Limits for Cumulative Input Processes to Queues

Ward Whitt[1]

AT&T Labs

August 2, 1999

Probability in the Engineering and Informational Sciences 14 (2000) 123–150

[1]AT&T Labs, Room A117, Shannon Laboratory, 180 Park Avenue, Florham Park, NJ 07932-0971; email: wow@research.att.com

Abstract

We establish functional central limit theorems (FCLTs) for a cumulative input process to a fluid queue from the superposition of independent on-off sources, where the on periods and off periods may have heavy-tailed probability distributions. Variants of these FCLTs hold for cumulative busy time and idle time processes associated with standard queueing models. The heavy-tailed on-period and off-period distributions can cause the limit process to have discontinuous sample paths, e.g., to be a non-Brownian stable process or more general Lévy process, even though the converging processes have continuous sample paths. Consequently, we exploit the Skorohod M_1 topology on the function space D of right-continuous functions with left limits. The limits here combined with the previously established continuity of the reflection map in the M_1 topology implies both heavy-traffic and non-heavy-traffic FCLTs for buffer-content processes in stochastic fluid networks.

Keywords: functional central limit theorems, invariance principles, heavy-traffic limit theorems, stochastic fluid networks, cumulative input processes, cumulative busy-time processes, heavy-tailed probability distributions, stable processes, Lévy processes, communication networks

1. Introduction

This is part of a series of papers devoted to obtaining approximations via limit theorems for stochastic fluid queues and stochastic fluid queueing networks with bursty input. Motivated by evolving communication networks, we represent the input at each queue (node) in the network as the input from a superposition of mutually independent on-off sources. Each source is alternately on and off for random periods of time. During on periods the source sends packets, represented as deterministic fluid, at constant rate; during off periods, the source is idle, not sending input. (We also consider the generalization in which the input during on periods is stochastic.)

We consider the case of a single-class fluid network. We let fluid be processed at each node in a first-come first-served (FCFS) manner at a constant rate and we stipulate that a proportion Q_{ij} of all fluid output from node i is immediately routed to queue j, where $Q \equiv (Q_{ij})$ is a substochastic matrix with $Q^n \to 0$ as $n \to \infty$; fluid not routed to another node leaves the network; see [18] for additional background.

Network measurements have revealed that the traffic carried on the communication networks is quite complex, exhibiting features such as long-range dependence, self-similarity and heavy-tailed probability distributions (having infinite variance); e.g., see [1], [8], [43]. This traffic complexity is evidently due, to a large extent, to the file sizes being transmitted over the networks having heavy-tailed probability distributions. We represent this phenomenon in our stochastic fluid network by allowing the source on periods to have heavy-tailed probability distributions. The on period represents the time a source is active, which will tend to be long when a large file is to be sent. Because of the fluid assumption, the cumulative input process from each source and the aggregate cumulative input process at each node have continuous sample paths, but the limit processes may have jumps due to the burstiness.

As a basis for developing useful approximations, we want to establish limit theorems for the buffer-content stochastic processes in these stochastic fluid networks. The limit theorems we have in mind are generalizations of heavy-traffic limit theorems for the same fluid models in which the on-periods do not have heavy-tailed distributions. Since evolving communication networks with bursty input may be required to operate far from the heavy-traffic regime, it is significant that our limit theorems do not require that the models be in the heavy-traffic regime. However, the heavy-traffic regime is a principal case. There are a variety of detailed assumptions that can be made about the distribution of the on-off stochastic processes and the scaling; e.g., see Konstantopoulos and Lin [20], Kurtz [21] and Taqqu, Willinger and Sherman [34]. For example, the number of sources can be allowed to go to infinity in the limit. Here we assume that there is a fixed number of sources, but we allow the individual sources to change in the limit process.

Even though the limits are not restricted to heavy-traffic, that is a useful reference case. Even with the usual independence condition, the heavy-tailed probability distributions have a dramatic impact on the heavy-traffic limiting behavior, making the limit become a reflected (non-Brownian) stable process or a more general reflected Lévy process instead of a reflected Brownian motion (RBM) as in Reiman [25]. Reflected stable and Lévy processes have independent increments; they arise when the on and off periods come from independent sequences of i.i.d. random variables or, more generally, under weak dependence. Limits with dependent increments, such as fractional stable processes [29], are also possible when there is more dependence.

In comparison with the usual heavy-traffic limits, the limits that we establish involve different scaling and have limit processes with different distributions. The limit processes also have sample paths with jumps, in contrast to the continuous sample paths of RBM. In order to obtain convergence of a sequence of stochastic processes with continuous sample paths to a limiting stochastic process with jumps, we need to replace the familiar Skorohod [30] J_1 topology on the function space $D \equiv D([0,\infty), \mathbb{R}^k)$ of right-continuous $\mathbb{R}^k$-valued functions with left limits with the Skorohod M_1 topology; [15, p. 301].

Thus, in [41] we established basic properties of the function space D with the Skorohod M topologies, and in [42] we showed that the multi-dimensional reflection map on D is Lipschitz continuous provided that a metric inducing the standard M_1 topology is used on the domain, while a metric inducing the weaker product M_1 topology is used on the range. We also applied that continuity result to establish functional central theorems (FCLTs) for the buffer-content processes in stochastic fluid networks. Those FCLTs show that a limit holds for the buffer-content process in the stochastic fluid network with a suitable scaling if a corresponding limit holds for the cumulative input process.

The purpose of this paper is to fill in the final step and establish FCLTs for the cumulative input process in the M_1 topology. Assuming that the different sources, at each node as well as at different nodes, are mutually independent, it suffices to establish a FCLT for the cumulative input process associated with a single source. (The sum of independent Lévy processes will be a new Lévy process.) When we consider more than one process, we use the product

M_1 topology on the product space $D \times \cdots \times D$. Convergence in the product M_1 topology extends to the standard M_1 topology on $D([0,\infty), \mathbb{R}^k)$ when the limit has discontinuities in only one coordinate at a time [41]. A sufficient condition is for the component processes to be independent without any fixed discontinuities.

Since we are considering on-off sources with fluid input, the cumulative input process of one source is essentially the same as the cumulative busy time process. Indeed, suppose that the input rate during on periods is λ and $B(t)$ and $C(t)$ are the cumulative busy time and input during the time interval $[0,t]$. Then $C(t) = \lambda B(t)$. Moreover, if $I(t)$ is the cumulative idle time in $[0,t]$, then $I(t) = t - B(t)$. Hence, FCLTs for $C(t)$, $B(t)$ and $I(t)$ are all essentially equivalent. We will focus on the cumulative busy time, $B(t)$. The results here thus also apply to busy-time and idle-time processes in other queueing models.

For a single queue, our results here for the case of heavy-tailed on-period distributions are very closely related to the non-Brownian limits in [40]. However, in that paper we considered a discrete-time model with heavy-tailed input distributions, for which it is possible to apply the familiar Skorohod J_1 topology throughout. Nevertheless, the same reflected stable processes and reflected Lévy processes are obtained as limit processes for the buffer content in our setting when we restrict attention to a single queue under independence conditions. The methods for calculating the probability distributions described there apply here as well. In particular, when the Laplace transforms can be characterized, numerical transform inversion can be used. However, those explicit results only apply to single queues. More work is needed to obtain explicit limiting distributions for stochastic fluid networks. For some results in this direction, see [17], [16].

Largely motivated by the traffic measurements, there has been growing interest in queues with heavy-tailed distributions. Thus there is a growing body of related work; see Boxma and Cohen [4], [5], Cohen [6], Furrer, Michna and Weron [13], Konstantopoulos and Lin [20], Kurtz [21], Resnick and Rootzén [26], Resnick and Samorodnitsky [27], Resnick and van den Berg [28] and Tsoukatos and Makowski [35], [36], [37]. The main contribution here is showing how to obtain results via the continuous mapping theorem exploiting the M_1 topology on D. Even though the M_1 topology was defined in 1956 by Skorohod [30], it has not received much attention.

Although our primary focus is on obtaining discontinuous limits in the M_1 topology, stemming from the heavy-tailed on-time distributions, we also discuss the standard case in which the cumulative-input limit process is Brownian motion in Sections 5 and 6. Then attention centers on identifying the variance constant.

2. Limits for the Cumulative Busy Time

Consider a queueing system in which there are alternating (necessarily positive) periods I_i and B_i in which the system is idle (off) and busy (on). We initially allow these random variables to be very general. In particular, we allow them to be mutually dependent and have infinite variance or even infinite mean. As a regularity condition, we assume that the number of busy cycles (idle period plus following busy period) in any finite interval $[0, t]$ is finite. We assume that the first idle period begins at time 0.

We now show how FCLTs for partial sums of the vectors (I_i, B_i) imply corresponding FCLTs for the cumulative busy-time process $B(t)$. To establish the FCLTs, we exploit results about the Skorohod M_1 topology in [30], [39], [41] and [24]. Let $\Rightarrow$ denote convergence in distribution and let $D \equiv D[0, \infty)$ denote the function space of right-continuous real-valued functions with left limits, endowed with the Skorohod M_1 topology. Let the σ-field on D be the Borel σ-field, which coincides with the usual Kolmogorov σ-field generated by the projection maps. Let $D^r \equiv (D, M_1)^r$ be the r-fold product space of D with itself, here always endowed with the product M_1 topology. Let C and C^r be the subsets of continuous functions, endowed with the relative topology, which corresponds to uniform convergence on all closed bounded intervals. When the limit processes belong to C^r, this becomes the familiar setting. For a random element of D, let $Disc(X)$ be the (random) set of discontinuities of X in $[0, \infty)$. Let $\stackrel{\mathrm{d}}{=}$ denote equality in distribution.

In general, we allow a sequence of models indexed by n, so that we start with a sequence of sequences $\{\{(I_{ni}, B_{ni}) : i \geq 1\} : n \geq 1\}$; I_{ni} is the i^{th} idle period in model n. Let $N_n(t)$ be the number of complete busy cycles (idle period plus the following busy period) in $[0, t]$ and let $B_n(t)$ be the cumulative busy time in $[0, t]$, both for model n. With a sequence of models it is possible to absorb the normalization constants into the processes, but we refrain from doing this, so that heavy-traffic limits for a single model are obtained by a direct application. We write $X_n(t) \Rightarrow X(t)$ as if we were talking about convergence of the marginal distributions in $\mathbb{R}$, but we establish the much stronger convergence in D. Weak convergence in D is indicated by "in D" written after the limit.

