CrossMark

Networks of $\cdot/G/\infty$ queues with shot-noise-driven arrival intensities

D. T. Koops[1] 🆔 · **O. J. Boxma[2]** · **M. R. H. Mandjes[1]**

Abstract We study infinite-server queues in which the arrival process is a Cox process (or doubly stochastic Poisson process), of which the arrival rate is given by a shot-noise process. A shot-noise rate emerges naturally in cases where the arrival rate tends to exhibit sudden increases (or shots) at random epochs, after which the rate is inclined to revert to lower values. Exponential decay of the shot noise is assumed, so that the queueing systems are amenable to analysis. In particular, we perform transient analysis on the number of jobs in the queue jointly with the value of the driving shot-noise process. Additionally, we derive heavy-traffic asymptotics for the number of jobs in the system by using a linear scaling of the shot intensity. First we focus on a one-dimensional setting in which there is a single infinite-server queue, which we then extend to a network setting.

Keywords Infinite-server queue · Stochastic arrival rate · Cox · Doubly stochastic Poisson · Shot noise · Network · Functional central limit theorem · Heavy traffic

Mathematics Subject Classification 60K25

1 Introduction

In the queueing literature, one has traditionally studied queues with Poisson input. The Poisson assumption typically facilitates explicit analysis, but it does not always align well with actual data; see, for example, [11] and references therein. More specifically,

✉ D. T. Koops
d.t.koops@uva.nl

[1] Korteweg-de Vries Institute, University of Amsterdam, Amsterdam, The Netherlands

[2] Eurandom and Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

🌀 Springer

statistical studies show that in many practical situations, Poisson processes underestimate the variability of the queue's input stream. This observation has motivated research on queues fed by arrival processes that better capture the burstiness observed in practice.

The extent to which burstiness takes place can be measured by the dispersion index, i.e., the ratio of the variance of the number of arrivals in a given interval, and the corresponding expected value. In arrival streams that display burstiness, the dispersion index is larger than unity (as opposed to Poisson processes, for which it is equal to unity), a phenomenon that is usually referred to as *overdispersion*. It is desirable that the arrival process of the queueing model takes the observed overdispersion into account. One way to achieve this is to make use of *Cox processes*, which are Poisson processes, conditional on the stochastic time-dependent intensity. It is an immediate consequence of the law of total variance that Cox processes *do* have a dispersion index larger than unity. Therefore, this class of processes makes for a good candidate to model overdispersed input processes.

In this paper we contribute to the development of queueing models fed by input streams that exhibit overdispersion. We analyze infinite-server queues driven by a particular Cox process, in which the rate is a (stochastic) shot-noise process.

The shot-noise process we use is one in which there are only upward jumps (or shots) that arrive according to a homogeneous Poisson process. Furthermore, we employ an exponential 'response' or 'decay' function, which encodes how quickly the process will decline after a jump. In this case, the shot-noise process is a Markov process; see [16, p. 393]. There are several variations on shot-noise processes; see, for example, [10] for a comprehensive overview.

It is not a novel idea to use a shot-noise process as stochastic intensity. For instance, in insurance mathematics, the authors of [5] use a shot-noise-driven Cox process to model the claim count. They assume that disasters happen according to a Poisson process, and each disaster can induce a cluster of arriving claims. The disaster corresponds to a shot upwards in the claim intensity. As time passes, the claim intensity process decreases, as more and more claims are settled.

Another example of shot-noise arrival processes is found in the famous paper [13], where it is used to model the occurrences of earthquakes. The arrival process considered in [13] has one crucial difference with the one used in this paper: it makes use of Hawkes processes [9], which do have a shot-noise structure, but have the special feature that they are *self-exciting*. More specifically, in Hawkes processes, an arrival induces a shot in the arrival rate, whereas in our shot-noise-driven Cox model these shots are merely exogenous. The Hawkes process is less tractable than the shot-noise-driven Cox process. A very recent effort to analyze $\cdot/G/\infty$ queues that are driven by a Hawkes process has been made in [8], where a functional central limit theorem is derived for the number of jobs in the system. In this model, obtaining explicit results (in a non-asymptotic setting), as we are able to do in the shot-noise-driven Cox variant, is still an open problem.

In order to successfully implement a theoretical model, it is crucial to have methods to estimate its parameters from data. The shot-noise-driven Cox process is attractive since it has this property. Statistical methods that filter the unobservable intensity process, based on Markov Chain Monte Carlo (MCMC) techniques, have been developed;

see [3] and references therein. By filtering, they refer to the estimation of the intensity process in a given time interval, given a realized arrival process. Subsequently, given this 'filtered path' of the intensity process, the parameters of the shot-noise process can be estimated by a Monte Carlo version of the expectation maximization (EM) method. Furthermore, the shot-noise-driven Cox process can also be easily simulated; see, for example, the thinning procedure described in [12].

In this paper we study networks of infinite-server queues with shot-noise-driven Cox input. We assume that the service times at a given node are i.i.d. samples from a general distribution. The output of a queue is routed to a next queue, or leaves the network. Infinite-server queues have the inherent advantage that jobs do not interfere with one another, which considerably simplifies the analysis. Furthermore, infinite-server systems are frequently used to produce approximations for corresponding finite-server systems. In the network setting, we can model queueing systems that are driven by correlated shot-noise arrival processes. With regards to applications, such a system could, for example, represent the call centers of a fire department and police department in the same town.

The contributions and organization of this paper are as follows. In this paper we derive exact and asymptotic results. The main result of the exact analysis is Theorem 4.6, where we find the joint Laplace transform of the numbers of jobs in the queues of a feedforward network, jointly with the shot-noise-driven arrival rates. We build up toward this result as follows. In Sect. 2 we introduce notation and we state the important Lemma 2.1 that we repeatedly rely on. Then we derive exact results for the single infinite-server queue with a shot-noise arrival rate in Sect. 3.1. Subsequently, in Sect. 3.2, we show that after an appropriate scaling the number of jobs in the system satisfies a functional central limit theorem (Theorem 3.4); the limiting process is an Ornstein–Uhlenbeck (OU) process driven by a superposition of a Brownian motion and an integrated OU process. We then extend the theory to a network setting in Sect. 4. Before we consider full-blown networks, we first consider a tandem system consisting of an arbitrary number of infinite-server queues in Sect. 4.1. Then it is argued in Sect. 4.2 that a feedforward network can be seen as a number of tandem queues in parallel. We analyze two different ways in which dependency can enter the system through the arrival process. Firstly, in Model (M1), parallel service facilities are driven by a multidimensional shot-noise process in which the *shots* are simultaneous (which includes the possibility that all shot-noise processes are equal). Secondly, in Model (M2), we assume that there is one shot-noise arrival intensity that generates simultaneous *arrivals* in all tandems. In Sect. 5 we finish with some concluding remarks.

2 Notation and preliminaries

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be a probability space, in which the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is such that $\Lambda(\cdot)$ is adapted to it. A shot-noise process is a process that has random jumps at Poisson epochs, and a deterministic 'response' or 'decay' function, which governs the behavior of the process. See [16, Section 8.7] for a brief account of shot-noise processes. The shot noise that we use in this paper has the following representation:
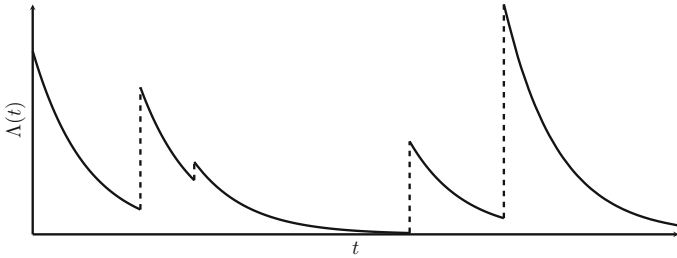
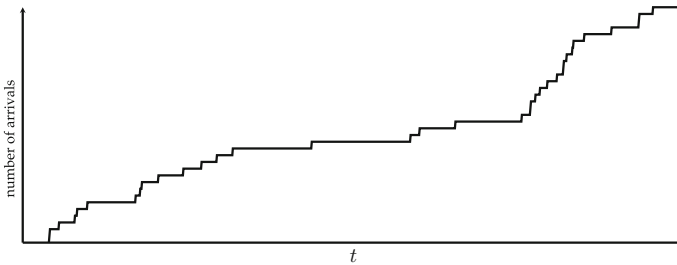**Fig. 1** Sample path of shot-noise process

Fig. 2 A realization of arrival process corresponding to the sample path of the arrival rate process in Fig. 1

$$\Lambda(t) = \Lambda(0)e^{-rt} + \sum_{i=1}^{P_B(t)} B_i e^{-r(t-t_i)}, \tag{1}$$

where the $B_i \geq 0$ are i.i.d. shots from a general distribution, the decay function is exponential with rate $r > 0$, P_B is a homogeneous Poisson process with rate ν, and the epochs of the shots, that arrived before time t, are labeled $t_1, t_2, \ldots, t_{P_B(t)}$.

