Shot-noise fluid queues and infinite-server systems with batch arrivals

W.F. de Graaf [a], W.R.W. Scheinhardt [b,*], R.J. Boucherie [b]

[a] *University of Utrecht, The Netherlands*
[b] *University of Twente, The Netherlands*

ABSTRACT

We show how a shot-noise fluid queue can be considered as the limiting case of a sequence of infinite-server queues with batch arrivals. The shot-noise queue we consider receives fluid amounts at the arrival times of a (time-inhomogeneous) Poisson process, the sizes of which are governed by some probability distribution that may also depend on time. The continuous rate at which fluid leaves the queue is proportional to the current content of the queue. Thus, intuitively, one can think of drops of fluid arriving in batches, which are taken into service immediately upon arrival, at an exponential service rate.

We show how to obtain the partial differential equation for (the Laplace–Stieltjes transform of) the queue content at time t, as well as its solution, from the corresponding infinite-server systems by taking appropriate limits. Also, for the special case of a time-homogeneous arrival process, we show that the scaled number of occupied servers in the infinite-server system converges as a process to the shot-noise queue content, implying that finite-dimensional distributions also converge.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Shot noise was first used in physics as a model for discontinuous noise, conveyed by pulses, as opposed to white noise, being the continuous noise conveyed by thermal fluctuations. Today shot noise is widely used in other fields as well, such as mathematical finance, insurance and queueing. For more on shot noise in a historical and physical setting, see e.g. Eliazar and Klafter [1].

A shot-noise fluid queue can be described as a fluid reservoir from which fluid is released continuously at a rate that is proportional to the current content $V(t)$ of the reservoir. Hence, at any time t, the output rate $r(t)$ satisfies $r(t) = rV(t)$ for a fixed constant $r > 0$. In the standard model, fluid arrives to the reservoir according to a compound Poisson process, i.e., quantities of fluid arrive instantaneously to the reservoir according to a Poisson process with rate λ. The sizes of these quantities are assumed to be i.i.d., according to some generic random variable B with distribution function $B(x) = \mathbb{P}(B \leq x)$, and independent of the Poisson process. In the main part of the paper we will look at the more general model with time-inhomogeneous input, where both the Poisson rate and the distribution of B are allowed to vary over time in a continuous manner, replacing λ and $B(x)$ by $\lambda(t)$ and $B(x, t)$.

Let us first consider the content process $V_i(t)$ of a shot-noise queue that starts empty at time $t = 0$, and to which just a single quantity of fluid arrives at time $t = t_i$ with size B_i. In the sequel we will simply call B_i the size of the *jump* at time t_i,
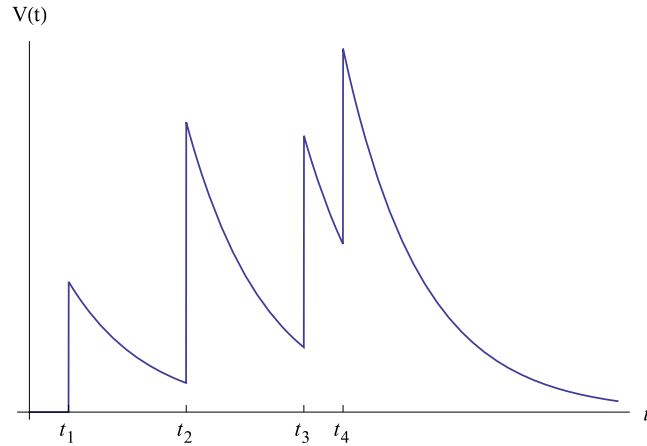
Fig. 1. The content process of a shot-noise queue.

for obvious reasons. Clearly the dynamics of the content process $V_i(t)$ are then given by the initial value problem

$$\frac{dV_i(t)}{dt} = -rV_i(t), \quad V_i(t_i) = B_i,$$

which can be solved to obtain

$$V_i(t) = B_i e^{-r(t-t_i)}, \qquad t \geq t_i.$$

The content process $V(t)$ of a general shot-noise queue can be viewed as a sum of independent processes $V_i(t)$, see also [2]:

$$V(t) = \sum_{i=1}^{N(t)} B_i e^{-r(t-t_i)} \mathbf{1}\{t \geq t_i\}, \tag{1.1}$$

where $N(t)$ denotes the number of jumps arriving up to time t, and $\mathbf{1}\{t \geq t_i\}$ is 1 when $t \geq t_i$ and 0 otherwise. In Fig. 1 a possible realization of the content process in a shot-noise queue is given.

For the standard case with time-homogeneous input (where λ and $B(x)$ do not vary over time) it was shown in [2] how to obtain the Laplace–Stieltjes Transform (LST) of the amount of fluid in the reservoir at any time t, and from that one can easily find the LST of the amount of fluid in stationarity, see the expressions below Theorem 2.2 in the sequel. In fact it is not difficult to generalize the time-dependent results to the case of time-inhomogeneous input (where $\lambda(t)$ and $B(x, t)$ may vary over time in a continuous way). We could work with (1.1) directly, as in [2], viewing $V(t)$ as a functional of a – now time-dependent – Poisson Process; the result can then be found by applying an extended version of Campbell's theorem, see Kingman [3, section 5.3]. Another approach would be to write down the integro-differential equation of Takács [4], similar as was done in e.g. [5] for a feedback fluid queue, or in Bekker [6] for queues with workload-dependent service rate. Then a partial differential equation (PDE) can be derived for the Laplace–Stieltjes Transform of $V(t)$, see (2.3) in Section 2.1, which can be solved straightforwardly. This solution is presented in Appendix A.1. A third approach is to consider (1.1) in the context of Levy processes [7], which could also yield the solution to generalized problems, e.g. involving a Brownian motion component. We outline this approach in Appendix A.2, just to show its efficiency and elegance.

The solution for time-inhomogeneous input has never appeared in the literature to the best of our knowledge. We present it in this paper, but our main goal is in fact not to show the result, but rather to show the way we find it, which is different from the approaches outlined before. Our approach is to establish a relation with infinite-server systems with batch arrivals. Roughly the idea is that in the shot-noise queue each 'drop' of fluid is assigned its own dedicated small server. In particular we introduce for any shot-noise queue a sequence of infinite-server systems, indexed by some k, with a corresponding PDE and its solution. We then show how these converge as k grows large to the PDE and its solution for the shot-noise queue.

The relation with infinite-server systems has been mentioned before in Kella and Whitt [8]. They consider networks of fluid queues which are linear in the sense that the workload process vector for all queues in the network is just the sum of all individual workload vectors that can be associated to the existing superposition of inputs. Due to this linearity property the authors are then able to come up with elegant results at considerable generality. In their introduction they mention that the workload process can be regarded as the limit as $n \to \infty$ of the scaled number of busy servers at each node in a network of infinite-server systems with batch arrivals, where the batch sizes in model n are the batch sizes in model 1 multiplied by n.

