12] D.P. Gaver and R.G. Miller (1962). *Limiting distributions for some storage problems*. In: K.J. Arrow, S. Karlin and H. Scarf (eds.), *Studies in Applied Probability and Management Science* (Stanford University Press, Stanford) pp. 110-126.
- [13] H. Kaspi and D. Perry (1989). *On a duality between a non-Markovian storage/production process and a Markovian dam process with state-dependent input and output*. J. Appl. Probab. **26**, 835-844.

- [14] H. Kaspi and M. Rubinovitch (1975). *The stochastic behaviour of a buffer with non-identical input lines*. Stochastic Processes and their Applications **3**, 73-88.
- [15] O. Kella and W. Whitt (1992). *A storage model with a two-state random environment*, Operations Research **S40**, S257-S262.
- [16] L. Kosten (1974). *Stochastic theory of a multi-entry buffer (I)*. Delft Progress Report **1**, 10-18.
- [17] L. Kosten (1974). *Stochastic theory of a multi-entry buffer (II)*. Delft Progress Report **1**, 44-50.
- [18] L. Kosten (1984). *Stochastic theory of data handling systems with groups of multiple sources*. In: W. Bux and H. Rudin (eds.), *Performance of Computer-Communication Systems* (North-Holland Publ. Cy., Amsterdam), pp. 321-331.
- [19] L. Kosten and O.J. Vrieze (1975). *Stochastic theory of a multi-entry buffer (II)*. Delft Progress Report **1**, 103-115.
- [20] V.G. Kulkarni (1997). *Fluid models for single buffer systems*. In: J.H. Dshalalow (ed.), *Frontiers in Queueing* (CRC Press, Boca Raton), pp. 321-338.
- [21] R.G. Miller (1963). *Continuous time stochastic storage processes with random linear inputs and outputs*. J. Math. Mech. **12**, 275-291.
- [22] D. Perry and B. Levikson (1989). *Continuous production/inventory model with analogy to certain queueing and dam models*. Adv. Appl. Probab. **21**, 123-141.
- [23] N.U. Prabhu (1980). *Stochastic Storage Processes* (Springer-Verlag, Berlin).
- [24] L.C.G. Rogers (1994). *Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains*. Annals Applied Probability **4**, 390-413.
- [25] M. Rubinovitch (1973). *The output of a buffered data communication system*. Stochastic Processes and their Applications **1**, 375-382.

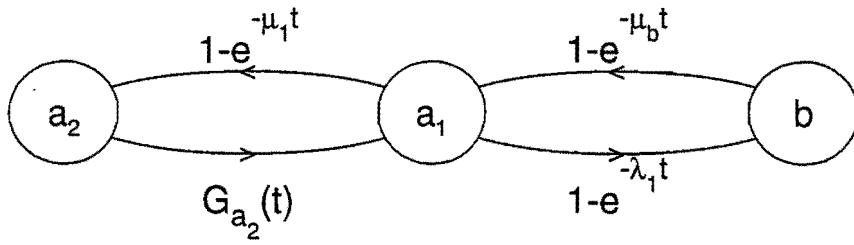


Figure 1: State-transition diagram for Case (i)

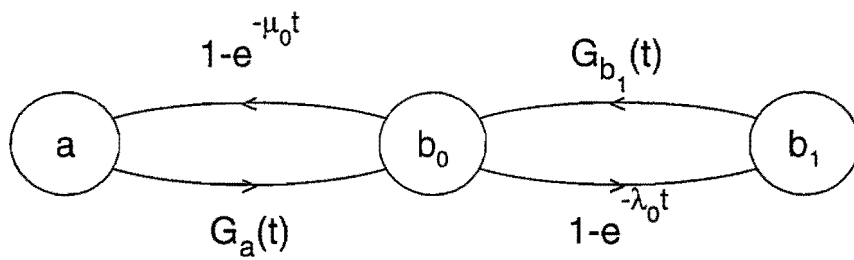


Figure 2: State-transition diagram for Case (ii)