We first obtain an FCLT with time scaling by n and space scaling by c_n, where $nc_n \to \infty$ (e.g., $c_n = n^{-q}$ for $0 < q < 1$) and then afterwards obtain FCLTs with the space scaling by c_n where $nc_n \not\to \infty$ (e.g., $c_n = n^{-q}$ for $q \geq 1$). The standard case involving Brownian motion limits is $c_n = n^{-1/2}$. The case $n^{-1} < c_n < n^{-1/2}$ typically arises when the distribution of B_{ni} or I_{ni} has finite mean and infinite variance.

Theorem 2.1. *If*

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} [(I_{ni}, B_{ni}) - (m_{n,1}, m_{n,2})] \Rightarrow [X_1(t), X_2(t)] \quad \text{in } (D, M_1)^2 \text{ as } n \to \infty \ , \tag{2.1}$$

where $nc_n \to \infty$ *and* $m_{n,i} \to m_i$ *as* $n \to \infty$ *for* $i = 1, 2$, *with* $0 < m_1 + m_2 < \infty$ *and*

$$P(Disc(X_1) \cap Disc(X_2) \neq \phi) = 0 \ , \tag{2.2}$$

then

$$c_n[N_n(nt) - \gamma_n nt, B_n(nt) - \xi_n nt] \Rightarrow (-\gamma[X_1(\gamma t) + X_2(\gamma t)], (1-\xi)X_2(\gamma t) - \xi X_1(\gamma t)) \tag{2.3}$$

in $(D, M_1)^2$ *as* $n \to \infty$, *where*

$$\xi_n \equiv \frac{m_{n,2}}{m_{n,1} + m_{n,2}} \to \xi \quad \text{and} \quad \gamma_n \equiv \frac{1}{m_{n,1} + m_{n,2}} \to \gamma > 0 \ . \tag{2.4}$$

Proof. The idea is to repeatedly apply the continuous mapping theorem and its variants, as in Theorem 5.1 of Billingsley [3]. In particular, we invoke the Skorohod representation theorem [30], which allows us to replace random elements of a separable metric space converging in distribution with corresponding random elements defined on a new sample space having the same distributions that converge with probability one. In [41] it is shown that the space (D, M_1) is metrizable as a complete separable metric space, so that Skorohod's [30] representation theorem can be applied. (Even if (D, M_1) were not topologically complete, the representation would still be valid, because Dudley [9] showed that topological completeness is not needed.)

First we observe that the cumulative busy-time processes can be closely approximated by appropriate random sums. In particular, let the centered approximating processes be

$$B_n^a(t) - \xi_n t \equiv (1 - \xi_n) \sum_{i=1}^{N_n^B(t)} (B_{n,i} - m_{n,1}) - \xi_n \sum_{i=1}^{N_n^I(t)} (I_{n,i} - m_{n,2}) \ , \tag{2.5}$$

where $N_n^B(t)$ and $N_n^I(t)$ are the numbers of complete busy periods and idle periods up to time t in model n. Note that

$$N_n(t) = N_n^B(t) \leq N_n^I(t) \leq N_n(t) + 1 \quad \text{for all} \quad t \ . \tag{2.6}$$

Also note that

$$B_n^a(\tau_{n,k}) = B_n(\tau_{n,k}) \quad \text{and} \quad B_n^a(\tau'_{n,k}) = B_n(\tau'_{n,k}) \quad \text{for all} \quad k \geq 0 \, , \tag{2.7}$$

where $\tau_{n,0} = 0$,

$$\tau_{n,k} = I_{n,1} + B_{n,1} + \cdots I_{n,k} + B_{n,k}, \;\; k \geq 1 \;, \tag{2.8}$$

and

$$\tau'_{n,k} = \tau_{n,k} + I_{n,k+1}, \;\; k \geq 0 \;. \tag{2.9}$$

Moreover $B_n^a(t)$ is piecewise constant, while $B_n(t)$ is piecewise linear in each of the intervals $[\tau_{n,k}, \tau'_{n,k}]$ and $[\tau'_{n,k}, \tau_{n,k+1}]$. After appropriate scaling, the busy-period counting processes $N_n^B(t)$ and $N_n^I(t)$ are asymptotically equivalent to inverse partial sum processes. The partial sum processes are

$$S_n(\lfloor t \rfloor) = \sum_{i=1}^{\lfloor t \rfloor} (I_{ni} + B_{ni}), \quad t \geq 0 \;, \tag{2.10}$$

and the inverse map is

$$x^{-1}(t) = \inf\{s > 0 : x(s) > t\}, \quad t \geq 0 \;. \tag{2.11}$$

Note that

$$|N_n(t) - S_n^{-1}(\lfloor t \rfloor)| \leq 1 \;, \tag{2.12}$$

so that

$$c_n([N_n(nt) - n\gamma_n t] - [S_n^{-1}(\lfloor nt \rfloor) - n\gamma_n t]) \Rightarrow 0 \tag{2.13}$$

in D as $n \to \infty$.

Starting with the assumed limit (2.1), we consider the sum to get

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} [(I_{ni} + B_{ni}) - (m_{n,1} + m_{n,2})] \Rightarrow X_1 + X_2 \tag{2.14}$$

jointly with the limits in (2.1), invoking (2.2) and the analog of Theorem 4.1 of [39] for the M_1 topology, which is contained in [41]. (Equivalently, condition (2.2) allows us to replace convergence in the product M_1 topology in (2.1) by convergence in the strong M_1 topology on $D([0,\infty), \mathbb{R}^2)$, see [41].) We next get a limit for $N_n(t)$. To do so, we apply the Skorohod representation theorem and replace the convergence in distribution with convergence with probability one. We then apply the inverse map in (2.11), using (2.13) and Theorem 7.5 of [39] with $c'_n \equiv n(m_{n,1} + m_{n,2})c_n$ playing the role of c_n there and

$$x_n(t) = \frac{1}{n(m_{n,1} + m_{n,2})} \sum_{i=1}^{\lfloor nt \rfloor} (I_{ni} + B_{ni}) = (n\gamma_n^{-1})^{-1} S_n(\lfloor nt \rfloor) \;. \tag{2.15}$$

We first get $c'_n(x_n - e) \to x$ as $n \to \infty$, where e is the identity map and $x = X_1 + X_2$. Then we get $c'_n(x_n^{-1} - e) \to -x$ as $n \to \infty$, where $x_n^{-1} = n^{-1} S_n^{-1} \circ n\gamma_n^{-1} e$, so that

$$c_n(N_n(nt) - \gamma_n nt) \Rightarrow -\gamma[X_1(\gamma t) + X_2(\gamma t)] \;, \tag{2.16}$$

again jointly with the limits above. As a consequence of (2.16), we get

$$n^{-1}(N^I(nt), N^B(nt)) \Rightarrow (\gamma t, \gamma t) \quad \text{in} \quad (D, M_1)^2 \tag{2.17}$$

jointly with the limits above. Applying the continuous mapping theorem with the composition map, using (2.14), (2.5) and (2.17), we get

$$\mathbf{B}_n^a(t) \equiv c_n[B_n^a(nt) - \xi_n nt] \Rightarrow L(t) \equiv (1-\xi)X_2(\gamma t) - \xi X_1(\gamma t) \quad \text{in} \quad (D, M_1) \tag{2.18}$$

jointly with the previous limits. We will use (2.18) to get the desired limit

$$\mathbf{B}_n(t) \equiv c_n[B_n(nt) - \xi_n nt] \Rightarrow L(t) \quad \text{in} \quad (D, M_1) \;. \tag{2.19}$$

We now apply the Skorohod representation theorem to replace the convergence in distribution by convergence w.p.1. From the special version of $\mathbf{B}_n^a$ in (2.18) we can directly construct the associated special version of $\mathbf{B}_n$ in (2.19). For each continuity point t of the limit function L, we obtain $\mathbf{B}_n(t) \to L(t)$ w.p.1 from (2.18). From (2.5)–(2.9), we are able to bound the M_1 oscillation function of $\mathbf{B}_n$ in (2.19) over any finite interval $[0, T]$ by the corresponding oscillation function of $\mathbf{B}_n^a$ in (2.18). In particular,

$$w_s(\mathbf{B}_n, \delta) \le w_s(\mathbf{B}_n^a, 2\delta)$$

for all suitably large n, where

$$w_s(x, \delta) = \sup_{0 \vee (t-\delta) \le t_1 < t_2 < t_3 \le (t+\delta) \wedge T} \{|x(t_2) - [x(t_1), x(t_3)]|\}$$

and $[a, b]$ is the segment joining a and b, i.e. $[a, b] = \{\alpha a + (1-\alpha) b : 0 \le \alpha \le 1\}$. We can thus apply Theorem 6.1(iv) in [41] to establish convergence w.p.1 of the special versions of $\mathbf{B}_n$, jointly with the other quantities. That in turn implies the convergence in distribution of the original versions.

Remark 2.1. As a consequence of the FCLT (2.3), we obtain the functional weak law of large numbers (FWLLN)

$$n^{-1}[N_n(nt), B_n(nt)] \Rightarrow (\gamma t, \xi t) \tag{2.20}$$

in $(D, M_1)^2$ as $n \to \infty$. Since the limit in (2.20) is continuous, the M_1 convergence is equivalent to uniform convergence in bounded intervals. Note that the limit of $n^{-1}B_n(nt)$ in (2.20), ξt, trivially increasing in ξ, as we would expect. However, when $X_1 = 0$, from (2.3), the limit of $c_n|B_n(nt) - \xi nt|$, $(1-\xi)|X_2(\gamma t)|$, is *decreasing* in ξ. Upon reflection, this is consistent with intuition as well. For example, suppose that I_n is a deterministic value for all n, so that $X_1(t) = 0$. Then, as $\xi \to 1$, $B_n(t)$ approaches t, and we should anticipate that the limit of $c_n[B_n(nt) - \xi_n nt]$ approaches 0 as first $n \to \infty$ and then $\xi \to 1$, as implied by (2.3).

Remark 2.2. When we can establish condition (2.1), even with a discontinuous limit, it will typically be possible to obtain convergence in the stronger Skorohod J_1 topology; e.g., by applying results in Jacod and Shiryaev [15]. However, we cannot as a consequence obtain the conclusion (2.3) in the J_1 topology, unless the limit process has continuous paths. Since the normalized cumulative-busy-time processes have continuous sample paths, the M_1 topology is needed in the final limit. The M_1 topology is also needed to get (2.16) for the counting processes via the inverse map. Theorem 7.4 of [39] shows that a limit does not hold in the J_1 topology for limit processes with discontinuities.