While in [8] the connection with (networks of) infinite-server systems appears to be mentioned 'only' as an intuitive explanation of why the presented results may be expected, it is our goal to actually show the convergence of infinite-server systems to shot-noise queues, and to use this relation to find the related PDE and its solution. We will limit ourselves to single

queues only, although much of this work can be generalized to networks as in [8]. In the main part of the paper we allow the input process to be time-inhomogeneous, while most of the results in [8] are for time-homogeneous input processes.

The paper is organized as follows. In Section 2 we give the models of the shot-noise queue and the Markovian infinite-server system, and preliminary results. Then in Section 3 we show how the PDE and solution for the sequence of infinite-server systems converge to the corresponding PDE and solution of the shot-noise queue for the case of time-inhomogeneous input. Finally in Section 4 we show convergence at the process level, i.e., we prove that all finite-dimensional distributions of the (properly scaled) number of active servers converge to those of the shot-noise workload process; for the analysis in Section 4 we need to restrict ourselves to the case of time-homogeneous input.

2. Models and preliminary results

In this section we first present the shot-noise queueing model, the equations that govern the dynamics of this system, as well as the form of the corresponding time-dependent solution in terms of Laplace–Stieltjes transforms (LSTs). We then continue to do the same for the Markovian infinite-server system model.

2.1. Shot-noise queue

We consider a shot-noise queue as introduced in Section 1 with time-dependent input. Thus, both the current Poisson arrival rate $\lambda(t)$ and the distribution function $B(x, t) = \mathbb{P}(B(t) \leq x)$ of the size of a jump $B(t)$ occurring at time t are functions of t; we assume these functions to be continuously differentiable. Allowing the queue to be nonempty at $t = 0$, we are interested in the time-dependent distribution function of the amount $V(t)$ of fluid present at time t, which is denoted as $V(x, t) = \mathbb{P}(V(t) \leq x)$.

We first derive the integro-differential equation of Takács for $V(x, t)$. For h small we have

$$V(x, t + h) = (1 - \lambda(t)h) V((1 + rh)x, t)$$
$$+ \lambda(t)h \int_0^x B(x - y, t) \mathrm{d}_y V(y, t) + o(h).$$

Taking the limit $h \downarrow 0$ and using the continuity of $B(x, t)$, gives the following result.

Lemma 2.1. *The distribution function $V(x, t)$ of the content of the shot-noise queue at time t satisfies*

$$\frac{\partial V(x, t)}{\partial t} - rx \frac{\partial V(x, t)}{\partial x} = -\lambda(t) \int_0^x (1 - B(x - y, t)) \mathrm{d}_y V(y, t), \qquad x > 0, t \geq 0. \tag{2.1}$$

Next we define the LSTs of $B(t)$ and $V(t)$ as

$$\beta(s, t) = \int_0^\infty \mathrm{e}^{-sx} \mathrm{d}_x B(x, t) \quad \text{and} \quad \Phi(s, t) = \int_0^\infty \mathrm{e}^{-sx} \mathrm{d}_x V(x, t), \qquad \mathrm{Re}(s) \geq 0. \tag{2.2}$$

The following theorem gives the PDE for $\Phi(s, t)$ and its solution.

Theorem 2.2. *The LST $\Phi(s, t)$ of the amount of fluid $V(t)$ in the shot-noise queue satisfies the following partial differential equation,*

$$\frac{\partial \Phi(s, t)}{\partial t} + rs \frac{\partial \Phi(s, t)}{\partial s} = -\lambda(t)(1 - \beta(s, t))\Phi(s, t). \tag{2.3}$$

Assuming that $\lambda(t)$ is a C^1 function, the unique solution is given by

$$\Phi(s, t) = \Phi(se^{-rt}, 0) \exp\left(-\int_0^t \lambda(\eta)(1 - \beta(se^{-r(t-\eta)}, \eta)) \mathrm{d}\eta\right). \tag{2.4}$$

Proof. The proof of the PDE in (2.3) is immediate from Lemma 2.1, by taking Laplace transforms using (2.2). The solution (2.4) can be derived in several ways, as mentioned in the Introduction, two of which are presented in Appendices A.1 and A.2. □

For the time-homogeneous case, with $\lambda(t) \equiv \lambda$ and $\beta(s, t) \equiv \beta(s)$, the solution (2.4) is (by a change of variables),

$$\Phi(s, t) = \Phi(se^{-rt}, 0) \exp\left(-\lambda \int_0^t (1 - \beta(se^{-r(\eta)})) \mathrm{d}\eta\right).$$

In that case we can also take the limit as $t \to \infty$ to find the LST of the stationary distribution,

$$\lim_{t \to \infty} \Phi(s, t) = \exp\left(-\lambda \int_0^\infty \left(1 - \beta(se^{-r(\eta)})\right) \mathrm{d}\eta\right).$$

For the special case of exponential jump sizes, with $\beta(s) = \nu/(\nu + s)$, this leads to $\Phi(s, \infty) = (\nu/(\nu + s))^{\lambda/r}$, i.e., a Gamma$(\nu, \lambda/r)$ distribution.

In Section 3 we will find the same results as in Theorem 2.2 by studying the relation with infinite-server systems with batch arrivals. Therefore we will first discuss such systems in the next subsection.

2.2. The infinite-server system with batch arrivals

In the book by Chaudhry and Templeton [9] the $M^X(t)/M/\infty$ system is investigated. In such a system customers arrive in batches according to a Poisson process with time-dependent arrival rate $\lambda(t)$. The system has an infinite number of homogeneous servers. Service times are exponentially distributed with parameter μ. We assume that all random variables involved are independent. Furthermore, we extend the model in [9] by allowing the distribution of the size $X(t)$ of a batch that arrives at time t to depend on t. We call this a $M^{X(t)}(t)/M/\infty$ system, and denote the probability generating function of $X(t)$ by

$$P_X(z, t) = \sum_{m=1}^{\infty} \mathbb{P}(X(t) = m)z^m, \qquad |z| \leq 1, t \geq 0.$$

We let $N(t)$ denote the number of customers (or busy servers) in the system at time t. Clearly $\{N(t)\}$ is a (nonhomogeneous) Markov process. The probability generating function of $N(t)$ is given by

$$P_N(z, t) = \sum_{n=0}^{\infty} \mathbb{P}(N(t) = n)z^n, \qquad |z| \leq 1, t \geq 0.$$

Theorem 2.3. *The probability generating function $P_N(z, t)$ of the number of busy servers $N(t)$ in the $M^{X(t)}(t)/M/\infty$ system satisfies the following partial differential equation,*

$$\frac{\partial P_N(z, t)}{\partial t} - \mu(1 - z)\frac{\partial P_N(z, t)}{\partial z} = -\lambda(t)(1 - P_X(z, t))P_N(z, t). \tag{2.5}$$