As explained in the introduction, the shot-noise process serves as a stochastic arrival rate to a queueing system. It is straightforward to simulate a shot-noise process; for an illustration of a sample path, consider Fig. 1. Using the thinning method for nonhomogeneous Poisson processes [12], and using the sample path of Fig. 1 as the arrival rate, one can generate a corresponding sample path for the arrival process, as is displayed in Fig. 2. Typically, most arrivals occur shortly after peaks in the shot-noise process in Fig. 1, as expected.

We write Λ (i.e., without argument) for a random variable with distribution equal to that of $\lim_{t\to\infty} \Lambda(t)$. We now present well-known transient and stationary moments of the shot-noise process, see Appendix 2 and, for example [16]: with B distributed as B_1,

$$\mathbb{E}\,\Lambda(t) = \Lambda(0)e^{-rt} + \frac{\nu\,\mathbb{E}\,B}{r}(1 - e^{-rt}), \quad \mathbb{E}\,\Lambda = \frac{\nu\,\mathbb{E}\,B}{r},$$

$$\mathbb{V}\text{ar}\,\Lambda(t) = \frac{\nu\,\mathbb{E}\,B^2}{2r}(1 - e^{-2rt}), \quad \mathbb{V}\text{ar}\,\Lambda = \frac{\nu\,\mathbb{E}\,B^2}{2r},$$

$$\mathbb{C}\text{ov}(\Lambda(t), \Lambda(t+\delta)) = e^{-r\delta}\,\mathbb{V}\text{ar}\,\Lambda(t). \tag{2}$$

We remark that, for convenience, we throughout assume $\Lambda(0) = 0$. The results can be readily extended to the case in which $\Lambda(0)$ is a nonnegative random variable, at the cost of a somewhat more cumbersome notation.

In the one-dimensional case, we denote $\beta(s) = \mathbb{E}\, e^{-sB}$, and in the multidimensional case, where $s = (s_1, s_2, \ldots, s_d)$, for some integer $d \geq 2$, now denotes a vector, we write

$$\beta(s) = \mathbb{E}\, e^{-\sum_i s_i B_i}.$$

The following lemma will be important for the derivation of the joint transform of $\Lambda(t)$ and the number of jobs in system, in both single- and multi-node cases.

Lemma 2.1 *Let $\Lambda(\cdot)$ be a shot-noise process. Let $f : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a function which is piecewise continuous in its first argument, with at most a countable number of discontinuities. Then it holds that*

$$\mathbb{E} \exp\left(\int_0^t f(u, z)\Lambda(u)\, \mathrm{d}u - s\Lambda(t)\right)$$

$$= \exp\left(v \int_0^t \left(\beta\left(se^{-r(t-v)} - e^{rv}\int_v^t f(u, z)e^{-ru}\, \mathrm{d}u\right) - 1\right) \mathrm{d}v\right).$$

Proof See Appendix 1. □

3 A single infinite-server queue

In this section we study the $M_S/G/\infty$ queue. This is a single infinite-server queue, of which the arrival process is a Cox process driven by the shot-noise process $\Lambda(\cdot)$, as defined in Sect. 2. First we derive exact results in Sect. 3.1, where we find the joint transform of the number of jobs in the system and the shot-noise rate, and derive expressions for the expected value and variance. Subsequently, in Sect. 3.2, we derive a functional central limit theorem for this model.

3.1 Exact analysis

We let J_i be the service requirement of the i-th job, where $J_1, J_2, \ldots$ are assumed to be i.i.d.; in the sequel J denotes a random variable that is equal in distribution to J_1. Our first objective is to find the distribution of the number of jobs in the system at time t, in the sequel denoted by $N(t)$. This can be found in several ways; because of the appealing underlying intuition, we here provide an argument in which we approximate the arrival rate on intervals of length Δ by a constant, and then let $\Delta \downarrow 0$.

This procedure works as follows. We let $\Lambda(t) = \Lambda(\omega, t)$ be an arbitrary sample path of the driving shot-noise process. Given $\Lambda(t)$, the number of jobs that arrived in the interval $[k\Delta, (k+1)\Delta)$ and are still in the system at time t has a Poisson distribution

with parameter $\mathbb{P}(J > t - (k\Delta + \Delta U_k)) \cdot \Delta\Lambda(k\Delta) + o(\Delta)$, where $U_1, U_2, \ldots$ are i.i.d. standard uniform random variables. Summing over k yields that the number of jobs in the system at time t has a Poisson distribution with parameter

$$\sum_{k=0}^{t/\Delta-1} \mathbb{P}(J > t - (k\Delta + \Delta U_k))\Delta\Lambda(k\Delta) + o(\Delta),$$

which converges, as $\Delta \downarrow 0$, to

$$\int_0^t \mathbb{P}(J > t - u)\Lambda(u)\,\mathrm{d}u. \tag{3}$$

The argument above is not new: a similar observation was mentioned in, for example [6], for deterministic rate functions. Since $\Lambda(\cdot)$ is actually a stochastic process, we conclude that the number of jobs has a mixed Poisson distribution, with the expression in Eq. (3) as random parameter. As a consequence, we find by conditioning on $\mathcal{F}_t$,

$$\xi(t, z, s) := \mathbb{E}\, z^{N(t)} e^{-s\Lambda(t)} = \mathbb{E}\left(e^{-s\Lambda(t)}\, \mathbb{E}\left(z^{N(t)} \mid \mathcal{F}_t \right) \right)$$

$$= \mathbb{E}\exp\left(\int_0^t (z-1)\mathbb{P}(J > t - u)\Lambda(u)\,\mathrm{d}u - s\Lambda(t) \right). \tag{4}$$

We have found the following result.

Theorem 3.1 *Let $\Lambda(\cdot)$ be a shot-noise process. Then*

$$\log\xi(t, z, s)$$
$$= v\int_0^t \left(\beta\left((1-z)e^{rv}\int_v^t \mathbb{P}(J > t - u)e^{-ru}\,\mathrm{d}u + se^{-r(t-v)} \right) - 1 \right)\mathrm{d}v. \tag{5}$$

Proof The result follows directly from Lemma 2.1 and Eq. (4). $\square$

In Theorem 3.1 we found that $N(t)$ has a Poisson distribution with the random parameter given in Eq. (3). This leads to the following expression for the expected value:

$$\mathbb{E}\,N(t) = \int_0^t \mathbb{E}\,\Lambda(u)\,\mathbb{P}(J > t - u)\,\mathrm{d}u. \tag{6}$$

In addition, by the law of total variance we find

$$\mathbb{V}\mathrm{ar}\,N(t) = \mathbb{V}\mathrm{ar}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J > t - u)\,\mathrm{d}u \right)$$

$$+ \mathbb{E}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J > t - u)\,\mathrm{d}u \right). \tag{7}$$

The latter expression we can further evaluate, using an approximation argument that resembles the one we used above. Using a Riemann sum approximation, we find

$$\mathbb{V}\text{ar}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J > t - u)\,du\right) = \lim_{\Delta \downarrow 0} \mathbb{V}\text{ar}\left(\sum_{i=0}^{t/\Delta-1} \Lambda(i\Delta)\,\mathbb{P}(J > t - i\Delta)\Delta\right)$$

$$= 2\lim_{\Delta \downarrow 0} \sum_{i=0}^{t/\Delta-1}\sum_{j>i}^{t/\Delta-1} \mathbb{C}\text{ov}(\Lambda(i\Delta)\,\mathbb{P}(J > t - i\Delta)\Delta, \Lambda(j\Delta)\,\mathbb{P}(J > t - j\Delta)\Delta)$$

$$= 2\int_0^t \int_v^t \mathbb{C}\text{ov}(\Lambda(u), \Lambda(v))\,\mathbb{P}(J > t - u)\,\mathbb{P}(J > t - v)\,du\,dv.$$

Assuming that $u \geq v$, we know that $\mathbb{C}\text{ov}(\Lambda(u), \Lambda(v)) = e^{-r(u-v)}\,\mathbb{V}\text{ar}\,\Lambda(v)$ (cf. Lemma 4.7). We thus have that (7) equals

$$2\int_0^t \int_v^t e^{-r(u-v)}\,\mathbb{V}\text{ar}\,\Lambda(v)\,\mathbb{P}(J > t - u)\,\mathbb{P}(J > t - v)\,du\,dv$$

$$+ \int_0^t \mathbb{E}\,\Lambda(u)\,\mathbb{P}(J > t - u)\,du.$$

We can make this more explicit using the corresponding formulas in (2).

