

Heavy traffic approximations of a queue with varying service rates and general arrivals ^{*†‡}

Robert Buche
North Carolina State University

Arka P. Ghosh
Iowa State University

Vladas Pipiras
University of North Carolina

July 9, 2009

Abstract

Heavy traffic limit theorems are established for a class of single server queueing models including those with heavy-tailed or long-range dependent arrivals and time-varying service rates. The models are motivated by wireless queueing systems for which there is an increasing evidence of the presence of heavy-tailed or long range dependent arrivals, and where the service rates vary with the changes in the wireless medium. Since changes in the wireless medium occur slowly, the service rate is assumed to vary slower than that of the arrivals.

The main focus of the paper is to obtain the different possible limit forms that can arise depending on the relationship between established scalings for both the arrival and departure processes. The limit forms obtained here are driven by either Brownian motion (when the contribution from the departure process dominates the limit) or the limits of properly scaled arrivals (when the contribution from the arrival process dominates the limit), typical examples being stable Lévy motion or fractional Brownian motion. In particular, for the case where arrival process is given by the infinite source Poisson process, this relationship, which determines the type of the limiting queue-length process, is a simple condition involving the heavy tail exponent, arrival rate and channel variation parameter in the wireless medium model. The limit forms are also affected by the assumption of a slower varying service rate.

To establish these limit results, two approaches are studied. In one approach, when the limit is driven by Brownian motion, the perturbed test function method is extended to incorporate reflection. In contrast, the second approach allows for non-Markovian driving processes in the limit. Both approaches involve averaging in the drift term arising from random service rates at the departures. In the second approach, this averaging is carried out directly and pathwise, thus sidestepping the assumption of driving Brownian motion used in the perturbed test function method. In the case of the infinite source Poisson model, when the limit is stable Lévy motion, it is also argued how the limit form can be deduced using infinitesimal operators.

*The authors were supported in part by the NSF grants DMS-0608669 at North Carolina State University, DMS-0608634 at Iowa State University, and DMS-0608663 at University of North Carolina at Chapel Hill.

[†]*AMS Subject classification.* Primary: 60F05, 60K25. Secondary: 60G18, 60G52, 90B18.

[‡]*Keywords and phrases:* heavy traffic limit theorems, long range dependence, heavy tails, fractional Brownian motion, stable Lévy motion, Brownian motion, averaging, perturbed test function method, reflection, wireless queueing systems.

1 Introduction

We discuss heavy traffic approximations of a queue with random service rates and general arrivals, with the focus on arrivals exhibiting heavy tails or long range dependence features. The basic model is briefly described as follows (see Section 2 for more detail). Arrivals to a single queue are characterized by a sequence of embedded processes denoted $A^\epsilon(t)$, $t \geq 0$, having a mean workload of $\beta_\epsilon \mu \epsilon^{-1} t$, where μ is a constant and β_ϵ denotes some normalization (or state-space scaling) to be specified later for various cases and $\epsilon \rightarrow 0$. We are interested in $A^\epsilon(t)$ exhibiting heavy tails (Resnick (2007), Samorodnitsky and Taqu (1994)) or long range dependence features (Doukhan, Oppenheim and Taqu (2003)). For example, $A^\epsilon(t)$ can be defined through a popular infinite source Poisson arrival model (Mikosch, Resnick, Rootzen and Stegeman (2002)) with heavy tailed “ON” times associated with each arrival. The corresponding departure sequence is denoted $D^\epsilon(t)$ and involves random service rates governed by an exogenous process $L^\epsilon(t)$, $t \geq 0$, which is assumed to take on a finite number of states indexed by $j \in J$. The variation of the process $L^\epsilon(t)$ is assumed “slower” than $A^\epsilon(t)$ ($= A(\epsilon^{-1}t)$) through specifying $L^\epsilon(t) = L(\epsilon^{-\nu}t)$ where $0 < \nu < 1$. We call ν the variation parameter. In the considered model, the service rate is given by $\epsilon^{-1}(r(j) + u^\epsilon(x^\epsilon(t), j))$, where u^ϵ allows for small deviations from $\epsilon^{-1}r(j)$ and depends on the queue length $x^\epsilon(t)$, and $r(j)$ denotes an average rate. Heavy traffic conditions refer to the balance in means between arrivals and departures, and $u^\epsilon(x, j)$ being small in a suitable sense (see Section 2.3, where heavy traffic conditions are described).

Under heavy traffic and other suitable assumptions (including the application of a state-space scaling β_ϵ), it is natural to expect that the queue-length process $x^\epsilon(t)$ can be approximated by a stochastic differential equation limit model, obtained as $\epsilon \rightarrow 0$. As with many examples of heavy traffic analysis, one expects that these approximations are simpler to analyze, serve as good models for physical systems even under conditions away from heavy traffic and are useful for control policy analysis (see, for example, Kushner (2001), Whitt (2002)). In the framework here, a range of limit models can be obtained depending on the interplay of normalizations for the arrival and departure processes. For example, in the case of infinite source Poisson arrival model, this translates into simple relationships among the heavy tail exponent α , source rate λ defined in Section 2, and variation parameter ν . In general, the possible queue limit models are driven by Brownian motion (Bm, in short) as the limit of the departure process, or the limit process arising from the arrival process. Common limits of the arrival processes, under heavy tails or long range dependence, are stable Lévy motion (sLm, in short) and fractional Brownian motion (fBm, in short). sLm and fBm are non-Brownian like, and fBm is non-Markovian.

In more technical terms, the queue-length process $x^\epsilon(t)$ in the prelimit can be written in the general form

$$x^\epsilon(t) = x^\epsilon(0) + \int_0^t G(x^\epsilon(s), L^\epsilon(s)) ds + \eta^\epsilon(t) + z^\epsilon(t), \quad (1.1)$$

where the integral represents a drift component, η^ϵ represents stochastic variations from a mean process, and z^ϵ is a reflection term ensuring that the queue length is nonnegative. In the scaled system (1.1), the process L^ϵ in the drift term operates on a faster time scale and can be “averaged out” in the limit as $\epsilon \rightarrow 0$. We therefore expect that, under suitable assumptions, $x^\epsilon(t)$ converges weakly to the limit $x(t)$ satisfying

$$x(t) = x(0) + \int_0^t \bar{G}(x(s)) ds + \eta(t) + z(t), \quad (1.2)$$

where $\eta(t)$ is the weak limit of $\eta^\epsilon(t)$, \overline{G} is a suitable average of G and $z(t)$ is a reflection term.

The convergence of systems of the type (1.1) to (1.2) has been widely studied in the cases where $\eta(t)$ is Brownian motion (more generally, Brownian diffusion) and there is no reflection term. The techniques include the so-called perturbed test function method, methods based on Markov structures or occupation measures (see Kushner (1984), Chapter 12 in Ethier and Kurtz (1986), Kurtz (1992), Section 6 in Fouque, Garnier, Papanicolaou and Sølna (2007), Section 5 in Skorokhod, Hoppensteadt and Salehi (2002), and numerous references and historical perspectives therein). In this work, we depart from previous results by considering non-Brownian-like limits $\eta(t)$ and incorporating an elementary (one-dimensional) form of reflection. For this, it is convenient to write the queue-length process in (1.1) as

$$\begin{cases} x^\epsilon(t) = (\Gamma y^\epsilon)(t), \\ y^\epsilon(t) = y^\epsilon(0) + \int_0^t G((\Gamma y^\epsilon)(s), L^\epsilon(s)) ds + \eta^\epsilon(t), \end{cases} \quad (1.3)$$

where Γ denotes the usual Skorokhod map on $[0, \infty)$ (Whitt (2002)), and the expected limit as

$$\begin{cases} x(t) = (\Gamma y)(t), \\ y(t) = y(0) + \int_0^t \overline{G}((\Gamma y)(s)) ds + \eta(t). \end{cases} \quad (1.4)$$

The main results of the paper are in the following directions:

- First, as mentioned above, the driving process in the limit is determined by the relationship between the normalizations $\beta_{a,\epsilon}$ and $\beta_{b,\epsilon}$ for the arrival and departure processes, respectively. (We use the notation $\beta_{b,\epsilon}$ instead of $\beta_{d,\epsilon}$ since we consider normalizations of the departure processes leading to Brownian motion limits.) For example, if $\beta_{b,\epsilon} \ll \beta_{a,\epsilon}$ (i.e., $\beta_{a,\epsilon}/\beta_{b,\epsilon} \rightarrow \infty$ as $\epsilon \rightarrow 0$), using the state space scaling $\beta_\epsilon = \beta_{b,\epsilon}$ leads to Bm as the driving process in the limit. The relationship is illustrated, in particular, for the infinite source Poisson arrival model where surprisingly simple relationship among heavy tail exponent α , source rate λ and variation parameter ν is obtained for various (namely, sLm, fBm or Bm) limits. When the function $u^\epsilon(x, j)$ above does not depend on j , similar conditions were derived in Buche, Ghosh and Pipiras (2007).
- Second, we extend a previous approach based on perturbed test functions (Kushner (1984) and others) by incorporating the reflection term. This presently works for

$$\eta^\epsilon(t) = \epsilon^{-\nu/2} \int_0^t F(L^\epsilon(s)) ds \quad (1.5)$$

in (1.1) with $EF(L^\epsilon(s)) = 0$ and Bm $\eta(t)$ in the limit (1.2). The basic idea is to use the representation (1.3) above, and extend earlier results to handle functional stochastic differential equations with progressively measurable (drift) coefficients. Basic martingale problem results for infinitesimal operators with progressively measurable coefficients (as, for example, in Stroock (1975), Karatzas and Shreve (1991)) are used. It should also be noted that, in the context of this work, the perturbed test function method is limited to the situation when the departure process dominates in the limit, and is not applicable to that when the arrival process dominates.

- Third, we provide a direct, pathwise averaging result for the convergence of (1.1) to (1.2), without using martingale, Markov or similar techniques. The result just assumes that η^ϵ converges weakly to η in the Skorokhod J_1 - or M_1 -topology, and that further technical

assumptions on the functions G and \overline{G} hold. This allows one to handle the cases of fBm and sLm in the limit, the M_1 -topology being necessary for the sLm limit. Close relatives of this result in the case of Bm can be found in Kushner (1984), though not stated as explicitly as we do here. See also Marty et al. (2004, 2005, 2008) discussed below.

- Fourth, in some cases, the infinitesimal operator of $y^\epsilon(t)$ in (1.3) can be computed directly and argued to converge to the desired limit. As shown below, this happens, for example, for the infinite source Poisson arrival model in the case of sLm limit. This approach sheds light on the way the limit process arises, and has not been used in connection to the infinite source Poisson model before.

Several other works also explore some of the directions outlined above. Constantini and Kurtz (2006) allow for reflection but their model is different from the one considered here. Reflection is also considered in Kushner (2001), for the case of a general model with short range dependence and light tails features. First results on heavy traffic with fBm in the limit can be found in Konstantopoulos and Lin (1996), Dębicki and Mandjes (2004), and Delgado (2007). The second relation in (1.4) viewed as a mapping of η appears in Reed and Ward (2004), Ward and Kumar (2008) where it is called a generalized regulator mapping. In a recent series of works, Marty (2004, 2005), Marty and Sølna (2008) study the convergence of the systems

$$x^\epsilon(t) = x^\epsilon(0) + \int_0^t \overline{G}(x^\epsilon(s))ds + \beta_\epsilon \int_0^t m(\epsilon^{-1}s) F(x^\epsilon(s))ds, \quad (1.6)$$

where m is a centered, Gaussian, stationary process with long range dependence, and β_ϵ is a suitable normalization. Their method uses the approach of rough paths and applies to multidimensional systems.

Part of the motivation for this study stems from wireless communication applications with random service rates modeling fluctuations in the wireless medium. With the increasing prevalence of high speed wireless systems capable of fast transmissions of large files (such as multimedia), there is a considerable literature analyzing wireless systems with heavy tails or long range dependence. Evidence of long range dependence and heavy tails in several wireless systems is reported in Kalden and Ibrahim (2004), Lee and Fapojuwo (2005), Jiang et al. (2001), and others. In Teymori and Zhuang (2007), an ON-OFF source arrival with heavy-tailed ON periods is considered in studying queueing behavior. In Shao and Madhow (2002), a renewal arrival process with heavy-tailed interarrival times is considered and scheduling algorithms are investigated. In Debernardi (2006), Poisson arrivals with heavy-tailed requests are analyzed for scheduling in wireless systems. Heavy tails and long range dependence are also emerging issues in planning for next generation wireless networks (Ribeiro and DaSilva (2002), Krendzel et al. (2002)). In the wireline context, the relevance of long range dependence and heavy tails was brought up in the seminal work by Taqqu, Willinger and Sherman (1997).

The specific model we consider can be viewed as an extension of the heavy traffic wireless models under light tails and short range dependence such as those considered in Buche and Kushner (2002) and Stolyar (2004). Our extension allows, in particular, to incorporate heavy-tailed and long range dependent arrivals. The exogenous process $L^\epsilon(t)$ models the wireless medium (channel state). As in the above references, the service rate depends on the channel state $L^\epsilon(t) = j$ and applied power, and is written by following the notation of Buche and Kushner (2002) as

$$\epsilon^{-1} \bar{\lambda}_i^d(j) (\bar{p}_i(j) + v_i^\epsilon(x^\epsilon(t), j)), \quad (1.7)$$

where $\bar{\lambda}_i^d(j)$ is the service rate per unit power, $\bar{p}_i(j)$ is the nominal power for balancing the mean arrival rates, and $v_i^\epsilon(x^\epsilon(t), j)$ is the reserve power for queue i at state j . The reserve power

$v_i^\epsilon(x^\epsilon(t), j)$ denotes a control policy, commonly trading off fairness and throughput for competing queues in the multidimensional case. The general model of Buche and Kushner (2002) allows for transfer of nominal power from empty queues. Here, we consider only the 1- D (one queue) case which can be easily extended to the multidimensional case without reallocation of nominal power. In the general case of power transfer, there are technical difficulties surrounding the reflection process. Further comments on these difficulties can be found in Section 5.