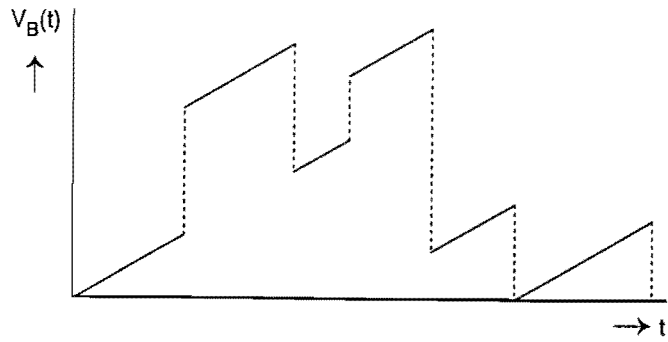
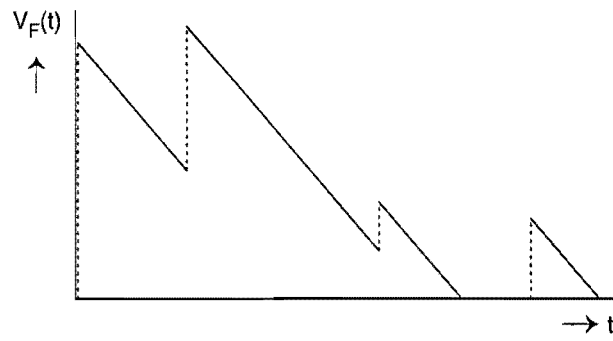


Figure 3: Case (i):  $X(t)$ ,  $V_F(t)$  and  $V_B(t)$

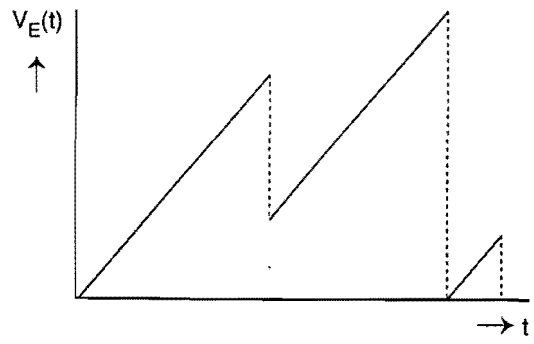
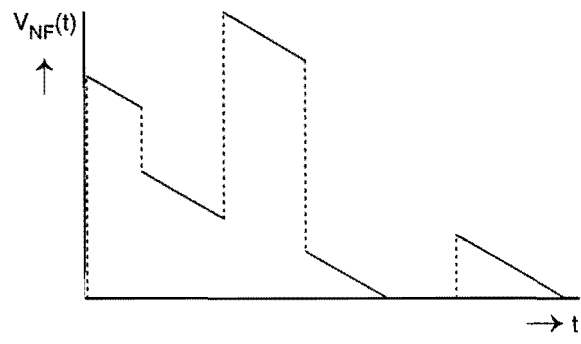
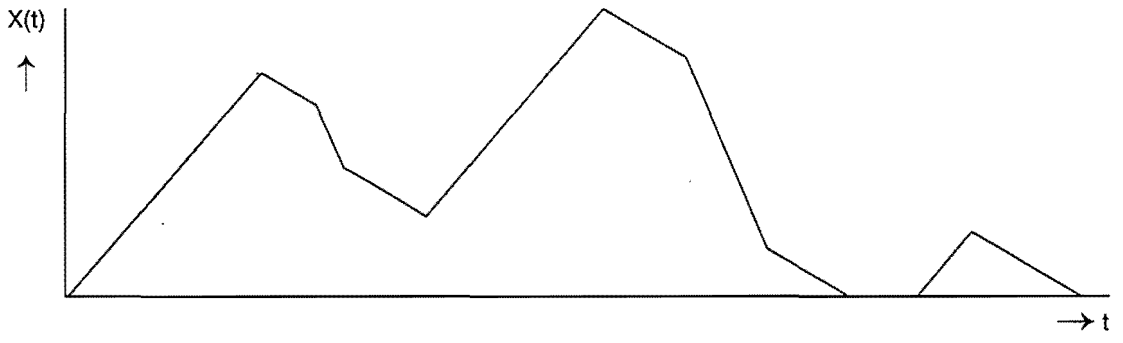


Figure 4: Case (ii):  $X(t)$ ,  $V_{NF}(t)$  and  $V_E(t)$