Example 3.2 (Exponential case) Consider the case in which J is exponentially distributed with mean $1/\mu$ and $\Lambda(0) = 0$. Then we can calculate the mean and variance explicitly. For $\mu \neq r$,

$$\mathbb{E}\,N(t) = \frac{\mathbb{E}\,\Lambda}{\mu}h_{r,\mu}(t),$$

where the function $h_{r,\mu}(\cdot)$ is defined by

$$t \mapsto \begin{cases} \dfrac{\mu(1 - e^{-rt}) - r(1 - e^{-\mu t})}{\mu - r} & \text{if } \mu \neq r, \\ 1 - e^{-rt} - rte^{-rt} & \text{if } \mu = r. \end{cases}$$

For the variance, we thus find for $\mu \neq r$

$$\mathbb{V}\text{ar}\,N(t)$$
$$= \frac{v\,\mathbb{E}\,B^2}{2r}\frac{r^2(1 - e^{-2\mu t}) + \mu^2(1 - e^{-2rt}) + \mu r(4e^{-t(\mu+r)} - e^{-2\mu t} - e^{-2rt} - 2)}{\mu(\mu - r)^2(\mu + r)}$$
$$+ \mathbb{E}\,N(t),$$

and for $\mu = r$

$$\mathbb{V}\text{ar}\,N(t) = \frac{v\,\mathbb{E}\,B^2}{4r^3}\left(1 - e^{-2rt} - 2rt(1 + rt)e^{-2rt}\right) + \mathbb{E}\,N(t).$$

3.2 Asymptotic analysis

This subsection focuses on deriving a functional central limit theorem (FCLT) for the model under study, after having appropriately scaled the shot rate of the shot-noise process. In the following, we assume that the service requirements are exponentially distributed with rate μ, and we point out how it can be generalized to a general distribution in Remark 3.6 below. We follow the standard approach to derive the FCLT for infinite-server queueing systems; we mimic the argumentation used in, for example [1,14]. As the proof has a relatively large number of standard elements, we restrict ourselves to the most important steps.

We apply a linear scaling to the shot rate of the shot-noise process, i.e., $\nu \mapsto n\nu$. It is readily checked that under this scaling, the steady-state level of the shot-noise process and the steady-state number of jobs in the queue blow up by a factor n. It is our objective to prove that, after appropriate centering and normalization, the process recording the number of jobs in the system converges to a Gaussian process.

In the n-th scaled model, the number of jobs in the system at time t, denoted by $N_n(t)$, has the following (obvious) representation: with $A_n(t)$ denoting the number of arrivals in $[0, t]$, and $D_n(t)$ the number of departures,

$$N_n(t) = N_n(0) + A_n(t) - D_n(t). \tag{8}$$

Here, $A_n(t)$ corresponds to a Cox process with a shot-noise-driven rate, and therefore we have, with $\Lambda_n(s)$ the shot-noise in the scaled model at time s and $S_A(\cdot)$ a unit-rate Poisson process,

$$A_n(t) = S_A \left(\int_0^t \Lambda_n(u) \mathrm{d}u \right);$$

in line with our previous assumptions, we put $\Lambda_n(0) = 0$. For our infinite-server model the departures $D_n(t)$ can be written as, with $S_D(\cdot)$ a unit-rate Poisson process (independent of $S_A(\cdot)$),

$$D_n(t) = S_D \left(\int_0^t \mu N_n(u) \, \mathrm{d}u \right).$$

We start by identifying the average behavior of the process $N_n(t)$. Following the reasoning of [1], assuming that $N_n(0)/n \Rightarrow \rho(0)$ (where '$\Rightarrow$' denotes weak convergence), $N_n(t)/n$ converges almost surely to the solution of

$$\rho(t) = \rho(0) + \int_0^t \mathbb{E}\,\Lambda(u) \, \mathrm{d}u - \int_0^t \mu\rho(u) \, \mathrm{d}u. \tag{9}$$

This equation is solved by $\rho(t) = \mathbb{E}\,N(t)$, with $\mathbb{E}\,N(t)$ provided in Example 3.2.

Now we move to the derivation of the FCLT. Following the approach used in [1], we proceed by studying an FCLT for the input rate process. To this end, we first define

$$\hat{\Lambda}_n(t) := \sqrt{n}\left(\frac{1}{n}\Lambda_n(t) - \mathbb{E}\,\Lambda(t)\right); \qquad \hat{K}_n(t) := \int_0^t \hat{\Lambda}(u)\,\mathrm{d}u.$$

The following lemma states that $\hat{K}_n(\cdot)$ converges to an integrated Ornstein–Uhlenbeck (OU) process, corresponding to an OU process $\hat{\Lambda}(\cdot)$ with a speed of mean reversion equal to r, long-run equilibrium level 0, and variance $\sigma_\Lambda^2 := \nu\,\mathbb{E}\,B^2/(2r)$.

Lemma 3.3 *Assume that for the shot sizes, distributed as B, it holds that $\mathbb{E}\,B$, $\mathbb{E}\,B^2 < \infty$. Then $\hat{K}_n(\cdot) \Rightarrow \hat{K}(\cdot)$ as $n \to \infty$, where*

$$\hat{K}(t) = \int_0^t \hat{\Lambda}(u)\,\mathrm{d}u, \tag{10}$$

in which $\hat{\Lambda}$ satisfies, with $W_1(\cdot)$ a standard Brownian motion,

$$\hat{\Lambda}(t) = \sigma_\Lambda W_1(t) - r\int_0^t \hat{\Lambda}(u)\,\mathrm{d}u. \tag{11}$$

Proof This proof is standard; for instance from [2, Prop. 3], by putting the λ_d in that paper to zero, it follows that $\hat{\Lambda}_n(\cdot) \Rightarrow \hat{\Lambda}(\cdot)$. This implies $\hat{K}_n(\cdot) \Rightarrow \hat{K}(\cdot)$, using (10) together with the continuous mapping theorem. $\qquad\square$

Interestingly, the above result entails that the arrival rate process displays mean-reverting behavior. This also holds for the job count process in standard infinite-server queues. In other words, the job count process in the queueing system we are studying can be considered as the composition of two mean-reverting processes. We make this more precise in the following.

From now on we consider the following centered and normalized version of the number of jobs in the system:

$$\hat{N}_n(t) := \sqrt{n}\left(\frac{1}{n}N_n(t) - \rho(t)\right).$$

We assume that $\hat{N}_n(0) \Rightarrow \hat{N}(0)$ as $n \to \infty$. To prove the FCLT, we rewrite $\hat{N}_n(t)$ in a convenient form. Mimicking the steps performed in [1] or [14], with $\bar{S}_A(t) := S_A(t) - t$, $\bar{S}_D(t) := S_D(t) - t$,

$$R_n(t) := \bar{S}_A\left(\int_0^t \Lambda_n(u)\mathrm{d}u\right) - \bar{S}_D\left(\mu\int_0^t N_n(u)\mathrm{d}u\right),$$

and using the relation (9), we eventually obtain

$$\hat{N}_n(t) = \hat{N}_n(0) + \frac{R_n(t)}{\sqrt{n}} + \hat{K}_n(t) - \mu\int_0^t \hat{N}_n(u)\mathrm{d}u.$$

Our next goal is to apply the martingale FCLT to the martingales $R_n(t)/\sqrt{n}$; see, for background on the martingale FCLT, for instance [7] and [17]. The quadratic variation equals

$$
\left[\frac{R_n}{\sqrt{n}}\right]_t = \frac{1}{n}\left(S_A\left(\int_0^t \Lambda_n(u)\mathrm{d}u\right) + S_D\left(\mu\int_0^t N_n(u)\mathrm{d}u\right)\right),
$$

which converges to $\int_0^t \mathbb{E}\,\Lambda(u)\mathrm{d}u + \mu\int_0^t \rho(u)\mathrm{d}u$. Appealing to the martingale FCLT, the following FCLT is obtained.

Theorem 3.4 *The centered and normalized version of the number of jobs in the queue satisfies an FCLT:* $\hat{N}_n(\cdot) \Rightarrow \hat{N}(\cdot)$ *as* $n \to \infty$, *where* $\hat{N}(t)$ *solves the stochastic integral equation*

$$
\hat{N}(t) = \hat{N}(0) + \int_0^t \sqrt{\mathbb{E}\,\Lambda(u) + \mu\rho(u)}\,\mathrm{d}W_2(u) + \hat{K}(t) - \mu\int_0^t \hat{N}(u)\mathrm{d}u,
$$

with $W_2(\cdot)$ *a standard Brownian motion that is independent of the Brownian motion* $W_1(\cdot)$ *we introduced in the definition of* $\hat{K}(\cdot)$.