Assuming that $\lambda(t)$ is a C^1 function, the unique solution is given by

$$P_N(z, t) = P_N(v(z, t), 0)\exp\left(-\int_0^t \lambda(\eta)\left(1 - P_X(v(z, t - \eta), \eta)\right)d\eta\right), \tag{2.6}$$

where

$$v(z, t) = 1 - (1 - z)e^{-\mu t}.$$

Proof. For the case in which the jump size distribution does not depend on t, a straightforward derivation of the PDE (2.5) can be found in [9], see (5.1.4) there. To keep the paper self-contained, we present the main idea here. Writing $p_m(t) = \mathbb{P}(X(t) = m)$ and $a_m(t) = \mathbb{P}(N(t) = m)$, $m = 0, 1, \ldots$ we have the following system of (backward) differential-difference equations for $p_n(t)$,

$$p_0'(t) = -\lambda(t)p_0(t) + \mu p_1(t),$$

$$p_n'(t) = -(\lambda(t) + n\mu)p_n(t) + \lambda(t)\sum_{m=1}^{n} a_m p_{n-m}(t) + (n + 1)\mu p_{n+1}(t), \quad n = 1, 2, \ldots.$$

Multiplying these equations by $1, z, z^2, \ldots$ and then summing leads to (2.5), using for the convolution that

$$\sum_{n=1}^{\infty} z^n \sum_{m=1}^{n} a_m p_{n-m}(t) = P_X(z, t)P_N(z, t).$$

For the solution (2.6) one could rewrite (5.1.9) in [9] while taking into account the time-dependent distribution of $B(t)$. In Appendix A.3, we present a proof of (2.6) by solving the PDE (2.5) straightforwardly. □

3. Relation with infinite-server systems with batch arrivals

In this section we show how for each shot-noise queue we can define a certain sequence of infinite-server systems that converges to it. We first define the sequence in Section 3.1 based on some shot-noise queue, and then show in Sections 3.2 and 3.3 how by taking appropriate limits the corresponding PDE's and solutions of Theorem 2.3 converge to the PDE and solution, as given in Theorem 2.2, of the shot-noise queue we started with.

3.1. Sequence of $M^{X(t)}(t)/M/\infty$ systems

Our starting point is a shot-noise queue with time-dependent input as in Section 2.1. The arrival rate (of jumps) is $\lambda(t)$ and the size of a jump arriving at time t is $B(t)$, with distribution function $B(x,t) = \mathbb{P}(B(t) \leq x)$ and LST $\beta(s,t) = \int_0^\infty e^{-sx} d_x B(x,t)$. Moreover, the amount of fluid in the buffer at time 0 is ze of a jump arriving at time t is $B(t)$, with distribution function $B(x,t) = \mathbb{P}(B(t) \leq x)$ and LST $\beta(s,t) = \int_0^\infty e^{-sx} d_x B(x,t)$.

The next step is to define a sequence of corresponding $M^{X(t)}(t)/M/\infty$ systems as in Section 2.2, indexed by $k = 1, 2, \ldots$. The arrival rate (of batches) is taken identical to $\lambda(t)$, while the size $X_k(t)$ of a batch arriving at time t in the kth model is determined as follows. The idea is that, as k increases, $X_k(t)$ scales with k, while simultaneously the amount of work per arriving customer scales as $1/k$, so that the total amount of work arriving in a batch will remain of the same order. Thus, we define $X_k(t)$ as the smallest integer larger than or equal to $kB(t)$, i.e.,

$$X_k(t) = \lceil kB(t) \rceil. \tag{3.1}$$

Now let the random variable $Z_k(t)$ be the total amount of work arriving at time t in the kth $M^{X(t)}(t)/M/\infty$ system. The service rate in the kth model is always μ, so we model $Z_k(t)$ as the sum of $X_k(t)$ customer sizes which are exponential with parameter k. To do so, and to facilitate later taking the limit $k \to \infty$, we first introduce random variables C_i as i.i.d. $\sim \exp(1)$, independent of everything else. The variable C_i represents a scaled version of the size (service time) of the ith customer of a batch, where the scaling is independent of the particular value $k = 1, 2, \ldots$. To retrieve the properly scaled size of customer i in model k, we simply rescale this C_i, dividing it by k, to end up with the following,

$$Z_k(t) = \sum_{i=1}^{X_k(t)} \frac{1}{k} C_i.$$

The LST of $Z_k(t)$, defined as $\tilde{Z}_k(s,t) = \mathbb{E}[e^{-sZ_k(t)}]$, can be found by conditioning on $X_k(t)$,

$$\tilde{Z}_k(s,t) = \mathbb{E}\left[\left(\frac{k}{k+s}\right)^{X_k(t)}\right] \equiv P_{X_k}\left(\frac{k}{k+s}, t\right). \tag{3.2}$$

To show that $Z_k(t)$ indeed behaves as expected we first write

$$Z_k(t) = \frac{X_k(t)}{k} \cdot \frac{1}{X_k(t)} \sum_{i=1}^{X_k(t)} C_i.$$

Since for any $b > 0$ it holds that $\lim_{k\to\infty} \lceil kb \rceil / k = b$, we conclude that $X_k(t)/k \to B(t)$ a.s. The second factor converges to $\mathbb{E}C = 1$ a.s. by the strong law of large numbers, noting that as $k \to \infty$ also $X_k \to \infty$ a.s. But then also the product converges a.s. to $B(t)$ (since for any two events, $P(E) = 1$ and $P(F) = 1$ together imply $P(E \cap F) = 1$). Thus we arrive at the following.

Lemma 3.1. *The amount of work $Z_k(t)$ arriving at time t in the kth $M^{X(t)}(t)/M/\infty$ system satisfies*

$$Z_k(t) \to B(t) \text{ a.s.,} \quad \text{as } k \to \infty.$$

In particular, its LST $\tilde{Z}_k(s,t)$ satisfies

$$\lim_{k\to\infty} \tilde{Z}_k(s,t) = \beta(s,t). \tag{3.3}$$

Note that the proof does not need the C_i's to be exponentially distributed, but the following does require this.

Turning to the amount of work $W_k(t)$ in the kth $M^{X(t)}(t)/M/\infty$ system, we can write similarly as above,

$$W_k(t) = \sum_{i=1}^{N_k(t)} \frac{1}{k} C_i,$$

with $N_k(t)$ the number of busy servers in model k, and the C_i distributed as before, using their memorylessness. For the LST $\tilde{W}_k(s,t) = \mathbb{E}[e^{-sW_k(t)}]$ it holds, by conditioning on $N_k(t)$, that

$$\tilde{W}_k(s,t) = \sum_{m=0}^{\infty} \left(\frac{k}{k+s}\right)^m \mathbb{P}(N_k(t) = m) = P_{N_k}\left(\frac{k}{k+s}, t\right). \tag{3.4}$$

In particular, the amount $V(0)$ of fluid present in the buffer at $t = 0$ can be discretized with a similar explicit construction as we used for the jump size distribution, since its distribution function $V(x,0)$ and LST $\Phi(s,0)$ are known, just as for the jump size the distribution function $B(x,t)$ and LST $\beta(s,t)$ are also known. Thus, analogous to (3.1)–(3.2), and replacing the random variables $X_k(t)$ and $Z_k(t)$ by $N_k(0)$ and $W_k(0)$, respectively, we first define $N_k(0) = \lceil kV(0) \rceil$ and then have

$$W_k(0) = \sum_{i=1}^{N_k(0)} \frac{1}{k} C_i,$$

with LST as in (3.4) with $t = 0$. Similarly to the arguments leading up to Lemma 3.1 we can observe that $N_k(0)/k \to V(0)$ a.s., and $\frac{1}{N_k(0)} \sum_{i=1}^{N_k(0)} C_i$ converges to $\mathbb{E}C = 1$ a.s. Thus we arrive at the following lemma.