Remark 2.3. The discontinuity condition $P(Disc(X_1) \cap Disc(X_2) \neq \phi) = 1$ in Theorem 2.1 is obviously automatically satisfied if one of the two limit processes X_1 and X_2 has continuous paths, i.e. if $P(X_i \in C) = 1$ for one i. In fact, both have continuous paths in the standard short-range-dependence finite-variance case, in which they are Brownian motions, which we consider in Section 3. Otherwise, the discontinuity condition is automatically satisfied if X_1 and X_2 are independent processes without fixed discontinuities.

Remark 2.4. If the scaling is not by $c_n = n^{-1/2}$, then often the busy periods will dominate the idle periods in the sense that $X_1(t) = 0$, $t \geq 0$, in (2.1). Then the discontinuity condition (2.2) is trivially satisfied and the limit process in (2.3) simplifies. Indeed, it is a simple time-and-space rescaling of the limit X_2 in (2.1). ■

Assuming that the limit process in Theorem 2.1 is continuous at t with probability one, we obtain the associated ordinary CLT (convergence of marginal distributions) in $\mathbb{R}^2$ by applying the continuous mapping theorem with the projection map $\pi_t : D^2 \to \mathbb{R}^2$, defined by $\pi_t(x_1, x_2) = (x_1(t), x_2(t))$.

The limit processes in (2.1) and (2.3) will often be self-similar, i.e., the finite-dimensional distributions will satisfy

$$[X(ct_1),\ldots,X(ct_n)] \stackrel{\mathrm{d}}{=} [c^H X(t_1,\ldots,c^H(t_n)] \tag{2.21}$$

for all n, $0 < t_1 < \cdots < t_n$ and $c > 0$ see [29, p. 311]. Then H is the self-similarity index. Indeed, in the case that there is only a single model, i.e., in which $\{(I_i, B_i)\}$ are not indexed by n, the limit process is necessarily self-similar. We summarize the observation.

Theorem 2.2. *If the conditions of Theorem* 2.1 *hold with* $c_n = n^{-q}$ *for a single model* $\{(I_i, B_i) : i \geq 1\}$*, then the limit process* $[X_1(t), X_2(t)]$ *in* (2.1) *is self-similar with index* q *and the limit process in* (2.3) *is distributed as*

$$(1-\xi)\gamma^q X_2(t) - \xi\gamma^q X_1(t)\ . \tag{2.22}$$

Proof. Consider the limits two ways, first replacing t by γt and then replacing n by γn in the time argument nt. ∎

We next consider the case in which the partial sums of the busy periods satisfies a FCLT with normalization c_n where $nc_n \to c \leq 1$, which corresponds to the occurrence of exceptionally long busy periods. When $q > 1$, the average busy period $\bar{B}_n = (B_{n1} + \cdots + B_{nn})/n$ is diverging to $+\infty$ as $n \to \infty$. We assume that the idle times satisfy a functional weak law of large numbers (FWLLN) with the usual normalization n^{-1}. The dual case involving large idle times and standard busy times is covered by just changing the names. We consider the case in which both idle and busy times are large afterwards.

Let e denote the identity map on $[0,\infty)$. Let C^k_m be the subset of functions in C^k with each coordinate function being monotone. Let $D^1_\uparrow$ be the subset of nondecreasing nonnegative functions in D^1.

Theorem 2.3. *If*

$$n^{-1}\sum_{i=1}^{\lfloor nt\rfloor} I_{ni} \Rightarrow m_1 t \quad \textit{in } (D, M_1) \textit{ as } n \to \infty \tag{2.23}$$

and

$$c_n \sum_{i=1}^{\lfloor nt\rfloor} B_{ni} \Rightarrow X_2(t) \quad \textit{in } (D, M_1) \textit{ as } n \to \infty\ , \tag{2.24}$$

where $nc_n \to c \leq 1$ *and* $P(X_2^{-1}(0) = 0) = 1$*, then*

$$n^{-1}(N_n(c_n^{-1}t), I_n(c_n^{-1}t), B_n(c_n^{-1}t) - c_n^{-1}t) \Rightarrow (Y_1(t), Y_2(t), -Y_2(t)) \quad \textit{in } (D, M_1)^3 \tag{2.25}$$

as $n \to \infty$, where

$$Y_1(t) = \begin{cases} X_2^{-1}, & c = 0 \\ (m_1 e + X_2)^{-1}(t), & c = 1 \end{cases} \tag{2.26}$$

and

$$Y_2(t) = m_1 Y_1(t) \ . \tag{2.27}$$

Proof. We modify the proof of Theorem 2.1. First, by Theorem 4.4 of Billingsley [3], since the limit in (2.23) is deterministic, the separate convergence in (2.23) and (2.24) implies joint convergence. We first assume that $c = 0$. Multiplying by nc_n in (2.23), we obtain

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} (I_{ni}, B_{ni}) \Rightarrow (0, X_2(t)) \quad \text{in } D^2 \text{ as } n \to \infty \ . \tag{2.28}$$

Then, adding, we get

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} (I_{ni} + B_{ni}) \Rightarrow X_2(t) \quad \text{in } D \text{ as } n \to \infty \ . \tag{2.29}$$

joint with (2.23) and (2.24). Using (2.12) and the inverse map, from (2.29) we get

$$n^{-1} N_n(c_n t) \Rightarrow X_2^{-1}(t) \quad \text{in } D \text{ as } n \to \infty \ , \tag{2.30}$$

again joint with the limits above. We use Lemma 2.1 of [24], which requires the assumed condition $P(X_2^{-1}(0) = 0) = 1$. Note that X_2 necessarily has nondecreasing sample paths since $I_{ni} \geq 0$ and $B_{ni} \geq 0$ for all n and i. The processes B and I can be treated at once because $B(t) = t - I(t)$. We apply composition with (2.23) and (2.30) to get

$$n^{-1} \sum_{i=1}^{N_n(c_n^{-1} t)} I_{ni} \Rightarrow m_1 X_2^{-1}(t) \quad \text{in } D \text{ as } n \to \infty \tag{2.31}$$

jointly with (2.30), using Theorem 9.1 of [41] with the limit $(x, y) \in C_m^k \times D_\uparrow^1$ there, i.e., exploiting the fact that $m_1 e$ is continuous and monotone. Next note that

$$\left| n^{-1} I_n(c_n^{-1} t) - n^{-1} \sum_{i=1}^{N_n(c_n^{-1} t)} I_{n,i} \right| \leq n^{-1} I_{n, N_n(c_n^{-1} t) + 1} \ . \tag{2.32}$$

By (2.23), $n^{-1} I_{n, \lfloor nt \rfloor} \Rightarrow 0$ in (D, M_1). That with (2.30) implies that

$$n^{-1} I_{n, N_n(c_n^{-1} t) + 1} \Rightarrow 0 \quad \text{in} \quad (D, M_1) \ . \tag{2.33}$$

Hence, we have established (2.25) in the case $c = 0$. When $c = 1$, (2.28) and (2.29) hold with the limits changed to $(m_1 t, X_2(t))$ and $m_1(t) + X_2(t)$, respectively. Thus, (2.30) holds with the limit changed to $(m_1 e + X_2)^{-1}(t)$. Similarly, (2.31) holds with the limit process changed to $m_1 (m_1 e + X_2)^{-1}(t)$, so that we have (2.25)–(2.27) when $q = 1$. ∎

Remark 2.5. Just as noted in Remark 2.2 about Theorem 2.1, we typically will be able to obtain convergence in the stronger J_1 topology in condition (2.24). However, we use the M_1 topology when we work with the inverse map. The inverse map is continuous in the J_1 topology if the limit is strictly increasing; see Theorem 7.2 of [39]. Since the processes in (2.24) are nondecreasing, the M_1 topology is equivalent to pointwise convergence on a dense subset.

The following lemma helps to apply Theorem 2.3.

Lemma 2.4. *Let X be an element of $D^1_\uparrow$. For all $s, t > 0$ such that $P(X(s) = t) = 0$,*

$$P(X^{-1}(t) \le s) = P(X(s) > t) \ .$$

Proof. From (2.11),

$$\{X(s) > t\} \subseteq \{X^{-1}(t) \le s\} \subseteq \{X(s) \ge t\} \ ,$$

so that

$$P(X(s) > t) \le P(X^{-1}(t) \le s) \le P(X(s) \ge t)$$

for all s, t. Under the extra condition, the two outer probabilities are equal. ■

Now we consider the case in which both the partial sums of I_{ni} and B_{ni} both have a nondegenerate limits without translation terms. In general, we have difficulty if the limit process X_1 has discontinuities. Hence the following theorem seems less useful.

Let $C_{\uparrow\uparrow}$ be the subset of strictly increasing nonnegative functions in C.

Theorem 2.5. *Suppose that*

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} (I_{ni}, B_{ni}) \Rightarrow (X_1(t), X_2(t)) \quad \text{in } (D, M_1)^2 \text{ as } n \to \infty \tag{2.34}$$

where $c_n \to 0$ as $n \to \infty$. (a) If X_1 and X_2 are independent processes without fixed discontinuities and $(X_1 + X_2)^{-1}(0) = 0$, then

$$n^{-1} N_n(c_n^{-1} t) \Rightarrow (X_1 + X_2)^{-1}(t) \quad \text{in } (D, M_1) \text{ as } n \to \infty \ . \tag{2.35}$$

(b) If, in addition,

$$P((X_2, (X_1 + X_2)^{-1}) \in (C_m \times D^1_\uparrow) \cup (D \times C_{\uparrow\uparrow})) = 1 \ , \tag{2.36}$$

then

$$(n^{-1} N_n(c n^{-1} t), c_n I_n(c_n^{-1} t)) \Rightarrow ((X_1 + X_2)^{-1}(t) \ , X_2 \circ (X_1 + X_2)^{-1})(t) \quad \text{in} \quad (D, M_1)^2 \tag{2.37}$$

as $n \to \infty$.