Remark 3.5 In passing, we have proven that the arrival process as such obeys an FCLT. With

$$
\hat{A}_n(t) := \sqrt{n}\left(\frac{1}{n}A_n(t) - \int_0^t \mathbb{E}\,\Lambda(u)\,\mathrm{d}u\right),
$$

we find that $\hat{A}_n(t) \Rightarrow \hat{A}(t)$ as $n \to \infty$, where

$$
\begin{aligned}
\hat{A}(t) &:= \int_0^t \sqrt{\mathbb{E}\,\Lambda(u) + \mu\rho(u)}\,\mathrm{d}W_2(u) + \hat{K}(t) \\
&= \int_0^t \sqrt{2\mu\rho(u) + \rho'(u)}\,\mathrm{d}W_2(u) + \hat{K}(t);
\end{aligned}
$$

the last equality follows from the fact that $\rho(\cdot)$ satisfies (9).

Remark 3.6 The FCLT can be extended to non-exponential service requirements, by making use of [15, Thm. 3.2]. Their approach relies on two assumptions:

- The arrival process should satisfy an FCLT;
- The service times are i.i.d. nonnegative random variables with a general c.d.f. independent of the arrival process.

As noted in Remark 3.5, the first assumption is satisfied for the model in this paper. The second assumption holds as well. In the non-exponential case, the results are less clean; in general, the limiting process can be expressed in terms of a Kiefer process, cf. for example [4].

4 Networks

Now that the reader is familiar with the one-dimensional setting, we extend this to networks. In this section, we focus on feedforward networks in which each node corresponds to an infinite-server queue. Feedforward networks are defined as follows.

Definition 4.1 (*feedforward network*) Let $G = (V, E)$ be a directed graph with nodes V and edges E. The nodes represent infinite-server queues and the directed edges between the facilities demonstrate how jobs move through the system. We suppose that there are no cycles in G, i.e., there is no sequence of nodes, starting and ending at the same node, with each two consecutive nodes adjacent to each other in the graph, consistent with the orientation of the edges.

We focus on feedforward networks to keep the notation manageable. In Theorem 4.6, we derive the transform of the numbers of jobs in all nodes, jointly with the shot-noise process(es) for feedforward networks. Nonetheless, we provide Example 4.5, to show that analysis is in fact possible if there is a loop, but at the expense of more involved calculations.

Since all nodes represent infinite-server queues, one can see that whenever a node has multiple input streams, it is equivalent to multiple infinite-server queues that work independently from each other, but have the same service speed and induce the same service requirement for arriving jobs. Consider Fig. 3 for an illustration. The reason why this holds is that different job streams move independently through the system, without creating waiting times for others. Therefore, merging streams do not increase the complexity of our network. The same holds for 'splits' in job streams. By this we mean that after jobs finished their service at a server, they move to server i with probability q_i (with $\sum_i q_i = 1$). Then, one can simply sample the entire path that the job will take through the system, at the arrival instance at its first server.

If one recognizes the above, then all feedforward networks reduce to parallel tandem systems in which the first node in each tandem system is fed by external input. The procedure to decompose a network into parallel tandems consists of finding all paths between nodes in which jobs either enter or leave the system. Each of these paths will subsequently be considered as a tandem queue, which are then set in parallel.
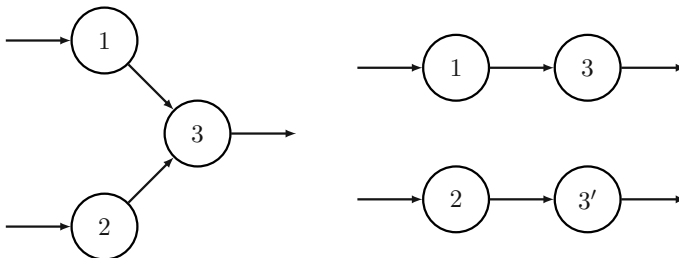

Fig. 3 Since the jobs are not interfering with each other, the network on the *left* is equivalent to the graph on the *right*. Node 3' is a copy of node 3: It works at the same speed and induces the same service requirements

To build up to the main result, we first study tandem systems in Sect. 4.1. Subsequently, we put the tandem systems in parallel in Sect. 4.2 and finally we present the main theorem and some implications in Sect. 4.3.

4.1 Tandem systems

As announced, we proceed by studying tandem systems. In Sect. 4.2 below, we study d parallel tandem systems. In this subsection, we consider the i-th of these tandem systems, where $i = 1, \ldots, d$.

Suppose that tandem i has S_i service facilities and the input process at the first node is Poisson, with a shot-noise arrival rate $\Lambda_i(\cdot)$. We assume that jobs enter node $i1$. When they finish service, they enter node $i2$, etc., until they enter node $i S_i$ after which they leave the system. We use ij as a subscript referring to node j in tandem system i and we refer to the node as node ij. Hence $N_{ij}(t)$ and J_{ij} denote the number of jobs in node ij at time t, and a copy of a service requirement, respectively, where $j = 1, \ldots, S_i$.

Fix some time $t > 0$. Again we derive results by splitting time into intervals of length Δ. Denote by $M_{ij}(k, \Delta)$ the number of jobs present in node ij at time t that have entered node $i1$ between time $k\Delta$ and $(k+1)\Delta$; as we keep t fixed we suppress it in our notation. Because jobs are not interfering with each other in the infinite-server realm, we can decompose the transform of interest:

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)}\right) = \lim_{\Delta \downarrow 0} \prod_{k=0}^{t/\Delta - 1} \mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{M_{ij}(k, \Delta)}\right). \tag{12}$$

Supposing that the arrival rate is a deterministic function of time $\lambda_i(\cdot)$, by conditioning on the number of arrivals in the k-th interval,

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{M_{ij}(k, \Delta)}\right) = \sum_{m=0}^{\infty} e^{-\lambda_i(k\Delta)\Delta} \frac{(\lambda_i(k\Delta)\Delta)^m}{m!} \left(f_i(k\Delta, z)\right)^m$$

$$= \exp\left(\Delta \lambda_i(k\Delta)(f_i(k\Delta, z) - 1)\right),$$

where

$$f_i(u, z) := p_i(u) + \sum_{j=1}^{S_i} z_{ij} p_{ij}(u), \tag{13}$$

in which $p_i(u)$ ($p_{ij}(u)$, respectively) denotes the probability that the job that entered tandem i at time u has already left the tandem (is in node j, respectively) at time t.

Note that

$$p_i(u) = \mathbb{P}\left(\sum_{\ell=1}^{S_i} J_{i\ell} < t - u\right),$$

$$p_{ij}(u) = \mathbb{P}\left(\sum_{\ell=1}^{j-1} J_{i\ell} < t - u, \sum_{\ell=1}^{j} J_{i\ell} > t - u\right). \tag{14}$$

Recognizing a Riemann sum and letting $\Delta \downarrow 0$, we conclude that Eq. (12) takes the following form:

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)}\right) = \exp\left(\int_0^t \lambda_i(u)(f_i(u, z) - 1)\, \mathrm{d}u\right).$$

In case of a stochastic rate process $\Lambda_i(\cdot)$, we obtain

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} \,\middle|\, \Lambda_i(\cdot)\right) = \exp\left(\int_0^t \Lambda_i(u)(f_i(u, z) - 1)\, \mathrm{d}u\right).$$

Therefore, it holds that

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s\Lambda_i(t)}\right) = \mathbb{E}\left(\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s\Lambda_i(t)} \,\middle|\, \Lambda_i(\cdot)\right)\right)$$

$$= \mathbb{E}\left(e^{-s\Lambda_i(t)} \mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} \,\middle|\, \Lambda_i(\cdot)\right)\right),$$

and we consequently find

$$\mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s\Lambda_i(t)}\right) = \mathbb{E}\exp\left(\int_0^t \Lambda_i(u)(f_i(u, z) - 1)\, \mathrm{d}u - s\Lambda_i(t)\right). \tag{15}$$

4.2 Parallel (tandem) systems

Now that the tandem case has been analyzed, the next step is to put the tandem systems as described in Sect. 4.1 in parallel. We assume that there are d parallel tandems. There are different ways in which dependence between the parallel systems can be created. Two relevant models are listed below, and illustrated in Fig. 4.
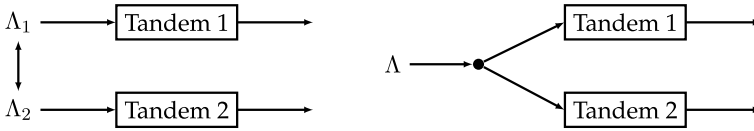
Fig. 4 Model (M1) is illustrated on the *left*, and Model (M2) is illustrated on the *right*. The *rectangles* represent tandem systems, which consist of an arbitrary number of nodes in series

(M1) Let $\Lambda \equiv \Lambda(\cdot)$ be a d-dimensional shot-noise process $(\Lambda_1, \ldots, \Lambda_d)$ where the shots in all Λ_i occur simultaneously (the shot distributions and decay rates may be different). The process Λ_i, for $i = 1, \ldots, d$, corresponds to the arrival rate of tandem system i. Each tandem system has an arrival process, in which the Cox processes are independent given their shot-noise arrival rates.