The outline of the paper is as follows. In Section 2, the model is given in more detail along with other preliminaries. In Section 3, the limit results for the considered model and various heavy traffic regimes are given. In Section 4, the two aforementioned approaches are outlined and proved for the convergence of (1.1) to (1.2), and for the result of Section 3. In Section 5, a brief summary along with a short discussion of open problems can be found. Appendix A contains discussion on convergence of infinitesimal operators for the infinite source Poisson arrival model when the limit is sLm, and the proof of an auxiliary result can be found in Appendix B.

2 Model and other preliminaries

We are interested in heavy traffic convergence and limits of a queue-length process for a single queue. As in a typical heavy traffic analysis, we consider a sequence of queueing systems indexed by ϵ , where $\epsilon \rightarrow 0$. The queue-length process $x^\epsilon(t)$ in the ϵ -th system is modeled as

$$x^\epsilon(t) = x^\epsilon(0) + A^\epsilon(t) - D^\epsilon(t), \quad t \geq 0, \quad (2.1)$$

where $x^\epsilon(0)$ is the initial length of the queue at time $t = 0$, and $A^\epsilon(t)$ and $D^\epsilon(t)$ represent the cumulative number of arrivals to and departures from the queue until time $t \geq 0$, respectively. We think of $x^\epsilon(t)$ (and $A^\epsilon(t), D^\epsilon(t)$) as already being properly normalized to converge to a limit, that is,

$$x^\epsilon(t) = \beta_\epsilon q^\epsilon(t), \quad (2.2)$$

where β_ϵ is a normalization and $q^\epsilon(t)$ is the unnormalized process. The choice of normalization will depend on both the arrival and departure processes. The latter are modeled according to the assumptions specified in the following subsections.

2.1 Arrival processes and their convergence

We are interested in general cumulative arrival processes $A^\epsilon(t)$, with the focus on those exhibiting features of heavy tails and long range dependence. We shall assume that

$$EA^\epsilon(t) = \beta_\epsilon \mu \epsilon^{-1} t, \quad (2.3)$$

where μ is a constant for convenience (but will be specified as part of a heavy traffic condition in (2.34)), and that

$$\beta_{a,\epsilon} \left(\frac{A^\epsilon(t)}{\beta_\epsilon} - \mu \epsilon^{-1} t \right) = \beta_{a,\epsilon} \frac{\tilde{A}^\epsilon(t)}{\beta_\epsilon} \xrightarrow{d} \tilde{A}(t) \quad (2.4)$$

in $D[0, \infty)$ with a suitable topology, where

$$\tilde{A}^\epsilon(t) = A^\epsilon(t) - \beta_\epsilon \mu \epsilon^{-1} t \quad (2.5)$$

is a centered arrival process. A number of such models could be considered depending on the application in mind.

Remark 2.1 We use the assumption of the form (2.4) to have the normalization β_ϵ free and flexible to match it to the necessary normalization arising from either the arrival or departure processes. Note also that we start above with a general sequence of scaled arrival models. This is perhaps in contrast to more common approaches that start with an unscaled model. However, as the examples below show, our formulation allows for a range of scaled arrival processes and hence a range of unscaled models.

Example 2.1 (Long range dependent point process arrivals) Consider the arrivals given by

$$A^\epsilon(t) = \beta_\epsilon N(\epsilon^{-1}t) \quad (2.6)$$

with

$$N(t) = \sum_{k=1}^{\infty} 1_{\{T_k \leq t\}}, \quad (2.7)$$

where $N(dt)$ is a stationary point process with arrivals (point masses) at $\{T_k\}$. Note that (2.3) holds by using (2.6) and since N is a renewal counting process. Suppose $N(t)$ is long range dependent in the sense that

$$\text{Var}(N(t)) \sim ct^{2H}, \quad 1/2 < H < 1, \quad (2.8)$$

as $t \rightarrow \infty$. The relation (2.8) takes place when interarrivals $\{\tau_k = T_{k+1} - T_k\}$ are i.i.d. and have heavy tails in the sense that

$$P(\tau_k > v) \sim c'v^{-\alpha}, \quad 1 < \alpha = 3 - 2H < 2, \quad (2.9)$$

or interarrivals have finite second moment but are dependent in a suitable sense. See Daley and Vesilo (1997), Daley, Rolski and Vesilo (2000), Kulik and Szekli (2001), Lowen and Teich (2005), and others. In wireless or internet traffic data, $\{T_k\}$ can be thought as arrival times of (equal size) packets for transmission.

Suppose $m = E\tau_k \in (0, \infty)$ is the mean interarrival time. Then, by Theorem 6.3.1 in Whitt (2002),

$$\beta_{a,\epsilon} (N(\epsilon^{-1}t) - m^{-1}\epsilon^{-1}t) = \beta_{a,\epsilon} \left(\frac{A^\epsilon(t)}{\beta_\epsilon} - m^{-1}\epsilon^{-1}t \right) = \beta_{a,\epsilon} \frac{\tilde{A}^\epsilon(t)}{\beta_\epsilon} \xrightarrow{d} \tilde{A}(t) \quad (2.10)$$

in $D[0, \infty)$ with the Skorokhod M_1 -topology if and only if

$$\beta_{a,\epsilon} (S^\epsilon(t) - m\epsilon^{-1}t) \xrightarrow{d} S(t) \quad (2.11)$$

in the same space, where $S^\epsilon(t) = \sum_{k=1}^{\lfloor \epsilon^{-1}t \rfloor} \tau_k$. Moreover, the limits are related through $\tilde{A}(t) = m^{-1}S(m^{-1}t)$.

For example, in the case of renewal processes with heavy-tailed interarrivals (2.9), the convergence (2.11) holds with $\beta_{a,\epsilon} = \epsilon^{1/\alpha}$ and α -sLm $S(t) = X_{\alpha,c,1}(t)$ by the stable FCLT (Theorem 4.5.3 in Whitt (2002)). The sLm $X_{\alpha,c,1}(t)$ has stationary, independent increments and is such that its characteristic function at time $t = 1$ is given by

$$Ee^{i\theta X_{\alpha,c,1}(1)} = \exp \left\{ -c|\theta|^\alpha \left(1 - i \text{sign}(\theta) \tan \frac{\pi\alpha}{2} \right) \right\}, \quad \theta \in \mathbb{R}, \quad (2.12)$$

where $c = c' C_\alpha^{-1}$ with the well-known constant

$$C_\alpha^{-1} = \frac{\Gamma(2 - \alpha) \cos(\alpha\pi/2)}{1 - \alpha} \quad (2.13)$$

(see, for example, Ibragimov and Linnik (1971)). In other words, $X_{\alpha,c,1}(1)$ is an α -stable random variable with skewness parameter $\beta = 1$, shift parameter $\mu = 0$ and scale parameter $c^{1/\alpha}$. Hence, the limit in (2.10) is also sLm.

Example 2.2 (Infinite source Poisson arrival model) Consider the arrivals

$$A^\epsilon(t) = \beta_\epsilon \frac{1}{\lambda_\epsilon} \int_0^{\epsilon^{-1}t} N^\epsilon(s) ds \quad (2.14)$$

with

$$N^\epsilon(s) = \sum_{k=-\infty}^{\infty} 1_{\{\Gamma_k^\epsilon \leq s < \Gamma_k^\epsilon + V_k\}}, \quad (2.15)$$

where $\{\Gamma_k^\epsilon\}_{k \in \mathbb{Z}}$ are Poisson arrivals with intensity $\lambda_\epsilon \uparrow \infty$, and $\{V_k\}_{k \in \mathbb{Z}}$ is a sequence of positive, i.i.d. random variables. The variables V_k are assumed to be heavy-tailed in the sense that

$$1 - F_V(v) = \bar{F}_V(v) = P(V > v) \sim c_V v^{-\alpha}, \quad 1 < \alpha < 2, \quad (2.16)$$

where V denotes the common distribution of V_k , and $c_V > 0$ is a constant. Since $\alpha > 1$, the variable V has a finite mean

$$\mu_V = EV. \quad (2.17)$$

One can show that $EN^\epsilon(s) = \lambda_\epsilon \mu_V$ and hence $EA^\epsilon(t) = \beta_\epsilon \mu_V \epsilon^{-1}t$. Up to normalization β_ϵ , the presence of the normalization factor λ_ϵ in (2.14) is thus to ensure that the effective arrival rate is of the order ϵ^{-1} . A physical interpretation of the infinite source Poisson model is that, at each time Γ_k^ϵ governed by the Poisson process, work is brought to the queue at a constant rate for duration V_k . In wireless or internet traffic data, Γ_k^ϵ are the request times of files, and V_k are sizes of transmitted files.

The asymptotics of the process $A^\epsilon(t)$ in (2.14) depends on the rate of growth of the intensity λ_ϵ . Two regimes are commonly distinguished, namely,

$$\text{slow growth} : \lambda_\epsilon \ll \epsilon^{1-\alpha}, \quad (2.18)$$

$$\text{fast growth} : \epsilon^{1-\alpha} \ll \lambda_\epsilon, \quad (2.19)$$

where $f(\epsilon) \ll g(\epsilon)$ stands for $g(\epsilon)/f(\epsilon) \rightarrow \infty$. The following result is taken from Mikosch et al. (2002). Recall that (standard) fractional Brownian motion (fBm, in short) $\{B_H(t)\}_{t \geq 0}$ is a zero mean, Gaussian process with the covariance function

$$EB_H(t)B_H(s) = \frac{1}{2} \{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}, \quad s, t \geq 0, \quad (2.20)$$

where $H \in (0, 1)$ is the so-called self-similarity parameter. fBm is known to be non-Markovian when $H \neq 1/2$.

Theorem 2.1 (a) (Theorem 1, Mikosch et al. (2002)) Under the slow growth condition,

$$\beta_{s,\epsilon} \frac{\tilde{A}^\epsilon(t)}{\beta_\epsilon} = \frac{1}{\lambda_\epsilon^{1/\alpha} \epsilon^{-1/\alpha}} \left(\int_0^{\epsilon^{-1}t} N^\epsilon(s) ds - \mu_V \lambda_\epsilon \epsilon^{-1}t \right) \xrightarrow{d} c_V^{1/\alpha} X_{\alpha,c,1}(t), \quad (2.21)$$

where $\beta_{s,\epsilon} = \lambda_\epsilon^{1-1/\alpha} \epsilon^{1/\alpha}$, $X_{\alpha,c,1}$ is an α -stable Lévy motion described in Example 2.1 with $c = C_\alpha^{-1}$, and the convergence takes place in $D[0, \infty)$ equipped with the M_1 -topology.

(b) (Theorem 3, Mikosch et al. (2002)) Under the fast growth condition,

$$\beta_{f,\epsilon} \frac{\tilde{A}^\epsilon(t)}{\beta_\epsilon} = \frac{1}{\lambda_\epsilon^{1/2} \epsilon^{(\alpha-3)/2}} \left(\int_0^{\epsilon^{-1}t} N^\epsilon(s) ds - \mu_V \lambda_\epsilon \epsilon^{-1}t \right) \xrightarrow{d} \sigma B_H(t), \quad (2.22)$$

where $\beta_{f,\epsilon} = \lambda_\epsilon^{1/2} \epsilon^{(3-\alpha)/2}$, B_H is a standard fractional Brownian motion with the self-similarity parameter $H = (3 - \alpha)/2$, $\sigma^2 = c_V^2(4 + \alpha(\mu_V - 2))/(3 - \alpha)(2 - \alpha)\mu_V$, and the convergence is in $D[0, \infty)$ with the Skorokhod J_1 -topology.

The subscripts s and f in $\beta_{s,\epsilon}$ and $\beta_{f,\epsilon}$ indicate the normalizations for sLm and fBm, respectively (as well as the normalizations under the slow and fast growth conditions, respectively).

Remark 2.2 The result (2.21) found in Mikosch et al. (2002) states that $c = 1$. The constant C_α , however, is necessary when going from the Lévy-Khintchine representation of stable variables (used implicitly in their work) and the form (2.12) of the characteristic function. The constant C_α also arises naturally in Appendix A below where, in effect, the limit sLm is derived through a method different from that in Mikosch et al. (2002).

Example 2.3 (Arrival model with short range dependence and light tails) For the sake of comparison and illustration of our general framework, we also include here the arrival model considered by Buche and Kushner (2002), which exhibits features of short range dependence and light tails. Let $\epsilon \Delta_k^{a,\epsilon}$, $k \geq 1$, be interarrival times of batches intended for a queue and $N^{a,\epsilon}(t)$ be the number of batches arrived until time t . Suppose the batch k consists of v_k^b packets. Then, the arrival process of packets (the “batch arrival process”) is defined as

$$A^\epsilon(t) = \beta_\epsilon \sum_{k=1}^{N^{a,\epsilon}(t)} v_k^b. \quad (2.23)$$

It is supposed in Buche and Kushner (2002) that the sequences of processes

$$\sqrt{\epsilon} \sum_{k=1}^{[\epsilon^{-1}t]} (v_k^b - \bar{v}^b), \quad \sqrt{\epsilon} \sum_{k=1}^{[\epsilon^{-1}t]} (\Delta_k^{a,\epsilon} - \bar{\Delta}^a) \quad (2.24)$$

are tight, where \bar{v}^b and $\bar{\Delta}^a$ refer to the corresponding means. Convergence of suitable normalized and centered arrival process is not explicitly stated in Buche and Kushner (2002) because the arrival variations get dominated by those of departures. But, under mild additional assumptions, one can think that the arrival process $A^\epsilon(t)$ satisfies (2.4) with $\beta_\epsilon = \epsilon^{1/2}$ and Brownian motion for $\tilde{A}(t)$ in the limit.