Proof. The argument is similar to that for Theorems 2.1 and 2.3. For both parts, the conditions on the limit processes X_1 and X_2 allow us to apply the continuous mapping theorem with addition to get

$$c_n \sum_{i=1}^{\lfloor nt \rfloor} (I_{ni} + B_{ni}) \Rightarrow X_1(t) + X_2(t) \tag{2.38}$$

joint with (2.34). Using (2.12) and the inverse map (2.11), from (2.38) we get (2.35), again joint with the limits above. We again use Lemma 2.1 of [24]. Note that the limit processes X_i in (2.34) necessarily have nondecreasing nonnegative sample paths because $I_{ni} \geq 0$ and $B_{ni} \geq 0$ for all n and i. Turning to (2.37) in part (b), we use the extra condition to justify applying the continuous mapping theorem with composition, using Theorem 9.1 of [41]. We use the argument in the proof of Theorem 2.1 to relate the cumulative busy time process to the random sum.

3. Random Input During On Periods

We have indicated that the limits for the cumulative busy time $B(t)$ in Section 2 translate immediately into corresponding limits for the cumulative input $C(t)$ for an on-off source when the input during the on periods is always at a constant rate λ; then $C(t) = \lambda B(t)$. In this section we consider the more general situation in which the input during on periods occurs randomly according to a stochastic process $\{\Lambda(t) : t \geq 0\}$ with nondecreasing sample paths. Now we assume that $\Lambda(0) = 0$ and

$$C(t) = \Lambda(B(t)), \quad t \geq 0 . \tag{3.1}$$

Definition (3.1) means that input for the source is generated from the stochastic process $\{\Lambda(t) : t \geq 0\}$ whenever the process is on, with successive increments from the same stochastic process $\{\Lambda(t) : t \geq 0\}$ being used whenever the source turns on. A simple case naturally covered by (3.1) is when $\{\Lambda(t) : t \geq 0\}$ and $\{B(t) : t \geq 0\}$ are independent stochastic processes with $\{\Lambda(t) : t \geq 0\}$ having stationary and independent increments. However, (3.1) can apply usefully in much more general situations.

A first general result is a direct consequence of the M_1 limit under a random time change in Theorem 11.2 of [41]. We use the M_1 topology in the condition because that is the mode of convergence obtained from Theorem 2.1. Note that the condition below for $B_n(t)$ corresponds to the conclusion of Theorem 2.1 here, in the situation considered there.

Theorem 3.1. *If*

$$c_n[B_n(nt) - \xi_n nt, \Lambda_n(nt) - \lambda_n nt] \Rightarrow [X_1(t), X_2(t)] \quad \text{in } (D, M_1)^2 \text{ as } n \to \infty \ , \tag{3.2}$$

where $nc_n \to \infty$, $\xi_n \to \xi$ *and* $\lambda_n \to \lambda$ *as* $n \to \infty$ *and*

$$P(Disc(X_2 \circ \xi e) \cap Disc(X_1) = \phi) = 1 \ , \tag{3.3}$$

then

$$c_n[C_n(nt) - \lambda_n \xi_n nt] \Rightarrow X_2(\xi t) + \lambda X_1(t) \quad \text{in } (D, M_1) \text{ as } n \to \infty \ . \tag{3.4}$$

Proof. Note that

$$\begin{aligned} c_n[C_n(nt) - \lambda_n \xi_n t] &= c_n[(\Lambda_n(nt) - \lambda_n nt) \circ n^{-1} B_n(nt) + \lambda_n (B_n(nt) - \xi_n nt)] \\ &\Rightarrow X_2(\xi t) + \lambda X_1(t) \end{aligned}$$

by Theorem 11.2 of [41]. ∎

There are two sources of variability in Theorem 3.1: the two processes Λ_n and B_n. When the nonstandard scaling with $c_n \neq n^{-1/2}$ occurs in condition (3.2) of Theorem 3.1, we should anticipate that the processes B_n and Λ_n typically will require different normalizations in order to have nondegenerate limits. Thus, we regard the case in which either X_1 or X_2 in (3.2) is the zero process as the common case with $c_n \neq n^{-1/2}$. That is fortunate because the limit (3.2) will then typically be easier to verify.

Moreover, Theorem 3.1 invites us to compare the two sources of variability in applications and determine which dominates, which could conceivably vary from situation to situation. However, if burstiness is primarily due to exceptionally long on periods, then we should anticipate that X_2 will be the zero process; i.e., the fluctuations in Λ_n should be asymptotically negligible compared to the fluctuations in B_n, so that $P(X_2(t) = 0) = 1$ in (3.2). Then we can treat the two components in (3.2) separately, applying Theorem 4.4 of Billingsley [3]. Moreover, it is not necessary to identify a nondegenerate limit for Λ_n, which necessarily must involve a different scaling. We may well be able to deduce that

$$c_n[\Lambda_n(nt) - \lambda_n nt] \Rightarrow 0 \quad \text{in } (D, M_1) \text{ as } n \to \infty \tag{3.5}$$

for quite general processes Λ_n (without requiring independent increments). Finally, in this case the limit process is the same as if $\Lambda(t) = \lambda t$, as assumed in Section 2.

We can combine Theorems 2.1 and 3.1 to show that the limit for the cumulative input processes C_n after appropriate normalization is just a deterministic scaling of the limit process

X_2 for the partial sums of the busy times when the idle times and the processes Λ_n are asymptotically negligible compared to the busy times.

Corollary 3.2. *If condition* (2.1) *in Theorem* 2.1 *holds with* $P(X_1(t) = 0) = 1$ *for all* t*, and if*

$$c_n[\Lambda_n(nt) - \lambda_n nt] \Rightarrow 0 \quad \text{in } (D, M_1) \text{ as } n \to \infty \ ,$$

where $\lambda_n \to \lambda$ *as* $n \to \infty$*, then*

$$c_n[C_n(nt) - \xi_n \lambda_n nt] \Rightarrow \lambda(1 - \xi)X_2(\gamma t) \quad \text{in } (D, M_1) \text{ as } n \to \infty,$$

where X_2*,* γ *and* ξ *are as in Theorem* 2.1.

4. Sufficient Conditions

The main remaining problem is to provide useful sufficient conditions for the conditions in the theorems in Sections 2 and 3, especially condition (2.1) in Theorem 2.1. As noted in Remark 2.2, this condition requires convergence in the M_1 topology, but we typically will be able to establish the required convergence in (2.1) in the stronger J_1 topology, even with heavy-tailed probability distributions and discontinuous sample paths. Indeed, Jacod and Shiryaev [15] give numerous sufficient conditions for convergence of the form (2.1), to processes with discontinuous sample paths, all in the J_1 topology.

The theorems in Sections 2 and 3 do not require any independence for the underlying random variables B_{ni} and I_{ni}, but that is an important special case. In particular, if we assume that the pairs (I_{ni}, B_{ni}) for $i \geq 1$ are i.i.d. for each n, then condition (2.1) falls into the classical setting of limits for triangular arrays of partial sums of i.i.d. random vectors, for which the limits are known to be Lévy processes. Such limits, with applications to the single-server queue are discussed in [40]. (However, there the summands have a different interpretation than busy and idle periods.)

The standard framework for heavy-traffic limit theorems for queues involves a sequence of queueing processes associated with a sequence of queueing models, which we take to be indexed by n. The condition of heavy-traffic is achieved by having the associated traffic intensities ρ_n approach 1, the critical level for stability, from below as $n \to \infty$. The queueing models can change quite generally with n, but it often suffices to consider essentially a single model, letting the n^{th} arrival (cumulative input) process be a simple time-scaling of a single reference arrival process. We can achieve the same simple scaling in our idle-busy cycles by letting the idle and busy periods in model n be obtained by simply scaling the idle and busy periods in a single reference system. In particular, suppose that I_{ni} and B_{ni} are defined in terms of I_i and B_i by letting

$$I_{ni} = \alpha_n I_i \quad \text{and} \quad B_{ni} = \beta_n B_i \tag{4.1}$$

where $\alpha_n \to \alpha$ and $\beta_n \to \beta$ as $n \to \infty$.

With the framework (4.1), we can easily establish the conditions in the theorems in Section 2 using limits for the single sequence $\{(I_i, B_i) : i \geq 1\}$. We state the elementary result for Theorem 2.1.

Theorem 4.1. *Suppose that* (4.1) *holds with* $\alpha + \beta > 0$,

$$c_n \left[\sum_{i=1}^{\lfloor nt \rfloor} (I_i, B_i) - (\hat{m}_1, \hat{m}_2) \right] \Rightarrow (Y_1, Y_2) \quad \textit{in } (D, M_1)^2 \textit{ as } n \to \infty, \tag{4.2}$$

where $nc_n \to \infty$ *as* $n \to \infty$, $\hat{m}_1 + \hat{m}_2 > 0$ *and*

$$P(Disc(Y_1) \cap Disc(Y_2) \neq \phi) = 0 \ . \tag{4.3}$$

Then the conditions and conclusions of Theorem 2.1 *hold with* $X_1 = \alpha Y_1$, $X_2 = \beta Y_2$, $m_{n1} = \alpha_n \hat{m}_1$, $m_{n2} = \beta_n \hat{m}_2$, $m_1 = \alpha \hat{m}_1$, $m_2 = \beta \hat{m}_2$ *and* $\gamma = (m_1 + m_2)^{-1} > 0$.

For the single-model framework in (4.1), it suffices to establish condition (4.2). If, as above, we assume that the pairs (I_i, B_i) for $i \geq 1$ are i.i.d., then we are in the restricted classical setting of limiting stable processes. We will elaborate since this seems to be the most relevant for treating heavy-tailed on periods.

We now discuss non-standard limits when there is essentially a single model with i.i.d. busy cycles, where the busy-period cdf has a heavy tail but the idle-period cdf does not. In that setting, it turns out that a nonstandard limit in one of Theorem 2.1 or 2.3 holds with $c_n = n^{-1/\alpha}$ if and only if the busy-period cdf has a power tail with decay rate $x^{-\alpha}$ for $0 < \alpha < 2$. (If we allow more general normalization constants, then the busy-period cdf tail can be regularly varying.) Under that condition, the limit process $X_2(t)$ becomes a *stable Lévy motion* (totally skewed to the right and centered), by which we mean that $X_2(0) = 0$, $\{X_2(t) : t \geq 0\}$ has stationary and independent increments, and $X(t) - X(s) \stackrel{\mathrm{d}}{=} S_\alpha(\sigma(t-s)^{1/\alpha}, 1, 0)$, where $S_\alpha(\sigma, \beta, \mu)$ denotes a stable probability law on $\mathbb{R}$ with index α $(0 < \alpha \leq 2)$, scale parameter σ, skewness parameter β $(-1 \leq \beta \leq 1)$ and location (or shift) parameter μ as in Samorodnitsky and Taqqu [29], to which we refer for background. In particular, the logarithmic characteristic function of an $S_\alpha(\sigma,\beta,\mu)$ variable X is

$$\log Ee^{i\theta X} = \begin{cases} -\sigma^\alpha|\theta|^\alpha(1-i\beta(\operatorname{sign}\theta)\tan(\pi\alpha/2)+i\mu\theta, & \alpha\neq 1 \\ -\sigma|\theta|(1+i\beta\frac{2}{\pi}(\operatorname{sign}\theta)\ln(|\theta|)+i\mu\theta, & \alpha=1\ , \end{cases} \tag{4.4}$$

where $\operatorname{sign}\theta=+1$, 0 or -1 for $\theta>0$, $\theta=0$ and $\theta<0$. The stable law is skewed totally to the right when $\beta=1$ and skewed totally to the left when $\beta=-1$; we are interested in the case $\beta=1$. It is centered when $\mu=0$.