Lemma 3.2. *The amount of work $W_k(0)$ present at time 0 in the kth $M^{X(t)}(t)/M/\infty$ system satisfies*

$$W_k(0) \to V(0) \text{ a.s.,} \quad \text{as } k \to \infty.$$

In particular, its LST $\tilde{W}_k(s, 0)$ satisfies

$$\lim_{k \to \infty} \tilde{W}_k(s, 0) = \Phi(s, 0).$$

In the next two subsections we insert the expressions (3.2) and (3.4) for general $t \geq 0$ into the partial differential equation for the $M^{X(t)}(t)/M/\infty$ system, and its solution, respectively. Then we take the limit $k \to \infty$ to show that both expressions converge.

3.2. Convergence of PDEs

Here we focus on the partial differential equation (2.5) applied to model k. From identities (3.2) and (3.4) we find

$$P_{X_k}(z, t) = \tilde{Z}_k(s, t) \quad \text{and} \quad P_{N_k}(z, t) = \tilde{W}_k(s, t), \tag{3.5}$$

where $s = k\left(z^{-1} - 1\right)$ of equivalently $z = k/(k + s)$. Using this, (2.5) is transformed to

$$\frac{\partial \tilde{W}_k(s, t)}{\partial t} + \mu s \left(1 + \frac{s}{k}\right) \frac{\partial \tilde{W}_k(s, t)}{\partial s} = -\lambda(t)(1 - \tilde{Z}_k(s, t))\tilde{W}_k(s, t).$$

Now let $\tilde{W}(s, t) = \lim_{k \to \infty} \tilde{W}_k(s, t)$, assuming for now that this limit exists (the proof of this is given below in Section 3.3). Then taking the limit $k \to \infty$ in the above and also using (3.3) we obtain

$$\frac{\partial \tilde{W}(s, t)}{\partial t} + \mu s \frac{\partial \tilde{W}(s, t)}{\partial s} = -\lambda(t)(1 - \beta(s, t))\tilde{W}(s, t).$$

This equation corresponds to the partial differential equation (2.3) in Theorem 2.2 for the shot-noise queue, with $r = \mu$. So when the service times in the $M^{X(t)}(t)/M/\infty$ system are exponentially distributed with parameter r, the PDE of the infinite-server system with batch arrivals converges to that of the shot-noise queue.

3.3. Convergence of solutions

Next, we focus on the solution (2.6) of the partial differential equation for model k. Using (3.5) again, this can be transformed, leading to

$$\tilde{W}_k(s, t) = \tilde{W}_k(\tau_k(s, t), 0) \exp\left(-\int_0^t \lambda(\eta)\left(1 - \tilde{Z}_k(\tau_k(s, t - \eta), \eta)\right) d\eta\right),$$

where

$$\tau_k(s, t) = k\left[\frac{1}{\upsilon(k/(k+s), t)} - 1\right] = e^{-\mu t} \frac{ks}{k + s(1 + e^{-\mu t})}.$$

To prove that $\tilde{W}(s, t) = \lim_{k \to \infty} \tilde{W}_k(s, t)$ exists, first note that $\tau_k(s, t) \to se^{-\mu t}$ as $k \to \infty$, from which we immediately have that $\tilde{W}_k(\tau_k(s, t), 0) \to \tilde{W}(se^{-\mu t}, 0)$, and also $\tilde{Z}_k(\tau_k(s, t - \eta), \eta) \to \beta(se^{-\mu(t-\eta)}, \eta)$ due to our assumption that $B(x, t)$, and hence $\beta(s, t)$, varies continuously over time t. Thus we arrive at

$$\tilde{W}(s, t) = \tilde{W}(se^{-\mu t}, 0) \exp\left(-\int_0^t \lambda(\eta)(1 - \beta(se^{-\mu(t-\eta)}, \eta)) d\eta\right).$$

This equation is exactly equal to the solution (2.4) for the shot-noise queue, just with $r = \mu$, and $\Phi(s, t)$ replaced by $\tilde{W}(s, t)$. Thus, the random variable $W(t)$ behaves exactly like $V(t)$, the content of the shot-noise queue at time t. So indeed the $M^{X(t)}(t)/M/\infty$ solution converges for $k \to \infty$ to the shot-noise solution when the service times in the $M^{X(t)}(t)/M/\infty$ system are exponentially distributed with parameter r.

3.4. Discussion

We conclude that a shot-noise queue with time-dependent arrival process can be interpreted as the limiting case of a sequence of $M^{X(t)}(t)/M/\infty$ system with output parameter equal to the service rate of the infinite-server systems.

It would be interesting to see whether a similar convergence could be shown for other arrival processes than (time-inhomogeneous) Poisson, e.g. Markov-Modulated (compound) Poisson processes, in which the arrival rate depends on some

background Markov process. We expect a similar convergence result holds for a sequence of Markov-modulated infinite-server systems with batch arrivals, where the batch size distribution depends on the same background Markov-process. The same may hold for models with state-dependent arrivals, and models with more than one queue.

Remark 1. In this section we considered the convergence based on workloads in the infinite-server systems (distributed as $Z_k(t)$ per arriving batch and as $W_k(t)$ in the system). Instead, we could also have considered the scaled number of customers (distributed as $X_k(t)/k$ per arriving batch and as $N_k(t)/k$ in the system). Even though the derivations would be more straightforward in that case, we chose to show the result for workloads since in our view they are the most natural candidates to convergence to the fluid analogues in the shot-noise queue. However, in the next section we will consider the (scaled) number of customers to show how this discrete process converges at the process level to the continuous shot-noise queue content.

4. Convergence at the process level

In this section we show convergence at the process level, i.e., we prove that all finite-dimensional distributions of the discrete infinite-server system converge to those of the shot-noise workload process. We apply the theory of convergence of Markov processes in Ethier and Kurtz [10], and in order to do so we need to restrict ourselves to the case of time-homogeneous input. Thus, in the sequel we assume that the arrival rate and the jump size distribution do not depend on time, writing λ and $B(x)$ rather than $\lambda(t)$ and $B(x, t)$, respectively. As a result, $\{V(t)\}$ is a time-homogeneous Markov process in continuous time with continuous state space $\mathbb{R}^+$. For the corresponding kth infinite server system, we choose the batch size as in (3.1), i.e.,

$$X_k = \lceil kB \rceil. \tag{4.1}$$

Instead of the (continuous) workload in the kth infinite-server system, we study $\{N_k(t)\}$, where $N_k(t)$ as before is the number of busy servers in the kth infinite-server system; see also Remark 1. Due to the time-homogeneous input, this is now also a Markov process, with discrete state space $\mathbb{N}$. Despite the different state spaces of $\{N_k(t)\}$ and $\{V(t)\}$, we will show convergence of $\{N_k(t)\}$ to $\{V(t)\}$ as a process, following Ethier and Kurtz [10]. Roughly, the idea is to show uniform convergence of the semigroups involved, by the equivalent statement that their generators converge uniformly. As a result we may then conclude convergence of the processes, under appropriate (converging) initial conditions.