2.2 Departure process and its convergence

The departures from the queue are assumed to be governed by varying service rates. Which service rate is used, is characterized by a state process $L(t)$, $t \geq 0$. The process $L(t)$ is assumed to be a stationary Markov process with a finite number of states $j \in J$ and right-continuous. Its stationary distribution is denoted by

$$\{\pi(j), j \in J\}. \quad (2.25)$$

In the ϵ -th system, the state process is assumed to be given by

$$L^\epsilon(t) = L(\epsilon^{-\nu}t), \quad \nu \in (0, 1). \quad (2.26)$$

The choice of variation parameter $\nu \in (0, 1)$ is to ensure that the rate of variations of the state process is slower than that of arrivals and departures (that is, a rate slower than $O(\epsilon^{-1})$). This assumption is motivated by a wireless context where $L(t)$ is known as a channel state process modeling random variations of a wireless medium. In a wireless context, the channel coherence times (how long the channel is in a given state) are typically much longer than the service times for packets. The same assumption is also made by Buche and Kushner (2002).

If the state process $L^\epsilon(t)$ is in the state $j \in J$, it is assumed that the queue is serviced at the rate

$$\epsilon^{-1}(r(j) + u^\epsilon(x^\epsilon(t), j)). \quad (2.27)$$

Here, $r(j)$ is thought as an average rate and u^ϵ is associated with stochastic variations from the mean. The factor ϵ^{-1} is to match the arrival rate. The more explicit form of u^ϵ is specified below in (2.35) as part of the heavy traffic assumption. In a wireless context, the rate (2.27) is written as in (1.7); that is, $\epsilon^{-1}\bar{\lambda}^d(j)(\bar{p}(j) + u^\epsilon(x^\epsilon(t), j))$, where $\bar{p}(j)$ is thought as a mean power allocated to the queue and u^ϵ is a reserve power. Under the above assumptions, the departure process $D^\epsilon(t)$ in (2.1) can be expressed as

$$D^\epsilon(t) = \beta_\epsilon \epsilon^{-1} \int_0^t \sum_{j \in J} \mathbf{1}_{\{L^\epsilon(s)=j\}} (r(j) + u^\epsilon(x^\epsilon(s), j)) \mathbf{1}_{\{x^\epsilon(s)>0\}} ds. \quad (2.28)$$

The term $u^\epsilon(x^\epsilon(s), j)$ in (2.28) will turn out to give rise to the drift in the limit model, and the complementary indicator $\mathbf{1}_{\{x^\epsilon(s)=0\}}$ of $\mathbf{1}_{\{x^\epsilon(s)>0\}}$ in (2.28) will enter into the reflection. Deviations from the mean in the departure process will be governed by the process

$$\tilde{D}^\epsilon(t) = \beta_\epsilon \epsilon^{-1} \int_0^t \sum_{j \in J} (\mathbf{1}_{\{L^\epsilon(s)=j\}} - \pi(j)) r(j) ds =: \beta_\epsilon \epsilon^{-1} \int_0^t F(L^\epsilon(s)) ds, \quad (2.29)$$

where

$$F(l) = \sum_{j \in J} (\mathbf{1}_{\{l=j\}} - \pi(j)) r(j) = r(l) - \sum_{j \in J} \pi(j) r(j). \quad (2.30)$$

(Note that F can be assumed continuous in l and that $EF(L^\epsilon(s)) = 0$.) Its asymptotics can be analyzed as a special case of the general framework (1.1) without reflection, where $G(x, u) = 0$ and $\eta^\epsilon(t)$ has a form (1.5). The following are assumptions and the convergence result which are adapted from Theorem 11 in Kushner (1984), Chapter 5. Assume that, for some $\sigma > 0$,

$$E \left| \int_s^\tau E_t F(L(u)) F(L(s)) du - \frac{\sigma^2}{2} \right| \rightarrow 0, \quad (2.31)$$

as $\tau - s, s - t \rightarrow \infty$, where $E_t(\cdot) = E(\cdot | L(u), u \leq t)$, and the family

$$\left\{ \sup_{\Delta < 1} \left| \int_{t+\Delta}^\tau E_{t+\Delta} F(L(u)) du \right|^2; t \leq \tau < \infty \right\} \quad (2.32)$$

is uniformly integrable.

Theorem 2.2 *Suppose that the conditions (2.31) and (2.32) hold. Then, in the uniform topology,*

$$\beta_{b,\epsilon} \frac{\tilde{D}^\epsilon(t)}{\beta_\epsilon} \xrightarrow{d} \sigma W(t), \quad (2.33)$$

where $W(t)$ is a standard Brownian motion and

$$\beta_{b,\epsilon} = \epsilon^{1-\nu/2}.$$

(See also the related discussion following Assumptions T1–T6 in Section 4.1.)

2.3 Heavy traffic assumptions and the queue process

One of our main goals is to establish different possible limits of the queue-length process $x^\epsilon(t)$, $t \geq 0$. As will be seen in the following sections, different choices of the normalization β_ϵ lead to different nontrivial limits. The following assumptions define multiple heavy traffic regimes, each depending on a choice of β_ϵ :

$$\mu = \sum_{j \in J} \pi(j)r(j), \quad (2.34)$$

$$\epsilon^{-1}\beta_\epsilon u^\epsilon(x, j) = -u(x, j), \quad x \geq 0, \quad (2.35)$$

for some function $u(x, j)$, where the negative sign in (2.35) is for notational convenience. The assumption (2.35) can be weakened to a convergence as $\epsilon \rightarrow 0$ but we will not pursue this for the sake of simplicity. The assumption (2.34) states that the queueing system is asymptotically balanced. The function $u(x, j)$, as usual in heavy traffic analysis, shows up as a drift in the limit models.

Note that, under the assumptions (2.34) and (2.35), the system equation (2.1) can be rewritten as

$$x^\epsilon(t) = x^\epsilon(0) + \tilde{A}^\epsilon(t) - \tilde{D}^\epsilon(t) + \int_0^t \sum_{j \in J} 1_{\{L^\epsilon(s)=j\}} u(x^\epsilon(s), j) ds + z^\epsilon(t), \quad (2.36)$$

where $\tilde{A}^\epsilon(t)$ is given in (2.5), $\tilde{D}^\epsilon(t)$ is given by (2.29), and

$$z^\epsilon(t) = \beta_\epsilon \epsilon^{-1} \int_0^t \sum_{j \in J} 1_{\{L^\epsilon(s)=j\}} r(j) 1_{\{x^\epsilon(s)=0\}} ds, \quad (2.37)$$

and we assume that $u(0, j) = 0$. The term $z^\epsilon(t)$ plays the role of the standard one-dimensional reflection. Note also that, as indicated in Section 1, the relation (2.36) has a general form

$$x^\epsilon(t) = x^\epsilon(0) + \int_0^t G(x^\epsilon(s), L^\epsilon(s)) ds + \eta^\epsilon(t) + z^\epsilon(t), \quad (2.38)$$

where $\eta^\epsilon(t) = \tilde{A}^\epsilon(t) - \tilde{D}^\epsilon(t)$, and the limit is expected to satisfy

$$x(t) = x(0) + \int_0^t \bar{G}(x(s)) ds + \eta(t) + z(t), \quad (2.39)$$

where \bar{G} is a suitable average of G , η is the weak limit of η^ϵ , and z is the reflection term. Convergence of x^ϵ to x is a well-studied problem (see the references in Section 1). The novelty here is that the focus is on long range dependent or heavy-tailed limits whereas previous works have considered Markovian models with light tails. Another important difference is that most of the previous works do not consider reflection in (2.38).

Remark 2.3 Because of the presence of $L^\epsilon(s)$ in (2.38), the process $x^\epsilon(t)$ cannot be expressed simply through a Skorokhod map (or its more general relatives as in Reed and Ward (2004), Ward and Kumar (2008)) of $\eta^\epsilon(t)$, in which case the convergence of $x^\epsilon(t)$ would just follow from that of $\eta^\epsilon(t)$ and continuity property of the Skorokhod map. With $L^\epsilon(s)$ present, a finer analysis for the convergence is necessary.

3 Heavy traffic regimes and statement of main result

The limit process $x(t)$ in (2.39) is characterized by the process $\eta(t)$ which in turn is the weak limit of $\eta^\epsilon(t)$ in (2.38). The convergence assumptions and results summarized in Sections 2.1 and 2.2 suggest that the state space scaling

$$\beta_\epsilon = \beta_{a,\epsilon}, \quad \text{or} \quad (3.1)$$

$$\beta_\epsilon = \beta_{b,\epsilon} \quad (3.2)$$

leads to a nontrivial limit of $x^\epsilon(t)$ in (2.36), where $\beta_{b,\epsilon} = \epsilon^{1-\nu/2}$ is given in Theorem 2.2 and $\beta_{a,\epsilon}$ depends on the arrivals at hand – some examples are given in Section 2.1. If

$$\beta_{a,\epsilon} \ll \beta_{b,\epsilon}, \quad (3.3)$$

one expects that, with $\beta_\epsilon = \beta_{a,\epsilon}$, the queue-length process $x(t)$ in the limit is driven by the process $\tilde{A}(t)$ in (2.4). If

$$\beta_{b,\epsilon} \ll \beta_{a,\epsilon}, \quad (3.4)$$

one expects that $\beta_\epsilon = \beta_{b,\epsilon}$ leads to the limit driven by Bm. The precise result about this is next in Theorem 3.1. The assumptions of the theorem are those of the general results in Section 4 as the proof of Theorem 3.1 follows directly from the results in Section 4.

Theorem 3.1 *Consider a single queue model described in Section 2, and its queue-length process $x^\epsilon(t)$ given by (2.36) which implicitly assumes that the heavy traffic conditions (2.34)–(2.35) hold and where $\tilde{A}^\epsilon(t)$ and $\tilde{D}^\epsilon(t)$ denote the arrival and departure processes. Suppose that the convergence (2.4) of the arrival process $\tilde{A}^\epsilon(t)$ holds in either Skorokhod J_1 - or M_1 -topology with the limit $\tilde{A}(t)$ satisfying Assumption P5 in Section 4.2 when substituted for $\eta(t)$, and that the convergence (2.33) of the departure process $\tilde{D}^\epsilon(t)$ holds in J_1 -topology. Suppose also that Assumption P4 holds for $x^\epsilon(0), x(0)$ substituting $y^\epsilon(0), y(0)$, and that the function $G(x, j) = u(x, j)$ satisfies Assumptions P2, P3 with*

$$\bar{G}(x) = \sum_{j \in J} \pi(j) r(j) u(x, j) \quad (3.5)$$

and $L(u)$ replacing $\xi(u)$ in Assumption P3. Then,

$$x^\epsilon \xrightarrow{d} x, \quad (3.6)$$

where x is given by (2.39), in the following two cases:

(a) If (3.3) holds: the convergence in (3.6) is in the Skorokhod J_1 - or M_1 -topology (depending on that in (2.4)), the driving process $\eta(t)$ in (2.39) is $\tilde{A}(t)$ and β_ϵ is given by (3.1).

(b) If (3.4) holds: the convergence in (3.6) is in the Skorokhod J_1 -topology, the driving process $\eta(t)$ in (2.39) is the Bm in (2.33) and β_ϵ is given by (3.2).

In case (b), the conclusion also holds if Assumptions P2–P5 are replaced by Assumptions T1–T6 in Section 4.1 where, in addition, the function $F(j)$ appearing in Assumptions T1, T3, T5 is given in (2.30).

We next illustrate Theorem 3.1 using the examples considered in Section 2.1.

Example 3.1 (Long range dependent point process arrivals) In the case of renewal process with heavy-tailed interarrivals (2.9), the normalization is $\beta_{a,\epsilon} = \epsilon^{1/\alpha}$. Comparing it to the normalization $\beta_{b,\epsilon} = \epsilon^{1-\nu/2}$ for departures, we obtain that the limit queue-length process is driven by sLm when $\alpha^{-1} > 1 - \nu/2$, and Bm when $\alpha^{-1} < 1 - \nu/2$. (Both α^{-1} and $1 - \nu/2$ belong to the interval $(1/2, 1)$ since $\alpha \in (1, 2)$ and $\nu \in (0, 1)$.)

Example 3.2 (Infinite source Poisson arrival model) By Theorem 2.1, the arrivals have the following two normalizations:

$$\beta_{a,\epsilon} = \beta_{s,\epsilon} = \lambda_\epsilon^{1-1/\alpha} \epsilon^{1/\alpha}, \quad \text{under slow growth condition,} \quad (3.7)$$

$$\beta_{a,\epsilon} = \beta_{f,\epsilon} = \lambda_\epsilon^{1/2} \epsilon^{(3-\alpha)/2}, \quad \text{under fast growth condition.} \quad (3.8)$$

For example, in the slow growth regime, if

$$\beta_{b,\epsilon} \ll \beta_{s,\epsilon}, \quad (3.9)$$

one expects that, with $\beta_\epsilon = \beta_{b,\epsilon}$, the queue-length process $x(t)$ obtained in the limit is driven by Bm.

Relation (3.9) characterizes a particular heavy traffic regime leading to a Brownian limit model. The following two results provide conditions for all such regimes in terms of parameters α, ν and the growth rate λ_ϵ . Since there are two normalizations for the arrivals, the comparison of various scalings is not as easy as in the example above. Let

$$\delta_1 = 1 - \frac{\alpha\nu}{2(\alpha-1)}, \quad (3.10)$$

$$\delta_2 = \alpha - 1 - \nu. \quad (3.11)$$

(Note that δ_1, δ_2 can be negative or positive.)