A random variable X distributed as $S_\alpha(\sigma,1,0)$ for $\alpha<2$ has a cdf with power upper tail decaying as $x^{-\alpha}$; in particular,

$$\lim_{x\to\infty} x^\alpha P(X>x) = K_\alpha\sigma^\alpha\ , \tag{4.5}$$

where

$$K_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \alpha\neq 1 \\ \frac{2}{\pi}, & \alpha=1 \end{cases} \tag{4.6}$$

and $\Gamma(x)$ is the gamma function. For $0<\alpha<1$, the stable law $S_\alpha(\sigma,1,0)$ is concentrated on the positive half line, and the associated stable process has nonnegative nondecreasing sample paths. In that case, the positively skewed stable Lévy motion is called a *stable subordinator.* For $1\le\alpha<2$, the skewed stable law $S_\alpha(\sigma,1,0)$ has support on the entire line, but it decays faster than exponentially. If X has the $S_\alpha(\sigma,1,0)$ law, then the logarithm of its Laplace transform (defined only for real positive s for $\alpha\ge 1$) is

$$\log Ee^{-sX} = \begin{cases} -\sigma^\alpha s^\alpha/\cos(\pi\alpha/2) & \text{if}\quad \alpha\neq 1 \\ 2\sigma s\ln(s)/\pi & \text{if}\quad \alpha=1\ , \end{cases} \tag{4.7}$$

for $\mathrm{Re}(s)>0$. Closed-form representations for stable pdf's and cdf's are available in only a very few cases, but numerical calculations can be done exploiting finite-interval integral representations in Section 2.2 of Zolotarev [44]. These integral representations have been applied to generate tables of pdf, cdf and factile values, as indicated in Section 1.6 of Samorodnitsky and Taqqu [29].

With this background, we can state the basic limit theorem. We omit the somewhat pathological boundary case of $\alpha=1$. Note that our assumptions are about I_1 and B_1 separately; we need not make any assumption about the joint distribution of I_1 and B_1.

Theorem 4.2. *Consider a single model with i.i.d. busy cycles in which $EI_1^2<\infty$.*

(a) The conditions of Theorem 2.1 *(or Theorem* 4.1*) with normalization constants $c_n=n^{-1/\alpha}$ hold for $1<\alpha<2$ if and only if there is a constant K for which*

$$\lim_{x\to\infty} x^\alpha P(B_1>x) = K\ , \tag{4.8}$$

in which case $m_1 = EI_1$, $m_2 = EB_1 < \infty$ *and the limit process* $[X_1(t), X_2(t)]$ *has* $X_1(t) = 0$ *and* $X_2(t)$ *stable Lévy motion with marginals distributed as* $S_\alpha(\sigma t^{1/\alpha}, 1, 0)$ *for* $\sigma = (K/K_\alpha)^{1/\alpha}$ *with* K *in* (4.8) *and* K_α *in* (4.5). *The limit process* $(1-\xi)X_2(\gamma t)$ *in* (2.3) *has one-dimensional marginals at* t *distributed as* $S_\alpha(\sigma t^{1/\alpha}, 1, 0)$, *where* $\sigma = (1-\xi)(K\gamma/K_\alpha)^{1/\alpha}$ *for* $\xi = \rho = EB_1/(EB_1 + EI_1)$ *and* γ *in* (2.4) *and* K *and* K_α *as above.*

(b) The conditions of Theorem 2.3 *with normalization constants* $c_n = n^{-1/\alpha}$ *hold for* $0 < \alpha < 1$ *if and only if* (4.8) *holds, in which case* $X_2(t)$ *is the stable subordinator, with marginals distributed as* $S_\alpha(\sigma t^{1/\alpha}, 1, 0)$ *for* $\sigma = (K/K_\alpha)^{1/\alpha}$ *with* K *in* (4.8) *and* K_α *in* (4.5). *The limit process* $-m_1 X_2^{-1}(t)$ *in* (2.26) *and* (2.27) *has marginal distribution*

$$P(-m_1 X_2^{-1}(t) \geq -x) = P(m_1 X_2^{-1}(t) \leq x) = P(X_2(x/m_1) > t) \tag{4.9}$$

with $X_2(x/m_1)$ *being distributed as* $S_\alpha(\sigma(x/m_1)^{1/\alpha}, 1, 0)$.

We omit the proof of Theorem 4.2 because it is contained in Theorems 3.3 and 3.8 of [40], which draws on Feller [12] and Jacod and Shiryaev [15]. See [40] for further discussion.

5. Limits for a Single Queue

We now combine the previous results to obtain FCLTs for a single-server fluid queue fed by the superposition of k independent on-off sources with heavy-tailed on periods. We construct a sequence of models indexed by n, letting ρ_n be the traffic intensity for model n.

As in [40], we consider a single-server queue with finite waiting space, which is defined by the two-sided reflection map R. The reflection map R takes elements $x \in D \equiv D([0,T], \mathbb{R})$ with $0 \leq x(0) \leq C$ into $(z, l, u) \equiv (\phi(x), \psi_l(x), \psi_u(x))$ in D^3, where

$$z(t) = x(t) + l(t) - u(t), \quad t \geq 0\ , \tag{5.1}$$

l and u have nondecreasing sample paths with $l(0) = u(0) = 0$, $l(t)$ increases only when $z(t) = 0$ and $u(t)$ increases only when $z(t) = C$, i.e.,

$$\int_0^\infty z(t)dl(t) = \int_0^\infty [C - z(t)]du(t) = 0\ . \tag{5.2}$$

As shown in [42], the reflection map on $D([0,\infty), \mathbb{R})$ into $D([0,\infty), \mathbb{R}^3)$ is continuous provided the product M_1 (PM_1) topology is used on the range. If we focus on the one-dimensional buffer content process z, the topologies on the domain and range become just M_1 on $D([0,\infty), \mathbb{R}^1)$.

We first establish a general limit in the setting of Section 2. Let $\{Z_n(t) : t \geq 0\}$ be the buffer-content stochastic process in model n with buffer capacity C_n and fluid processing rate r_n. Let e be the identity map on $[0, \infty)$. We obtain the following from Theorem 2.1 by applying the continuous mapping theorem with the reflection map above.

Theorem 5.1. *If the conditions of Theorem* 2.1 *hold with* $C_n = c_n^{-1}C$, $c_n Z_n(0) \Rightarrow Z(0)$ *as* $n \to \infty$ *in* $\mathbb{R}$ *and*

$$nc_n(\lambda\xi_n - r_n) \to c \tag{5.3}$$

as $n \to \infty$, *then*

$$Z_n \Rightarrow \phi(\lambda(1-\xi)X_2 \circ \gamma e - \lambda\xi X_1 \circ \gamma e + ce) \quad in \quad (D, M_1) \quad as \quad n \to \infty\,, \tag{5.4}$$

where ϕ *is the first* (*content*) *component of the reflection map.*

Proof. By (5.3),

$$\begin{aligned} c_n Z_n(nt) &= \phi(\{c_n[\lambda B(nt) - r_n nt]\}) \\ &= \phi(\{c_n[\lambda B(nt) - \lambda\xi_n nt] + (\lambda\xi_n - r_n)n^{1-q}t\}) \\ &\to \phi(\{\lambda(1-\xi)X_2(\gamma t) - \lambda\xi X_1(\gamma t) + ct\}) \end{aligned}$$

by the continuity of the reflection map. This two-sided reflection map with the M_1 topology is discussed in Section 10 of [42]. ■

In the standard heavy-traffic applications of Theorem 5.1, $\xi_n \to \xi > 0$ and $r_n \to r > 0$, so that $\lambda\xi_n - r_n \to 0$ and $\rho_n \equiv \lambda\xi_n/r_n \to 1$. However, we can have non-heavy-traffic applications by having $nc_n\lambda\xi_n \to a > 0$ and $nc_n r_n \to b > 0$, so that $c = a - b$ and $\rho_n \equiv \lambda\xi_n/r_n \to a/b$, where a/b can be any positive value. Then $\xi = 0$ and the limit in (5.4) simplifies.

Corollary 5.2. *If, in addition to conditions of Theorem* 5.1, $nc_n\lambda\xi_n \to a > 0$ *and* $nc_n r_n \to b > 0$ *as* $n \to \infty$. *Then*

$$Z_n \Rightarrow \phi(\lambda X_2 \circ \gamma e + ce) \quad in \quad (D, M_1) \quad as \quad n \to \infty\ ,$$

where X_2 *has nondecreasing sample paths, but* $\lambda X_2 \circ \gamma e + ce$ *need not.*

It is natural to treat the non-heavy-traffic case in Corollary 5.2 without using the scaling in Theorem 2.1. We can then consider fixed capacity C and fluid processing rate r. The idea is to define the sequence of models so that $\{B_n(t) : t \geq 0\}$ directly converges to a limiting process, say $\{B(t) : t \geq 0\}$. (The limit B corresponds to $\lambda X_2 \circ \gamma e$ in Corollary 5.2.) By the continuity of the reflection map, the associated buffer-content stochastic processes $\{Z_n(t) : t \geq 0\}$ converge to $\phi(B - re)$, where ϕ is the first component of the reflection map. If the sequence $\{(B_{ni}, I_{ni}) : i \geq 1\}$ is i.i.d. for each n, then the limit process X_2 in Theorem 2.1 can be a general Lévy process with nondecreasing sample paths (subordinator), so that $\phi(B - ce)$ is a reflected Lévy process. If we work in the single-model framework of Theorem 4.1, we still do not need to be in heavy traffic, but the possible limit processes are more restricted; then the limit processes B and $\phi(B - ce)$ become a stable process with nondecreasing sample paths (stable subordinator) and a reflected stable process, respectively.