(M2) Let $\Lambda \equiv \Lambda(\cdot)$ be the shot-noise rate of a Cox process. The corresponding Poisson process generates *simultaneous* arrivals in all tandems.

Remark 4.2 The model in which there is essentially one shot-noise process that generates arrivals for all queues independently is a special case of Model (M1). This can be seen by setting all components of $\Lambda = (\Lambda_1, \ldots, \Lambda_d)$ equal, by letting the shots and decay rate be identical.

In Model (M1), correlation between the shot-noise arrival rates induces correlation between the numbers of jobs in the different queues. In Model (M2), correlation clearly appears because all tandem systems have the same input process. Of course, the tandem systems will not behave identically because the jobs may have different service requirements. In short, correlation across different tandems in Model (M1) is due to linked arrival rates, and correlation in Model (M2) is due to simultaneous arrival epochs. We feel that both versions are relevant, depending on the application, and hence we analyze both.

Analysis of (M1)— Suppose that the dependency is of the type as in Model (M1). This means that the shots, in each component of Λ, occur simultaneously. Recall the definition of f_i as stated in Eq. (13). It holds that

$$\mathbb{E}\left(\prod_{i=1}^{d} e^{-s_i \Lambda_i(t)} \prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} \right) = \mathbb{E}\left(\prod_{i=1}^{d} \mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s_i \Lambda_i(t)} \,\middle|\, \Lambda_i(\cdot) \right) \right)$$

$$= \mathbb{E}\left(\prod_{i=1}^{d} \left(e^{-s_i \Lambda_i(t)} \mathbb{E}\left(\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} \,\middle|\, \Lambda_i(\cdot) \right) \right) \right)$$

$$= \mathbb{E}\exp\left(\sum_{i=1}^{d} \left(\int_0^t \Lambda_i(u)(f_i(u, z) - 1)\, \mathrm{d}u - s_i \Lambda_i(t) \right) \right),$$

(16)

where the last equality holds due to (15).

Analysis of (M2)— Now suppose that the dependency in this model is of type (M2), i.e., there is one shot-noise process that generates simultaneous arrivals in the parallel tandem systems.

First we assume a deterministic arrival rate function $\lambda(\cdot)$. Let $M_{ij}(k, \Delta)$ be the number of jobs present in tandem system i at node j at time t that have arrived in the system between $k\Delta$ and $(k+1)\Delta$. Note that

$$\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)}\right) = \lim_{\Delta\downarrow 0}\prod_{k=0}^{t/\Delta - 1}\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{M_{ij}(k,\Delta)}\right).$$

To further evaluate the right-hand side of the previous display, we observe that we can write

$$\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{M_{ij}(k,\Delta)}\right) = \sum_{m=0}^{\infty} e^{-\lambda(k\Delta)\Delta}\frac{\lambda(k\Delta)\Delta)^m}{m!}(f(k\Delta, z))^m = e^{\Delta\lambda(k\Delta)(f(k\Delta,z)-1)},$$

where

$$f(u, z) := \sum_{j=1}^{d}\sum_{\ell_j=1}^{S_{j+1}} p_{\ell_1,\dots,\ell_d}\prod_{i=1}^{d} z_{i\ell_i}; \tag{17}$$

in this definition $p_{\ell_1,\dots,\ell_d} \equiv p_{\ell_1,\dots,\ell_d}(u)$ equals the probability that a job that arrived at time u in tandem i is in node ℓ_i at time t [cf. Eq. (14)]. The situation that $\ell_i = S_i + 1$ means that the job left the tandem system; we define $z_{i,S_i+1} = 1$.

In a similar fashion as before, we conclude that

$$\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)}e^{-s\Lambda(t)}\right) = \mathbb{E}\exp\left(\int_0^t \Lambda(u)(f(u, z) - 1)\,\mathrm{d}u - s\Lambda(t)\right), \tag{18}$$

with f defined in Eq. (17).

Example 4.3 (Two-node parallel system) In the case of a parallel system of two infinite-server queues, $f(u, z)$ simplifies to

$$f(u, z_{11}, z_{21}) = \sum_{\ell_1=1}^{2}\sum_{\ell_2=1}^{2} z_{1\ell_1} z_{2\ell_2} p_{\ell_1,\ell_2} = z_{11}z_{21}p_{11} + z_{21}p_{21} + z_{11}p_{12} + p_{22}.$$

Remark 4.4 (Routing) Consider a feedforward network with routing. As argued in the beginning of this section, the network can be decomposed as a parallel tandem system. In case there is splitting at some point, then one decomposes the network as a

parallel system, in which each tandem i receives the job with probability q_i, such that $\sum q_i = 1$. This can be incorporated simply by adjusting the probabilities contained in f_i in Eq. (16), which are given in Eq. (14), so that they include the event that the job joined the tandem under consideration. For instance, the expression for $p_i(u)$ in the left equation in (14) would become

$$\mathbb{P}\left(Q = i, \sum_{\ell=1}^{S_i} J_{i\ell} < t - u\right),$$

where Q is a random variable with a generalized Bernoulli (also called 'categorical') distribution, where

$$\mathbb{P}(\text{job is assigned to tandem } i) = \mathbb{P}(Q = i) = q_i, \quad \text{for } i = 1, \ldots d,$$

with $\sum q_i = 1$; the right equation in (14) is adjusted similarly. Other than that the analysis is the same for the case of splits.

Remark 4.5 (Networks with loops) So far we have only considered feedforward networks. Networks with loops can be analyzed as well, but the notation becomes quite cumbersome. To show the method in by which networks with loops and routing can be analyzed, we consider a specific example. Suppose that arrivals enter node one, after which they enter node two. After they have been served in node two, they go back to node one with probability η, or leave the system with probability $1 - \eta$. In this case, with similar techniques as before, we can find

$$\mathbb{E}\, z_1^{N_1(t)} z_2^{N_2(t)} = \exp\left(\int_0^t \Lambda(u)(f(u, z_1, z_2) - 1)\,\mathrm{d}u\right),$$

with

$$f(u, z_1, z_2) = \mathbb{P}(\text{job}(u) \text{ left system}) + \sum_{i=1}^{2} z_i\, \mathbb{P}(\text{job}(u) \text{ is in node } i),$$

in which job(u) is the job that arrived at time u and we are examining the system at time t. Now, if we denote service times in the j-th node by $J^{(j)}$, then, at a specific time t,

$$\mathbb{P}(\text{job}(u) \text{ left system}) = \sum_{k=0}^{\infty} \mathbb{P}\left(\sum_{i=1}^{k+1}(J_i^{(1)} + J_i^{(2)}) \leq t - u\right) \eta^k (1 - \eta).$$

Analogously, $\mathbb{P}(\text{job}(u)$ is in node $1)$ equals, by conditioning on the job having taken k loops,

$$\sum_{k=0}^{\infty} \eta^k \, \mathbb{P}\left(J_{k+1}^{(1)} + \sum_{i=1}^{k} \left(J_i^{(1)} + J_i^{(2)} \right) > t - u, \sum_{i=1}^{k} \left(J_i^{(1)} + J_i^{(2)} \right) \leq t - u \right);$$

likewise, $\mathbb{P}(\text{job}(u)\text{ is in node } 2)$ equals

$$\sum_{k=0}^{\infty} \eta^k \, \mathbb{P}\left(\sum_{i=1}^{k+1} \left(J_i^{(1)} + J_i^{(2)} \right) > t - u, \ J_{k+1}^{(1)} + \sum_{i=1}^{k} \left(J_i^{(1)} + J_i^{(2)} \right) \leq t - u \right).$$

For example, in the case where all $J_i^{(j)}$ are independent and exponentially distributed with mean $1/\mu$, we can calculate those probabilities explicitly. Indeed, if we denote by Y a Poisson process with rate μ, then, for example

$$\mathbb{P}\left(\sum_{i=1}^{k+1} \left(J_i^{(1)} + J_i^{(2)} \right) > t - u, \ J_{k+1}^{(1)} + \sum_{i=1}^{k} \left(J_i^{(1)} + J_i^{(2)} \right) \leq t - u \right)$$
$$= \mathbb{P}(Y(t - u) = 2k + 1)$$
$$= e^{-\mu(t-u)} \frac{(\mu(t - u))^{2k+1}}{(2k + 1)!}$$

and thus

$$\mathbb{P}(\text{job}(u)\text{ is in node } 2) = \sum_{m=0}^{\infty} \eta^m e^{-\mu(t-u)} \frac{(\mu(t - u))^{2m+1}}{(2m + 1)!}.$$

A similar calculation can be done for the probability that the job is in node one. Recalling that a sum of exponentials has a Gamma distribution, we can write

$$f(u, z_1, z_2) = z_1 \sum_{m=0}^{\infty} \eta^m e^{-\mu(t-u)} \frac{(\mu(t - u))^{2m}}{(2m)!} + z_2 \sum_{m=0}^{\infty} \eta^m e^{-\mu(t-u)} \frac{(\mu(t - u))^{2m+1}}{(2m + 1)!}$$
$$+ \sum_{m=0}^{\infty} \eta^m (1 - \eta) F_{\Gamma(2m+2,\mu)}(t - u)$$
$$= z_1 e^{-\mu(t-u)} \cosh\left(\mu\sqrt{\eta}(t - u) \right) + z_2 \frac{e^{-\mu(t-u)}}{\sqrt{\eta}} \sinh\left(\mu\sqrt{\eta}(t - u) \right)$$
$$+ (1 - \eta) \sum_{m=0}^{\infty} \eta^m F_{\Gamma(2m+2,\mu)}(t - u),$$

where $F_{\Gamma(2m+2,\mu)}$ denotes the distribution function of a Γ-distributed random variable with rate μ and shape parameter $2m + 2$.