We start off by determining the relevant semigroups and generators, which are operators[1] on appropriate Banach spaces of test functions. For the Markov processes $\{N_k(t)\}$ we take this space to be $\hat{\mathcal{B}}(\mathbb{N}) = \{f \in \mathcal{B}(\mathbb{N}) : \lim_{x\to\infty} |f(x)| = 0\}$, where $\mathcal{B}(\mathbb{N})$ (also known as ℓ_∞) is the space of all bounded functions on $\mathbb{N}$ with norm $\|f\| = \sup_{x\in\mathbb{N}} |f(x)|$. The 'hat' notation here reflects the vanishing-at-infinity aspect. Now appending a subindex k to the test functions and operators that correspond to the process $\{N_k(t)\}$, we define its semigroup, and give its first order expansion (where $o(h)/h$ vanishes as $h \downarrow 0$), as follows. For all $f_k \in \hat{\mathcal{B}}(\mathbb{N})$,

$$T_k(h)f_k(x) = \mathbb{E}[f_k(N_k(h))|N_k(0) = x] = \sum_{y=0}^{\infty} f_k(y)\mathbb{P}(N_k(h) = y|N_k(0) = x)$$

$$= (1 - \lambda h - x\mu h)f_k(x) + x\mu h f_k(x - 1) + \lambda h \sum_{m=1}^{\infty} a_m^{(k)} f_k(x + m) + o(h),$$

where $a_m^{(k)}$ is the probability that a batch size X_k equals m, and is given by, cf. (4.1),

$$a_m^{(k)} = \mathbb{P}(X_k = m) = B(m/k) - B((m - 1)/k), \quad m = 1, 2, \ldots.$$

The generator A_k of $\{N_k(t)\}$ is given by

$$A_k f_k(x) = \lim_{h\to 0} \frac{T_k(h)f_k(x) - f_k(x)}{h} = -\lambda f_k(x) - \mu x f_k(x) + \mu x f_k(x - 1) + \lambda \sum_{m=1}^{\infty} a_m^{(k)} f_k(x + m). \tag{4.2}$$

The domain of A_k is $\mathcal{D}(A_k) = \{f \in \hat{\mathcal{B}}(\mathbb{N}) : x(f(x) - f(x - 1)) \in \hat{\mathcal{B}}(\mathbb{N})\}$, to ensure that A_k is an operator into $\hat{\mathcal{B}}(\mathbb{N})$.

Now turning to the process $\{V(t)\}$, we take the Banach space of test functions to be $\hat{\mathcal{C}}(\mathbb{R}^+) = \{f \in \mathcal{C}(\mathbb{R}^+) : \lim_{x\to\infty} |f(x)| = 0\}$, where $\mathcal{C}(\mathbb{R}+)$ is the space of bounded and continuous functions on $\mathbb{R}^+$ with norm $\|f\| = \sup_{x\in\mathbb{R}^+} |f(x)|$. Again, the 'hat' notation reflects the vanishing-at-infinity aspect. The semigroup of $\{V(t)\}$ can now be given as

$$T(h)f(x) = \mathbb{E}[f(V(h))|V(0) = x] = \int_0^\infty f(y)\mathrm{d}_y\mathbb{P}(V(h) = y|V(0) = x)$$

$$= f(x(1 - rh))(1 - \lambda h) + \lambda h \int_x^\infty f(y)\mathrm{d}_y B(y - x) + o(h),$$

---

[1] In fact the semigroup is a whole family of operators indexed by the time parameter t.

for all $f \in \hat{\mathcal{C}}(\mathbb{R}^+)$, and the corresponding generator A is

$$Af(x) = \lim_{h \to 0} \frac{T(h)f(x) - f(x)}{h} = -\lambda f(x) + \lim_{h \to 0} \frac{f(x(1 - rh)) - f(x)}{h} + \lambda \int_x^\infty f(y)\mathrm{d}_y B(y - x)$$

$$= -\lambda f(x) - rxf'(x) + \lambda \int_x^\infty f(y)\mathrm{d}_y B(y - x). \tag{4.3}$$

The domain of A is given by

$$\mathcal{D}(A) = \{f \in \hat{\mathcal{C}}(\mathbb{R}^+) : f \text{ differentiable}, xf'(x) \in \hat{\mathcal{C}}(\mathbb{R}^+)\},$$

which ensures that A is an operator into $\hat{\mathcal{C}}(\mathbb{R}^+)$. Note that we need not assume that f'' exists, but we do need f' to be continuous.

Since we work with the scaled number of busy servers $N_k(t)/k$, we define the operator $\pi_k : \hat{\mathcal{C}}(\mathbb{R}^+) \to \hat{\mathcal{B}}(\mathbb{N})$ to be $\pi_k f = f(\cdot/k)$ for $k \geq 1$; this simply 'samples' the continuous function $f \in \hat{\mathcal{C}}(\mathbb{R}^+)$ at distances of $1/k$, resulting in an element $f_k \in \hat{\mathcal{C}}(\mathbb{N})$ with $f_k(i) = f(i/k)$.

In the remainder we will use the following results from Ethier and Kurtz [10], see Theorems 1.6.1 and 4.2.5. We write them down for our situation at hand.

Theorem 4.1. *Let D be a core of A. The following are equivalent:*

(i) *For each $f \in \hat{\mathcal{C}}(\mathbb{R}^+)$, $T_k(h)\pi_k f$ converges to $T(h)f$ for all $h \geq 0$, uniformly, that is*

$$\lim_{k \to \infty} \sup_{x \in \mathbb{N}} |T_k(h)\pi_k f(x) - \pi_k T(h)f(x)| = 0$$

(ii) *For each $f \in D$ there exists a $f_k \in \mathcal{D}(A_k)$ for each $k \geq 1$, such that f_k converges uniformly to f, that is*

$$\lim_{k \to \infty} \sup_{x \in \mathbb{N}} |f_k(x) - \pi_k f(x)| = 0,$$

and $A_k f_k$ converges uniformly to Af, that is

$$\lim_{k \to \infty} \sup_{x \in \mathbb{N}} |A_k f_k(x) - \pi_k Af(x)| = 0.$$

Furthermore, if the initial distribution of $N_k(0)/k$ converges to the initial distribution of $V(0)$, and either (i) *or* (ii) *above holds (and hence both hold), then*

$$\frac{N_k}{k} \to V \qquad \text{for } k \to \infty$$

as a process. In particular this implies that all finite dimensional distributions coincide.