We now describe the standard heavy-traffic limit in more detail, allowing stochastic input during on periods as in Section 3. We now choose measuring units so that the constant fluid processing rate is 1 for all n. The net-input process in model n is

$$X_n(t) = C_n(t) - t \ , \tag{5.5}$$

with the cumulative input process being

$$C_n(t) = \sum_{i=1}^{k} \Lambda_n^i(B_n^i(t)), \quad t \geq 0 \ , \tag{5.6}$$

corresponding to the superposition of k independent on-off sources, where the i^{th} source submits fluid according to the stochastic process $\Lambda_n^i(t)$ when it is on.

As in Corollary 3.2, we will assume that

$$c_n[\Lambda_n^i(nt) - \lambda_n^i nt] \Rightarrow 0 \quad \text{in} \quad (D, M_1) \quad \text{as} \quad n \to \infty \tag{5.7}$$

for $1 \leq i \leq k$, where $\lambda_n^i \to \lambda^i > 0$ as $n \to \infty$. where $n^{-1/\alpha}$ is the final normalization we will use to obtain a nondegenerate limit.

For source i in model n, $B_n^i(t)$ is the cumulative busy time in $[0, t]$. It is determined by on periods B_{nj}^i and off periods I_{nj}^i. As in Section 4, we assume that we have essentially a single model, i.e.,

$$B_{nj}^i = \beta_n^i B_j^i \quad \text{and} \quad I_{nj}^i = \alpha_n^i I_j^i \tag{5.8}$$

for $1 \leq i \leq k$, $j \geq 1$ and $n \geq 1$, where β_n^i and α_n^i are constants satisfying $\beta_n^i \to \beta^i$ and $\alpha_n^i \to \alpha^i$ as $n \to \infty$, where $\alpha^i + \beta^i > 0$. Without loss of generality, we assume that $EB_j^i = EI_j^i = 1$, so that $EB_{nj}^i = \beta_n^i$ and $EI_{nj}^i = \alpha_n^i$. We assume that $\{B_j^i : j \geq 1\}$ and $\{I_j^i : j \geq 1\}$ for $1 \leq i \leq k$ are $2k$ mutually independent sequences of i.i.d. random variables.

We will consider the case in which the first j of the k on periods have power tails with exponent $x^{-\alpha}$, while the rest are asymptotically negligible. Let $Z_n(t)$ be the buffer content at time t in model n.

Theorem 5.3. *Consider the fluid queue model with k independent on-off sources, each with independent sequences of i.i.d. on periods and off periods as specified by* (5.6)–(5.8)*, where $\alpha^i + \beta^i > 0$ for $1 \leq i \leq k$. Suppose that $E[(I_1^i)^2] < \infty$ for $1 \leq i \leq k$,*

$$\lim_{x \to \infty} x^{\alpha} P(B_1^i > x) = K^i \tag{5.9}$$

for $1 \leq i \leq j$ and

$$\lim_{x \to \infty} x^{\alpha} P(B_1^i > x) = 0 \tag{5.10}$$

for $j+1 \leq i \leq k$ and $1 < \alpha < 2$. Then

$$n^{-1/\alpha}[C_n(nt) - \zeta_n nt] \Rightarrow S^{\alpha}(t) \quad \textit{in } (D, M_1) \textit{ as } n \to \infty \ , \tag{5.11}$$

where

$$\zeta_n = \sum_{i=1}^{k} \lambda_n^i \xi_n^i \tag{5.12}$$

with λ_n^i determined by (5.7)*,*

$$\xi_n^i = \frac{\beta_n^i}{\alpha_n^i + \beta_n^i} \to \xi^i \quad \textit{as} \quad n \to \infty \tag{5.13}$$

as determined by (5.8)*, $S^{\alpha}(t)$ is a Stable Lévy motion with marginal distribution $S_{\alpha}(\sigma t^{1/\alpha}, 1, 0)$ and*

$$\sigma = \left(\sum_{i=1}^{j} (\lambda^i (1 - \xi^i) \beta^i)^{\alpha} \gamma_i \frac{K_i}{K_{\alpha}} \right)^{1/\alpha} \ , \tag{5.14}$$

with K_i in (5.9)*, K_{α} in* (4.6)*, $\lambda_n^i \to \lambda^i$, $\beta_n^i \to \beta^i$ and*

$$\gamma_n^i = \frac{1}{\alpha_n^i + \beta_n^i} \to \frac{1}{\alpha^i + \beta^i} = \gamma^i > 0 \quad \textit{as} \quad n \to \infty \ . \tag{5.15}$$

If, in addition,

$$n^{1-\alpha^{-1}}(\zeta_n - 1) \to c \quad \textit{as} \quad n \to \infty, \quad -\infty < c < \infty \ ,$$

then the net-input processes satisfy

$$n^{-1/\alpha} X_n(nt) \Rightarrow ct + S^{\alpha}(t) \quad \textit{in } (D, M_1) \textit{ as } n \to \infty \ . \tag{5.16}$$

If, in addition, the capacity in model n is $n^{1/\alpha} C$ and $n^{-1/\alpha} Z_n(0) \Rightarrow Z(0)$, then

$$n^{-1/\alpha} Z_n(nt) \Rightarrow Z(t) \quad \textit{in } (D, M_1) \textit{ as } n \to \infty \ , \tag{5.17}$$

where $Z = \phi(S^\alpha + ce)$ *for* S^α *in* (5.11) *and* e *is the identity map. Then*

$$\lim_{t\to\infty} P(R(S^\alpha + ce)(t) \le x) = \frac{H(x)}{H(C)}, \quad 0 \le x \le C \ , \tag{5.18}$$

where H *is a cdf with pdf* h *with Laplace transform*

$$\hat{h}(s) \equiv \int_0^\infty e^{-sx} h(x) dx = \frac{1}{1 + (vs)^{\alpha-1}} \tag{5.19}$$

and scaling constant ν *defined by*

$$v^{\alpha-1} = \frac{-\sigma^\alpha}{c\cos(\pi\alpha/2)} > 0 \tag{5.20}$$

for σ *in* (5.14)*.*

Proof. By previous results, the cumulative input process from the i^{th} source, $1 \le i \le j$, has the limit

$$n^{-1/\alpha}[C_n^i(nt) - \lambda_n^i \xi_n^i nt] \Rightarrow \lambda^i (1 - \xi^i) \gamma_i^{1/\alpha} \beta^i X_2^i(t) \ , \tag{5.21}$$

as $n \to \infty$, where D is endowed with the M_1 topology, X_2^i is stable Lévy motion with $X_2^i(t)$ having marginal distribution $S_\alpha(\sigma t^{1/\alpha}, 1, 0)$ and $\sigma = (K/K_\alpha)^{1/\alpha}$ for K in (4.8) and K_α in (4.6). Thus the limit has marginal distribution $S_\alpha(\sigma_i t^{1/\alpha}, 1, 0)$, where

$$\sigma_i = \lambda^i (1 - \xi^i) \gamma_i^{1/\alpha} \beta^i \left(\frac{K}{K_\alpha}\right)^{1/\alpha} \tag{5.22}$$

see (1.2.3) on p. 11 of [29]. Since the k sources are mutually independent and the last $k - j$ are asymptotically negligible, we can add over the first j sources, using (5.22) and 1.2.1 of [29], to obtain the limit (5.11) with σ in (5.14). The net-input limit in (5.16) differs only by a deterministic translation. Finally, we obtain (5.17) by applying the continuous mapping theorem with the reflection map. The steady-state distribution of $R(S^\alpha + ce)$ is classic; see [40] and references therein. ▪

Remark 5.1. As illustrated in [40], numerical values of the steady-state distribution of the limiting reflected stable process is Theorem 5.3 are easily obtained by numerical transform inversion. ▪

6. The Standard Case: Brownian Limits

The standard case involves a single model with short-range dependence and finite variances for the variables I_i and B_i. Then the basic limit process $[X_1(t), X_2(t)]$ in Theorem 2.1 should be the Wiener process or Brownian motion, here denoted by $W(t)$ to distinguish it from the cumulative busy process $B(t)$. Since the Wiener process has continuous sample paths, the M_1 convergence in the Theorems of Section 2 is equivalent to uniform convergence on compact intervals. In this section we give further results for this case.

Theorem 6.1. *If*

$$n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} [(I_i, B_i) - (m_1, m_2)] \Rightarrow [W_1(t), W_2(t)] \quad \text{in} \quad D^2 \quad \text{as} \quad n \to \infty \;, \tag{6.1}$$

where $[W_1(t), W_2(t)]$ *is a centered* (0*-drift*) *two-dimensional Brownian motion or Wiener process with covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix} \;, \tag{6.2}$$

then

$$n^{-1/2}[B(nt) - \xi nt] \Rightarrow \sigma W(t) \;, \tag{6.3}$$

where ξ *and* γ *are as in* (2.4)*,* $W(t)$ *is a centered Wiener process and*

$$\sigma^2 = \gamma[\xi^2 \sigma_1^2 + (1-\xi)^2 \sigma_2^2 - 2\xi(1-\xi)\sigma_{12}^2] \;. \tag{6.4}$$

Proof. We apply Theorem 2.1 with $c_n = n^{-1/2}$ to obtain the limit process

$$(1-\xi)\gamma^{1/2} X_2(t) - \xi \gamma^{1/2} X_1(t) \tag{6.5}$$

where $X_i(t) = W_i(t)$. Next note that this linear combination of centered Wiener processes is again a centered Wiener process with variance parameter σ^2 in (6.4). ∎

To apply Theorem 6.1, we need to verify that the assumed FCLT in (6.1) holds and identify the five parameters m_1, m_2, σ_1^2, σ_2^2, σ_{12}^2. We now consider additional assumptions under which the conditions of Theorem 6.1 are satisfied and we can identify the parameters.