4.3 Main result

In this subsection, we summarize and conclude with the following main result. Recall Definition 4.1 of a feedforward network. In the beginning of Sect. 4 we argued that we can decompose a feedforward network into parallel tandems. In Sect. 4.2 we studied exactly those systems, leading up to the following results.

Theorem 4.6 *Suppose we have a feedforward network of infinite-server queues, where the input process is a Poisson process with shot-noise arrival rate. Then the network can be decomposed into parallel tandem systems. In Model (M1), it holds that*

$$\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s\Lambda_i(t)}\right) = \mathbb{E}\exp\left(\sum_{i=1}^{d}\left(\int_0^t \Lambda_i(u)(f_i(u,z)-1)\,\mathrm{d}u - s_i\Lambda_i(t)\right)\right)$$

$$= \exp\left(v\int_0^t \left(\beta(g(s,v))-1\right)\mathrm{d}v\right),$$

with $f_i(\cdot,\cdot)$ as defined in Eq. (13) and where $g(s,v)$ is a vector-valued function in which component i is given by

$$s_i e^{-r_i(t-v)} - e^{r_i v}\int_v^t (f_i(u,z)-1)e^{-r_i u}\,\mathrm{d}u.$$

Furthermore, in Model (M2),

$$\mathbb{E}\left(\prod_{i=1}^{d}\prod_{j=1}^{S_i} z_{ij}^{N_{ij}(t)} e^{-s\Lambda(t)}\right) = \mathbb{E}\exp\left(\int_0^t \Lambda(u)(f(u,z)-1)\,\mathrm{d}u - s\Lambda(t)\right)$$

$$= \exp\left(v\int_0^t \left(\beta\left(se^{-r(t-v)} - e^{rv}\int_v^t (f(u,z)-1)e^{-ru}\,\mathrm{d}u\right)-1\right)\mathrm{d}v\right),$$

with $f(\cdot,\cdot)$ as defined in Eq. (17).

Proof These are Eqs. (16) and (18) to which we applied Lemma 2.1. $\qquad\square$

Next we calculate covariances between nodes in tandem and parallel thereafter.

Covariance in Tandem System— Consider a tandem system consisting of two nodes and we want to analyze the covariance between the number of jobs in the nodes. Dropping the index of the tandem system, denote by $N_1(\cdot)$ and $N_2(\cdot)$ the number of jobs in nodes 1 and 2, respectively. Using Eq. (16), differentiation yields

$$\mathbb{E}\,N_2(t) = \int_0^t \mathbb{P}(J_1 < t-u, J_1+J_2 > t-u)\,\mathbb{E}\,\Lambda(u)\,\mathrm{d}u$$

and

$$
\begin{aligned}
&\mathbb{E}\, N_1(t)N_2(t) \\
&= \mathbb{E}\left(\int_0^t \mathbb{P}(J_1 < t-u,\, J_1+J_2 > t-u)\Lambda(u)\,\mathrm{d}u \int_0^t \mathbb{P}(J_1 > t-v)\Lambda(v)\,\mathrm{d}v\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}(N_1(t), N_2(t)) \\
&= \mathbb{C}\mathrm{ov}\left(\int_0^t \mathbb{P}(J_1 < t-u,\, J_1+J_2 > t-u)\Lambda(u)\,\mathrm{d}u,\, \int_0^t \mathbb{P}(J_1 > t-v)\Lambda(v)\,\mathrm{d}v\right) \\
&= 2\int_0^t \int_v^t \mathbb{P}(J_1 < t-u,\, J_1+J_2 > t-u)\,\mathbb{P}(J_1 > t-v)\,\mathbb{C}\mathrm{ov}(\Lambda(u), \Lambda(v))\,\mathrm{d}u\,\mathrm{d}v \\
&= 2\int_0^t \int_v^t \mathbb{P}(J_1 < t-u,\, J_1+J_2 > t-u)\,\mathbb{P}(J_1 > t-v)e^{-r(u-v)}\,\mathbb{V}\mathrm{ar}\,\Lambda(v)\,\mathrm{d}u\,\mathrm{d}v,
\end{aligned}
$$

cf. Eq. (2) for the last equality.

Covariance parallel (M1)— Consider a parallel system consisting of two nodes only. In order to study covariance in the parallel (M1) case, we need a result about the covariance of the corresponding shot-noise process.

Lemma 4.7 *Let $\Lambda_1(\cdot)$, $\Lambda_2(\cdot)$ be shot-noise processes of which the jumps occur simultaneously according to a Poisson arrival process with rate v. Let the decay be exponential with rate r_1, r_2, respectively. Then it holds that, for $\delta > 0$,*

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}(\Lambda_1(t), \Lambda_2(t+\delta)) &= e^{-r_2\delta}\,\mathbb{C}\mathrm{ov}(\Lambda_1(t), \Lambda_2(t)) \\
&= e^{-r_2\delta}\frac{v\,\mathbb{E}\,B_{11}B_{12}}{r_1+r_2}(1 - e^{-(r_1+r_2)t}), \quad (19)
\end{aligned}
$$

which, in the case $\Lambda_1 = \Lambda_2$, reduces to

$$
\mathbb{C}\mathrm{ov}(\Lambda_i(t), \Lambda_i(t+\delta)) = e^{-r_i\delta}\,\mathbb{V}\mathrm{ar}\,\Lambda_i(t), \quad for\, i = 1, 2,
$$

corresponding to [16, p. 394] and Eq. (2).

Proof See Appendix 2. $\qquad\square$

By making use of Eq. (16), we find

$$
\mathbb{E}\, N_1(t)N_2(t) = \mathbb{E}\left(\int_0^t \Lambda_1(u)\,\mathbb{P}(J_1 > t-u)\,\mathrm{d}u \int_0^t \Lambda_2(v)\,\mathbb{P}(J_2 > t-v)\,\mathrm{d}v\right).
$$

This implies

$$\mathbb{C}\mathrm{ov}(N_1(t), N_2(t)) = \mathbb{C}\mathrm{ov}\left(\int_0^t \Lambda_1(u)\,\mathbb{P}(J_1 > t - u)\,\mathrm{d}u, \int_0^t \Lambda_2(v)\,\mathbb{P}(J_2 > t - v)\,\mathrm{d}v\right)$$

$$= 2\int_0^t \int_v^t \mathbb{C}\mathrm{ov}(\Lambda_1(u), \Lambda_2(v))\,\mathbb{P}(J_1 > t - u)\,\mathbb{P}(J_2 > t - v)\,\mathrm{d}u\,\mathrm{d}v$$

$$= 2\int_0^t \int_v^t \frac{v\,\mathbb{E}\,B_{11}B_{12}}{r_1 + r_2}\left(1 - e^{-(r_1+r_2)v}\right)e^{-r_2(u-v)}\,\mathbb{P}(J_1 > t - u)\,\mathbb{P}(J_2 > t - v)\,\mathrm{d}u\,\mathrm{d}v,$$

where we made use of the fact that, for $u \geq v$,

$$\mathbb{C}\mathrm{ov}(\Lambda_1(u), \Lambda_2(v)) = \frac{v\,\mathbb{E}\,B_{11}B_{12}}{r_1 + r_2}\left(1 - e^{-(r_1+r_2)v}\right)e^{-r_2(u-v)},$$

cf. Lemma 4.7.