Proposition 3.1 *Under the slow growth condition $\lambda_\epsilon \ll \epsilon^{1-\alpha}$, we have for*

$$\lambda_\epsilon \ll \epsilon^{\delta_1} : \beta_{s,\epsilon} \ll \beta_{b,\epsilon}, \quad (3.12)$$

$$\epsilon^{\delta_1} \ll \lambda_\epsilon : \beta_{b,\epsilon} \ll \beta_{s,\epsilon}. \quad (3.13)$$

Under the fast growth condition $\epsilon^{1-\alpha} \ll \lambda_\epsilon$, we have for

$$\lambda_\epsilon \ll \epsilon^{\delta_2} : \beta_{f,\epsilon} \ll \beta_{b,\epsilon}, \quad (3.14)$$

$$\epsilon^{\delta_2} \ll \lambda_\epsilon : \beta_{b,\epsilon} \ll \beta_{f,\epsilon}. \quad (3.15)$$

PROOF: The result follows from the following two facts:

$$\frac{\beta_{s,\epsilon}}{\beta_{b,\epsilon}} = \frac{\lambda_\epsilon^{1-1/\alpha} \epsilon^{1/\alpha}}{\epsilon^{1-\nu/2}} = \left(\frac{\lambda_\epsilon}{\epsilon^{\delta_1}} \right)^{(\alpha-1)/\alpha}, \quad \frac{\beta_{f,\epsilon}}{\beta_{b,\epsilon}} = \frac{\lambda_\epsilon^{1/2} \epsilon^{(3-\alpha)/2}}{\epsilon^{1-\nu/2}} = \left(\frac{\lambda_\epsilon}{\epsilon^{\delta_2}} \right)^{1/2}. \quad \square$$

Corollary 3.1 *In the considered single queue model and with infinite source Poisson arrival model, the limit queue-length process is driven by $\eta = \{\eta(t)\}_{t \geq 0}$ which is sLm, fBm or Bm according to the following conditions. If*

$$\alpha < \frac{\nu}{2} + 1, \quad (3.16)$$

then under

$$\lambda_\epsilon \ll \epsilon^{1-\alpha} : \eta \text{ is sLm,} \quad (3.17)$$

$$\epsilon^{1-\alpha} \ll \lambda_\epsilon \ll \epsilon^{\delta_2} : \eta \text{ is fBm,} \quad (3.18)$$

$$\epsilon^{\delta_2} \ll \lambda_\epsilon : \eta \text{ is Bm.} \quad (3.19)$$

If

$$\alpha > \frac{\nu}{2} + 1, \quad (3.20)$$

then under

$$\lambda_\epsilon \ll \epsilon^{\delta_1} : \eta \text{ is sLm}, \quad (3.21)$$

$$\epsilon^{\delta_1} \ll \lambda_\epsilon \ll \epsilon^{1-\alpha} : \eta \text{ is Bm}, \quad (3.22)$$

$$\epsilon^{1-\alpha} \ll \lambda_\epsilon : \eta \text{ is Bm}. \quad (3.23)$$

The limiting behavior in different parameter regimes defined by Corollary 3.1 are summarized in Figure 1.

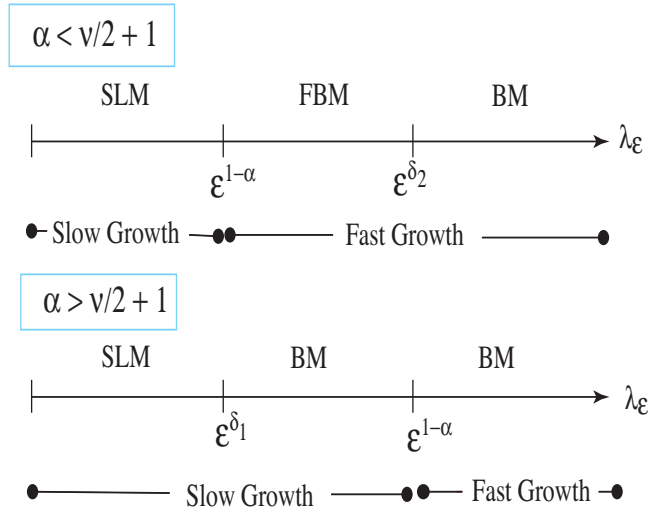


Figure 1: Summary of heavy traffic regimes and driving processes.

PROOF: Consider, for example, the case of (3.16)–(3.19). Relation $\lambda_\epsilon \ll \epsilon^{1-\alpha}$ corresponds to slow growth. The inequality (3.16) is equivalent to $\delta_1 < 1 - \alpha$ and hence $\epsilon^{1-\alpha} \ll \epsilon^{\delta_1}$. In this case, the driving process is sLm by (3.12). The cases of $\epsilon^{1-\alpha} \ll \lambda_\epsilon \ll \epsilon^{\delta_2}$ and $\epsilon^{\delta_2} \ll \lambda_\epsilon$ follow directly from (3.14) and (3.15). (Note that $1 - \alpha > \delta_2$ is equivalent to (3.16) so that $\epsilon^{1-\alpha} \ll \epsilon^{\delta_2}$ holds.) The arguments for the case of (3.20)–(3.23) follow in a similar manner. \square

Example 3.3 (Arrival model with short range dependence and light tails) Since $\beta_{b,\epsilon} = \epsilon^{1-\nu/2} \ll \epsilon^{1/2} = \beta_{a,\epsilon}$, the queue-length process $x(t)$ in the limit is always driven by Bm. This is also the situation reported in Buche and Kushner (2002).

4 Approaches to establishing convergence

We establish here several general results allowing one to prove the convergence of (2.36) or (2.38) to (2.39). These general results are used directly in obtaining the main Theorem 3.1 of Section

3. For notational simplicity, we shall replace the model (2.38) by

$$x^\epsilon(t) = x^\epsilon(0) + \int_0^t G(x^\epsilon(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t) + z^\epsilon(t), \quad (4.1)$$

where $z^\epsilon(t)$ is the one-dimensional reflection at 0, and

$$\xi^\epsilon(t) = \xi\left(\frac{t}{\epsilon}\right) \quad (4.2)$$

with bounded, right-continuous, stationary $\xi(s)$ such that $|\xi(s)| \leq K$. Note that the process L in (2.38) has become ξ in (4.1), and $\epsilon^{-\nu}$ in L^ϵ is now replaced by ϵ^{-1} in ξ^ϵ . The assumptions on ξ are obviously inherited from those on L . We shall suppose that $\epsilon \rightarrow 0$, and expect the limit to have the general form

$$x(t) = x(0) + \int_0^t \bar{G}(x(s)) ds + \eta(t) + z(t), \quad (4.3)$$

where $z(t)$ is the usual one-dimensional reflection at 0, \bar{G} is a suitable average of G and η is the weak limit of η^ϵ .

Various approaches are considered below for establishing the convergence of x^ϵ to x . In Section 4.1, the perturbed test function method is used. In the context of this work, this method is limited to the case (3.4), that is, when the departure process dominates in the limit. A pathwise, direct averaging result is shown in Section 4.2. In these two sections, it is convenient to write (4.1) as

$$x^\epsilon(t) = y^\epsilon(t) + z^\epsilon(t) = (\Gamma y^\epsilon)(t), \quad (4.4)$$

$$y^\epsilon(t) = y^\epsilon(0) + \int_0^t G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t), \quad (4.5)$$

where $(\Gamma y)(t) = y(t) - (\inf\{y(s) : 0 \leq s \leq t\} \wedge 0)$ denotes the usual one-dimensional Skorokhod map, $y^\epsilon(0) = x^\epsilon(0)$, and similarly for the expected limit

$$x(t) = (\Gamma y)(t), \quad y(t) = y(0) + \int_0^t \bar{G}((\Gamma y)(s)) ds + \eta(t). \quad (4.6)$$

See also Appendix A for a related discussion for an infinite source Poisson arrival model.

4.1 Perturbed test function method

The perturbed test function method applied below will establish the convergence of (4.1) with

$$\eta^\epsilon(t) = \frac{1}{\sqrt{\epsilon}} \int_0^t F(\xi^\epsilon(s)) ds \quad (4.7)$$

or, more generally,

$$\eta^\epsilon(t) = \frac{1}{\sqrt{\epsilon}} \int_0^t F(\xi^\epsilon(s)) ds + \zeta^\epsilon(t), \quad (4.8)$$

where $EF(\xi(s)) = 0$ and $\zeta^\epsilon \rightarrow_d 0$ in a suitable topology. The limit process in (4.3) will be driven by

$$\eta(t) = \text{Brownian motion} \quad (4.9)$$

(see Theorem 2.2). This corresponds to the case $\eta^\epsilon(t) = \tilde{A}^\epsilon(t) - \tilde{D}^\epsilon(t)$ in (2.38) where the departure process $\tilde{D}^\epsilon(t) = \beta_\epsilon \epsilon^{-1} \int_0^t F(L^\epsilon(s)) ds$ in (2.29) dominates over the arrival process $\tilde{A}^\epsilon(t)$ in (2.5),

and to the normalization $\beta_\epsilon = \beta_{b,\epsilon} = \epsilon^{1-\nu/2}$. (The term ζ^ϵ in (4.8) represents the arrivals \tilde{A}^ϵ which are dominated in the limit.) NonBrownian-like limits in the case when arrivals \tilde{A}^ϵ dominate over departures \tilde{D}^ϵ cannot be considered using the perturbed test function method and are dealt with in Section 4.2. In what follows, we focus on the case of (4.7). The more general case (4.8) is then immediate using, for example, Lemma 5 in Kushner (1984), Chapter 3, p. 50.

To establish the convergence using (4.7), we follow the perturbed test function method as developed in Kushner (1984, 1990) and others. The method will be applied to show the convergence of y^ϵ given by (4.5). The convergence of x^ϵ in (4.4) will follow from the continuity of the Skorokhod map. A small, yet important, difference from Kushner (1984) is that due to the Skorokhod map Γ , the drift term

$$G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) \quad (4.10)$$

depends on the whole past of y^ϵ through $y^\epsilon(u)$, $0 \leq u \leq s$. The resulting SDE in (4.6) is of the functional SDE form. As with (4.3), we shall assume that the drift term is progressively measurable.

The perturbed test function method relies on the martingale problem formulation and infinitesimal operators which we describe next, adapted to our context. As (4.10) is progressively measurable, we consider the following martingale problem (Stroock (1975), Karatzas and Shreve (1991)). Let \widehat{C}_0^2 denote the space of twice continuously differentiable functions with compact support. Define the operator \mathcal{A}_t as

$$(\mathcal{A}_t f)(y) = \overline{G}(t, y) f'(y(t)) + \frac{1}{2} \sigma^2(t, y) f''(y(t)), \quad f \in \widehat{C}_0^2, \quad (4.11)$$

where both $\overline{G}(t, y)$, $\sigma^2(t, y) : \mathbb{R} \times D[0, \infty) \rightarrow \mathbb{R}$ are progressively measurable functionals. In the case considered here, $\sigma^2(t, y)$ will be taken to be a constant σ^2 .

Definition 4.1 Process $\{y(t)\}_{t \geq 0}$ solves the martingale problem for $\{\mathcal{A}_t\}_{t \geq 0}$, defined in (4.11), if for every $f \in \widehat{C}_0^2$,

$$f(y(t)) - f(y(0)) - \int_0^t (\mathcal{A}_s f)(y) ds \quad (4.12)$$

is a \mathcal{F}_t -martingale, where $\mathcal{F}_t = \sigma\{y(s), 0 \leq s \leq t\}$.

The solution $y(t)$ to the martingale problem in Definition 4.1 is part of the weak solution to the functional SDE given by

$$y(t) = y(0) + \int_0^t \overline{G}(s, y) ds + \int_0^t \sigma(s, y) dW(s). \quad (4.13)$$

The Brownian motion W is another part of the weak solution and can be constructed as in Karatzas and Shreve (1991), Proposition 4.6, p. 315, which uses the martingale representation theorem.

Associated with y^ϵ in (4.5) is a type of infinitesimal operator (Kushner (1984), p. 38) given by

$$\widehat{\mathcal{A}}^\epsilon f(t) = p\text{-}\lim_{\delta \downarrow 0} \frac{E_t^\epsilon f(t + \delta) - f(t)}{\delta}, \quad (4.14)$$

where $E_t^\epsilon(\cdot) = E(\cdot | \mathcal{F}_t^\epsilon)$ with $\mathcal{F}_t^\epsilon = \sigma\{y^\epsilon(s), s \leq t\}$, and $f(t)$ is \mathcal{F}_t^ϵ -measurable. The meaning of $p\text{-}\lim_\delta$ in (4.14) is given by the following. Writing $p\text{-}\lim_\delta f_\delta(t) = f(t)$ stands for $\sup_{t,\delta} E|f_\delta(t)| < \infty$ and $\lim_\delta E|f_\delta(t) - f(t)| = 0$ for each t . Random function f in (4.14) belongs to $\mathcal{D}(\widehat{\mathcal{A}}^\epsilon)$, the domain of $\widehat{\mathcal{A}}^\epsilon$, which requires that $\sup_t |f(t)| < \infty$, and the limit in (4.14)

exists with $\sup_t |\widehat{\mathcal{A}}^\epsilon f(t)| < \infty$. Operator $\widehat{\mathcal{A}}^\epsilon$ is applied to the perturbed test function acting on the prelimit process y^ϵ . As will be seen, the perturbed test function has a form which provides a convenient cancelation among some terms in expansions of (4.14), which aides in the averaging and convergence analysis to a process with infinitesimal operator in the form of (4.11).

We shall use the following technical result which extends Theorem 8 in Kushner (1984), p. 127, to the case of progressively measurable coefficients using parallel notation. We make the following assumptions.

Assumption A1: The functions $\overline{G}(t, y)$ and $\sigma(t, y)$ in (4.11) are continuous.

Assumption A2: The martingale problem in Definition 4.1 has a unique solution in $D[0, \infty)$ for each initial condition.