Proof. See Ethier and Kurtz [10], Theorems 1.6.1 and 4.2.5. The condition that the semigroups involved are Feller semigroups is proven in Appendix A.4. □

We use this theorem by showing (ii), simply choosing the core D to be the domain $\mathcal{D}(A)$ of A. For any function $f \in \mathcal{D}(A)$ we then define the function f_k as $f_k = \pi_k f$. Then f_k converges uniformly to f in a trivial manner. For the generator A_k we have from (4.2) that

$$A_k f_k(x) = -\lambda f(x/k) - \mu x \left(f(x/k) - f((x-1)/k)\right) + \lambda \sum_{m=1}^\infty a_m^{(k)} f((x+m)/k),$$

where we have used that $f_k(x) = \pi_k f(x) = f(x/k)$. For the generator A it follows from (4.3) that

$$\pi_k Af(x) = (Af)(x/k) = -\lambda f(x/k) - \frac{rx}{k}f'(x/k) + \lambda \int_{x/k}^\infty f(y)\mathrm{d}B(y - x/k).$$

Hence to show (ii) above, it suffices to show that for any $f \in \mathcal{D}(A)$,

$$\lim_{k \to \infty} \sup_{x \in \mathbb{N}} \left| -\frac{\mu x}{k} \frac{f\left(\frac{x}{k}\right) - f\left(\frac{x-1}{k}\right)}{1/k} + \frac{rx}{k}f'\left(\frac{x}{k}\right) \right| = 0$$

and

$$\lim_{k \to \infty} \sup_{x \in \mathbb{N}} \left| \sum_{m=1}^\infty a_m^{(k)} f((x+m)/k) - \int_{x/k}^\infty f(v)\mathrm{d}_v B(v - x/k) \right| = 0.$$

Choosing $r = \mu$ and letting $y = x/k$, the correctness of these statements follows from the following lemma.

Lemma 4.2. *For any $f \in \mathcal{D}(A)$ the following two statements of (uniform) convergence hold.*

$$\lim_{k \to \infty} \sup_{y \geq 1/k} \left| y f'(y) - y \frac{f(y) - f(y - 1/k)}{1/k} \right| = 0 \tag{4.4}$$

and

$$\lim_{k \to \infty} \sup_{y > 0} \lim_{N \to \infty} \left| \sum_{m=1}^{kN} f\left(y + \frac{m}{k}\right) \left[B\left(\frac{m}{k}\right) - B\left(\frac{m-1}{k}\right) \right] - \int_0^N f(y+v) dB(v) \right| = 0. \tag{4.5}$$

Moreover, $\mathcal{D}(A)$ lies dense in $\hat{\mathcal{C}}(\mathbb{R}^+)$, so $\mathcal{D}(A)$ is a proper core of A.

Proof. Uniform convergence of (4.4) follows from the fact that $xf'(x)$ is bounded when $f \in \mathcal{D}(A)$.

For (4.5) we argue that the given Riemann–Stieltjes integral in it is approximated arbitrarily closely by the summation in front of it. That the convergence is uniform follows from the fact that f is continuous and f' is bounded.

Finally, to show that $\mathcal{D}(A)$ lies dense in $\hat{\mathcal{C}}(\mathbb{R}^+)$, take an arbitrary $f \in \hat{\mathcal{C}}(\mathbb{R}^+)$ outside of $\mathcal{D}(A)$. Then either f is not differentiable at one or more points, or $\lim_{x \to \infty} |xf'(x)| \neq 0$, or both. We indicate how a sequence of $f_n \in \mathcal{D}(A)$ can be constructed with $f_n \to f$ uniformly. When f is not differentiable at some x_0, this can be solved in a standard manner, letting the f_n be closer and closer to f near x_0 as $n \to \infty$, while maintaining that all f_n (for finite n) are in $\mathcal{D}(A)$. On the other hand, when $xf'(x)$ does not vanish at infinity, while f itself does, we can e.g. choose $f_n(x)$ to be a smoothed version (near $x = n$) of $f(x)\mathbf{1}\{x \leq n\}$, so that all f_n are in $\mathcal{D}(A)$ when f is differentiable, and clearly $xf'_n(x)$ vanishes at infinity. When f features both problems, we can combine both approaches to find a suitable sequence f_n. $\quad\square$

Finally we choose the initial distribution of $N_k(t)$ as before,

$$N_k(0) = \lceil kV(0) \rceil, \tag{4.6}$$

where $V(0)$ is the initial content of the shot-noise queue we started with. Clearly $N_k(0)/k$ converges a.s. to $V(0)$ as $k \to \infty$, as already mentioned above Lemma 3.2. Therefore, we are now ready to apply Theorem 4.1 and conclude the following.

Corollary 4.3. *If we choose the batch size distribution in the $M^X/M/\infty$ system according to (4.1), the initial distribution according to (4.6), and choose $\mu = r$, then N_k/k converges to V as a process.*

In particular this implies that all finite dimensional distributions for the infinite-server system, i.e. $\mathbb{P}(N_k(t_1) \leq kx_1, \ldots, N_k(t_n) \leq kx_n)$, converge in distribution as $k \to \infty$ to those for the shot-noise queue, i.e. to $\mathbb{P}(V_k(t_1) \leq x_1, \ldots, V_k(t_n) \leq x_n)$. Obviously this is a stronger statement then those in Section 3, where we only considered the distribution at one point in time.

Appendix. Proofs of some theorems

In this appendix we present some proofs that are not central to the line of reasoning in our paper. For the solution in Theorem 2.2 we present two proofs, illustrating the fact that the solution (2.4) can be found in a number of ways, see also the discussion in Section 1; in Appendix A.1 we give a full proof including uniqueness, while Appendix A.2 is mainly included for its elegance. Furthermore, Appendix A.3 presents the proof of (the solution in) Theorem 2.3 and Appendix A.4 presents the Feller properties of the semigroup $\{T(h)\}$, in the context of the proof of Theorem 4.1.

A.1. Proof of Theorem 2.2

The linear first order partial differential equation (2.3) with initial condition $g(s) := \Phi(s, 0)$ can be solved by the method of characteristics, see e.g. [11]. For the partial differential operator L we have that

$$L = rs \frac{\partial}{\partial s} + \frac{\partial}{\partial t}$$

and for the inhomogeneous term ψ that

$$\psi(s, t; \Phi(s, t)) = -\lambda(t)(1 - \beta(s, t))\Phi(s, t).$$

Note that ψ is at least a C^1 function of $\Phi(s, t)$, s and t. Furthermore we define the set S to be $S := (s, 0)$.