Covariance parallel (M2)— Extracting the mixed moment from the transform in Eq. (18), we derive directly that

$$\mathbb{E}\,N_1(t)N_2(t) = \mathbb{E}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J_1 > t - u, J_2 > t - u)\,\mathrm{d}u\right)$$

$$+ \mathbb{E}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J_1 > t - u)\,\mathrm{d}u \int_0^t \Lambda(u)\,\mathbb{P}(J_2 > t - u)\,\mathrm{d}u\right).$$

This implies

$$\mathbb{C}\mathrm{ov}(N_1(t), N_2(t)) = \mathbb{C}\mathrm{ov}\left(\int_0^t \Lambda(u)\,\mathbb{P}(J_1 > t - u)\,\mathrm{d}u, \int_0^t \Lambda(u)\,\mathbb{P}(J_2 > t - u)\,\mathrm{d}u\right)$$

$$+ \int_0^t \mathbb{E}\,\Lambda(u)\,\mathbb{P}(J_1 > t - u, J_2 > t - u)\,\mathrm{d}u.$$

The following proposition compares the correlations present in Model (M1) and (M2). In the proposition, we refer to the number of jobs in queue j in Model (Mi) at time t as $N_j^{(i)}(t)$, for $i = 1, 2$. We find the anticipated result that the correlation in Model (M2) is stronger than in Model (M1).

Proposition 4.8 *Let $\Lambda(\cdot)$ be the shot-noise process that generates simultaneous arrivals in both queues and let $\Lambda_1(\cdot), \Lambda_2(\cdot)$ be processes that have simultaneous jumps and generate arrivals in both queues independently. Suppose that $\Lambda_1(t) \overset{\mathrm{d}}{=} \Lambda_2(t) \overset{\mathrm{d}}{=} \Lambda(t)$, for $t \geq 0$. Then, for any $t \geq 0$,*

$$\mathbb{C}\mathrm{orr}\left(N_1^{(1)}(t), N_2^{(1)}(t)\right) \leq \mathbb{C}\mathrm{orr}\left(N_1^{(2)}(t), N_2^{(2)}(t)\right).$$

Proof Because of the assumption $\Lambda_1(t) \stackrel{d}{=} \Lambda_2(t) \stackrel{d}{=} \Lambda(t)$, we have that, for all combinations $i, j \in \{1, 2\}$, the $N_i^{(j)}(t)$ are equal in distribution. Therefore, it is sufficient to show that

$$\mathbb{Cov}\left(N_1^{(1)}(t), N_2^{(1)}(t)\right) \leq \mathbb{Cov}\left(N_1^{(2)}(t), N_2^{(2)}(t)\right).$$

The expressions for the covariances, which are derived earlier in this section, imply that

$$\mathbb{Cov}\left(N_1^{(2)}(t), N_2^{(2)}(t)\right) - \mathbb{Cov}\left(N_1^{(1)}(t), N_2^{(1)}(t)\right)$$
$$= \mathbb{E}\int_0^t \Lambda(u)\,\mathbb{P}(J_1 > t - u, J_2 > t - u)\,\mathrm{d}u,$$

which is nonnegative, as desired. □

5 Concluding remarks

We have considered networks of infinite-server queues with shot-noise-driven Coxian input processes. For the single queue, we found explicit expressions for the Laplace transform of the joint distribution of the number of jobs and the driving shot-noise arrival rate, as well as a functional central limit theorem of the number of jobs in the system under a particular scaling. The results were then extended to a network context: we derived an expression for the joint transform of the numbers of jobs in the individual queues, jointly with the values of the driving shot-noise processes.

We included the functional central limit theorem for the single queue, but it is anticipated that a similar setup carries over to the network context, albeit at the expense of considerably more involved notation. Our future research will include the study of the departure process of a single queue; the output stream should remain Coxian, but of another type than the input process.

Appendix 1: Proof of Lemma 2.1

There are various ways to prove this result; we here include a procedure that intensively relies on the probabilistic properties of the shot-noise process involved. Observe that,

recognizing a Riemann sum,

$$\int_0^t f(u, z) \Lambda(u) \, du = \lim_{\Delta \downarrow 0} \Delta \sum_{k=1}^{t/\Delta} f(k\Delta, z) \Lambda(k\Delta). \tag{20}$$

With $P_B(t)$ a Poisson process with rate ν and the U_i i.i.d. samples from a uniform distribution on $[0, 1]$, it holds that

$$\Lambda(k\Delta) = \sum_{\ell=1}^k \sum_{i=P_B((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i e^{-r\Delta U_i} e^{-r(k-\ell)\Delta}.$$

We thus obtain that the expression in (20) equals (where the equality follows by interchanging the order of the summations)

$$\lim_{\Delta \downarrow 0} \Delta \sum_{k=1}^{t/\Delta} f(k\Delta, z) \sum_{\ell=1}^k \sum_{i=P_B((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i e^{-r\Delta U_i} e^{-r(k-\ell)\Delta}$$

$$= \lim_{\Delta \downarrow 0} \Delta \sum_{\ell=1}^{t/\Delta} \sum_{i=P_B((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i e^{-r\Delta U_i} \sum_{k=\ell}^{t/\Delta} f(k\Delta, z) e^{-r(k-\ell)\Delta},$$

which behaves as

$$\lim_{\Delta \downarrow 0} \sum_{\ell=1}^{t/\Delta} e^{r\ell\Delta} \int_{\ell\Delta}^t f(u, z) e^{-ru} \, du \sum_{i=P_B((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i e^{-r\Delta U_i}.$$

Furthermore, we have the representation

$$\Lambda(t) = \lim_{\Delta \downarrow 0} \sum_{\ell=1}^{t/\Delta} \sum_{i=P((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i e^{-r(t-\ell\Delta+\Delta U_i)}.$$

We conclude that $\mathbb{E} z^{N(t)} e^{-s\Lambda(t)}$ equals

$$\lim_{\Delta \downarrow 0} \mathbb{E} \exp\left(\sum_{\ell=1}^{t/\Delta} \sum_{i=P_B((\ell-1)\Delta)+1}^{P_B(\ell\Delta)} B_i \left(e^{-r\Delta U_i} e^{r\ell\Delta} \int_{\ell\Delta}^t f(u, z) e^{-ru} \, du - s e^{-r(t-\ell\Delta+\Delta U_i)} \right) \right). \tag{21}$$

Conditioning on the values of $P_B(\ell\Delta) - P_B((\ell-1)\Delta)$, for $\ell = 1, \ldots, t/\Delta$, and using that the B_i are i.i.d., we find that the expression in Eq. (21)) is equal to the limit, as $\Delta \downarrow 0$, of the following expression

$$\prod_{\ell=1}^{t/\Delta} e^{-\nu\Delta} \sum_{k_\ell=0}^{\infty} \frac{(\nu\Delta)^{k_\ell}}{k_\ell!} \left(\mathbb{E}\exp\left(B_1\left(e^{-r\Delta U_i} e^{r\ell\Delta} \int_{\ell\Delta}^{t} f(u,z) e^{-ru}\,\mathrm{d}u - s e^{-r(t-\ell\Delta+\Delta U_i)}\right)\right)\right)^{k_\ell},$$

which, in the limit $\Delta \downarrow 0$, equals

$$\prod_{\ell=1}^{t/\Delta} e^{-\nu\Delta} \exp\left(\nu\Delta\, \mathbb{E}\exp\left(B_1\left(e^{-r\Delta U_i} e^{r\ell\Delta} \int_{\ell\Delta}^{t} f(u,z) e^{-ru}\,\mathrm{d}u - s e^{-r(t-\ell\Delta+\Delta U_i)}\right)\right)\right),$$

which can be written as

$$\lim_{\Delta\downarrow 0}\exp\left(\nu\Delta \sum_{\ell=1}^{t/\Delta} \left(\beta\left(s e^{-r(t-\ell\Delta+\Delta U_i)} - e^{-r\Delta U_i} e^{r\ell\Delta} \int_{\ell\Delta}^{t} f(u,z) e^{-ru}\,\mathrm{d}u\right) - 1\right)\right).$$

The lemma now follows from continuity of the exponent and the definition of the Riemann integral.

Appendix 2: Proof of Lemma 4.7

Let $P_B(\cdot)$ be the Poisson process with rate ν, corresponding to the occurrences of shots, and let $\mathcal{E}_{t,\delta}(n)$ be the event that $P_B(t+\delta) - P_B(t) = n$. By conditioning on the number of shots in the interval $(t, t+\delta]$, we find

$$\mathbb{E}\,\Lambda_1(t)\Lambda_2(t+\delta) = \sum_{n=0}^{\infty} \mathbb{E}(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{E}_{t,\delta}(n))\, \mathbb{P}(\mathcal{E}_{t,\delta}(n))$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{E}_{t,\delta}(n))\, \frac{(\delta\nu)^n}{n!} e^{-\nu\delta}.$$

We proceed by rewriting the conditional expectation as

$$\mathbb{E}(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{E}_{t,\delta}(n)) = \frac{1}{\delta^n} \int_t^{t+\delta} \cdots \int_t^{t+\delta} \mathbb{E}(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{F}_{t_1,\dots,t_n,\delta}(n))\,\mathrm{d}t_1\dots\mathrm{d}t_n,$$

denoting by $\mathcal{F}_{t_1,\dots,t_n,\delta}(n)$ the event $\mathcal{E}_{t,\delta}(n)$ and the arrival epochs are $t_1, \dots, t_n$. Note that we have, due to Eq. (1), conditional on $\mathcal{F}_{t_1,\dots,t_n,\delta}(n)$, the distributional equality

$$\Lambda_2(t+\delta) = \Lambda_2(t) e^{-r_2\delta} + \sum_{i=1}^{n} B_{i2} e^{-r_2(t+\delta-t_i)}, \tag{22}$$

and consequently

$$\mathbb{E}\left(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{F}_{t_1,\ldots,t_n,\delta}(n)\right)$$

$$= \mathbb{E}\,\Lambda_1(t)\Lambda_2(t)e^{-r_2\delta} + \mathbb{E}\,\Lambda_1(t)\sum_{i=1}^{n}\mathbb{E}\,B_{i2}e^{-r_2(t+\delta-t_i)}. \tag{23}$$

Note that for all $i = 1, \ldots, n$ we have

$$\int_{t}^{t+\delta}\cdots\int_{t}^{t+\delta}\int_{t}^{t+\delta} e^{-r_2(t+\delta-t_i)}\,\mathrm{d}t_1\,\mathrm{d}t_2\ldots\mathrm{d}t_n = \frac{1}{r_2}(1-e^{-r_2\delta})\delta^{n-1}. \tag{24}$$

After unconditioning Eq. (23) with respect to the arrival epochs by integrating over all t_i from t to $t + \delta$ and dividing by δ^n, we thus obtain

$$\mathbb{E}(\Lambda_1(t)\Lambda_2(t+\delta) \mid \mathcal{E}_{t,\delta}(n)) = \mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2(t)e^{-r_2\delta} + \mathbb{E}\,\Lambda_1(t)\frac{1}{r_2\delta}(1-e^{-r_2\delta})n\,\mathbb{E}\,B_{12}$$

and hence, denoting $\Lambda_i := \lim_{t\to\infty}\Lambda_i(t)$ for $i = 1, 2$,

$$\mathbb{E}\,\Lambda_1(t)\Lambda_2(t+\delta) = \sum_{n=0}^{\infty}\left(\mathbb{E}\,\Lambda_1(t)\Lambda_2(t)e^{-r_2\delta} + \mathbb{E}\,\Lambda_1(t)\frac{1}{r_2\delta}(1-e^{-r_2\delta})n\,\mathbb{E}\,B_{12}\right)\frac{(\delta\nu)^n}{n!}e^{-\nu\delta}$$

$$= \mathbb{E}\,\Lambda_1(t)\Lambda_2(t)e^{-r_2\delta} + (1-e^{-r_2\delta})\,\mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2$$

$$= \mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2 + e^{-r_2\delta}\left(\mathbb{E}\,\Lambda_1(t)\Lambda_2(t) - \mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2\right),$$

where we made use of $\mathbb{E}\,\Lambda_i = \nu\,\mathbb{E}\,B_{1i}/r_i$. It follows that

$$\mathbb{C}\mathrm{ov}(\Lambda_1(t), \Lambda_2(t+\delta)) = \mathbb{E}\,\Lambda_1(t)\Lambda_2(t+\delta) - \mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2(t+\delta)$$

$$= \mathbb{E}\,\Lambda_1(t)\Lambda_2(t)e^{-r_2\delta} + (1-e^{-r_2\delta})\,\mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2$$

$$- \mathbb{E}\,\Lambda_1(t)\left(\mathbb{E}\,\Lambda_2(t)e^{-r_2\delta} + (1-e^{-r_2\delta})\,\mathbb{E}\,\Lambda_2\right),$$

where the equality $\mathbb{E}\,\Lambda_2(t+\delta) = \mathbb{E}\,\Lambda_2(t)e^{-r_2\delta} + (1-e^{-r_2\delta})\,\mathbb{E}\,\Lambda_2$ is used, which can be directly checked using the expressions for the mean in Eq. (2). This proves the first equality in Eq. (19). The proof of the second equality follows from $\mathbb{C}\mathrm{ov}(\Lambda_1(t), \Lambda_2(t)) = \mathbb{E}\,\Lambda_1(t)\Lambda_2(t) - \mathbb{E}\,\Lambda_1(t)\,\mathbb{E}\,\Lambda_2(t)$, in which

$$\mathbb{E}\,\Lambda_1(t)\Lambda_2(t) = \mathbb{E}\left[\left(\sum_{i=1}^{N(t)}B_{i1}e^{-r_1(t-U_i)}\right)\left(\sum_{j=1}^{N(t)}B_{j2}e^{-r_2(t-U_j)}\right)\right]$$

$$= \sum_{n=0}^{\infty}e^{-\nu t}\frac{(\nu t)^n}{n!}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(B_{i1}B_{j2}e^{-r_1(t-U_i)}e^{-r_2(t-U_j)})$$

$$= \sum_{n=0}^{\infty}e^{-\nu t}\frac{(\nu t)^n}{n!}\left(\frac{n}{t}\,\mathbb{E}(B_{11}B_{12})\int_{0}^{t}e^{-(r_1+r_2)(t-u)}\,\mathrm{d}u\right)$$

$$+ \mathbb{E} B_{11} \mathbb{E} B_{12} \frac{n(n-1)}{t^2} \int_0^t e^{-r_1(t-u)} \, \mathrm{d}u \int_0^t e^{-r_2(t-v)} \, \mathrm{d}v \Big)$$

$$= \frac{\nu^2 \mathbb{E} B_{11} \mathbb{E} B_{12}(1-e^{-r_1 t})(1-e^{-r_2 t})}{r_1 r_2} + \frac{\nu \mathbb{E} B_{11} B_{12}}{r_1 + r_2}(1-e^{-(r_1+r_2)t}),$$

and $\mathbb{E} \Lambda_i(t)$, for $i = 1, 2$, is given in Eq. (2).

References

1. Anderson, D., Blom, J., Mandjes, M., Thorsdottir, H., de Turck, K.: A functional central limit theorem for a Markov-modulated infinite-server queue. Method. Comput. Appl. Probab. **18**, 153–168 (2016)
2. Bar-Lev, S., Boxma, O., Mathijsen, B., Perry, D.: A blood bank model with perishable blood and impatience. Eurandom reports (2015)
3. Centanni, S., Minozzo, M.: A Monte Carlo approach to filtering for a class of marked doubly stochastic Poisson processes. J. Am. Stat. Assoc. **101**(476), 1582–1597 (2006)
4. Csörgő, M., Révész, P.: Strong Approximations in Probability and Statistics. Academic Press, New York (1981)
5. Dassios, A., Jang, J.-W.: Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. J. Financ. Stoch. **7**, 73–95 (2003)
6. Eick, S., Massey, W., Whitt, W.: The physics of the $M_t/G/\infty$ queue. Manag. Sci. **39**(2), 241–252 (1993)
7. Kurtz, T., Ethier, S.: Markov Processes: Characterization and Convergence. Wiley, New York (1986)
8. Gao, X., Zhu, L.: A functional central limit theorem for stationary Hawkes processes and its application to infinite-server queues. (2016). https://arxiv.org/abs/1607.06624
9. Hawkes, A.G.: Point spectra of some mutually exciting point processes. J. R. Stat. Soc. **33**, 438–443 (1971)
10. Iksanov, A., Jurek, Z.: Shot noise distributions and selfdecomposability. Stoch. Anal. Appl. **21**(3), 593–609 (2003)
11. Kim, S.-H., Whitt, W.: Are call center and hospital arrivals well modeled by nonhomogenous Poisson processes? Manuf. Serv. Oper. Manag. **16**(3), 464–480 (2014)
12. Lewis, P.A.: Simulation of non-homogenous Poisson processes by thinning. Naval Res. Logist. Q. **26**(3), 403–413 (1979)
13. Ogata, Y.: Statistical models for earthquake occurences and residual analysis for point processes. J. Am. Stat. Assoc. **83**, 9–27 (1988)
14. Pang, G., Talreja, R., Whitt, W.: Martingale proofs of many-server heavy-traffic limits for Markovian queues. Probab. Surv. **4**, 193–267 (2007)
15. Pang, G., Whitt, W.: Two-parameter heavy-traffic limits for infinite-server queues. Queueing Syst. **65**, 325–364 (2010)
16. Ross, S.M.: Stochastic Processes. Wiley, New York (1996)
17. Whitt, W.: Proofs of the Martingale FCLT. Probab. Surv. **4**, 268–302 (2007)