Assumption A3: For each $T < \infty$, $f \in \widehat{C}_0^2$ and $\widehat{\mathcal{A}}^\epsilon$ given in (4.14), there is $f^\epsilon \in \mathcal{D}(\widehat{\mathcal{A}}^\epsilon)$ such that

$$E |f^\epsilon(t) - f(y^\epsilon(t))| \rightarrow 0, \quad t \leq T, \quad \text{and} \quad (4.15)$$

$$\sup_{\epsilon > 0, t \leq T} E \left| \widehat{\mathcal{A}}^\epsilon f^\epsilon(t) \right| < \infty. \quad (4.16)$$

Assumption A4: There exists $\delta_\epsilon \rightarrow 0$ such that, for any non-decreasing sequence $t_\epsilon \leq T$,

$$E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon \left(\widehat{\mathcal{A}}^\epsilon f^\epsilon(u) - (\mathcal{A}_u f)(y^\epsilon) \right) du \right| \rightarrow 0. \quad (4.17)$$

Lemma 4.1 *Suppose that Assumptions A1-A4 above hold. If $\{y^\epsilon\}$ is tight in the uniform topology and $y^\epsilon(0) \rightarrow_d y(0)$, then (in the uniform topology) $y^\epsilon \rightarrow_d y$, where y is the solution to the martingale problem associated with $\{\mathcal{A}_t\}$ and initial condition $y(0)$.*

The proof of Lemma 4.1 can be found in Appendix B. We can now turn to the main result of the section. We shall make the following assumptions, adapted from Theorem 11 in Kushner (1984), p. 134. See also Theorem 4.3 in Kushner (1990), Chapter 7, p. 169.

Assumption T1: The functions $G(\cdot, \cdot)$, $\overline{G}(\cdot)$, $F(\cdot)$ are continuous.

Assumption T2:

$$E \left| \frac{1}{\tau} E_t \int_t^{t+\tau} G(y, \xi(u)) du - \overline{G}(y) \right| \rightarrow 0, \quad (4.18)$$

as $\tau \rightarrow \infty$ and $t \rightarrow \infty$, for each fixed y , where $E_t(\cdot) = E(\cdot | \xi(u), u \leq t)$.

Assumption T3:

$$E \left| \int_s^\tau E_t F(\xi(s)) F(\xi(u)) du - \frac{\sigma^2}{2} \right| \rightarrow 0, \quad (4.19)$$

as $\tau - s \rightarrow \infty$ and $s - t \rightarrow \infty$.

Assumption T4: $y^\epsilon(0) \xrightarrow{d} y(0)$.

Assumption T5: The family

$$\left\{ \sup_{\Delta < 1} \left| \int_{t+\Delta}^\tau E_{t+\Delta} F(\xi(u)) du \right|^2; t \leq \tau < \infty \right\} \quad (4.20)$$

is uniformly integrable.

Assumption T6: The solution to the martingale problem with the operator \mathcal{A}_t given in (4.11) and $\overline{G}(t, y) = \overline{G}((\Gamma y)(t))$, $\sigma^2(t, y) = \sigma^2$, is unique in $D[0, \infty)$ for each initial condition $y(0)$.

Remark 4.1 Assumption T6 is satisfied, for example, for Lipschitz functions $\overline{G}(y)$. Indeed, by Theorem 1.1 in Stroock (1975), $\{y(t)\}$ is the solution to the martingale problem in Assumption T6 if and only if

$$m(t) \equiv y(t) - \int_0^t \overline{G}((\Gamma y)(s)) ds \quad (4.21)$$

is the solution to the martingale problem with the operator

$$(\tilde{\mathcal{A}}_s f)(x) = \frac{\sigma^2}{2} f''(x(s)). \quad (4.22)$$

The uniqueness of the latter martingale problem is well-established. If $\overline{G}(y)$ is Lipschitz, Lemma 1 in Reed and Ward (2004) (see also Ward and Kumar (2008)) ensures that (4.21) defines a one-to-one correspondence between m and y . Therefore the uniqueness in Assumption T6 follows. Note also that Theorem 11 in Kushner (1984), p. 135, would suggest an additional assumption that $\int_s^\tau EF(\xi(s))F(\xi(u))du \rightarrow \sigma^2/2$ as $\tau - s \rightarrow \infty$. This assumption, however, is implied by Assumption T3 (in the same way, (8.22) follows from (8.23) in Kushner (1984), p. 135).

Theorem 4.1 *Let $\{y^\epsilon(t)\}_{t \geq 0}$ be defined through (4.5) and (4.7). Suppose Assumptions T1-T6 above hold. Then $\{y^\epsilon(t)\}_{t \geq 0}$ is tight and the limit of any weakly convergent (in uniform topology) subsequences of y^ϵ satisfies the martingale problem for $\{\mathcal{A}_t\}$ specified in Assumption T6.*

Remark 4.2 The model considered in Kushner (1984), Chapter 5, p. 134, is more general than (4.7) and assumes state dependence through

$$\eta^\epsilon(t) = \frac{1}{\sqrt{\epsilon}} \int_0^t F(x^\epsilon(s), \xi^\epsilon(s)) ds. \quad (4.23)$$

We do not include dependence on the state $x^\epsilon(t)$ in F since there are differentiability issues surrounding the reflection process and applying the perturbed test function method. More specifically, a natural perturbation (4.24) in this case is given by

$$f_1(y^\epsilon, t) = \frac{1}{\sqrt{\epsilon}} f'(y^\epsilon(t)) \int_t^T E_t^\epsilon F((\Gamma y^\epsilon)(t), \xi^\epsilon(s)) ds$$

and computation of $\widehat{\mathcal{A}}^\epsilon f_1(y^\epsilon, t)$ involves the derivative of $(\Gamma y^\epsilon)(t)$ which does not exist in the usual sense. We intend to address the case of (4.23) in a future work.

Remark 4.3 A very general result on tightness of systems involving reflection can be found in Theorem 6.1 of Kushner (2001), p. 130. This result, in particular, applies to establish tightness of processes considered here. However, as here and also for the models considered by Kushner (2001), the result does not give a limit form of the considered system.

PROOF: As in Kushner (1984), we suppose that G, \overline{G} have a compact support in x . If this is not the case, then a truncation argument should be used. The x -compactness of G , in particular, yields tightness of $\{y^\epsilon(t), \epsilon > 0, t \leq T\}$, $T < \infty$. (The term $\{\eta^\epsilon(t), \epsilon > 0, t \leq T\}$ in (4.7) is tight using Assumption T5.) Let $f \in \widehat{C}_0^2$ and define the *perturbation*

$$f_1(y^\epsilon, t) = \frac{1}{\sqrt{\epsilon}} f'(y^\epsilon(t)) \int_t^T E_t^\epsilon F(\xi^\epsilon(s)) ds, \quad (4.24)$$

and the *perturbed test function*

$$f^\epsilon(t) = f(y^\epsilon(t)) + f_1(y^\epsilon, t). \quad (4.25)$$

We shall compute next $\widehat{\mathcal{A}}^\epsilon f^\epsilon(t)$. Considering the perturbation term $f_1(y^\epsilon, t)$ first, observe that

$$\begin{aligned} E_t^\epsilon f_1(y^\epsilon, t + \delta) - f_1(y^\epsilon, t) &= \frac{1}{\sqrt{\epsilon}} E_t^\epsilon f'(y^\epsilon(t + \delta)) \int_{t+\delta}^T E_{t+\delta}^\epsilon F(\xi^\epsilon(s)) ds \\ &\quad - \frac{1}{\sqrt{\epsilon}} f'(y^\epsilon(t)) \int_t^T E_t^\epsilon F(\xi^\epsilon(s)) ds \\ &= -\frac{1}{\sqrt{\epsilon}} f'(y^\epsilon(t)) \int_t^{t+\delta} E_t^\epsilon F(\xi^\epsilon(s)) ds \\ &\quad + \frac{1}{\sqrt{\epsilon}} f''(y^\epsilon(t)) E_t^\epsilon (y^\epsilon(t + \delta) - y^\epsilon(t)) \int_{t+\delta}^T E_{t+\delta}^\epsilon F(\xi^\epsilon(s)) ds + o(\delta), \end{aligned}$$

where $o(\delta)$ follows from $f \in \widehat{C}_0^2$, the definition of y^ϵ , continuity of F and G , and the assumed x -compactness of G . This yields

$$\begin{aligned} \widehat{\mathcal{A}}^\epsilon f_1(y^\epsilon, t) &= -\frac{1}{\sqrt{\epsilon}} f'(y^\epsilon(t)) F(\xi^\epsilon(t)) \\ &\quad + \frac{1}{\sqrt{\epsilon}} f''(y^\epsilon(t)) \left(G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) + \frac{1}{\sqrt{\epsilon}} F(\xi^\epsilon(t)) \right) \int_t^T E_t^\epsilon F(\xi^\epsilon(s)) ds. \end{aligned} \quad (4.26)$$

Similarly, we have

$$\widehat{\mathcal{A}}^\epsilon f(y^\epsilon(t)) = f'(y^\epsilon(t)) \left(G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) + \frac{1}{\sqrt{\epsilon}} F(\xi^\epsilon(t)) \right). \quad (4.27)$$

Combining (4.26) and (4.27) leads to

$$\begin{aligned} \widehat{\mathcal{A}}^\epsilon f^\epsilon(t) &= f'(y^\epsilon(t)) G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) \\ &\quad + \frac{1}{\sqrt{\epsilon}} f''(y^\epsilon(t)) \left(G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) + \frac{1}{\sqrt{\epsilon}} F(\xi^\epsilon(t)) \right) \int_t^T E_t^\epsilon F(\xi^\epsilon(s)) ds \\ &= f'(y^\epsilon(t)) G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) + f''(y^\epsilon(t)) F(\xi^\epsilon(t)) \int_{t/\epsilon}^{T/\epsilon} E_{t/\epsilon} F(\xi(s)) ds \\ &\quad + O(\sqrt{\epsilon}) \int_{t/\epsilon}^{T/\epsilon} E_{t/\epsilon} F(\xi(s)) ds, \end{aligned} \quad (4.28)$$

where $O(\sqrt{\epsilon})$ follows from $f \in \widehat{C}_0^2$, and continuity and x -compactness of G .

The convergence in the theorem statement follows from applying Lemma 4.1 above to the case considered here. The following highlights how Lemma 4.1 is applied. In Assumption A1, $\overline{G}(t, y) = \overline{G}((\Gamma y)(t))$, which is continuous by Assumption T1, and $\sigma(t, y)$ is a constant σ . The operator \mathcal{A}_t specified in Assumption T6 is a particular case satisfying Assumption A2.

Tightness of $\{y^\epsilon\}$ in Lemma 4.1 follows from Theorem 4 in Kushner (1984), Chapter 3, p. 48, as long as

$$E \sup_{t \leq T} |f^\epsilon(t) - f(y^\epsilon(t))|^2 = E \sup_{t \leq T} |f_1(y^\epsilon, t)|^2 \rightarrow 0, \quad (4.29)$$

as $\epsilon \rightarrow 0$, and $\{\widehat{\mathcal{A}}^\epsilon f^\epsilon(t), \epsilon > 0, t \leq T\}$ is uniformly integrable. To show (4.29), observe that, after a change of variables,

$$\begin{aligned}
E \sup_{t \leq T} |f_1(y^\epsilon, t)|^2 &= \frac{1}{\epsilon} E \sup_{t \leq T} \left| f'(y^\epsilon(t)) \int_t^T E_t^\epsilon F(\xi^\epsilon(s)) ds \right|^2 \leq C \epsilon E \sup_{t \leq T} \left| \int_{t/\epsilon}^{T/\epsilon} E_{t/\epsilon} F(\xi(u)) du \right|^2 \\
&\leq C \epsilon L^2 + C \epsilon E \sup_{t/\epsilon \leq T/\epsilon} \left| \int_{t/\epsilon}^{T/\epsilon} E_{t/\epsilon} F(\xi(u)) du \right|^2 \mathbf{1}_{\left\{ \left| \int_{t/\epsilon}^{T/\epsilon} E_{t/\epsilon} F(\xi(u)) du \right| \geq L \right\}} \leq C' \epsilon \\
&+ C' \epsilon \frac{T}{\epsilon} \sup_{t/\epsilon \leq T/\epsilon} E \sup_{\Delta < 1} \left| \int_{t/\epsilon + \Delta}^{T/\epsilon} E_{t/\epsilon + \Delta} F(\xi(u)) du \right|^2 \mathbf{1}_{\left\{ \sup_{\Delta < 1} \left| \int_{t/\epsilon + \Delta}^{T/\epsilon} E_{t/\epsilon + \Delta} F(\xi(u)) du \right| \geq L \right\}}, \quad (4.30)
\end{aligned}$$

where the last bound is obtained by using $E \sup_{t/\epsilon \leq T/\epsilon} X(t/\epsilon) \leq \sum_{k=1}^{T/\epsilon} E \sup_{\Delta < 1} X(t_k/\epsilon + \Delta) \leq (T/\epsilon) \sup_{t/\epsilon \leq T/\epsilon} E \sup_{\Delta < 1} X(t/\epsilon + \Delta)$ for a suitable sequence t_k . The right-hand side of (4.30) is arbitrarily small as $\epsilon \rightarrow 0$ for large L using Assumption T5. The uniform integrability of $\{\widehat{\mathcal{A}}^\epsilon f^\epsilon(t), \epsilon > 0, t \leq T\}$ follows from (4.28) using $f \in \widehat{C}_0^2$, continuity of G and F , Assumption T5 and the assumed x -compactness of G . The conditions in Assumption A3 are weaker than what is proved for tightness above.