The parametrized characteristic curves $\gamma = (\gamma_1, \gamma_2)^T$ for L can be computed to be

$$\frac{d\gamma_1(w)}{dw} = r\gamma_1(w) \Rightarrow \gamma_1(w) = \gamma_1(0)e^{rw},$$

$$\frac{d\gamma_2(w)}{dw} = 1 \Rightarrow \gamma_2(w) = w + \gamma_2(0).$$

Then, by means of the method of characteristics, we find that

$$\frac{\partial \Phi(\gamma(w))}{\partial w} = \psi(\gamma(w), \Phi(\gamma(w)))$$
$$= -\lambda(w + \gamma_2(0))(1 - \beta(\gamma_1(0)e^{rw}, w + \gamma_2(0)))\Phi(\gamma_1(0)e^{rw}, w + \gamma_2(0)).$$

Next we define for every $\xi \in S$, γ_ξ in such a way that $\gamma(0) = \xi$. This condition comes down to

$$\gamma_1(0) = \xi \quad \text{and} \quad \gamma_2(0) = 0.$$

So γ_ξ is defined as follows,

$$\gamma_{\xi 1}(w) = \xi e^{rw} \quad \text{and} \quad \gamma_{\xi 2}(w) = w.$$

Next step is to define the function ψ_ξ,

$$\psi_\xi(w, y) = \psi(\gamma_\xi(w), y) = -\lambda(w)(1 - \beta(\xi e^{rw}, w))y.$$

According to the method of characteristics the function $v(w) = \Phi(\gamma(w))$ has to be a solution of the initial value problem

$$\frac{dv(w)}{dw} = \psi_\xi(w, v(w)) = -\lambda(w)(1 - \beta(\xi e^{rw}, w))v(w), \quad v(0) = g(\xi).$$

Integrating both sides of the equation gives that

$$\int_{v(0)}^{v(w)} \frac{1}{v(w)} dv(w) = -\int_0^w \lambda(\eta)(1 - \beta(\xi e^{r\eta}, \eta))d\eta,$$

which results in

$$v(w) = -g(\xi) \exp\left(\int_0^w \lambda(\eta)(1 - \beta(\xi e^{r\eta}, \eta))d\eta\right).$$

Substituting $\xi = e^{-rw}s$, since $\xi e^{rw} = s$, and $t = w$ into the above solution we get

$$\Phi(s, t) = \Phi(se^{-rt}, 0) \exp\left(-\int_0^t \lambda(\eta)(1 - \beta(se^{-r(t-\eta)}, \eta))d\eta\right).$$

Since the function ψ is at least one time continuously differentiable, the solutions γ_ξ and ψ_ξ are at least C^1 functions. This implies that the solution is uniquely determined on the union of characteristic curves which intersect S.

A.2. Alternative proof of Theorem 2.2

In this section we choose to present an elegant method due to [7], in which the process is viewed as a (time-dependent) Lévy process, rather than solving (2.3) directly.

It is well-known that the stochastic behavior of a Lévy process $J(t)$ is characterized by its exponent $\xi(s)$, for which we have $\mathbb{E}e^{-sJ(t)} = e^{-t\xi(s)}$. A similar result holds when $J(t)$ is a time-dependent Lévy process:

$$\mathbb{E}e^{-s(J(t)-J(t_0))} = e^{-\int_{t_0}^t \xi(s, u)du},$$

where $\xi(s, u)$ is the current Lévy exponent at time u. We first present the general derivation, supposing that we are interested in $\mathbb{E}e^{-sX(t)}$, where for some function $h(u)$,

$$X(t) = \int_0^t h(u)dJ(u). \tag{A.1}$$

We may now subdivide the interval $(0, t]$ into n equal subintervals $(t_i, t_{i+1}]$, with $t_i = ti/n$, $i = 0, \ldots, n$, and then let $n \to \infty$ to obtain

$$\mathbb{E}e^{-s\int_0^t h(u)dJ(u)} = \lim_{n\to\infty} \mathbb{E}e^{-s\sum_{i=0}^{n-1} h(t_i)(J(t_{i+1})-J(t_i))}$$

$$= \lim_{n\to\infty} \mathbb{E}\prod_{i=0}^{n-1} e^{-sh(t_i)(J(t_{i+1})-J(t_i))}$$

$$= \lim_{n\to\infty} \prod_{i=0}^{n-1} \mathbb{E}e^{-sh(t_i)(J(t_{i+1})-J(t_i))}$$

$$= \lim_{n\to\infty} \prod_{i=0}^{n-1} \exp\left(-\int_{t_i}^{t_{i+1}} \xi(sh(t_i), u)du\right)$$

$$= \lim_{n \to \infty} \exp\left(-\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \xi(sh(t_i), u)du\right)$$

$$= \exp\left(-\int_0^t \xi(sh(u), u)du\right).$$

Next we apply this general result to our situation at hand. Just as a compound Poisson process with rate λ and jump size LST $\beta(s)$ is a Lévy process with exponent $\xi(s) = \lambda(1 - \beta(s))$, a time-dependent compound Poisson process is obtained by considering a time-dependent Lévy process with exponent $\xi(s, u) = \lambda(u)(1 - \beta(s, u))$. Furthermore we choose the functional $h(\cdot)$ of this process as $h(u) = e^{-r(t-u)}$, so that the process $X(t)$ defined in (A.1) coincides with our shot-noise content process $V(t)$, provided that we have $V(0) = 0$. Inserting the two equations above, we find the solution (2.4) as in Theorem 2.2, apart from the factor $\Phi(se^{-rt}, 0)$ which is 1 when $V(0) = 0$.

When $V(0) > 0$, the resulting additional remaining content $V(0)e^{-rt}$ is stochastically independent of $X(t)$ as defined in (A.1), so by multiplication of LSTs we end up with

$$\mathbb{E}[e^{-sV(t)}] = \mathbb{E}[e^{-s(V(0)e^{-rt}+X(t))}] = \Phi(se^{-rt}, 0)\mathbb{E}[e^{-sX(t)}],$$

leading to (2.4).

A.3. Proof of Theorem 2.3

The linear first order partial differential equation (2.5) with initial condition $g(z) := P_N(z, 0)$ can be solved by the method of characteristics, as is also used in the proof of Theorem 2.2 in Appendix A.1. We repeat the arguments to let this section be self-contained.

For the partial differential operator L we have that

$$L = -\mu(1 - z)\frac{\partial}{\partial z} + \frac{\partial}{\partial t},$$

and for the inhomogeneous term ψ that

$$\psi(z, t; P_N(z, t)) = -\lambda(t)(1 - P_X(z, t))P_N(z, t).$$

Note that ψ is at least a C^1 function of $P_N(z, t)$, z and t. Furthermore we define the set S to be $S := (z, 0)$.