Theorem 6.2. *If the successive pairs* (I_i, B_i) *are iid, and in addition have finite second moments, then the condition of Theorem 6.1 holds and the parameters in* (2.4) *and* (6.4) *can be identified as*

$$m_1 = EI, \quad m_2 = EB \tag{6.6}$$

$$\sigma_1^2 = Var\, I, \quad \sigma_2^2 = Var\, B \quad \text{and} \quad \sigma_{12}^2 = Cov(I,B) \;. \tag{6.7}$$

Proof. The limit (6.1) holds with the parameters as in (6.6) and (6.7) by the two-dimensional version of Donsker's FCLT; see Chapter 7 of Ethier and Kurtz [11]. ∎

In the setting of Theorem 6.2, if I_i and B_i are also mutually independent, then $\sigma_{12}^2 = Cov(I, B) = 0$. Then we can characterize variability in terms of the squared coefficient of variation (SCV, variance divided by the square of the mean) of the individual variables I and B. The variance parameter in (6.4) becomes

$$\begin{aligned} \sigma^2 &= \lambda[(1-\xi)^2\sigma_1^2 + \xi^2\sigma_2^2] \\ &= \frac{1}{(m_1+m_2)^3}[m_2^2\sigma_1^2 + m_1^2\sigma_2^2] \\ &= \frac{m_1^2 m_2^2}{(m_1+m_2)^3}[c_1^2 + c_2^2] \end{aligned} \tag{6.8}$$

where c_1^2 and c_2^2 are the SCV's of I and B.

We now consider special cases in which the busy and idle periods are associated with a queueing system. It is possible that the environment of the fluid model could involve a queue, but now we are thinking of simply developing limits for the cumulative busy time in the interval $[0, t]$ in a queueing model. Many queueing systems have Poisson arrival processes. Then I is exponentially distributed and independent of B. If, as in the M/GI/s/r queueing model with finite waiting room, I is exponentially distributed and independent of B, then $\sigma_1^2 = m_1^2$ and the variance parameter in (6.8) becomes

$$\sigma^2 = \frac{m_1^2 m_2^2}{(m_1+m_2)^3}[1 + c_2^2] \ . \tag{6.9}$$

From (6.9), we see that for the M/G/s/r model there are only two unspecified parameters — the mean and SCV of the busy period. Of course, these are well known for the M/GI/1 queue; e.g., see [6, p. 251]. The result in this case has a long history; see [31, Theorem 4 on p. 115].

Theorem 6.3. *For the M/GI/1/∞ model where the service time has mean* 1 *and SCV* c_s^2,

$$m_1 = \frac{1}{\rho}, \quad m_2 = \frac{1}{1-\rho}, \quad \frac{1}{m_1+m_2} = \rho(1-\rho) \ , \tag{6.10}$$

$$\sigma_2^2 = \frac{c_s^2+\rho}{(1-\rho)^3}, \quad c_2^2 = \frac{c_s^2+\rho}{1-\rho} \ , \tag{6.11}$$

and the variance parameter in (6.9) *becomes*

$$\sigma^2 = \rho(c_s^2 + 1) \ . \tag{6.12}$$

We now consider the M/M/1/r model with r extra waiting spaces. For the M/M/1/r queue, we can apply results for its busy period in Section 3 of [2]. For this model, we can have any traffic intensity ρ.

Theorem 6.4. *For the M/M/1/r model with service rate* 1 *and arrival rate* ρ*,*

$$m_2 \equiv E[B] = \begin{cases} r+1 & \text{if} \quad \rho = 1 \\ \frac{1-\rho^{r+1}}{1-\rho} & \text{if} \quad \rho \neq 1 \end{cases} \tag{6.13}$$

$$(c_2^2+1)m_2^2 = E[B^2] = \begin{cases} [8r^2+46r+39]/48 & \text{if} \quad \rho = 1 \\ \frac{2}{(1-\rho)^3}\{1-(2r+3)\rho^{r+1}(1-\rho)-\rho^{2r+3}\} & \text{if} \quad \rho \neq 1 \ . \end{cases} \tag{6.14}$$

Another relatively elementary example is the M/M/∞ queue. The following comes from Dupius and Guillemin [10].

Theorem 6.5. *For the M/M/∞ queue with individual service rate 1 and arrival rate* λ*,*

$$m_2 \equiv EB = (e^\lambda - 1)/\lambda \tag{6.15}$$

and

$$(c_2^2+1)m_2^2 \equiv E[B^2] = 2e^\lambda \sum_{n=1}^{\infty} \lambda^{n-1}/n(n!) \ . \tag{6.16}$$

Proof. From p. 61 of Dupius and Guillemin [10], the Laplace transform of the busy period (at least one server is busy) is

$$\hat{b}(s) \equiv Ee^{-sB} = \frac{\lambda+s}{\lambda} - \frac{e^\lambda s}{\lambda(1+s\hat{\alpha}(s))} \ , \tag{6.17}$$

where

$$\hat{\alpha}(s) = \sum_{n=1}^{\infty} \lambda^n/(s+n)n! \ . \quad \blacksquare \tag{6.18}$$

A generalization of Theorems 6.4 and 6.5 arises whenever the queue-content process evolves as a Markov chain (MC). Then the idle period and busy period are independent, and the idle period has an exponential (geometric) distribution if the MC evolves in continuous (discrete) time, so that we are in the setting of Theorem 6.3. The busy period then is a first passage time, whose moments can be readily computed; e.g., see Chapter III of Kemeny and Snell [19], especially p. 51. For larger state spaces, care needs to be given in the computation; see Heyman and O'Leary [14].

For extensions to M/G/1/r systems with finite waiting room, see Chapter 5 of Takagi [33]. For busy-period results in corresponding finite-population M/G/1 queues, see Chapter 4 of [33].

7. Busy and Idle Periods from the G/G/1/∞ Queue

In this section we consider the special case in which the successive busy and idle periods are associated with the general infinite-capacity single-server queue. For the G/G/1/∞ queue, we can apply asymptotics associated with the one-dimensional reflection map; as in [38] and Section 6 of [39]. For this purpose we assume that the interarrival times T_i and service times S_i satisfy a joint FCLT. When the busy and idle periods are associated with a queueing model, it is natural to regard the sequence $\{(T_i, S_i) : i \geq 1\}$ as the basic model data instead of the sequence $\{(I_i, B_i) : i \geq 1\}$ considered in Sections 2 and 3.

Let $A(t)$ count the number of arrivals in $[0, t]$. In this setting the cumulative busy time $B(t)$ is closely related to the total input of work, $X(t)$, where

$$X(t) = \sum_{i=1}^{A(t)} S_i, \quad t \geq 0 . \tag{7.1}$$

Indeed, the two processes are identical whenever the system is empty. Hence it should come as no surprise that their limit processes are identical. With Brownian limits, they have the same distribution and thus the same variance parameters. Let $Z(t)$ be the buffer content (workload) in the queue. The following is a generalization of results in [38].

As in Theorem 2.1, we consider a sequence of models indexed by n. Let $\{T_{ni}, S_{ni} : i \geq 1\}$ be the sequence for model n. We scale time so that the mean service time is 1.

Theorem 7.1. *If*

$$c_n \left[\sum_{i=1}^{\lfloor nt \rfloor} T_{ni} - \rho_n^{-1} nt, \quad \sum_{i=1}^{\lfloor nt \rfloor} S_{ni} - nt \right] = [Y_1(t), Y_2(t)] \quad in\ (D, M_1)^2\ as\ n \to \infty \tag{7.2}$$

where $nc_n \to \infty$ and $\rho_n \to \rho$ as $n \to \infty$, $0 < \rho \leq 1$ and $P(Disc(Y_1) \cap Disc(Y_2) \neq \phi) = 1$, then

$$c_n \left[A_n(nt) - \rho_n nt, \sum_{i=1}^{A_n(nt)} S_{ni} - \rho_n nt \right] \Rightarrow [-\rho Y_1(\rho t), Y_2(\rho t) - \rho Y_1(\rho t)] \tag{7.3}$$

in $(D, M_1)^2$ as $n \to \infty$. If, in addition, $\rho < 1$, then

$$c_n[Z_n(nt), B_n(nt) - (\rho_n)nt] \Rightarrow (0, Y_2(\rho t) - \rho Y_1(\rho t)) \quad in\ (D, M_1)^2\ as\ n \to \infty . \tag{7.4}$$

If, instead, $\rho = 1$ and $nc_n(1 - \rho_n) \to c$, then

$$c_n[Z_n(nt), I_n(nt)] \Rightarrow (Z(t), -\inf\{Z(s) : 0 \leq s \leq t\}) \tag{7.5}$$

in $(D, M_1)^2$ as $n \to \infty$, where $Z = \phi(Y)$ for the first component ϕ of the reflection map in (5.1) and (5.2) without upper barrier and $Y(t) = Y_2(\rho t) - \rho Y_1(\rho t)$, $t \geq 0$.

Proof. The proof of (7.3) is a minor modification of the proof of Theorem 2.1. We apply the inverse map to the arrival process to get the first component of (7.3). We then apply composition with addition to get the second component of (7.3). Then (7.4) follows from Theorem 6.3(ii) of [39]. The limit (7.5) is the heavy-traffic limit, as in [38]. We apply the continuous mapping theorem with the reflection and supremum maps; see Whitt (1999b).

Remark 7.1. The general triangular-array limit in Theorem 7.1 is used by Kurtz [21] to obtain a FCLT for the total input process with a fractional Brownian motion (FBM) limit process. He has scaling exponent $q = 1/2$, but general self-similarity index H. See Willinger, Taqqu, Sherman and Wilson [43] for another FBM limit. Fractional Brownian motion has continuous sample paths, which implies that the discontinuity condition in Theorem 2.1 is automatically satisfied; see [29, Exercise 10.1, pp. 490, 551].

We now consider the special case of a Brownian motion limit. The next result follows from Theorem 7.1 just like Theorem 6.1 follows from Theorem 2.1.

Theorem 7.2. *If the condition of Theorem 7.1 holds with $c_n = n^{-1/2}$ and $[Y_1(t), Y_2(t)]$ two-dimensional zero-drift Brownian motion with covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_a^2 & \sigma_{as}^2 \\ \sigma_{as}^2 & \sigma_s^2 \end{pmatrix} , \tag{7.6}$$

then the limit in the second term of (7.4) is distributed as $\sigma W(t)$, where

$$\sigma^2 = \rho[\rho^2\sigma_a^2 - 2\rho\sigma_{as}^2 + \sigma_s^2] \tag{7.7}$$

where $W(t)$ is standard (drift 0, diffusion 1) Brownian motion.

We next consider the GI/GI/1 queue, which combines conditions for (T_n, S_n) assumed for (I_n, B_n) in Theorem 6.2 and (6.8). The next result follows by the same reasoning. The ordinary CLT version is an early result; again see Takàcs [31, 32].