Assumption A4 holds when substituting $\widehat{\mathcal{A}}^\epsilon f^\epsilon(t)$ in (4.28) and \mathcal{A}_u as specified in Assumption T6. The term with $O(\sqrt{\epsilon})$ in (4.28) is negligible using Assumption T5. For the ‘‘diffusion part’’ of Assumption A4, observe that

$$\begin{aligned}
&E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon \left(f''(y^\epsilon(u)) F(\xi^\epsilon(u)) \int_{u/\epsilon}^{T/\epsilon} E_{u/\epsilon} F(\xi(s)) ds - \frac{\sigma^2}{2} f''(y^\epsilon(u)) \right) du \right| \\
&= E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} f''(y^\epsilon(t_\epsilon)) \left(\int_{u/\epsilon}^{T/\epsilon} E_{t_\epsilon/\epsilon} F(\xi^\epsilon(u)) F(\xi(s)) ds - \frac{\sigma^2}{2} \right) \right. \\
&\quad \left. + E_{t_\epsilon}^\epsilon (f''(y^\epsilon(u)) - f''(y^\epsilon(t_\epsilon))) \left(F(\xi^\epsilon(u)) \int_{u/\epsilon}^{T/\epsilon} E_{u/\epsilon} F(\xi(s)) ds - \frac{\sigma^2}{2} \right) du \right| \\
&\leq C \frac{1}{\delta_\epsilon} \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E \left| \int_{u/\epsilon}^{T/\epsilon} E_{t_\epsilon/\epsilon} F(\xi(u/\epsilon)) F(\xi(s)) ds - \frac{\sigma^2}{2} \right| du \\
&+ C \frac{1}{\delta_\epsilon} \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E |f''(y^\epsilon(u)) - f''(y^\epsilon(t_\epsilon))| \left(\left| \int_{u/\epsilon}^{T/\epsilon} E_{u/\epsilon} F(\xi(s)) ds \right| + 1 \right) du. \quad (4.31)
\end{aligned}$$

The first term on the right-hand side of (4.31) converges to 0 using Assumption T3. The second term converges to 0 using, in particular, Assumption T5 and the fact $f \in \widehat{C}_0^2$.

For the ‘‘drift part’’ of Assumption A4, we need to consider

$$E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon (f'(y^\epsilon(u)) G((\Gamma y^\epsilon)(u), \xi^\epsilon(u)) - f'(y^\epsilon(u)) \overline{G}((\Gamma y^\epsilon)(u))) du \right|.$$

Arguing as in (4.31) above, it is enough to show that

$$E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon (G((\Gamma y^\epsilon)(u), \xi^\epsilon(u)) - \overline{G}((\Gamma y^\epsilon)(u))) du \right| \quad (4.32)$$

converges to zero. To show this convergence, we invoke the Skorokhod representation theorem, under which we can assume that $y^\epsilon \rightarrow y$ and hence $\Gamma y^\epsilon \rightarrow \Gamma y$ a.s. in uniform topology. Since G and \bar{G} are continuous on their compact supports, they are also absolutely continuous. This implies, in particular, that the expression (4.32) is asymptotically equivalent to

$$E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon (G((\Gamma y^\epsilon)(t_\epsilon), \xi^\epsilon(u)) - \bar{G}((\Gamma y^\epsilon)(t_\epsilon))) du \right|. \quad (4.33)$$

Suppose now that, more specifically, $G(x, u) = 0$ for $|x| \geq N + 1$. Let $B_q, q = 1, \dots, Q$, be disjoint intervals such that $\cup_{q=1}^Q B_q = [-(N + 1), N + 1]$, and $a_q \in B_q$. By taking small enough (but fixed) B_q 's, and using absolute continuity, the expression (4.33) is asymptotically equivalent to

$$\begin{aligned} & E \frac{1}{\delta_\epsilon} \left| \int_{t_\epsilon}^{t_\epsilon + \delta_\epsilon} E_{t_\epsilon}^\epsilon \sum_{q=1}^Q 1_{\{(\Gamma y^\epsilon)(t_\epsilon) \in B_q\}} (G(a_q, \xi^\epsilon(u)) - \bar{G}(a_q)) du \right| \\ &= E \frac{\epsilon}{\delta_\epsilon} \left| \sum_{q=1}^Q 1_{\{(\Gamma y^\epsilon)(t_\epsilon) \in B_q\}} \int_{t_\epsilon/\epsilon}^{(t_\epsilon + \delta_\epsilon)/\epsilon} E_{t_\epsilon/\epsilon} (G(a_q, \xi(z)) - \bar{G}(a_q)) dz \right| \\ &\leq \sum_{q=1}^Q E \frac{\epsilon}{\delta_\epsilon} \left| \int_{t_\epsilon/\epsilon}^{(t_\epsilon + \delta_\epsilon)/\epsilon} E_{t_\epsilon/\epsilon} (G(a_q, \xi(z)) - \bar{G}(a_q)) dz \right|, \end{aligned} \quad (4.34)$$

where we used the change of variables $u = \epsilon z$ above. The right-hand side of (4.34) converges to 0 by Assumption T2. \square

4.2 Pathwise, direct averaging approach

In this section, we will establish the convergence of (4.1) under quite general assumptions on the function G and the processes η^ϵ . The limit process in (4.3) will not necessarily be driven by Brownian motion η . As in Section 4.1, since the Skorokhod map Γ is continuous in J_1 - and M_1 -topologies, it is enough to establish the convergence of y^ϵ in (4.5). We make the following assumptions.

Assumption P1: $\eta^\epsilon \xrightarrow{d} \eta$ in $D[0, \infty)$ with either J_1 - or M_1 -topology.

Assumption P2: The functions $G(\cdot, \cdot), \bar{G}(\cdot)$ are continuous.

Assumption P3: For any fixed x , as $t, \tau \rightarrow \infty$,

$$\frac{1}{\tau} \int_t^{t+\tau} G(x, \xi(u)) du \rightarrow \bar{G}(x), \quad \text{a.s.} \quad (4.35)$$

Assumption P4: $y^\epsilon(0) \xrightarrow{d} y(0)$.

Assumption P5: $P(\eta(t) = \eta(t-)) = 1$ for each fixed t .

We shall use below the following notation. For $z \in D([0, T], \mathbb{R}^k)$, let

$$w_{M_1}(z, \gamma) = \sup_{0 \leq t_1 \leq t \leq t_2 \leq T, 0 \leq t_2 - t_1 \leq \gamma} M(z(t_1), z(t), z(t_2)) \quad (4.36)$$

be the oscillation function in M_1 -topology, where

$$M(z_1, z_2, z_3) = \text{distance from } z_2 \text{ to the segment } [z_1, z_3]. \quad (4.37)$$

Let also

$$w_U(z, \gamma) = \sup_{0 \leq t_1, t_2 \leq T, |t_2 - t_1| \leq \gamma} |z(t_1) - z(t_2)| \quad (4.38)$$

be the usual uniform oscillation function. The following elementary lemma will be used below.

Lemma 4.2 *Let w_{M_1} , M and w_U be defined as (4.36), (4.37) and (4.38), respectively. Then, for any x_i, y_i ,*

$$M(x_1 + y_1, x_2 + y_2, x_3 + y_3) \leq M(x_1, x_2, x_3) + |y_1 - y_2| + |y_2 - y_3| \quad (4.39)$$

and, for any z_i ,

$$w_{M_1}(z_1 + z_2, \gamma) \leq w_{M_1}(z_2, \gamma) + 2w_U(z_1, \gamma). \quad (4.40)$$

PROOF: Note that $M(x_1, x_2, x_3) \leq M(x_1^*, x_2, x_3) + |x_1 - x_1^*|$, $M(x_1, x_2, x_3) \leq M(x_1, x_2, x_3^*) + |x_3 - x_3^*|$ for any x_i, x_i^* . Then, $M(x_1 + y_1, x_2 + y_2, x_3 + y_3) \leq M(x_1 + y_2, x_2 + y_2, x_3 + y_3) + |y_1 - y_2| \leq M(x_1 + y_2, x_2 + y_2, x_3 + y_2) + |y_1 - y_2| + |y_2 - y_3| = M(x_1, x_2, x_3) + |y_1 - y_2| + |y_2 - y_3|$ which yields (4.39). The inequality (4.40) is immediate from (4.39). \square

The following is the main result of the section.

Theorem 4.2 *With the model (4.1)–(4.2) and under Assumptions P1–P5 above, $\{y^\epsilon\}$ is tight and $y^\epsilon \rightarrow_d y$ in respective topologies such that $y(t)$ satisfies (4.6).*

PROOF: We only consider the more difficult case of the space $D[0, \infty)$ equipped with M_1 -topology. By using Assumption P5, it is enough to show the convergence in $D[0, T]$ for fixed T . We use a truncation argument in the spirit of, for example, Kushner (1984), p. 122. Similarly to (4.4)–(4.5), define

$$y^{\epsilon, N}(t) = y^{\epsilon, N}(0) + \int_0^t G^N((\Gamma y^{\epsilon, N})(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t), \quad (4.41)$$

where

$$G^N(x, \xi) = G(x, \xi)q_N(x), \quad q_N(x) = \begin{cases} 1, & |x| \leq N, \\ 0, & |x| \geq N + 1, \end{cases} \quad (4.42)$$

and $0 \leq q_N(x) \leq 1$ and $q_N \in C^\infty(\mathbb{R})$. Similarly, let

$$y^N(t) = y^N(0) + \int_0^t \bar{G}^N((\Gamma y^N)(s)) ds + \eta(t), \quad (4.43)$$

where

$$\bar{G}^N(x) = \bar{G}(x)q_N(x). \quad (4.44)$$

It is then enough to show that

$$y^{\epsilon, N} \xrightarrow{d} y^N, \quad \text{as } \epsilon \rightarrow 0, \quad (4.45)$$

$$y^N \xrightarrow{d} y, \quad \text{as } N \rightarrow \infty, \quad (4.46)$$

$$\limsup_{N \rightarrow \infty} \limsup_{\epsilon \rightarrow 0} P(d(y^{\epsilon, N}, y^\epsilon) \geq \delta) = 0, \quad (4.47)$$

for any $\delta > 0$, where d denotes the distance in M_1 -topology (Whitt (2002)). The convergence (4.45), in turn, would follow from the following two facts:

$$\{y^{\epsilon, N}\}_{\epsilon > 0} \text{ and, more generally, } \{(y^{\epsilon, N}, \eta^\epsilon)\}_{\epsilon > 0} \text{ is tight, and} \quad (4.48)$$

the limit of any convergent subsequence of $\{y^{\epsilon, N}\}_{\epsilon > 0}$ satisfies (4.43). (4.49)

Showing (4.48): We show the stronger statement in (4.48) as it is needed below. Let $z^\epsilon = (y^{\epsilon, N}, \eta^\epsilon)$. For simplicity of notation, we drop dependence on the index N . Then, $G(x, u) = G^N(x, u)$ can be supposed to be bounded in x . By Theorem 12.12.3 in Whitt (2002), p. 426, it is enough to show that, for every $\delta > 0$, there is $c > 0$ such that

$$P\left(\sup_{s \in [0, T]} |z^\epsilon(s)| \geq c\right) \leq \delta, \quad \text{for all } \epsilon > 0, \quad (4.50)$$

and, for every $\delta > 0$, $\eta > 0$, there is $\gamma > 0$ such that

$$P(w_{M_1}(z^\epsilon, \gamma) \geq \eta) \leq \delta, \quad \text{for all } \epsilon > 0. \quad (4.51)$$

The relation (4.41) can be written as $y^\epsilon(t) = y^\epsilon(0) + \tilde{y}^\epsilon(t) + \eta^\epsilon(t)$, where

$$\tilde{y}^\epsilon(t) = \int_0^t G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds. \quad (4.52)$$

Take for simplicity $y^\epsilon(0) = 0$. Then, $z^\epsilon = (\tilde{y}^\epsilon, 0) + (\eta^\epsilon, \eta^\epsilon) =: z_1^\epsilon + z_2^\epsilon$. Using Lemma 4.2 above,

$$w_{M_1}(z^\epsilon, \gamma) \leq w_{M_1}(z_2^\epsilon, \gamma) + 2w_U(z_1^\epsilon, \gamma). \quad (4.53)$$

By Assumption P1, $\{\eta^\epsilon\}_{\epsilon > 0}$ tight in M_1 , and hence so is $\{z_2^\epsilon\}_{\epsilon > 0}$ whose both coordinates are equal to η^ϵ . This shows that (4.51) is satisfied with $w_{M_1}(z_2^\epsilon, \gamma)$ replacing $w_{M_1}(z^\epsilon, \gamma)$. Since G is supposed bounded in both variables, the first coordinate $\{\tilde{y}^\epsilon\}_{\epsilon > 0}$ of z_1^ϵ is tight in uniform topology. Hence, (4.51) also holds with $w_U(z_1^\epsilon, \gamma)$ replacing $w_{M_1}(z^\epsilon, \gamma)$. This and (4.53) show that the tightness condition (4.51) holds. The condition (4.50) is even easier to prove using the assumptions.

Showing (4.49): We continue using the notation where the index N is suppressed. Suppose that the sequence $\{(y^\epsilon, \eta^\epsilon)\}$ itself converges weakly. By the Skorokhod representation theorem, we may suppose that $(y^\epsilon, \eta^\epsilon) \rightarrow (y, \eta)$ a.s. in M_1 -topology. It is then sufficient to argue that

$$\int_0^t G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds \rightarrow \int_0^t \overline{G}((\Gamma y)(s)) ds \quad (4.54)$$

a.s. in uniform topology because the latter convergence implies that in M_1 -topology. For simplicity, we suppress the indication ‘‘a.s.’’ below. Since G, \overline{G} have a compact support, they are absolutely continuous. Therefore, for arbitrary $\Delta > 0$, we may suppose

$$\sup_{z, u} |G(z, u) - G(z^\Delta, u)| \leq \Delta, \quad (4.55)$$

$$\sup_z |\overline{G}(z) - \overline{G}(z^\Delta)| \leq \Delta, \quad (4.56)$$

where

$$z^\Delta = \sum_{q=1}^Q z_q 1_{\{z \in B_q\}} \quad (4.57)$$

with disjoint intervals $B_q = B_q(\Delta)$ such that $\cup_{q=1}^Q B_q = [-(N+1), N+1]$, and $a_q \in B_q$. We may also suppose that $Leb\{s : (\Gamma y)(s) \in \partial B_q\} = 0$ for any $q = 1, \dots, Q$, where Leb indicates the Lebesgue measure.