The parametrized characteristic curves $\gamma = (\gamma_1, \gamma_2)^T$ for L can be computed to be

$$\frac{d\gamma_1(w)}{dw} = -\mu(1 - \gamma_1(w)) \Rightarrow \gamma_1(w) = (\gamma_1(0) - 1)e^{\mu w} + 1,$$

$$\frac{d\gamma_2(w)}{dw} = 1 \Rightarrow \gamma_2(w) = w + \gamma_2(0).$$

Then, by means of the method of characteristics, we find that

$$\frac{\partial P_N(\gamma(w))}{\partial w} = \psi(\gamma(w), P_N(\gamma(w)))$$

$$= -\lambda(w + \gamma_2(0))(1 - P_X((\gamma_1(0) - 1)e^{\mu w} + 1, w + \gamma_2(0)))$$

$$\times P_N((\gamma_1(0) - 1)e^{\mu w} + 1, w + \gamma_2(0)).$$

Next we define for every $\xi \in S$, γ_ξ in such a way that $\gamma(0) = \xi$. This condition comes down to

$$\gamma_1(0) = \xi \quad \text{and} \quad \gamma_2(0) = 0.$$

So γ_ξ is defined as follows,

$$\gamma_{\xi 1}(w) = (\xi - 1)e^{\mu w} + 1 \quad \text{and} \quad \gamma_{\xi 2}(w) = w.$$

Next step is to define the function ψ_ξ,

$$\psi_\xi(w, y) = \psi(\gamma_\xi(w), y) = -\lambda(w)(1 - P_X((\xi - 1)e^{\mu w} + 1, w))y.$$

According to the method of characteristics the function $v(w) = P_N(\gamma(w))$ has to be a solution of the initial value problem

$$\frac{dv(w)}{dw} = \psi_\xi(w, v(w)) = -\lambda(w)(1 - P_X((\xi - 1)e^{\mu w} + 1, w))v(w), \quad v(0) = g(\xi).$$

Integrating both sides of the equation gives that

$$\int_{v(0)}^{v(w)} \frac{1}{v(w)}dv(w) = -\int_0^w \lambda(\eta)(1 - P_X((\xi - 1)e^{\mu \eta} + 1, \eta))d\eta,$$

which results in

$$v(w) = g(\xi) \exp\left(-\int_0^w \lambda(\eta)(1 - P_X((\xi - 1)e^{\mu\eta} + 1, \eta))d\eta\right).$$

Substituting $\xi = e^{-\mu w}(z - 1) + 1$, since $(\xi - 1)e^{\mu w} + 1 = z$, and $t = w$ into the above solution, yields that

$$P_N(z, t) = P_N(v(z, t), 0) \exp\left(-\int_0^t \lambda(\eta)(1 - P_X(v(z, t - \eta), \eta))d\eta\right),$$

where $v(z, t) = 1 - (1 - z)e^{-\mu t}$. Since the function ψ is at least one time continuously differentiable, the solutions γ_ξ and ψ_ξ are at least C^1 functions. This implies that the solution is uniquely determined on the union of characteristic curves which intersect S.

A.4. Proof of Theorem 4.1

We need to prove that the semigroup $\{T(h)\}$ is a Feller semigroup on $\hat{\mathcal{C}}(\mathbb{R}_+)$, and that for any k the semigroup $\{T_k(h)\}$ is a Feller semigroup on $\hat{\mathcal{B}}(\mathbb{N})$. Since both proofs are very similar we only show the former, leaving the latter for the reader to verify. Firstly observe that $\{T(h)\}$ is a strongly continuous contraction semigroup on $\hat{\mathcal{C}}(\mathbb{R}_+)$, since

$$\lim_{h \downarrow 0} T(h)f(x) = f(x)$$

and since

$$\begin{aligned}
\|T(h)f\| &= \sup_{x \in \mathbb{R}_+} |\mathbb{E}[f(V(h))|V(0) = x]| \\
&\leq \sup_{x \in \mathbb{R}_+} |f(x)| \\
&= \|f\|
\end{aligned}$$

for all $h \geq 0$ and all $f \in \hat{\mathcal{C}}(\mathbb{R}_+)$. In order for $\{T(h)\}$ to be a Feller semigroup, it should also be positive on $\hat{\mathcal{C}}(\mathbb{R}_+)$ and its generator A conservative. It is positive since $T(h)$ is a positive operator for all $h \geq 0$. This is seen from the fact that for nonnegative functions f

$$T(h)f(x) = \mathbb{E}[f(V(h))|V(0) = x] \geq 0$$

for all $h \geq 0$. The generator $A : \hat{\mathcal{C}}(\mathbb{R}_+) \to \hat{\mathcal{C}}(\mathbb{R}_+)$ is conservative since it clearly contains $(1, 0)$, i.e. when $f(x) = 1$ is chosen, we have that

$$A1 = -\lambda + \lambda \int_x^\infty dB(y - x) = -\lambda + \lambda = 0.$$

References

[1] I. Eliazar, J. Klafter, On the nonlinear modeling of shot noise, Proc. Natl. Acad. Sci. USA 102 (2005) 13779–13782.
[2] S.M. Ross, Stochastic Processes, second ed., in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, vol. 14, John Wiley & Sons Inc., New York, 1996. Lectures in Mathematics.
[3] J.F.C. Kingman, Poisson Processes, in: Oxford Studies in Probability, vol. 3, The Clarendon Press, Oxford University Press, Oxford Science Publications, New York, 1993.
[4] L. Takács, Investigation of waiting time problems by reduction to Markov processes, Acta Math. Acad. Sci. Hungar. 6 (1955) 101-129.
[5] W. Scheinhardt, N. Van Foreest, M. Mandjes, Continuous feedback fluid queues, Oper. Res. Lett. 33 (2005) 551-559.
[6] R. Bekker, S. Borst, O. Boxma, O. Kella, Queues with workload-dependent arrival and service rates, Queueing Syst. 46 (2004) 537-556.
[7] O. Kella, Personal discussion, 2008.
[8] O. Kella, W. Whitt, Linear stochastic fluid networks, J. Appl. Probab. 36 (1999) 244–260.
[9] M. Chaudhry, J. Templeton, A First Course in Bulk Queues, John Wiley & Sons, Inc., 1983, p. xiv+694.
[10] S. Ethier, T. Kurtz, Markov Processes, Characterization and Convergence, in: Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc., Canada, 1986.
[11] L. Evans, Partial Differential Equatons, in: Graduate studies in mathematics, American Mathematical Society, 1998.

Wilfred F. de Graaf obtained his M.Sc. degrees in Mathematics (2007) and Theoretical Physics (2010) at Utrecht University, and is currently a Ph.D. student in History of Mathematics and Astronomy at the Mathematical Institute of Utrecht University.

Werner R.W. Scheinhardt is an assistant professor at the University of Twente. He obtained his M.Sc. degree (1994) and Ph.D. degree (1998) in Applied Mathematics at the same institute. He also held a Postdoc position at Eindhoven University of Technology and was affiliated to the Centre for Mathematics and Computer Science (CWI) in Amsterdam. His research interests are mainly in queueing theory and stochastic processes.

Richard J. Boucherie received M.Sc. degrees in 1988 in Mathematics (Stochastic Operations Research) and Theoretical Physics (Statistical Physics) from the Universiteit Leiden, and received the Ph.D. degree in Econometrics in 1992 from the Vrije Universiteit, Amsterdam, followed by Postdocs at INRIA Sophia Antipolis, CWI Amsterdam, and Universiteit van Amsterdam. Since 2000 he is with the Department of Applied Mathematics of the University of Twente, where he was appointed in 2003 as Full Professor of Stochastic Operations Research. He is co-founder and co-chair of the University of Twente Center for Healthcare Operations Improvement and Research (CHOIR) in the area of healthcare logistics. In 2014 he co-founded the spin-off company Rhythm, that carries out actual implementations of healthcare logistics solutions in healthcare organisations. His research interests are in queueing theory, Petri nets, and random walks with application areas including healthcare.