Theorem 7.3. *In the standard GI/GI/1 queue if S_n and T_n have means $ES_n = 1$, $ET_n = \rho^{-1}$ and finite second moments, then the conditions of Theorem 7.2 hold with $\sigma_a^2 = Var(T_n)$, $\sigma_s^2 = Var(S_n)$ and $\sigma_{as}^2 = 0$. Moreover, the variance parameter in (7.7) is*

$$\sigma^2 = \rho[c_a^2 + c_s^2]|, \tag{7.8}$$

where c_a^2 and c_s^2 are the SCV's of an interarrival time T and a service time S, respectively.

Note that the GI/GI/1 result in Theorem 7.3 is consistent with the M/GI/1 result in Theorem 6.3. It remains to relate the GI/GI/1 results in Theorems 7.3 and 6.2.

We next observe that for more general G/G/1 models than GI/GI/1 we can further identify some of the parameters, but it remains to relate the GI/GI/1 results in Sections 6 and 7.

Theorem 7.4. *Assume that (S_{ni}, T_{ni}) is a stationary ergodic sequence. Then*

$$\frac{EB_{n1}}{EI_{n1} + EB_{n1}} = \rho_n \quad \text{and} \quad EI_{n1} = \left(\frac{\rho_n}{1-\rho_n}\right) EB_{n1} \ . \tag{7.9}$$

If, in addition, condition (2.1) *of Theorem 2.1 holds, then*

$$c_n[B_n(nt) - \rho_n nt] \Rightarrow (1-\rho)X_2(\gamma t) - \rho X_1(\gamma t) \quad \text{in} \quad (D, M_1) \tag{7.10}$$

as $n \to \infty$. If in addition the conditions of Theorem 6.2 *hold, then the variability parameter σ^2 there can also be expressed as*

$$\sigma^2 = \gamma[\rho^2\sigma_1^2 + (1-\rho)^2\sigma_2^2 - 2\rho(1-\rho)\sigma_{12}^2] \ . \tag{7.11}$$

Remark 7.2. Combining (7.8) and (7.11), we see that in the GI/GI/1 queue there are two expressions for the variability parameter σ^2, which provides a relationship among the parameters:

$$\sigma^2 = \rho[c_a^2 + c_s^2] = \left(\frac{1}{EI + EB}\right)\left(\rho^2\sigma_I^2 + (1-\rho)^2\sigma_B^2 - 2\rho(1-\rho)\sigma_{I,B}^2\right) \tag{7.12}$$

Given (7.9), it suffices to learn one of EI and EB. Given ρ, c_a^2, c_s^2 and (7.12), it suffices to learn two of σ_I^2, σ_B^2 and $\sigma_{I,B}^2$.

Remark 7.3. Expressions for the variability parameter of the total input processes are available in the literature. For example, the asymptotic variance of the batch Markovian arrival process (BMAP), also known as the versatile Markovian point process, is given in Theorem 5.4.1 on p. 284 of Neuts [23].

Remark 7.4. It may also be useful to consider non-normal approximations when the basic variables (T_n, S_n) are iid. For example, we may want to allow for infinite variances. That case is discussed in [40].

8. Conclusions

We have obtained FCLTs for a cumulative input process associated with on-off sources feeding a fluid queue. Since the cumulative input process is closely related to cumulative busy-time and idle-time processes, we also obtained results for those processes. Since the limit involves a sequence of processes with continuous sample paths converging to a limiting process that may have jumps, we used the Skorohod (1956) M_1 topology on D, using recent results about the inverse and composition maps in [24], [41]. As shown in [42] the limits here combined with the reflection map yield FCLTs for buffer-content processes in single-class stochastic fluid networks, where the on-off sources may have heavy-tailed on-period distributions. Under independence assumptions, the limiting processes are reflected-stable-processes or more general reflected Lévy processes. However, the general limiting results apply without the independence assumptions. We illustrated the heavy-traffic FCLTs in Section 5 by establishing a reflected stable process limit for a single fluid queue with finite waiting space in which the on-period distributions have power tails. The results in that one-dimensional case are more tractable than for the multidimensional stochastic fluid networks, because we can explicitly calculate the steady-state distribution of the limiting reflected stable process, exploiting numerical transform inversion.

Our main focus was on the case in which the busy periods have heavy-tailed probability distributions. However, in Sections 6 and 7 we also obtained results for the standard case in which the normalized cumulative input processes converge to Brownian motion. Then the goal is to identify the variance constant.

References

[1] P. Barford and M. Crovella, Generating representative web workloads for network and server performance evaluation. *Proc. 1998 ACM Sigmetrics*, 1998, 151–160.

[2] A. W. Berger and W. Whitt, The Brownian approximation for rate control throttles and the G/G/1/C queue. *J. Discrete Event Dynamic Systems*, 2 (1992) 7–60.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[4] O. J. Boxma and J. W. Cohen, Heavy-traffic analysis for the GI/G/1 queue with heavy-tailed distributions. *Queueing Systems*, to appear.

[5] O. J. Boxma and J. W. Cohen, The M/G/1 queue: heavy tails and heavy traffic. In *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger (eds.), Wiley, New York, 1999, to appear.

[6] J. W. Cohen, *The Single-Server Queue*, North-Holland, Amsterdam, 1982.

[7] J. W. Cohen, A heavy-traffic theorem for the GI/G/1 queue with a Pareto-type service time distribution. *J. Applied Math. Stoch. Analysis*, special issue dedicated to R. Syski, 11 (1998), 247–254.

[8] M. E. Crovella and A. Bestavros, Self-similarity in World Wide Web traffic — evidence and possible causes. *Proc. ACM Sigmetrics* '96 (1996) 160–169.

[9] R. M. Dudley, Distances of probability measures and random variables. *Ann. Math. Statist.* 39, 1563–1572.

[10] A. Dupius and F. Guillemin, *Etude des Phenomenes Transitores de Congestion Lies an Multiplexage Statistique en Boucle Ouverte sur un Lien ATM*, France Telecom CNET Report, December 1997.

[11] S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley, New York, 1986.

[12] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II, second ed., Wiley, New York, 1971.

[13] H. Furrer, Z. Michna and A. Weron, Stable Lévy motion approximation in collective risk theory. *Insurance: Math and Econ.*, 20 (1997) 97–114.

[14] D. P. Heyman and D. P. O'Leary, What is fundamental for Markov chains: first passage times, fundamental matrices and group generalized inverses, Chapter 10, pp. 151–161 in *Computations with Markov Chains*, ed. W. J. Stewart, Kluwer, Boston, 1995.

[15] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, New York, 1987.

[16] O. Kella, Parallel and tandem fluid networks with dependent Lévy inputs. *Ann. Appl. Prob.*, 3 (1993) 682–695.

[17] O. Kella and W. Whitt, A tandem fluid network with Lévy input. In *Queues and Related Models*, ed. I. Basawa and U. Bhat. Oxford University Press, Oxford, 1992, pp. 112–128.

[18] O. Kella and W. Whitt, Stability and structural properties of stochastic storage networks. *J. Appl. Prob.*, 33 (1996) 1169–1180.

[19] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand, Princeton, New York, 1960.

[20] T. Konstantopoulos and S.-J. Lin, Macroscopic models for long-range dependent network traffic. *Queueing Systems*, 28 (1998) 215–243.

[21] T. G. Kurtz, Limit theorems for workload input models, in *Stochastic Networks, Theory and Applications*, eds. F. P. Kelly, S. Zachary and I. Ziedins, Clarendon Press, Oxford, 1996.

[22] W. E. Leland, M. Taqqu, W. Willinger and D. V. Wilson, On the self-similar nature of ethernet traffic. *IEEE/ACM Trans. Networking*, 2 (1994), 1–15.

[23] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, 1989.

[24] A. A. Puhalskii and W. Whitt, Functional large deviation principles for first-passage-time processes. *Ann. Appl. Prob.*, 7 (1997) 362–381.

[25] M. I. Reiman, Open queueing networks in heavy traffic. *Math. Oper. Res.*, 9 (1984) 441–458.

[26] S. Resnick and H. Rootzén, Self-similar communication models and very heavy tails, Cornell University, 1998.

[27] S. Resnick and G. Samorodnitsky, A heavy traffic limit theorem for workload processes with heavy tailed service requirements, Cornell University, 1998.

[28] S. Resnick and E. van den Berg, Weak convergence of high-speed network traffic models, Cornell University, 1999.

[29] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York, 1994.

[30] A. V. Skorohod, Limit theorems for stochastic processes. *Theor. Probab. Appl.*, 1 (1956) 261–290.

[31] L. Takács, *Combinatorial Methods in the Theory of Stochastic Processes*, Wiley, New York, 1967.

[32] L. Takács, Letter to the Editor. *J. Appl. Prob.*, 8 (1971) 848–849.

[33] H. Takagi, *Queueing Analysis, Vol. 2: Finite Systems*, North-Holland, Amsterdam, 1993.

[34] M. S. Taqqu, W. Willinger and R. Sherman, Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review* 27 (1997), 5–22.

[35] K. P. Tsoukatos and A. M. Makowski, Heavy-traffic analysis of a multiplexer driven by $M/GI/\infty$ input processes. *Teletraffic Contributions for the Information Age, Proceedings of ITC 15*, V. Ramaswami and P. E. Wirth (eds.), Elsevier, Amsterdam, 1997, 497–506.

[36] K. P. Tsoukatos and A. M. Makowski, Heavy traffic limits associated with $M/GI/\infty$ input processes. *Queueing Systems*, to appear.

[37] K. P. Tsoukatos and A. M. Makowski, Interpolation approximations for $M/GI/\infty$ arrival processes, Electrical Engineering Department, University of Maryland, 1999.

[38] W. Whitt, Weak convergence theorem for priority queues: preemptive-resume discipline. *J. Appl. Prob.*, 8 (1971) 74–94.

[39] W. Whitt, Some useful functions for functional limit theorems, *Math. Oper. Res.*, 5 (1980) 67–85.

[40] W. Whitt, Non-Brownian FCLTs for the single-server queue, AT&T Labs, 1998.

[41] W. Whitt, On the Skorohod M topologies, AT&T Labs, 1999a.

[42] W. Whitt, The reflection map is Lipschitz with appropriate Skorohod M_1 metrics, AT&T Labs, 1999b.

[43] W. Willinger, M. S. Taqqu, R. Sherman and D. V. Wilson, Self-similarity through high variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Trans. Networking* 5 (1997) 71–86.

[44] V. M. Zolotarev, *One-Dimensional Stable Distributions*, American Math. Society, vol. 65, Providence, RI, 1986.