Denote $((\Gamma y^\epsilon)(s))^\Delta = (\Gamma y^\epsilon)^\Delta(s)$, and consider the following four terms:

$$\begin{aligned}
I_1(t) &= \int_0^t \left(G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) - G((\Gamma y^\epsilon)^\Delta(s), \xi^\epsilon(s)) \right) ds, \\
I_2(t) &= \int_0^t \left(G((\Gamma y^\epsilon)^\Delta(s), \xi^\epsilon(s)) - \overline{G}((\Gamma y^\epsilon)^\Delta(s)) \right) ds, \\
I_3(t) &= \int_0^t \left(\overline{G}((\Gamma y^\epsilon)^\Delta(s)) - \overline{G}((\Gamma y^\epsilon)(s)) \right) ds, \\
I_4(t) &= \int_0^t \left(\overline{G}((\Gamma y^\epsilon)(s)) - \overline{G}((\Gamma y)(s)) \right) ds.
\end{aligned} \tag{4.58}$$

By using (4.55) and (4.56), $\sup_t |I_k(t)|$, $k = 1, 3$, are arbitrarily small. It is then enough to show that $\sup_t |I_k(t)| \rightarrow 0$, $k = 2, 4$.

For convergence of $I_4(t)$, first note that $\Gamma y^\epsilon \rightarrow \Gamma y$ in M_1 -topology by Theorem 13.5.1 in Whitt (2002), and that Γy^ϵ is uniformly bounded. The desired convergence then follows using Theorem 12.5.1, (iv) in Whitt (2002). For convergence of $I_2(t)$, choose δ_ϵ such that $\delta_\epsilon \rightarrow 0$ and $\delta_\epsilon/\epsilon \rightarrow \infty$. Setting

$$\tilde{G}^\epsilon(v) = \frac{1}{\delta_\epsilon} \int_{k\delta_\epsilon}^{(k+1)\delta_\epsilon} \left(G((\Gamma y^\epsilon)^\Delta(s), \xi^\epsilon(s)) - \overline{G}((\Gamma y^\epsilon)^\Delta(s)) \right) ds, \quad v \in [k\delta_\epsilon, (k+1)\delta_\epsilon), \tag{4.59}$$

it is enough to show the convergence for

$$\tilde{I}_2(t) = \int_0^t \tilde{G}^\epsilon(v) dv. \tag{4.60}$$

Observe now that

$$\tilde{G}^\epsilon(v) = \frac{1}{\delta_\epsilon} \int_{k\delta_\epsilon}^{(k+1)\delta_\epsilon} \sum_{q=1}^Q \mathbf{1}_{\{(\Gamma y^\epsilon)(s) \in B^q\}} \left(G(a_q, \xi^\epsilon(s)) - \overline{G}(a_q) \right) ds. \tag{4.61}$$

By using $Leb\{(\Gamma y)(s) \in \partial B_q\} = 0$ and Theorem 12.5.1, (v), in Whitt (2002), for a.e. v (in fact, for any v , at which $(\Gamma y^\epsilon)(v)$ is continuous) and small enough δ_ϵ , the indicator function in (4.61) is equal to 1 for some $q = q_0$ for all $s \in [k\delta_\epsilon, (k+1)\delta_\epsilon)$. Then, for such a.e. v , $\tilde{G}^\epsilon(v) \rightarrow 0$ by Assumption P3. Since \tilde{G}^ϵ is bounded, the dominated convergence theorem implies that $\int_0^T |\tilde{G}^\epsilon(v)| dv \rightarrow 0$ and hence that $\sup_t |\tilde{I}_2(t)| \rightarrow 0$.

Showing (4.46): This follows from

$$P(y^N \neq y \text{ on } [0, T]) \leq P\left(\sup_{s \in [0, T]} |y(s)| > N/2 \right) \rightarrow 0,$$

as $N \rightarrow \infty$.

Showing (4.47): By using Lemma 13.5.1 in Whitt (2002),

$$P(d(y^{\epsilon, N}, y^\epsilon) \geq \delta) \leq P\left(\sup_{s \in [0, T]} |y^{\epsilon, N}(s)| > N/2 \right).$$

Then, by (4.45),

$$\limsup_{\epsilon \rightarrow 0} P\left(\sup_{s \in [0, T]} |y^{\epsilon, N}(s)| > N/2 \right) = P\left(\sup_{s \in [0, T]} |y^N(s)| > N/2 \right)$$

and, by (4.46),

$$\limsup_{N \rightarrow \infty} P\left(\sup_{s \in [0, T]} |y^N(s)| > N/2 \right) = 0. \quad \square$$

5 Conclusions

This work raises a number of open problems concerning more complex models. The present approach based on perturbed test functions does not allow for state dependence in the “diffusion term” (see Remark 4.2 following Theorem 4.1). Though state dependence is absent from the model of interest here, it is an important feature elsewhere and should be taken into account in developing more general results. Extensions of the model (1.6) using the rough paths approach are also natural to consider. For example, as in the model (1.1) considered here, the function \bar{G} in (1.6) could include the noise term and a reflection term could also be present. All these extensions should be considered in the multidimensional context.

Extending to higher dimensions is important for obtaining control policies in limit models when there are competing queues. In this work, only a one-dimensional model is considered and the focus is on the averaging techniques for the noise process L^ϵ or ξ^ϵ in the convergence analysis. We are not concerned with control design in this work though the term $u(x, j)$ in our limit models can be thought as a control component and even in the one-dimensional case, the control policy design can entail power usage objectives. Extending to higher dimensions can be readily done if the preallocated power for queue i , or $r_i(j)$ in analogy to (2.27), is *not* reallocated from the empty queue i to the other non-empty queues. This is a practical approach and results in orthogonal reflection directions on the “faces” (i.e. when only one queue is empty at a time). On the “edges” (i.e. more than one queue empty at a time) the reflection direction is a convex combination of the reflection directions on the faces making up the edge. Under these conditions, the technical issues under reallocation, discussed below, are not encountered.

When allowing $r_i(j)$ to be reallocated when queue i is empty, one can achieve more efficient operation. For example, when queue i empties, $r_i(j)$ could be reallocated proportionally to the other nonempty queues according to their queue size or preallocated power. This reallocation occurs under varying conditions, that is, depending on the state $L^\epsilon(t) = j$. A resulting technical difficulty is determining the reflection directions for z^ϵ in higher-dimensional analogues of (4.1) (Buche and Kushner (2002)). Furthermore, it is possible for the reflection directions to be such that the Lipschitz continuity of the reflection mappings, which is used to obtain existence and uniqueness of the limit process, no longer holds. This is in contrast to early wireline models with driving Brownian motion and constant drift (see, for example, Dupuis (1993), Reiman and Williams (1993)). In this case, constant reflection directions can be chosen which give the desired continuity properties of the Skorokhod map. In our models, under reallocation, it is not clear whether the existence and uniqueness of the limit forms as in (4.3) can be established under the most general conditions.

A Convergence of infinitesimal operators for infinite source Poisson arrival model

Related to Sections 4.1 and 4.2, but in another direction for proving convergence of (4.5), there are instances where convergence of the infinitesimal operator can be established directly without using perturbed test functions. We illustrate this with the infinite source Poisson arrival model in Example 2.2 when the arrival limit process is sLm, that is, under the slow growth condition (see Theorem 2.1,(a)). Since we work with specific example, the discussion below will not be completely rigorous.

As proved in Mikosch et al. (2002), the part of the arrival process that contributes to the sLm

limit is given by

$$\eta^\epsilon(t) = \beta_\epsilon(\widehat{\eta}^\epsilon(t) - \mu_V \lambda_\epsilon \epsilon^{-1}t), \quad (\text{A.1})$$

where $\beta_\epsilon = \lambda_\epsilon^{-1/\alpha} \epsilon^{1/\alpha}$ and

$$\widehat{\eta}^\epsilon(t) = \sum_{k=-\infty}^{\infty} V_k 1_{\{(\Gamma_k^\epsilon, V_k) \in R(\epsilon^{-1}t)\}} \quad (\text{A.2})$$

with $R(s) = \{(\gamma, v) : 0 < \gamma, v \leq s, \gamma + v \leq s\}$, or

$$\eta^\epsilon(t) = \sum_{k=-\infty}^{\infty} (\beta_\epsilon V_k) 1_{\{(\Gamma_k^\epsilon, V_k) \in R(\epsilon^{-1}t)\}} - \beta_\epsilon \mu_V \lambda_\epsilon \epsilon^{-1}t. \quad (\text{A.3})$$

Ignoring the terms that do not contribute to the limit (these are defined over the complement region of $R(s)$), we suppose that $\eta^\epsilon(t)$ in (4.5) is given by (A.1) or equivalently (A.3). Note that $\eta^\epsilon(t)$ has independent increments.

With η^ϵ given by (A.3), the infinitesimal operator $\widehat{\mathcal{A}}^\epsilon$ in (4.14) of y^ϵ in (4.5) can be computed directly and shown to converge to the desired limit. Suppose that $y^\epsilon(0) = 0$ for simplicity. Observe that

$$E_t^\epsilon f(y^\epsilon(t + \delta)) = E_t^\epsilon f\left(\int_0^{t+\delta} G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t + \delta)\right) =: I_1 + I_2, \quad (\text{A.4})$$

where

$$\begin{aligned} I_1 &= E_t^\epsilon f\left(\int_0^{t+\delta} G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t + \delta)\right) \\ &\quad - E_t^\epsilon f\left(\int_0^t G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t + \delta)\right), \\ I_2 &= E_t^\epsilon f\left(\int_0^t G((\Gamma y^\epsilon)(s), \xi^\epsilon(s)) ds + \eta^\epsilon(t + \delta)\right). \end{aligned}$$

For the term I_1 , one expects under suitable assumptions that

$$I_1 = \delta f'(y^\epsilon(t)) G((\Gamma y^\epsilon)(t), \xi^\epsilon(t)) + o_p(\delta). \quad (\text{A.5})$$

For the term I_2 , observe that, by independence of increments,

$$I_2 = E_t^\epsilon f(y^\epsilon(t) + \eta^\epsilon(t + \delta) - \eta^\epsilon(t)) = E_t^\epsilon f(x + \eta^\epsilon(t + \delta) - \eta^\epsilon(t)) \Big|_{x=y^\epsilon(t)} =: I_2(x) \Big|_{x=y^\epsilon(t)}. \quad (\text{A.6})$$

Using the Poisson structure of η^ϵ in (A.3), we further deduce that

$$I_2(x) = E f(x + \beta_\epsilon V - \beta_\epsilon \mu_V \lambda_\epsilon \epsilon^{-1} \delta) p_\delta - f(x - \beta_\epsilon \mu_V \lambda_\epsilon \epsilon^{-1} \delta) (1 - p_\delta) + o(\delta), \quad (\text{A.7})$$

where

$$p_\delta = P\left(\text{exactly one Poisson point lies in } R(\epsilon^{-1}(t + \delta)) \setminus R(\epsilon^{-1}t)\right) \quad (\text{A.8})$$

(and $(1 - p_\delta)$ in (A.7) is for the approximate probability that there are no Poisson points). Since the intensity (mean) measure of the Poisson random measure with point masses at $\{(\Gamma_k^\epsilon, V_k)\}$ is

$m(du, dv) = \lambda_\epsilon du F_V(dv)$, the probability p_δ in (A.8) is

$$\begin{aligned}
p_\delta &= m(R(\epsilon^{-1}(t+\delta)) \setminus R(\epsilon^{-1}t)) \\
&= \lambda_\epsilon \int_0^{\epsilon^{-1}t} du \int_{\epsilon^{-1}t-u}^{\epsilon^{-1}(t+\delta)-u} F_V(dv) + \lambda_\epsilon \int_{\epsilon^{-1}t}^{\epsilon^{-1}(t+\delta)} du \int_0^{\epsilon^{-1}(t+\delta)-u} F_V(dv) \\
&= \lambda_\epsilon \int_0^{\epsilon^{-1}t} du \left(F_V(\epsilon^{-1}(t+\delta) - u) - F_V(\epsilon^{-1}t - u) \right) + \lambda_\epsilon \int_{\epsilon^{-1}t}^{\epsilon^{-1}(t+\delta)} du F_V(\epsilon^{-1}(t+\delta) - u) \\
&= \lambda_\epsilon \int_{\epsilon^{-1}t}^{\epsilon^{-1}(t+\delta)} dv F_V(v) = \lambda_\epsilon F_V(\epsilon^{-1}t) \epsilon^{-1} \delta + o(\delta),
\end{aligned} \tag{A.9}$$

where we supposed that F_V is continuous. Hence,

$$\begin{aligned}
I_2(x) &= E f(x + \beta_\epsilon V) \lambda_\epsilon F_V(\epsilon^{-1}t) \epsilon^{-1} \delta + f(x) \\
&\quad - f(x) \lambda_\epsilon F_V(\epsilon^{-1}t) \epsilon^{-1} \delta - f'(x) \beta_\epsilon \mu_V \lambda_\epsilon \epsilon^{-1} \delta + o(\delta).
\end{aligned} \tag{A.10}$$

By using (4.14) and gathering the results above, we obtain that

$$\widehat{\mathcal{A}}^\epsilon f(y^\epsilon(t)) = J_{1,\epsilon}(y)|_{y=y^\epsilon} + J_{2,\epsilon}(x)|_{x=y^\epsilon(t)}, \tag{A.11}$$

where

$$J_{1,\epsilon}(y) = f'(y(t)) G((\Gamma y)(t), \xi^\epsilon(t)), \tag{A.12}$$

$$J_{2,\epsilon}(y) = E f(x + \beta_\epsilon V) \lambda_\epsilon F_V(\epsilon^{-1}t) \epsilon^{-1} - f(x) \lambda_\epsilon F_V(\epsilon^{-1}t) \epsilon^{-1} - f'(x) \beta_\epsilon \mu_V \lambda_\epsilon \epsilon^{-1}. \tag{A.13}$$

As in Sections 4.1 and 4.2, when $\epsilon \rightarrow 0$, the term J_1 averages out to $f'(y(t)) \overline{G}((\Gamma y)(t))$. By the choice of β_ϵ , $\lambda_\epsilon \epsilon^{-1} P(\beta_\epsilon V \in dr)$ converges vaguely (in the sense of, for example, Resnick (1987)) to the measure $c_V \alpha dr / r^{\alpha+1}$ on $(0, \infty)$, as $\epsilon \rightarrow 0$. Hence, as $\epsilon \rightarrow 0$,

$$J_{2,\epsilon}(x) \rightarrow c_V \int_0^\infty (f(x+r) - f(x) - f'(x)r) \frac{\alpha dr}{r^{\alpha+1}}$$

and the limit infinitesimal operator is expected as

$$(\mathcal{A}_t f)(y) = f'(y(t)) \overline{G}((\Gamma y)(t)) + c_V \int_0^\infty (f(x+r) - f(x) - f'(x)r) \frac{\alpha dr}{r^{\alpha+1}}. \tag{A.14}$$

This is precisely the infinitesimal operator of $y(t)$ in (4.6) with sLm $\eta(t)$ as described in Theorem 2.1. The latter fact follows from the form of infinitesimal operators for sLm found in Applebaum (2004), and connections between various characterizations of stable laws in Sato (1999).

B Proof of Lemma 4.1

We provide the proof for the reader's convenience and since an analogous proof in Kushner (1984) is not included for the continuous-time case considered here. We suppose that $\{y^\epsilon(t), \epsilon > 0, t \leq T\}$ is bounded. (If this is not the case, a truncation argument can be used as in Kushner (1984).) Directly from Theorem 1 in Kushner (1984), p. 39, we have the martingale property:

$$E \left(f^\epsilon(t+s) - f^\epsilon(t) - \int_t^{t+s} \widehat{\mathcal{A}}^\epsilon f^\epsilon(u) du \middle| \mathcal{F}_t^\epsilon \right) = 0. \tag{B.1}$$

Define the so-called ‘‘averaging intervals’’ δ_ϵ where $\delta_\epsilon \rightarrow 0$ and $\delta_\epsilon/\epsilon \rightarrow \infty$. It follows from (B.1) that, for bounded continuous $h : \mathbb{R}^k \rightarrow \mathbb{R}$, $s_0 \leq s_1 \leq \dots \leq s_k \leq t$,

$$Eh(y^\epsilon(s_i), i \leq k) \left(f^\epsilon(t+s) - f^\epsilon(t) - \sum_{k=\lfloor \frac{t}{\delta_\epsilon} \rfloor}^{\lfloor \frac{t+s}{\delta_\epsilon} \rfloor - 1} \int_{k\delta_\epsilon}^{(k+1)\delta_\epsilon} E_{k\delta_\epsilon}^\epsilon \widehat{\mathcal{A}}^\epsilon f^\epsilon(u) du \right) = O(\delta_\epsilon), \quad (\text{B.2})$$

where the ‘‘error’’ $O(\delta_\epsilon)$ arises from the discretization across the endpoints t and $t+s$ and using (4.16) in Assumption A3.

Consider now

$$\tilde{G}^\epsilon(v) = \frac{1}{\delta_\epsilon} \int_{k\delta_\epsilon}^{k\delta_\epsilon + \delta_\epsilon} E_{k\delta_\epsilon}^\epsilon \widehat{\mathcal{A}}^\epsilon f^\epsilon(u) du, \quad v \in [k\delta_\epsilon, (k+1)\delta_\epsilon]. \quad (\text{B.3})$$

Then, (B.2) can be written as

$$Eh(y^\epsilon(s_i), i \leq k) \left(f^\epsilon(t+s) - f^\epsilon(t) - \int_t^{t+s} \tilde{G}^\epsilon(v) dv \right) = O(\delta_\epsilon). \quad (\text{B.4})$$

Let

$$\tilde{\tilde{G}}^\epsilon(v) = \frac{1}{\delta_\epsilon} \int_{k\delta_\epsilon}^{k\delta_\epsilon + \delta_\epsilon} E_{k\delta_\epsilon}^\epsilon (\mathcal{A}_u f)(y^\epsilon) du, \quad v \in [k\delta_\epsilon, (k+1)\delta_\epsilon]. \quad (\text{B.5})$$

For fixed v , Assumption A4 ensures that

$$E \left| \tilde{G}^\epsilon(v) - \tilde{\tilde{G}}^\epsilon(v) \right| \rightarrow 0. \quad (\text{B.6})$$

Moreover, by (4.16) in Assumption A3, boundedness of $\{y^\epsilon(t), \epsilon > 0, t \leq T\}$ and continuity of the coefficients for $\mathcal{A}_u f$, we have

$$\sup_{\epsilon > 0, v \leq T} E \left| \tilde{G}^\epsilon(v) \right|, E \left| \tilde{\tilde{G}}^\epsilon(v) \right| < \infty. \quad (\text{B.7})$$

By the dominated convergence theorem and (4.15) in Assumption A3, one can conclude from (B.4) that

$$Eh(y^\epsilon(s_i), i \leq k) \left(f(y^\epsilon(t+s)) - f(y^\epsilon(t)) - \int_t^{t+s} (\mathcal{A}_u f)(y^\epsilon) du \right) = o(1), \quad (\text{B.8})$$

as $\epsilon \rightarrow 0$. Now take a weakly convergent subsequence (still indexing using ϵ) with limit y . Relation (B.8) and continuity of the coefficients for $\mathcal{A}_u f$ yields, as $\epsilon \rightarrow 0$,

$$Eh(y(s_i), i \leq k) \left(f(y(t+s)) - f(y(t)) - \int_t^{t+s} (\mathcal{A}_u f)(y) du \right) = 0, \quad (\text{B.9})$$

so that y solves the desired martingale problem. The uniqueness in Assumption A2 and the fact that the limit does not depend on the subsequence taken yields the statement of the theorem.

References

- Applebaum, D. (2004), *Lévy Processes and Stochastic Calculus*, Vol. 93 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.
- Buche, R. & Kushner, H. J. (2002), ‘Control of mobile communications with time-varying channels in heavy traffic’, *IEEE Transactions on Automatic Control* **47**(6), 992–1003. Special issue on systems and control methods for communication networks.
- Buche, R., Ghosh, A. & Pipiras, V. (2007), ‘Heavy traffic limits in a wireless queueing model with long range dependence’, *Decision and Control, 2007 46th IEEE Conference on* pp. 4447–4452.
- Costantini, C. & Kurtz, T. G. (2006), ‘Diffusion approximation for transport processes with general reflection boundary conditions’, *Mathematical Models & Methods in Applied Sciences* **16**(5), 717–762.
- Daley, D. J. & Vesilo, R. (1997), ‘Long range dependence of point processes, with queueing examples’, *Stochastic Processes and their Applications* **70**(2), 265–282.
- Daley, D. J., Rolski, T. & Vesilo, R. (2000), ‘Long-range dependent point processes and their Palm-Khinchin distributions’, *Advances in Applied Probability* **32**(4), 1051–1063.
- Debernardi, N. (2006), *Scheduling schemes adapted to heavy-tailed traffic on wireless high speed shared channels*, Masters Degree Project, Royal Institute of Technology, Stockholm, Sweden. Radio Communication Systems Laboratory, Department of Signals, Sensors, and Systems.
- Dębicki, K. & Mandjes, M. (2004), ‘Traffic with an fBm limit: convergence of the stationary workload process’, *Queueing Systems. Theory and Applications* **46**(1-2), 113–127.
- Delgado, R. (2007), ‘A reflected fBm limit for fluid models with ON/OFF sources under heavy traffic’, *Stochastic Processes and their Applications* **117**(2), 188–201.
- Doukhan, P., Oppenheim, G. & Taqqu, M. S., eds (2003), *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston, MA.
- Dupuis, P. (1993), ‘Sdes with oblique reflection directions on non-smooth domains’, *Annals of Probability* **21**, 554–580.
- Ethier, S. N. & Kurtz, T. G. (1986), *Markov Processes. Characterization and Convergence*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York.
- Fouque, J.-P., Garnier, J., Papanicolaou, G. & Sølna, K. (2007), *Wave Propagation and Time Reversal in Randomly Layered Media*, Vol. 56 of *Stochastic Modelling and Applied Probability*, Springer, New York.
- Ibragimov, I. A. & Linnik, Y. V. (1971), *Independent and Stationary Sequences of Random Variables*, Wolters-Nordhoff, The Netherlands.
- Jiang, M., Nikolic, M., Hardy, S. & Trajkovic, L. (2001), ‘Impact of self-similarity on wireless data network performance’, *Communications, 2001. ICC 2001. IEEE International Conference on* **2**, 477–481.
- Kalden, R. & Ibrahim, S. (2004), ‘Searching for self-similarity in GPRS’, *Proceedings of the 5th annual Passive and Active Measurement Workshop, Antibes Juan-les-Pins, France*.
- Karatzas, I. & Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, Vol. 113 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York.
- Konstantopoulos, T. & Lin, S.-J. (1996), Fractional Brownian approximations of queueing networks, in ‘Stochastic networks (New York, 1995)’, Vol. 117 of *Lecture Notes in Statist.*, Springer, New York, pp. 257–273.
- Krendzel, A., Koucheryavy, Y., Harju, J. & Lopatin, S. (2002), Network planning problems in 3G/4G wireless systems, Technical document (04)003, COST 290: Wi-QoST: Traffic and QoS Management in Wireless Multimedia Networks.

- Kulik, R. & Szekli, R. (2001), ‘Sufficient conditions for long-range count dependence of stationary point processes on the real line’, *Journal of Applied Probability* **38**(2), 570–581.
- Kurtz, T. G. (1992), Averaging for martingale problems and stochastic approximation, in ‘Applied stochastic analysis (New Brunswick, NJ, 1991)’, Vol. 177 of *Lecture Notes in Control and Inform. Sci.*, Springer, Berlin, pp. 186–209.
- Kushner, H. J. (1984), *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press Series in Signal Processing, Optimization, and Control, 6, MIT Press, Cambridge, MA.
- Kushner, H. J. (1990), *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Vol. 3 of *System & Control: Foundations & Applications*, Birkäuser, Boston, MA.
- Kushner, H. J. (2001), *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*, Vol. 47 of *Applications of Mathematics (New York)*, Springer-Verlag, New York. Stochastic Modelling and Applied Probability.
- Lee, I. & Fapojuwo, A. (2005), ‘Estimating heavy-tails in long-range dependent wireless traffic’, *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st* **4**, 2132–2136.
- Lowen, S. B. & Teich, M. C. (2005), *Fractal-Based Point Processes*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ.
- Marty, R. (2004), ‘Théorème limite pour une équation différentielle à coefficient aléatoire à mémoire longue’, *Comptes Rendus Mathématique. Académie des Sciences. Paris* **338**(2), 167–170.
- Marty, R. (2005), ‘Asymptotic behavior of differential equations driven by periodic and random processes with slowly decaying correlations’, *ESAIM. Probability and Statistics* **9**, 165–184 (electronic).
- Marty, R. & Sølna, K. (2008), Acoustic waves in long range random media, Preprint.
- Mikosch, T., Resnick, S., Rootzén, H. & Stegeman, A. (2002), ‘Is network traffic approximated by stable Lévy motion or fractional Brownian motion?’, *The Annals of Applied Probability* **12**(1), 23–68.
- Reed, J. & Ward, A. R. (2004), A diffusion approximation for a generalized jackson network with reneging, in ‘Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing’.
- Reiman, M. I. & Williams, R. J. (1993), ‘A boundary property of semimartingale reflecting Brownian motion’, *Annals of Probability* **21**, 554–580.
- Resnick, S. I. (1987), *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York.
- Resnick, S. I. (2007), *Heavy-Tail Phenomena*, Springer Series in Operations Research and Financial Engineering, Springer, New York. Probabilistic and statistical modeling.
- Ribeiro, L. & DaSilva, L. (2002), ‘A framework for the dimensioning of broadband mobile networks supporting wireless Internet services’, *Wireless Communications, IEEE [see also IEEE Personal Communications]* **9**(3), 6–13.
- Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable non-Gaussian Random Processes*, Stochastic Modeling, Chapman & Hall, New York. Stochastic models with infinite variance.
- Sato, K.-i. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Vol. 68 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge.
- Shao, Z. & Madhow, U. (2002), ‘A QoS framework for heavy-tailed traffic over the wireless internet’, *MILCOM 2002. Proceedings* **2**, 1201–1205.
- Skorokhod, A. V., Hoppensteadt, F. C. & Salehi, H. (2002), *Random Perturbation Methods with Applications in Science and Engineering*, Vol. 150 of *Applied Mathematical Sciences*, Springer-Verlag, New York.
- Stoylar, A. L. (2004), ‘Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic’, *The Annals of Applied Probability* **14**(1), 1–52.

- Stroock, D. W. (1975), ‘Diffusion processes associated with Lévy generators’, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32**(3), 209–244.
- Taqqu, M. S., Willinger, W. & Sherman, R. (1997), ‘Proof of a fundamental result in self-similar traffic modeling’, *Computer Communications Review* **27**(2), 5–23.
- Teymori, S. & Zhuang, W. (2007), ‘Queue analysis and multiplexing of heavy-tailed traffic in wireless packet data networks’, *Mobile Networks and Applications* **12**(1), 31–41.
- Ward, A. R. & Kumar, S. (2008), ‘Asymptotically optimal admission control of a queue with impatient customers’, *Mathematics of Operations Research* **33**(1), 167–202.
- Whitt, W. (2002), *Stochastic-Process Limits*, Springer Series in Operations Research, Springer-Verlag, New York. An introduction to stochastic-process limits and their application to queues.

Robert Buche
Department of Mathematics
North Carolina State University
Campus Box 8205
Raleigh, NC 27696, USA
rtbuche@unity.ncsu.edu

Arka P. Ghosh
Statistics Department
Iowa State University
Ames, IA 50011, USA
apghosh@iastate.edu

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Smith Bldg.
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu