DEPARTURES FROM MANY QUEUES IN SERIES

by

Peter W. Glynn

Department of Operations Research
Stanford University
Stanford, CA 94305-4022

and

Ward Whitt

AT&T Bell Laboratories
Room 2C-178
Murray Hill, NJ 07974-2070

May 21, 1990

Revision: December 20, 1990

To appear in the *Annals of Applied Probability*

Abstract

We consider a series of n single-server queues, each with unlimited waiting space and the first-in first-out service discipline. Initially, the system is empty; then k customers are placed in the first queue. The service times of all the customers at all the queues are i.i.d. with a general distribution. We are interested in the time $D(k, n)$ required for all k customers to complete service from all n queues. In particular, we investigate the limiting behavior of $D(k, n)$ as $n \rightarrow \infty$ and/or $k \rightarrow \infty$. There is a duality implying that $D(k, n)$ is distributed the same as $D(n, k)$ so that results for large n are equivalent to results for large k. A previous heavy-traffic limit theorem implies that $D(k, n)$ satisfies an invariance principle as $n \rightarrow \infty$, converging after normalization to a functional of k-dimensional Brownian motion. We use the subadditive ergodic theorem and a strong approximation to describe the limiting behavior of $D(k_n, n)$ where $k_n \rightarrow \infty$ as $n \rightarrow \infty$. The case of $k_n = \lfloor xn \rfloor$ corresponds to a hydrodynamic limit.

AMS 1980 subject classifications. primary 60K25, 60F17; secondary 90B22, 60J60, 60F15.

Key words and phrases. tandem queues, queues in series, queueing networks, departure process, transient behavior, reflected Brownian motion, limit theorems, invariance principle, strong approximation, subadditive ergodic theorem, large deviations, interacting particle systems, hydrodynamic limit.

1. Introduction and Summary

In this paper we consider a queueing model that could be used to represent the start-up behavior of a long production line or the transient flow of messages over a long path in a communication network. In particular, we consider a series of n single-server queues, each with unlimited waiting space and the first-in first-out service discipline. Initially, the system is empty; then k customers are placed in the first queue. The service times of all the customers at all the queues are i.i.d. with a general distribution having mean 1 and finite positive variance σ^2. Let $D(k, n)$ be the departure time of customer k from queue n. Our object is to describe the distribution of $D(k, n)$ as n gets large. We may have k constant (independent of n) or $k \equiv k_n \to \infty$ as $n \to \infty$.

Our primary focus is on the early departures from a large number of queues. For example, the customer index k_n associated with n queues might be k, $\sqrt{n}$ or n. However, there is a duality discussed in Section 2 that makes our results also applicable to a large number of departures from relatively few queues. In particular, under our i.i.d. assumption for the service times,

$$\{D(i, j) : 1 \leq i \leq k,\ 1 \leq j \leq n\} \overset{d}{=} \{D(j, i) : 1 \leq j \leq n,\ 1 \leq i \leq k\}\ , \tag{1.1}$$

where $\overset{d}{=}$ denotes equality in distribution. In fact, the dual case with $k > n$ seems to have more applied relevance; many production lines periodically (e.g., daily) produce many items on relatively small number of machines.

Since $D(1, n)$ is just the sum of n service times,

$$[D(1, n) - n]/\sqrt{n} \Rightarrow N(0, \sigma^2) \text{ as } n \to \infty\ , \tag{1.2}$$

where $\Rightarrow$ denotes convergence in distribution or weak convergence, as in Billingsley (1968), and $N(m, \sigma^2)$ denotes a normal random variable with mean m and variance σ^2. The duality mentioned above and a previous heavy-traffic limit theorem by Iglehart and Whitt (1970) imply that

$$[D(k,\, n) - n]/\sqrt{n} \Rightarrow \sigma \hat{D}_k(1) \text{ as } n \to \infty \text{ for each } k\,, \tag{1.3}$$

where $\hat{D}_k(1)$ is a functional of k-dimensional standard Brownian motion. It is significant that the limit (1.3) depends on the service-time distribution only through its first two moments, and then in only a relatively trivial way; i.e., the mean and variance only appear via elementary scaling. Hence, the limit $\hat{D}_k(1)$ in (1.3) has wide applicability. Unfortunately, however, our analysis does not provide much information about $\hat{D}_k(1)$. Hence, simulation has been applied by Greenberg, Schlunk and Whitt (1990) to gain further insight.

To see how $D(k,\, n)$ behaves when k and n are both large, we consider the limiting behavior of $\hat{D}_k(1)$ in (1.3) as $k \to \infty$. We apply the subadditive ergodic theorem as on p. 277 of Liggett (1985) to show that

$$\hat{D}_k(1)/\sqrt{k} \Rightarrow \alpha \text{ as } k \to \infty\,, \tag{1.4}$$

α is a constant. In Greenberg et al. (1990) it is conjectured that $\alpha = 2$.

From (1.3) and (1.4), we see that $[D(k,\, n) - n]/\sqrt{nk}$ approaches α in the iterated limit as first $n \to \infty$ and then $k \to \infty$. We also show that this limit holds when $k \to \infty$ and $n \to \infty$ simultaneously. In particular, we establish a strong approximation result implying that

$$[D(k_n,\, n) - n]/\sqrt{nk_n} \Rightarrow \alpha \text{ as } n \to \infty \tag{1.5}$$

when

$$k_n/n^{1-\varepsilon} \to x \text{ as } n \to \infty\,, \text{ where } 0 < x < \infty\,, \tag{1.6}$$

for any ε, $0 < \varepsilon < 1$.

The essence of (1.6) is that $k_n \to \infty$ as $n \to \infty$ but that k_n be suitably less than n. In fact, we establish a different limit when $k_n = \lfloor xn \rfloor$, where $\lfloor x \rfloor$ is the integer part of x. In particular, if the service-time distribution has an exponential tail, then

$$D(\lfloor xn \rfloor, n)/n \to \gamma(x) \quad \text{w.p.1 as } n \to \infty \, , \tag{1.7}$$

where $\gamma(x)$ is a deterministic function of x (with $\gamma(1)$ being different from α) that may depend on the service-time distribution.

This paper was largely motivated by Srinivasan (1989), who applied results of Rost (1981), Andjel (1982), Andjel and Kipnis (1984), Kipnis (1986), and Benassi and Fouque (1987) about interacting particle systems (in particular, the zero-range process and the asymmetric simple exclusion process) to describe the *hydrodynamic limit* for our model in the special case of *exponential service times* (still with mean 1). Roughly speaking, the hydrodynamic limit says that the average queue length among the first $[xt]$ queues at time t is asymptotically almost surely (a.s.) equal to $(2 - \sqrt{x})/\sqrt{x}$ as $t \to \infty$. Consequently, the average queue length among queues in the neighborhood of queue $[xt]$ is asymptotically a.s. $(1 - \sqrt{x})/\sqrt{x}$ as $t \to \infty$. (Note that the total number of customers in the first $\lfloor xt \rfloor$ queues is $(2\sqrt{x} - x)t + o(t)$; then differentiate with respect to x. In the unsaturated case with external arrival process having rate $\lambda < 1$, asymptotically a.s. the first $(1 - \lambda)^2 t$ queues reach equilibrium at time t as $t \to \infty$, but the rest of the density profile remains the same.)

It is easy to apply Srinivasan's hydrodynamic limit in the saturated case (with i.i.d. exponential service times having mean 1) to deduce that the departure time of customer $\lfloor xn \rfloor$ from queue n is asymptotically a.s. $(1 + \sqrt{x})^2 n + o(n)$; i.e., for exponential service times $\gamma(x) = (1 + \sqrt{x})^2$; see Section 6. Thus the departure times of customers 1 and n from queue n are a.s. $n + o(n)$ and $4n + o(n)$, respectively. To put this result in perspective, if customer n only had to wait at the first queue (as would be the case if all queues after the first had infinitely many servers), then the departure time for customer n from queue n would be a.s. $2n + o(n)$. Hence, the additional delay experienced by customer n in the last $n - 1$ queues is approximately equal to his delay in the first queue plus the sum of his service times.

Our limit in (1.7) extends Srinivasan (1989) by establishing a hydrodynamic limit for *general service-time distributions.* As suggested by the discussion above, limits for the average queue length among the first $\lfloor xt \rfloor$ queues at time t as $t \to \infty$ are equivalent to limits for $n^{-1} D(\lfloor xn \rfloor , n)$ as $n \to \infty$, so (1.7) yields a hydrodynamic limit in the sense of Srinivasan (1989) for general service-time distributions. With regard to the interacting particle system literature, our result is interesting because the associated vector queue-length process depicting the number of customers at each queue (including the one in service, if any) is not Markov here. We treat this case by applying the subadditive ergodic theorem. However, we have not yet identified the limit $\gamma(x)$ in (1.7) for general service-time distributions.

We also complement Srinivasan (1989) by describing in more detail what happens at the front of the ‘‘wave’’ of customers passing through the network. Of course, the first customer departs from queue n at time n with a deviation of order $\sqrt{n}$, as indicated in (1.2). The limit in (1.3) reveals that the first k interdeparture times from queue n after the first departure are each asymptotically of order $\sqrt{n}$ as $n \to \infty$. Consequently, by the time customer k has reached queue n for large n, customer k rarely has to wait.

The model we consider has no external arrival process, but the same model can be interpreted as starting out empty with an external arrival process. Simply interpret the departure process from the first queue as the external arrival process. Of course, the assumption that the service times be all i.i.d. implies that the interarrival-time distribution must then be exactly the same as each service-time distribution. However, this is not required for the limits (1.3) and (1.5). These limits remain unchanged if the service-time distributions at an initial finite set of queues are different. (The stated results cover this generalization.)

The rest of this paper is organized as follows. In Section 2 we review a convenient representation for the departure process that facilitates its study. In particular, we exploit the fact that the departure time of customer k from queue n can be represented as the maximum partial sum of service times along nondecreasing paths of length $k + n - 1$ in a $k \times n$ lattice of service times. From this representation, the duality in (1.1) is immediate.

In Section 3 we establish (1.3) and in Section 4 we establish the strong approximation needed for (1.5). In Section 5 we establish stochastic order relations among the interdeparture times, which are of interest in their own right, but also help us describe the limit $\hat{D}_k(1)$ and establish (1.7). In Section 6 we obtain our hydrodynamic limit, i.e., we establish (1.7). In Section 7 we establish (1.4) and (1.5).

In Section 8 we consider a modification of the model in which each customer has the same service time at all queues, as occurs in packet communication networks with variable packet sizes; see Pinedo and Wolff (1982) and Wolff (1982). We see that quite different asymptotic behavior occurs in this case. Finally, in Section 9 we make some concluding remarks.

We end this introduction by mentioning some additional references that provide background or treat somewhat related problems: Chapter 6 of Disney and Kiessler (1987), Kelly (1982, 1984), Suresh and Whitt (1990) and Vere-Jones (1968).

2. The Basic Recursion for the Departure Epochs

Let $V(k, n)$ be the service time and $D(k, n)$ the departure time for customer k at queue n. Our starting point is a basic recursion for the departure times,

$$D(k, n) = \max \{ D(k - 1), n), D(k, n - 1) \} + V(k, n) \tag{2.1}$$

for $k \geq 1$ and $n \geq 1$, with $D(k, 0) = 0$ for all k and $D(0, n) = 0$ for all n, which can be taken as the definition. (At this point, we do not assume that the service times are i.i.d.)

We can easily express $D(k, n)$ more directly in terms of the service times. To do so, let $\Pi(k, n)$ be the set of all ‘‘nondecreasing continuous paths’’ of length $k + n - 1$ from (1,1) to (k, n) in the set of ordered pairs $3 \equiv \{(i, j) : 1 \le i \le k, 1 \le j \le n\}$; i.e., $\pi \in \Pi(k, n)$ if π is a subset of 3 of cardinality $k + n - 1$ containing (1,1) and either $(i + 1, j)$ or $(i, j + 1)$, but not both, whenever it contains (i, j). Since successive ordered pairs in any such path π increase in the first component exactly $k - 1$ times, there are $\begin{bmatrix} k + n - 2 \\ k - 1 \end{bmatrix}$ paths in $\Pi(k, n)$.

From (2.1), we easily establish the following by induction.

Proposition 2.1. For all $k \ge 1$ and $n \ge 1$,

$$D(k, n) = \max_{1 \le l \le n} \{D(k - 1, l) + \sum_{j=l}^{n} V(k, j)\} \tag{2.2}$$

$$= \max_{1 \le l \le k} \{D(l, n - 1) + \sum_{i=l}^{k} V(i, n)\} \tag{2.3}$$

$$= \max_{\pi \in \Pi(k, n)} \{ \sum_{(i, j) \in \pi} V(i, j)\} \ . \tag{2.4}$$

Evidently Proposition 2.1 is quite well known; e.g., formulas (2.1) and (2.3) appear as (1), (2) and (16) of Tembe and Wolff (1974). A variant of (2.4) for queues without extra waiting space appears in Muth (1979). As Muth observes, (2.4) implies that the departure times $D(k, n)$ are unchanged if we reverse the order of the queues and the order of the service times at each queue. Let superscripts index different models.

Corollary 1. If $V^2(i, j) = V^1(k - i, n - j)$ for $1 \le i \le k$, $1 \le j \le n$, then $D^2(k, n) = D^1(k, n)$.

Formula (2.4) also implies a certain duality, i.e., symmetry in k and n. Let $\overset{d}{=}$ denote equality in distribution.

Corollary 2. If $\{V^1(i, j) : 1 \le i \le k, 1 \le j \le n\} \overset{d}{=} \{V^2(j, i) : 1 \le i \le k, 1 \le j \le n\}$, then

$$\{D^1(i, j) : 1 \le i \le k,\ 1 \le j \le n\} \overset{d}{=} \{D^2(j, i) : 1 \le i \le k,\ 1 \le j \le n\}\ .$$

As an immediate consequence of Corollary 2, we obtain the following result in the i.i.d. setting which is of primary interest to us.

Corollary 3. If $V(i, j)$, $1 \le i \le k$, $1 \le j \le n$, are i.i.d., then (1.1) holds.

Corollaries 2 and 3 can be used to obtain limit theorems as $k \to \infty$ for fixed n from the limit theorems we establish as $n \to \infty$ for fixed k. Corollaries 2 and 3 also allow us to relate the interdeparture times of primary interest to us to associated sojourn times. The k^{th} *interdeparture time* from queue n is

$$\Delta(k, n) = D(k+1, n) - D(k, n) \tag{2.5}$$

with $D(0, n) = 0$, for $k \ge 0$ and $n \ge 1$. The *sojourn time* of customer k at queue n is

$$S(k, n) = D(k, n) - D(k, n-1)\ . \tag{2.6}$$

Corollary 4. Under the assumption of Corollary 3,

$$\{\Delta(i, j) : 0 \le i \le k-1,\ 1 \le j \le n\} \overset{d}{=} S(j, i) : 1 \le j \le n,\ 1 \le i \le k\} \tag{2.7}$$

Remark. (2.1) The function mapping $\{V(i, j) : 1 \le i \le k,\ \ 1 \le j \le n\}$ into $\{D(i, j) : 1 \le i \le k,\ 1 \le j \le n\}$ is obviously nondecreasing and convex, so that stochastic order relations for service times carry over to departure times; see Stoyan (1983). The function is also Lipschitz, i.e., for each path π

$$\left| \sum_{(i,j) \in \pi} V^1(i, j) - \sum_{(i,j) \in \pi} V^2(i, j) \right| \le \sum_{(i,j) \in \pi} \left| V^1(i, j) - V^2(i, j) \right|$$

and

$$\max_{\substack{1 \le i \le k \\ 1 \le j \le n}} \left\{ \left| D^1(i, j) - D^2(i, j) \right| \right\} \le (k+n-1) \max_{\substack{1 \le i \le k \\ 1 \le j \le n}} \left\{ \left| V^1(i, j) - V^2(i, j) \right| \right\}$$

so that there is model stability; see Whitt (1974).

3. The Functional Central Limit Theorem

We now apply (2.2) to show that $\{ D(k, n) \}$ satisfies a functional central limit theorem (FCLT) as $n \to \infty$ when $\{ V(k, n) \}$ does. (We do not assume that $\{ V(k, n) \}$ is i.i.d. here.) For this purpose, let $D[0, \infty)$ be the space of right-continuous real-valued functions on the interval $[0, \infty)$ with limits from the left, endowed with the usual Skorohod (1956) J_1 topology; see Ethier and Kurtz (1986) or Whitt (1980). Let $D[0, \infty)^\infty$ be the product space endowed with the product topology.

Let V_n and D_n be random elements of $D[0, \infty)^\infty$ defined as follows:

$$\begin{aligned} V_n &= (V_{1n}, V_{2n}, \ldots) \\ D_n &= (D_{1n}, D_{2n}, \ldots) \\ V_{kn}(t) &= n^{-\alpha} \left[\sum_{j=1}^{[nt]} V(k, j) - nt \right], \ t \geq 0 \ , \\ D_{kn}(t) &= n^{-\alpha} (D(k, [nt]) - nt) \ , \ t \geq 0 \ , \end{aligned} \tag{3.1}$$

for $\alpha > 0$.

Theorem 3.1. If $V_n \Rightarrow \hat{V}$ in $D[0, \infty)^\infty$ as $n \to \infty$ where $\hat{V}$ has continuous paths w.p.1, then $D_n \Rightarrow \hat{D}$ in $D[0, \infty)^\infty$ as $n \to \infty$, where $\hat{D} = f(\hat{V})$ with $f : D[0, \infty)^\infty \to D[0, \infty)^\infty$ defined by

$$f_1(x)(t) = x_1(t)$$

and

$$\begin{aligned} f_k(x)(t) &= \sup_{0 \leq s \leq t} \{ f_{k-1}(s) + x_k(t) - x_k(s) \} \\ &= x_k(t) - \inf_{0 \leq s \leq t} \{ x_k(s) - f_{k-1}(s) \} \end{aligned} \tag{3.2}$$

for all $k \geq 2$ and $t \geq 0$.

Proof. First, from (2.2) and (3.1) it is immediate that $D_{1n} = V_{1n}$. Next,

$$
\begin{aligned}
D_{kn}(t) &= n^{-\alpha}(D(k, [nt] - nt) \\
&= n^{-\alpha}\left[\max_{1 \le l \le [nt]} \{D(k-1, l) + \sum_{j=l}^{[nt]} V(k, j)\} - nt\right] \\
&= n^{-\alpha}\left[\sup_{0 \le s \le t} \{D(k-1, [ns]) - ns + \sum_{j=[ns]}^{[nt]} V(k, j) - n(t-s)\}\right] \\
&= \sup_{0 \le s \le t} \{D_{k-1,n}(s) + V_{kn}(t) - V_{kn}(s) + n^{-\alpha} V(k, [ns])\} \ .
\end{aligned}
$$

However, since $V_{kn} \Rightarrow \hat{V}_k$ where $\hat{V}_k$ has continuous paths,

$$
\sup_{0 \le s \le t} n^{-\alpha} \ V(k, [ns]) \Rightarrow 0 \quad \text{in} \quad D[0, \infty) \ ;
$$

i.e., the maximum jump functional is continuous. Hence, by the convergence-together theorem (Theorem 4.1 of Billingsley) and induction, $(D_{1n}, \ldots, D_{kn})$ converges if $(f_1(V_n), \ldots, f_k(V_n))$ converges. However, it is easy to see (e.g., by Section 6 of Whitt (1980) and induction) that $(f_1, \ldots, f_k) : D[0, \infty)^\infty \to D[0, \infty)^k$ is continuous for each k. Since we are using the product topology, that implies that f itself is continuous. Hence, the desired convergence holds by the continuous mapping theorem (Theorem 5.1 of Billingsley). ■.

Remark (3.1) By the duality in Corollaries 2-4 to Proposition 2.1, Theorem 3.1 can also be regarded as a direct consequence of previous heavy-traffic limit theorems for the sojourn times of the first $\lfloor nt \rfloor$ customers at the first k queues; see Iglehart and Whitt (1970), Harrison (1973), and Reiman (1984). For the sojourn times, the case we consider corresponds to having the traffic intensity at queue i be $\rho_i = 1$ for all i. As in previous heavy-traffic limit theorems, we could let the service-time distributions change in the limit. ■

We can obtain a representation for the limit process $\hat{D}$ in Theorem 3.1 paralleling the representation of $D(k, n)$ as the maximal partial sum of the service times over all paths in $\Pi(k, n)$ in (2.4). For this purpose let $T_k(t)$ be the set of nondecreasing $(k+1)$-tuples $(t_0, t_1, \ldots, t_k)$ with $t_0 = 0$ and $t_k = t$. The following is deduced from (3.2) by induction on k.

Corollary. The limit process $\hat{D} \equiv \{\hat{D}_k : k \geq 1\} \equiv f(\hat{V}) \equiv \{f_k(\hat{V}_1, \ldots, \hat{V}_k) : k \geq 1\}$ can be represented as

$$\hat{D}_k(t) = \sup\{\sum_{i=1}^{k} [\hat{V}_i(t_i) - \hat{V}_i(t_{i-1})] : (t_0, t_1, \ldots, t_k) \in T_k(t)\} \tag{3.3}$$

for all $k \geq 1$.

The standard case has normalization exponent $\alpha = 1/2$ in (3.1) and service-time limit process $\hat{V}$ being Brownian motion (BM), i.e., a vector of independent one-dimensional BMs. The resulting limit process $\hat{\Delta}$ for the interdeparture-time process is then an infinite-dimensional reflected Brownian motion (RBM) on the infinite-dimensional orthant. Such infinite-dimensional RBMs can be constructed by extending corresponding *k*-dimensional RBMs on the *k*-dimensional orthant; see p. 83 of Neveu (1965). The *k*-dimensional RBMs in turn coincide with those considered by Harrison (1978), Harrison and Reiman (1981a,b), Reiman (1984) and Harrison and Williams (1987a,b).

Let $\hat{B} = (\hat{B}_1, \hat{B}_2, \ldots)$ be a standard BM on $D[0, \infty)^\infty$, by which we mean a vector of independent standard (drift 0, diffusion coefficient 1) BMs. To obtain the standard limiting case, we assume that the service times are i.i.d. However, in order to cover the case of a general external arrival process, we exclude finitely many queues in the condition.

Theorem 3.2. If there exists a finite m such that $\{V(k, n) : k \geq 1,\ n \geq m\}$ is i.i.d. with $E\ V(1, m) = 1$ and Var $V(1, m) = \sigma^2 < \infty$, then the condition of Theorem 3.1 holds with $\hat{V} = \sigma\hat{B}$ where $\hat{B}$ is a standard BM. Then $\hat{D} = \sigma f(\hat{B})$ for f in (3.2). The associated interdeparture-time limit process $\hat{\Delta}$, defined by $\hat{\Delta}_k = \hat{D}_{k+1} - \hat{D}_k$, $k \geq 1$, and $\hat{\Delta}_0 = \hat{D}_1$, can be represented as

$$\hat{\Delta}_0 = \sigma \hat{B}_1 \ , \ \hat{Y}_k = \sigma \hat{B}_{k+1} - \sum_{i=0}^{k-1} \hat{\Delta}_i$$

$$\hat{\Delta}_k(t) = \hat{Y}_k(t) - \inf_{0 \le s \le t} \hat{Y}_k(s) \equiv \hat{Y}_k(t) + \hat{I}_k(t) \ , \ k \ge 1 \ . \tag{3.4}$$

Then $[(\hat{\Delta}_1, \ldots, \hat{\Delta}_k), (\hat{I}_1, \ldots, \hat{I}_k)]$ are the unique pair of k-dimensional processes so that $\hat{\Delta}_i(t) = \hat{Y}_i(t) + \hat{I}_i(t)$, $\hat{\Delta}_i(t) \ge 0$, $\hat{I}_i(t)$ is nondecreasing with $\hat{I}_i(0) = 0$ and

$$\int_0^t 1_{\{\hat{\Delta}_i(s) \ne 0\}} d\hat{I}_i(s) = 0$$

for $1 \le i \le k$ and $t \ge 0$. Moreover, for each k, $(\hat{\Delta}_1, \ldots, \hat{\Delta}_k)$ is a k-dimensional RBM as in Harrison and Reiman (1981a,b) generated by a zero-drift BM with covariance matrix Σ having elements $\Sigma_{ii} = 2\sigma^2$, $1 \le i \le k$, $\Sigma_{i,i+1} = \Sigma_{i+1,i} = -\sigma^2$, $1 \le i \le k-1$, and $\Sigma_{ij} = 0$ otherwise, and reflection matrix $R = I - Q$, where $Q_{i,i+1} = 1$ for $1 \le i \le k-1$ and $Q_{ij} = 0$ otherwise.

Proof. By Theorems 3.2, 4.1 and 16.1 of Billingsley (1968), $V_n \Rightarrow \sigma \hat{B}$. By induction, $f(\sigma x) = \sigma f(x)$ for f in (3.2). Hence, $\hat{D} = f(\sigma \hat{B}) = \sigma f(\hat{B})$. The representation (3.4) is an easy consequence of (3.2). The characterization of the pair $[(\hat{\Delta}_1, \ldots, \hat{\Delta}_k), (\hat{I}_1, \ldots, \hat{I}_k)]$ follows from repeated application of the one-dimensional characterization of the reflection map on p. 19 of Harrison (1985) (sometimes called Skorohod's lemma (1961)), and induction. The characterization of $(\hat{\Delta}_1, \ldots, \hat{\Delta}_k)$ as an RBM follows by the arguments of Harrison (1978) and Harrison and Reiman (1981a,b) or directly from those papers, after exploiting the duality in Corollaries 2-4 of Proposition 2.1. The RBM structure is easy to see in this case of an acyclic network by writing (3.4) in differential form. Then

$$d\hat{\Delta}_0 = d\hat{B}_1$$

$$d\hat{\Delta}_k = d\hat{B}_{k+1} - \sum_{i=0}^{k-1} d\hat{\Delta}_i + d\hat{I}_k \ . \tag{3.5}$$

By induction, (3.5) can be rewritten as

$$\begin{aligned} d\hat{\Delta}_0 &= d\hat{B}_1 \\ d\hat{\Delta}_1 &= d\hat{B}_2 - d\hat{B}_1 + d\hat{I}_1 \\ d\hat{\Delta}_k &= d\hat{B}_{k+1} - d\hat{B}_k - d\hat{I}_{k-1} + d\hat{I}_k \ , \ k \geq 2 \ . \end{aligned} \tag{3.6}$$

This is the differential form for the RBM plus Δ_0; i.e., from (3.6) we obtain $Z = X + YR$ as in Harrison and Reiman (1981a,b), where $Z = (\hat{\Delta}_1 , \ldots , \Delta_k)$, X is the BM with components $X_i = \hat{B}_{i+1} - \hat{B}_i$ and $Y = (\hat{I}_1 , \ldots , \hat{I}_k)$. ■

Remarks (3.2) Additional characterizations of the departure RBM such as the generator and a generalized Itô's formula follow from Harrison and Reiman (1981a,b). Since the BMs $\hat{B}_i$ in the construction have zero drift, the departure RBM does not have a proper stationary distribution.

(3.3) We do not know much about the joint distribution of $(\hat{\Delta}_1 (1) , \ldots , \hat{\Delta}_k (1))$, but simulation results appear in Grember, Schlunk and Whitt (1990). Since $\hat{\Delta}_1 = \sigma\hat{B}_2 - \sigma\hat{B}_1$, $\hat{\Delta}_1 \overset{d}{=} \sqrt{2}\ \sigma\ |B_1|$. Hence, $\hat{\Delta}_1 (1)$ has a positive normal distribution with $E[\hat{\Delta}_1 (1)] = 2\sigma/\sqrt{\pi}$ and $E[\hat{\Delta}_1 (1)^2] = 2\sigma^2$. In Section 5 we show that $\hat{\Delta}_k (t)$ is stochastically decreasing in *k* and stochastically increasing in *t*.

4. The Strong Approximation

Under the assumptions of Theorem 3.2, we know that the interdeparture times of the k^{th} customer from the n^{th} queue are asymptotically of order $\sqrt{n}$ as $n \to \infty$ for any *k*. We now want to say what happens if the customer index increases with *n*. For this purpose, we establish a strong approximation result, drawing on Komlós, Major and Tusnády (1975, 1976); see p. 107 of Csörgő′ and Révész (1981). Strong approximations have been used previously and/or concurrently to study queueing models by Csörgő′, Deheuvels and Horvath (1987), Glynn and Whitt (1991), Hanqin, Guanghui and Rongxin (1990) and Horvath (1990). These papers obtain rates of convergence via the strong approximations. We also obtain rates of convergence, but our motivation is different.

We show that the error in the diffusion approximation in Theorem 3.2 is $O(n^{(a\,-\,½)}\log n)$ when the largest customer index k is n^a. We state the result below in an equivalent unnormalized form; to obtain the stated bound, divide through by $\sqrt{n}$.

Theorem 4.1. If, in addition to the assumption of Theorem 3.2, all service times are independent and there exist positive constants K and λ such that $P(V(k,\,j) > x) \le Ke^{-\lambda x}$ for all $k,\,j$ and x, then there exists a probability space supporting the departure times $D(k,\,j)$ and the limit process $\hat{D} = \sigma f(\hat{B})$ such that, for any $a > 0$,

$$\max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le j \le n}} \{|D(k,\,j) - j - \sqrt{n}\,\hat{D}_k(j/n)|\} = O(n^a \log n) \;\; \text{a.s.}$$

Remarks. (4.1) Theorem 4.1 helps establish (1.5) under (1.6). To determine the order of magnitude of $D(k_n,\,n)$ for $k_n = \lfloor xn^a \rfloor$ for $0 < a < 1$, we have thus reduced the problem to determining how $\hat{D}_k(1)$ behaves as $k \to \infty$, which we discuss in Section 7.

(4.2) In Theorem 4.1 we focus on the departure times, but a corresponding result holds for the interdeparture times $\Delta(k,\,n)$ in (2.5) by applying the triangle inequality. In particular, as an immediate consequence of Theorem 4.1,

$$\max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le j \le n}} \{|\Delta(k,\,j) - \sqrt{n}\,[\hat{D}_{k+1}(j/n) - \hat{D}_k(j/n)]|\} = O(n^a \log n) \;\; \text{a.s.} \tag{4.1}$$

Theorem 4.1 is proved by combining Lemmas 4.4 and 4.5 below. Lemmas 4.1–4.3 below are used to prove Lemma 4.4.

Lemma 4.1. If $\{U_k : k \ge 1\}$ is a sequence of independent random variables and there exist positive constants K and λ such that $P(U_k > x) \le Ke^{-\lambda x}$ for all $x > 0$, then for any $a > 0$

$$\max_{1 \le k \le \lfloor n^a \rfloor} \{U_k\} = O(\log n) \quad \text{a.s.}$$

Proof. For any x_n,

$$P\left[\max_{1 \le k \le \lfloor n^a \rfloor} \{U_k\} > x_n\right] \le 1 - (1 - Ke^{-\lambda x_n})^{n^a} \ .$$

Hence, for $x_n = (a + 2) \log n/\lambda$,

$$\begin{aligned} P(A_n) \equiv P(\max_{1 \le k \le \lfloor n^a \rfloor} \{U_k\} > \frac{a+2}{\lambda} \log n) &\le 1 - (1 - Kn^{-(a+2)})^{n^a} \\ &\le 1 - \exp(\log[(1 - Kn^{-(a+2)})^{n^a}]) \\ &\le 1 - \exp(n^a \log [1 - Kn^{-(a+2)}]) \\ &\le - n^a \log (1 - Kn^{-(a+2)}) \\ &\le 2Kn^{-2} \quad \text{for } n \text{ sufficiently large} \end{aligned}$$

using $e^{-x} \ge 1 - x$ in the second to last step and $\log(1 - x) = - x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$ for $0 < x < 1$ in the last step. Since $\sum_{n=1}^{\infty} P(A_n) < \infty$, $P(A_n$ infinitely often$) = 0$ by the Borel-Cantelli lemma. Hence, there are positive random variables X_1 and X_2 such that

$$\max_{1 \le k \le \lfloor n^a \rfloor} \{U_k\} \le X_1 + X_2 \log n \ \text{ for all } n \ge 1 \quad \text{a.s.} \quad \blacksquare$$

We now extend a strong approximation result of Komlós, Major and Tusnády (1975, 1976), p. 107 of Csörgó′ and Révész (1981).

Lemma 4.2. Under the assumptions of Theorem 4.1, there is a probability space supporting independent standard BMs $\hat{B}_k$ and the service times so that

$$\max_{\substack{1 \le k < \lfloor n^a \rfloor \\ 1 \le l \le n}} \left\{ \left| \sum_{j=1}^{l} V(k, j) - l - \sigma \hat{B}_k(l) \right| \right\} = O(\log n) \quad \text{a.s.}$$

Proof. The service times of all customers at all queues after the first m are i.i.d., but we do not have identical distributions at earlier queues. However, by Lemma 4.1 and the assumption of Theorem 4.1, without loss of generality it suffices to assume that all the service times are i.i.d.

To support half this claim, note that $\sigma \hat{B}_k(l)$ is normally distributed with mean 0 and variance $l\sigma^2$, so that these variables satisfy the same tail condition imposed on the service times for $1 \le l \le m$ (possibly with different constants K and λ). Hence, it suffices to assume that $\{V(k, j)\}$ is i.i.d., with the distribution of $V(1, m)$, and we do. By Komlós, Major and Tusnády (1975, 1976), for each k there is a probability space containing a BM $\hat{B}_k$ such that

$$P\{ \max_{1 \le l \le n} | \sum_{j=1}^{l} V(k, j) - l - \sigma \hat{B}_k(l) | > C \log n + x \} < K e^{-\lambda x}$$

for positive constants C, K and λ depending on the distribution of $V(1, m)$. Hence, using a product space, we can achieve

$$P(A_n) \equiv P \left\{ \max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le l \le n}} \left| \sum_{j=1}^{l} V(k, j) - l - \sigma \hat{B}_k(l) \right| > C \log n + x_n \right\}$$

$$\le 1 - (1 - K e^{-\lambda x_n})^{\lfloor n^a \rfloor} .$$

As in Lemma 4.1, choose $x_n = (a + 2) \log n / \lambda$ to obtain $P(A_n) \le 2Kn^{-2}$ for n sufficiently large. By Borel-Cantelli, $P(A_n \text{ infinitely often}) = 0$. Hence, there exist random variables X_1 and X_2 such that

$$\max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le l \le n}} \left\{ \left| \sum_{j=1}^{l} V(k, j) - l - \sigma \hat{B}_k(l) \right| \right\} \le X_1 + X_2 \log n \quad \text{a.s.} \quad \blacksquare$$

For the next lemma, we specify some quantities associated with a real-valued function defined on the positive integers, say y. Let

$$y^{\uparrow}(n) = \max_{1 \le k \le n} y(k) \quad \text{and} \quad \| y \|_n = |y|^{\uparrow}(n) , \ n \ge 1 . \qquad (4.2)$$

The following elementary lemma can be viewed as a special case of Theorem 6.1 of Whitt (1980).

Lemma 4.3. For all $n \geq 1$, $\| y_1^{\uparrow} - y_2^{\uparrow} \|_n \leq \| y_1 - y_2 \|_n$.

Let $D^*(k, n)$ be the following function of the limiting BM $\hat{B}$,

$$\begin{aligned} D^*(1, n) &= \sigma \hat{B}_1(n) \\ D^*(k, n) &= \sigma \hat{B}_k(n) - \min_{1 \leq j \leq n} \{ \sigma \hat{B}_k(j) - D^*(k-1, j) \} \\ &= \max_{1 \leq j \leq n} \{ D^*(k-1, j) + \sigma \hat{B}_k(n) - \sigma \hat{B}_k(j) \} \end{aligned} \tag{4.3}$$

for $n \geq 1$ and $k \geq 2$. Let e denote the identity function, i.e., $e(t) = t, t \geq 0$.

Lemma 4.4. Under the assumptions of Theorem 4.1, for any $a > 0$ there exists a probability space supporting the departure times $D(k, j)$ and the process D^* in (4.3) such that

$$\max_{\substack{1 \leq k \leq \lfloor n^a \rfloor \\ 1 \leq j \leq n}} \{ | D(k, j) - j - D^*(k, j) | \} = O(n^a \log n) \text{ a.s.}$$

Proof. Note that (4.3) is not quite the same function of $\sigma \hat{B}_k(n)$ as $D(k, n)$ is of $\sum_{j=1}^{n} [V(k, j) - j]$ in (2.2). The exactly corresponding function is

$$\begin{aligned} D'(1, n) &= \sigma \hat{B}_1(n) \\ D'(k, n) &= \max_{1 \leq j \leq n} \{ D(k-1, j) + \sigma \hat{B}_k(n) - \sigma \hat{B}_k(j-1) \} . \end{aligned} \tag{4.4}$$

However, by (2.2), (4.1), (4.2), (4.4) and Lemmas 4.1 and 4.3, there are random variables X_1 and X_2 such that

$$\begin{aligned} \| D'(k, \cdot) - D^*(k, \cdot) \|_n &\leq \| D'(k-1, \cdot) - D^*(k-1, \cdot) \|_n \\ &\quad + \max_{1 \leq j \leq n} \{ | \sigma \hat{B}_k(j) - \sigma \hat{B}_k(j-1) | \} \\ &\leq \| D'(k-1, \cdot) - D^*(k-1, \cdot) \|_n + X_1 + X_2 \log n \qquad (4.5) \\ &\leq k(X_1 + X_2 \log n) \quad \text{for } k \leq \lfloor n^a \rfloor , \end{aligned}$$

where we have used the fact that $\sigma \hat{B}_k(j) - \sigma \hat{B}_k(j-1)$ are i.i.d. normal random variables in the second to last step and induction in the last step. In particular, by Lemma 4.1,

$$\max_{\substack{1 \le k \le \rfloor n^a \rfloor \\ 1 \le j \le n}} \{ |\sigma \hat{B}_k(j) - \sigma \hat{B}_k(j-1)| \} = O(\log n) \text{ a.s.}$$

Hence,

$$\max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le j \le n}} \{ |D'(k, j) - D^*(k, j)| \} \le n^a (X_1 + X_2 \log n)$$

and it suffices to do the proof with D' in (4.4) instead of D^* in (4.3). By an argument just like (4.5), using Lemma 4.2 now, there exists a probability space supporting the processes D and D' and finite random variables X_1 and X_2 such that

$$\| D(1,\cdot) - e(\cdot) - D'(1,\cdot) \|_n \le X_1 + X_2 \log n \quad \text{a.s.}$$

and

$$\| D(k,\cdot) - e(\cdot) - D'(k, \cdot) \|_n \le X_1 + X_2 \log n + \| D(k-1,\cdot) - e(\cdot) - D'(k-1,\cdot) \|_n , \quad \text{a.s.}$$

for $1 \le k \le \lfloor n^a \rfloor$. Hence,

$$\| D(k, \cdot) - e(\cdot) - D'(k, \cdot) \|_n \le \lfloor n^a \rfloor (X_1 + X_2 \log n) = O(n^a \log n) \quad \text{a.s.}$$

for all $k \le \lfloor n^a \rfloor$. ■

To prove our next lemma we want a continuous analog of (4.2). For a real-valued function of a real-variable, say y, let

$$y^{\uparrow}(t) = \sup_{0 \le s \le t} y(s) \quad \text{and} \quad \| y \|_t = |y|^{\uparrow}(t) , \; t \ge 0 \ . \tag{4.6}$$

Paralleling Lemma 4.3,

$$\| y_1^{\uparrow} - y_2^{\uparrow} \|_t \le \| y_1 - y_2 \|_t \ . \tag{4.7}$$

Lemma 4.5. For any $a > 0$

$$\max_{\substack{1 \le k \le \lfloor n^a \rfloor \\ 1 \le j \le n}} \{ |D^*(k, j) - \sqrt{n} \overrightarrow{D}_k(j/n)| \} = O(n^a \log n) \quad \text{a.s.}$$

Proof. Note that

$$\{\sqrt{n}\overline{\hat{D}}(t/n) : t \geq 0\} \overset{d}{=} \{\hat{D}(t) : t \geq 0\}$$

and

$$\{D^*(k, [t]) : k \geq 1, t \geq 0\} \overset{d}{=} \{\hat{D}_k([t]) : k \geq 1, t \geq 0\} .$$

Hence, what we want to show is

$$\sup_{\substack{1 \leq k \leq \lfloor n^a \rfloor \\ 0 \leq t \leq n}} \{|\hat{D}_k(t) - \hat{D}_k([t])|\} = O(n^a \log n) \quad \text{a.s.} \tag{4.8}$$

By (4.7),

$$\|\hat{D}_k(\cdot) - \hat{D}_k([\cdot])\|_n \leq \|\hat{D}_{k-1}(\cdot) - \hat{D}_{k-1}([\cdot])\|_n + \|\sigma\hat{B}_k(\cdot) - \sigma\hat{B}_k([\cdot])\|_n . \tag{4.9}$$

However,

$$\begin{aligned}\max_{1 \leq k \leq \lfloor n^a \rfloor} \|\sigma\hat{B}_k(\cdot) - \sigma\hat{B}_k([\cdot])\|_n &\leq \sigma \max_{\substack{1 \leq k \leq \lfloor n^a \rfloor \\ 1 \leq j \leq n}} \{ \sup_{j < s < j+1} \{|\hat{B}_k(s) - \hat{B}_k(j)|\}\} \qquad (4.10)\\ &\leq X_1 + X_2 \log n \quad \text{a.s.}\end{aligned}$$

for finite random variables X_1 and X_2, by Lemma 4.1. Combining (4.9) and (4.10) gives (4.8). ■

5. Stochastic Order for the Interdeparture Times

In this section we establish stochastic comparisons for the interdeparture times $\Delta(k, n)$ in (2.5). We say that a random element X_1 is stochastically less than or equal to another random element X_2, and write $X_1 \leq_{st} X_2$, if $Eh(X_1) \leq Eh(X_2)$ for all nondecreasing bounded measurable real-valued functions h; see Kamae, Krengel and O'Brien (1977). We are interested in the case X_i is an array of real-valued random variables. As before, let $\overset{d}{=}$ denote equality in distribution.

Theorem 5.1. Suppose that the service times $V(k, n)$ are all mutually independent.

(a) If $V(k, n) \overset{d}{=} V(1, n)$ for all $k \geq 1$ and $n \geq 1$, then

$$\{\Delta(k+1,n) : k \geq 1, n \geq 1\} \leq_{st} \{\Delta(k,n) : k \geq 1, n \geq 1\} ,$$

so that

$$V(1,n) \leq_{st} \Delta(k+1,n) \leq_{st} \Delta(k,n) \quad \text{for } k \geq 1 \text{ and } n \geq 1 .$$

(b) If $V(k,n) \overset{d}{=} V(k,1)$ for all $k \geq 1$ and $n \geq 1$, then

$$\{\Delta(k,n) : k \geq 1, n \geq 1\} \leq_{st} \{\Delta(k,n+1) : k \geq 1, n \geq 1\} ,$$

so that

$$V(k+1,1) \leq_{st} \Delta(k,n) \leq_{st} \Delta(k,n+1) \text{ for } k \geq 1 \text{ and } n \geq 1 .$$

Proof. We do only part (a) because the proof of (b) is similar. We construct a process $\{\tilde{\Delta}(k+1,n) : k \geq 1, n \geq 1\}$ with the same finite-dimensional distributions as $\{\Delta(k+1,n) : k \geq 1, n \geq 1\}$ such that

$$\tilde{\Delta}(k+1,n) \leq \Delta(k,n) \quad \text{a.s. for all } k \geq 1 \text{ and } n \geq 1 . \tag{5.1}$$

For this purpose, we use service times $\tilde{V}(k,n)$ defined by $\tilde{V}(k+1,n) = V(k,n)$ for all k and n. By our assumptions, $\{\tilde{V}(k,n) : k \geq 1, n \geq 1\}$ is distributed the same as $\{V(k,n) : k \geq 1, n \geq 1\}$. We define $\tilde{\Delta}(k,n)$ and $\tilde{D}(k,n)$ just like $\Delta(k,n)$ and $D(k,n)$ but using the service times $\tilde{V}(k,n)$ instead of $V(k,n)$.

By (2.1),

$$\begin{aligned} \Delta(k,n) &= \max\{D(k,n), D(k+1,n-1)\} + V(k+1,n) - D(k,n) \\ &= [D(k+1,n-1) - D(k,n)]^{+} + V(k+1,n) . \end{aligned} \tag{5.2}$$

Hence,

$$\tilde{\Delta}(k+1,n) = [\tilde{D}(k+2,n-1) - \tilde{D}(k+1,n)]^{+} + V(k+1,n) . \tag{5.3}$$

From (5.2) and (5.3), we see that (5.1) holds if

$$\tilde{D}(k+2, n-1) - \tilde{D}(k+1, n) \leq D(k+1, n-1) - D(k, n) \tag{5.4}$$

for all $k \geq 1$ and $n \geq 1$. We establish (5.4) by induction on the sum of the indices ($m = k + n$ in $D(k, n)$). Note that

$$\begin{aligned}
&\tilde{D}(k+2, n-1) - \tilde{D}(k+1, n) = \\
&\qquad \max\{\tilde{D}(k+1, n-1), \tilde{D}(k+2, n-2)\} + \tilde{V}(k+2, n-1) \\
&\qquad - \max\{\tilde{D}(k, n), \tilde{D}(k+1, n-1)\} - \tilde{V}(k+1, n) \\
&\quad = [\tilde{D}(k+2, n-2) - \tilde{D}(k+1, n-1)]^{+} + \tilde{V}(k+2, n-1) \\
&\qquad - [\tilde{D}(k, n) - \tilde{D}(k+1, n-1)]^{+} - \tilde{V}(k+1, n) \\
&\quad = [\tilde{D}(k+2, n-2) - \tilde{D}(k+1, n-1)]^{+} + V(k+1, n-1) \\
&\qquad\ + [\tilde{D}(k+1, n-1) - \tilde{D}(k, n)]^{-} - V(k, n) \\
&\quad \leq [D(k+1, n-2) - D(k, n-1)]^{+} + V(k+1, n-1) \\
&\qquad + [D(k, n-1) - D(k-1, n)]^{-} - V(k, n) \equiv Z
\end{aligned}$$

by the induction hypothesis, where

$$\begin{aligned}
Z &= \max\{D(k+1, n-2), D(k, n-1)\} - D(k, n-1) + V(k+1, n-1) \\
&\quad - \max\{D(k-1, n), D(k, n-1)\} + D(k, n-1) - V(k, n) \\
&= D(k+1, n-1) - D(k, n) .
\end{aligned}$$

To start the induction, note that, for $n = 1$ and any k,

$$\begin{aligned}
\tilde{D}(k+2, n-1) - \tilde{D}(k+1, n) &= -\tilde{D}(k+1, 1) \\
= -(\tilde{V}(1, 1) + \cdots + \tilde{V}(k+1, 1)) &= -\tilde{V}(1, 1) - [V(1, 1) + \cdots + V(k, 1)] \\
\leq -D(k, 1) &= D(k+1, 0) - D(k, 1) .
\end{aligned}$$

Hence, (5.4) is established and the proof is complete. ■

Corollary 1. Under the conditions of Theorem 5.1(a), for each $n \geq 1$ there exists a proper stochastic process $\{\tilde{\Delta}(k, n) : k \geq 1\}$ with $\tilde{\Delta}(k, n) \geq_{st} V(1, n)$ such that

$$\{\Delta(k+j, n) : k \geq 1\} \Rightarrow \{\tilde{\Delta}(k, n) : k \geq 1\} \text{ in } R^{\infty} \text{ as } j \to \infty .$$

We can apply Theorem 5.1 to obtain a stochastic comparison for the limit process in

Theorem 3.1. We actually focus on the associated interdeparture-time limit process $\hat{\Delta}_k(t) = \hat{D}_{k+1}(t) - \hat{D}_k(t)$.

Corollary 2. Suppose that the service times $V(k, n)$ are all independent and the FCLT $V_n \Rightarrow \hat{V}$ holds as required for Theorem 3.1.

(a) If $V(k, n) \overset{d}{=} V(1, n)$ for all $k \geq 1$ and $n \geq 1$, then

$$\{\hat{\Delta}_{k+1}(t) : k \geq 1, t \geq 0\} \leq_{st} \{\hat{\Delta}_k(t) : k \geq 1, t \geq 0\}$$

for all $k \geq 1$.

(b) If $V(k, n) \overset{d}{=} V(k, 1)$ for all $k \geq 1$ and $n \geq 1$, then

$$\{\hat{\Delta}_k(t) : k \geq 1, t \geq 0\} \leq_{st} \{\hat{\Delta}_k(t + u) : k \geq 1, t \geq 0\}$$

for all $u > 0$.

Proof. Use the fact that stochastic order is preserved under weak convergence. ■

6. The Hydrodynamic Limit: The Case $\mathbf{k_n} = O(\mathbf{n})$

In this section we describe the behavior of $D(k_n, n)$ (and, equivalently, $D(n, k_n)$) when k_n is of order n. We first apply the hydrodynamic limit of Rost (1981) as discussed in Section 4.2 of Srinivasan (1989) to treat the special case of exponential service times.

Theorem 6.1. If all the service times are i.i.d. with an exponential distribution having mean 1, then

$$\lim_{n \to \infty} n^{-1} D(\lfloor xn \rfloor, n) = (1 + \sqrt{x})^2 \quad \text{a.s.} \quad \text{for any } x > 0 .$$

Proof. By Section 4.2 of Srinivasan (1989), the average queue length among the first $\lfloor xt \rfloor$ queues at time t is asymptotically a.s. $(2 - \sqrt{x})/\sqrt{x}$ as $t \to \infty$. Hence, for $x > 1$, the average queue length among the first n and $\lfloor x^2 n \rfloor$ queues at time $x^2 n$ are asymptotically a.s. $(2 - x^{-1})/x^{-1} = 2x - 1$ and 1, respectively, as $n \to \infty$. Hence asymptotically a.s. there are

$x^2 n + o(n)$ customers in queues 2 through $x^2 n$ and $(2x - 1)n + o(n)$ customers in queues 2 through n. Hence, asymptotically a.s. $(x^2 - 2x + 1)n + o(n)$ customers have departed from queue n, and the departure time for customer $(x^2 - 2x + 1)n$ from queue n is $x^2 n + o(n)$. Now do a change of variables, replacing $(x - 1)^2$ by x.

To treat $x < 1$, note that $n^{-1} D(\lfloor x^2 n \rfloor, n) \overset{d}{=} n^{-1} D(n, \lfloor x^2 n \rfloor)$. Let $n' = x^2 n$. Then $n^{-1} D(n, \lfloor x^2 n \rfloor) = (x^2/n) D(\lfloor n/x^2 \rfloor, n) + o(1)$. From the previous argument,

$$(x^2/n) D(\lfloor n/x^2 \rfloor, n) \to x^2 \left[\frac{1}{x} + 1 \right]^2 = (x+1)^2 \quad \text{a.s. as } n \to \infty .$$

For $x = 1$, consider the average queue lengths among the first n and $4n$ queues, and reason similarly. ■

We now establish the existence of a limit for a general service-time distribution having an exponential tail. First, recall that under the conditions of Theorem 4.1,

$$\max_{\substack{1 \le i \le \lfloor xn \rfloor \\ 1 \le j \le n}} \{V(i, j)\} = O(\log n) \quad \text{a.s.}$$

by Lemma 4.1, so that

$$D(\lfloor xn \rfloor, n) \le O(n \log n) \quad \text{a.s.}$$

However, we will show that $D(\lfloor xn \rfloor, n)$ is actually $O(n)$.

For this purpose, we exploit a stochastic comparison involving associated random variables; see p. 29 of Barlow and Proschan (1975). Recall that a family of random variables are *associated* if all pairs of nondecreasing bounded real-valued functions of the random variables have nonnegative correlation.

Lemma 6.1. If the service times $\{V(i, j) : 1 \le i \le k,\ 1 \le j \le n\}$ are independent or just associated, then the partial sums $\sum_{(i,j) \in \pi} V(i, j)$ for the $\left[\begin{matrix} k + n - 2 \\ k - 1 \end{matrix} \right]$ paths π in $\Pi(k, n)$ are associated random variables.

Proof. The partial sums are all nondecreasing functions of the *kn* service times. ■

Theorem 6.2. If the service times $V(i, j)$ are all independent, then

$$D(k, n) \leq_{st} \max\{S_\pi : \pi \in \Pi(k, n)\}$$

where S_π, $\pi \in \Pi(k, n)$, are mutually independent with

$$S_\pi \overset{d}{=} \sum_{(i, j) \in \pi} V(i, j)$$

for each path π.

Proof. Apply Theorem 3.2, p. 33 of Barlow and Proschan (1975). ■

We now use this stochastic bound to develop a bound and heuristic estimate for $\lim_{n \to \infty} n^{-1} D(\lfloor xn, n \rfloor)$ for a general service-time distribution. We call this the *path-independence bound.* We also use this bound together with the subadditive ergodic theorem to show that the limit exists.

Theorem 6.3. If all the service times are i.i.d. with $EV(1, 1) = 1$ and there exist positive constants K and λ such that $P(V(1, 1) > x) \leq Ke^{-\lambda x}$ for all $x > 0$, then there exists a deterministic strictly increasing concave function $\gamma(x)$ with $\gamma(x) \geq 1$, $\gamma(x + y) - \gamma(x) \geq y$ and $\gamma(x) = x\gamma(x^{-1})$ for all $x, y > 0$ such that

$$\lim_{n \to \infty} n^{-1} D(\lfloor nx \rfloor, n) = \gamma(x) \quad \text{a.s.} \tag{6.1}$$

and

$$\lim_{n \to \infty} E|n^{-1} D(\lfloor nx \rfloor, n) - \gamma(x)| = 0 \tag{6.2}$$

for all $x > 0$. Moreover,

$$\gamma(x) \ge \begin{cases} 1 + xE(\max\{V(1,2), V(2,1)\}) \ge 1 + x \,, \; 0 < x \le 1 \\ \\ x + E(\max\{V(1,2), V(2,1)\}) \ge 1 + x \,, \; 1 \le x \,, \end{cases} \tag{6.3}$$

and

$$\gamma(x) \le (1 + a^*)(1 + x) \;, \tag{6.4}$$

where

$$a^* \equiv a^*(x) = \inf\{a > 0 : (1 + x)h(a) > (1 + x) \log (1 + x) - x \log x\} \tag{6.5}$$

and

$$h(a) = \sup_{\theta} \{\theta a - \log E e^{\theta[V(1,m) - 1]}\} \;. \tag{6.6}$$

Proof. We first establish the upper bound in (6.4). Using Stirling's formula, p. 52 of Feller (1968), we see that the number of paths in $\Pi(\lfloor xn \rfloor, n)$ is $\phi(x, n) = e^{n\psi(x) + o(n)}$, where

$$\psi(x) = (1 + x) \log (1 + x) - x \log x \;.$$

Let π_n be a path in $\Pi(\lfloor xn \rfloor, n)$ and let S_{π_n} be the partial sum of all service times on path π_n. By the Cramér (1938)–Chernoff (1952) theorem,

$$P(S_{\pi_n} > (1 + a)(1 + x)n) = e^{-(1 + x)nh(a) + o(n)}$$

where h is defined in (6.6); see Vanderbei and Weiss (1988) or pp. 3,7 of Varadhan (1984). Using the path-independence bound established in Theorem 6.2, we have

$$P(D(\lfloor xn \rfloor, n) > (1 + a)(1 + x)n) \le 1 - (1 - e^{-n(1 + x)h(a) + o(n)})^{\phi(x, n)} \;. \tag{6.7}$$

The critical case $a = a^*(x)$ given by (6.5). If $h(a) > h(a^*)$, then the probability in (7.7) converges to 0, whereas if $h(a) < h(a^*)$, then the probability converges to 1. In particular, for any $a > a^*$,

$$\sum_{n=1}^{\infty} P(D(\lfloor xn \rfloor, n) > (1+a)(1+x)n) < \infty ,$$

so that we can apply Borel-Cantelli to deduce that

$$\overline{\lim_{n \to \infty}} \; n^{-1} D(\lfloor xn \rfloor, n) \leq (1+a)(1+x) \quad \text{a.s.}$$

Hence, we have the claimed upper bound in (6.4).

Now we apply the subadditive ergodic theorem on p. 277 of Liggett (1985) to establish the existence of the limit. We first consider the limit of $n^{-1} D(kn, ln)$ for k and l integer. We let $X_{0,0} = 0$,

$$-X_{0,n} = D(kn, ln) - V(1, 1)$$

and $-X_{m,n}$ be $-X_{0,n-m}$ applied to the shifted service times $V'(i, j) = V(i + km, j + lm)$; e.g., for $k = l = 1$, $X_{n-1,n} = 0$ and $-X_{n-2,n} = V(n, n) + \max\{V(n-1, n), V(n, n-1)\}$. With this definition, X is subadditive, i.e., $X_{0,n} \leq X_{0,m} + X_{m,n}$ for $0 \leq m \leq n$. Moreover, X satisfies the other conditions of the subadditive ergodic theorem; in particular, $\{X_{(n-1)k,nk} : n \geq 1\}$ is a stationary process for each k, $\{X_{m,m+k} : k \geq 0\} \overset{d}{=} \{X_{m+1,m+k+1} : k \geq 0\}$ for each m, and $E(X_{01}^{+}) < \infty$. Hence, $n^{-1} D(kn, ln)$ has a deterministic limit $\eta(k, l)$ in the sense of (6.1) and (6.2) for all integers k and l. When n is a multiple of l,

$$\frac{1}{n} D\left[\frac{kn}{l}, n\right] = \frac{1}{l} \, \frac{l}{n} \, D\left[k\frac{n}{l}, l\frac{n}{l}\right] \to \frac{1}{l} \eta(k, l) \text{ as } n \to \infty \quad \text{a.s.}$$

More generally,

$$D\left[\frac{k}{l} \, l\lfloor n/l \rfloor, l\lfloor n/l \rfloor\right] \leq D(\lfloor kn/l \rfloor, n) \leq D\left[\frac{k}{l} \, l(\lfloor n/l \rfloor + 1), l([n/l] + 1)\right] ,$$

where

$$n^{-1}\, D\left[\frac{k}{l}\, l\lfloor n/l \rfloor,\, l\lfloor n/l \rfloor\right] \;=\; \frac{\lfloor n/l \rfloor}{n}\; \frac{1}{\lfloor n/l \rfloor}\, D(k\lfloor n/l \rfloor)\;,\, l\lfloor n/l \rfloor) \;\to\; \frac{1}{l}\eta(k,\, l)$$

and

$$n^{-1}\, D\left[\frac{k}{l}\; l(\lfloor n/l \rfloor + 1),\, l(\lfloor n/l \rfloor + 1)\right] \;=\; \frac{\lfloor n/l \rfloor + 1}{n}\; \frac{1}{\lfloor n/l \rfloor + 1}\; D(k(\lfloor n/l \rfloor + 1),\, l(\lfloor n/l \rfloor + 1))$$

$$\to\; \frac{1}{l}\;\eta(k,\, l) \quad \text{as } n \to \infty \quad \text{a.s.}$$

Hence, (6.1) holds for all positive rational x. A similar argument applies to (6.2).

To treat irrational x, we apply Theorem 5.1a to deduce that $\gamma(x + y) - \gamma(x)$ is decreasing in x through rational x for each rational y. Hence, γ is nondecreasing and concave restricted to the rationals. Since γ is nondecreasing overall, γ is nondecreasing and concave, and thus continuous, overall. Hence, the limits (6.1) and (6.2) extend to irrational x. (Note that $n^{-1} D(\lfloor nx \rfloor,\, n)$ is sandwiched between corresponding averages for rationals that converge. This implies the existence of convergent subsequences as $n \to \infty$. The continuity of γ on the rationals then implies that all limits of convergent subsequences converge to a common limit, implying convergence for the full sequence.)

To see that $\gamma(x + y) - \gamma(x) \ge y$, so that γ is strictly increasing, use the fact that

$$D(\lfloor (x + y) n \rfloor,\, n) \;\ge\; D(\lfloor xn \rfloor,\, n) \;+ \sum_{i\, =\, \lfloor xn \rfloor + 1}^{i\, =\, \lfloor (x+y) n \rfloor} V(i,\, n)\;\; .$$

By considering only paths through $(2k,\, 2k)$ for all k, $1 \le k \le \min\{xn,\, n\}$, we easily obtain the lower bound in (6.3). To see that $\gamma(x) \ge 1$ for all $x > 0$, note that, for all x, $D(\lfloor xn \rfloor,\, n) \;>\; D(1,\, n)$ for all n sufficiently large. Since $D(k,\, n) \overset{d}{=} D(n,\, k)$ for all n and k, we see that

$$\gamma(x) = \lim_{n \to \infty} n^{-1} D(\lfloor xn \rfloor, n) = \lim_{n \to \infty} \frac{\lfloor xn \rfloor}{n} \frac{1}{\lfloor xn \rfloor} D(\lfloor xn \rfloor, \lfloor xn \rfloor / x)$$
$$= \lim_{n \to \infty} \frac{\lfloor xn \rfloor}{n} \frac{1}{\lfloor xn \rfloor} D(\lfloor xn \rfloor / x, \lfloor xn \rfloor) = x \gamma(x^{-1}) \ . \quad \blacksquare$$

To illustrate the path-independence bound, in (6.4)–(6.6) suppose that the service times have an exponential distribution as in Theorem 6.1. Then

$$E e^{\theta [V(1,1) - 1]} = (1 - \theta)^{-1} e^{-\theta} \quad \text{and} \quad h(a) = a - \log(1 + a) \ . \tag{6.8}$$

From (6.5) and (6.8), we obtain a^* by solving

$$(1 + x)[a - \log(1 + a)] = (1 + x) \log (1 + x) - x \log x \ ,$$

which for the case $x = 1$ is

$$a^* - \log(1 + a^*) = \log 2 \ ,$$

yielding $a^* = 1.68$ and $\overline{\lim}_{n \to \infty} n^{-1} D(n, n) \le 5.36$. From Theorem 6.1 we see that this is indeed an upper bound, which seems to be not a terrible approximation. Evidently, there is enough dependence among the paths to reduce this estimate by a factor of 0.746.

Example 6.1. It is possible that the infimum in (6.5) is not attained as an equality. For example, suppose that $V(i, j)$ is Bernoulli, assuming the values 0 and 2 each with probability $1/2$. Then $h(a) = [(1 + a) \log (1 + a) + (1 - a) \log (1 - a)]/2$, $0 \le a < 1$, and $h(a) = \infty$ for $a \ge 1$. For $x = 1$, $\lim_{a \to 1-} h(a) = \log 2$, so that $a^*(1) = 1$, which yields $\gamma(1) \le 4$. Simulation suggests that $\gamma(1) = 3.63$; see Greenberg et al. (1990).

7. More Properties of the Limit Process

We established the strong approximation in Section 4 in order to deduce more about the departure times of customer k_n from queue n when $k_n \to \infty$ as $n \to \infty$. We now establish a limit for the components $\hat{D}_k(1)$ grow as $k \to \infty$ that enables us to conclude that the average of the first $\lfloor xn^a \rfloor$ interdeparture times from queue n after the first departure is of order $n^{(1 - a)/2}$ for any $x > 0$ and any a satisfying $0 < a < 1$. The limit is obtained by applying the subadditive ergodic theorem once more.

Theorem 7.1. Let $\hat{D} = f(\hat{B})$. Then there exists a constant α such that

$$\lim_{n \to \infty} n^{-1} \hat{D}_{\lfloor xn \rfloor}(n) = \alpha \sqrt{x} \quad \text{a.s.} \tag{7.1}$$

so that

$$n^{-1/2} \hat{D}_{\lfloor xn \rfloor}(1) \Rightarrow \alpha \sqrt{x} \text{ as } n \to \infty \tag{7.2}$$

for each $x > 0$.

Proof. As in the proof of Theorem 6.3, we apply the subadditive ergodic theorem on p. 277 of Liggett (1985). We first establish the limit for $n^{-1} \hat{D}_{jn}(kn)$ for j and k integer. We let $-X_{0,n} = \hat{D}_{jn}(n)$ and $-X_{m,n}$ be $-X_{0,n-m}$ applied to the shifted process $B_i'(t) = B_{i+km}(t + lm) - B_{i+km}(lm)$. With this definition, X is subadditive, i.e., $X_{0n} \le X_{0m} + X_{m,n}$ for $0 \le m \le n$, and X satisfies the other conditions of the subadditive ergodic theorem, except possibly for the bound. To establish the bound, we consider a related discrete problem. We consider the $kn \times ln$ integer lattice. We associate with the point (i,j) in this lattice the random variable

$$W(i, j) = \sup_{j-1 \le t < j} \{ | B_i(t) - B_i(j-1) | \} . \tag{7.3}$$

It is easy to see that

$$\hat{D}_{jn}(kn) \le \sup_{\pi \in \Pi(jn, kn)} \{ \sum_{(i,j) \in \pi} W(i,j) + W(jn, kn) \} . \tag{7.4}$$

For each path π, the random variables $W(i, j)$ for $(i, j) \in \pi$ are i.i.d. Moreover, for different paths, the partial sums are associated. Hence, we have a path-independence bound for the right side of (7.4) paralleling Theorem 6.2. Using the known tail behavior of $W(i, j)$ in (7.3), we have the required bound for the subadditive ergodic theorem. Hence, $n^{-1}\,\hat{D}_{jn}(kn)$ converges a.s. to a proper limit as $n \to \infty$. Let this proper limit be denoted by $\hat{\gamma}(x)$. As in the proof of Theorem 6.3, we use this result to deduce that (7.1) holds for each rational x. We then apply Corollary 2 to Theorem 5 to deduce that $\hat{\gamma}(x + y) - \hat{\gamma}(x)$ is decreasing in x through rational x for each rational y. Hence $\hat{\gamma}$ is nondecreasing and concave restricted to the rationals. Since $\hat{\gamma}$ is nondecreasing overall, $\hat{\gamma}$ is nondecreasing and concave overall. Hence, (7.1) extends to irrational x. Since $n^{-\frac{1}{2}}\,\hat{D}\lfloor xn\rfloor(1) \overset{d}{=} n^{-1}\hat{D}_{\lfloor xn\rfloor}(n)$ for all n, we obtain (7.2) directly from (7.1). From (7.2), we see that $\hat{\gamma}(x) = \alpha\sqrt{x}$ for some constant α. ■

Remark 7.1. Simulation by Greenberg et al. (1990) suggests that $\alpha = 2$ in (7.2). Moreover, simulation strongly suggests that the variance of $\hat{D}_n(1)$ *unnormalized* converges to 0 as $n \to \infty$.

We now apply Theorem 7.1 to obtain a limit for the average of the first $\lfloor n^a \rfloor$ departure times for $0 < a < 1$. The following shows that this average is asymptotically of order $n^{(1-a)/2}$.

Theorem 7.2. Under the assumptions of Theorem 4.1,

$$\frac{D(\lfloor xn^a \rfloor, n) - D(1, n)}{n^{(1+a)/2}} \Rightarrow \alpha\sqrt{x}$$

for α in Theorem 7.1 and $0 < a < 1$.

Proof. Note that

$$\frac{D(\lfloor xn^a \rfloor, n) - D(1, n)}{n^{(1+a)/2}} = \left[\frac{\sqrt{n}\,\overrightarrow{D}_{\lfloor xn^a \rfloor}(1)}{n^{(1+a)/2}}\right] + \left[\frac{D(\lfloor xn^a \rfloor, n) - n - \sqrt{n}\,\overrightarrow{D}_{\lfloor xn^a \rfloor}(1)}{n^{(1+a)/2}}\right] - \left[\frac{D(1, n) - n}{n^{(1+a)/2}}\right]. \tag{7.5}$$

By (7.2) in Theorem 7.1, the first term on the right in (7.5) converges in probability to $\alpha\sqrt{x}$. By Theorem 4.1, the second term on the right converges in probability to 0. By (1.2), the third term on the right converges to 0. ■

8. Common Service Times

We now consider the case in which $V(k, n) = V(k, 1)$ w.p.1 for all k and n, with $\{V(k, 1) : k \geq 1\}$ being i.i.d., just as in Section 2 of Pinedo and Wolff (1982). As before, let $EV(k, 1) = 1$. From Proposition 2.1, we easily obtain the following result, from which we can establish the limiting behavior as $k \to \infty$ and/or $n \to \infty$.

Theorem 8.1. For each k and n,

$$D(k, n) = (n - 1) M_k + S_k , \tag{8.1}$$

where

$$M_k = \max\{V(i, 1) : 1 \leq i \leq k\} \quad \text{and} \quad S_k = V(1, 1) + \ldots + V(k, 1) . \tag{8.2}$$

The limiting behavior of $D(k, n)$ as $k \to \infty$ depends on the limiting behavior of M_k and S_k. For M_k, this is the classical extreme value theory, e.g., see Leadbetter, Lindgren and Rootzén (1983). From Theorem 8.1 we easily obtain the following.

Corollary 8.1. If Var $V(1, 1) = \sigma^2$, $0 < \sigma^2 < \infty$, then

$$[D(k, n) - k]/\sqrt{k} \Rightarrow N(0, \sigma^2) \text{ as } k \to \infty .$$

Proof. Apply (8.1). By the SLLN, $k^{-1} \sum_{i=1}^{k} V(i, 1)^2 \to \beta < \infty$ a.s. as $k \to \infty$, so that $k^{-1} V(k, 1)^2 \to 0$ a.s. as $k \to \infty$, which implies that $k^{-1/2} M_k \to 0$ a.s. as $k \to \infty$. The CLT implies that $k^{-1/2} (S_k - k) \Rightarrow N(0, \sigma^2)$ as $k \to \infty$. ■

A more interesting case arises if k and n both go to infinity.

Corollary 8.2. Suppose that Var $V(1, 1) = \sigma^2$, $0 < \sigma^2 < \infty$ and there exist sequences of positive constants a_k and b_k such that $[M_k - a_k]/b_k \Rightarrow M^*$ as $k \to \infty$.

(a) If $n_k b_k / \sqrt{k} \to 1$ as $k \to \infty$, then

$$[D(k, n_k) - k - a_k n_k]/\sqrt{k} \Rightarrow M^* + N(0, \sigma^2) \quad \text{as } k \to \infty ,$$

where M^* and $N(0, \sigma^2)$ are independent.

(b) If $n_k b_k / \sqrt{k} \to 0$ as $k \to \infty$, then

$$[D(k, n_k) - k - a_k n_k]/\sqrt{k} \Rightarrow N(0, \sigma^2) \quad \text{as } k \to \infty .$$

(c) If $n_k b_k / \sqrt{k} \to \infty$ as $k \to \infty$, then

$$[D(k, n_k) - k - a_k n_k]/n_k b_k \Rightarrow M^* \quad \text{as } k \to \infty .$$

Proof. As in the proof of Corollary 8.1, first note that $b_k = o(\sqrt{k})$ since $E[V(1, 1)^2] < \infty$. Then note that

$$\left[\frac{M_k - a_k}{b_k} , \frac{S_k - k}{\sqrt{k}} \right] \Rightarrow (M^*, N(0, \sigma^2)) \quad \text{as } k \to \infty , \tag{8.3}$$

where M^* and $N(0, \sigma^2)$ are independent; see §4.5 of Resnick (1986). To directly establish (8.3), note that the maximum M_k is negligible in the normalized partial sum since $b_k = o(\sqrt{k})$. Thus, we prove (8.3) by showing that $S_k - M_k$ is asymptotically independent of M_k. Let F be the cdf of $V(1, 1)$. Given that $M_k = m_k$, $S_k - M_k$ is distributed as the sum of $(k - 1)$ i.i.d. variables with cdf $F(x)/F(m_k)$. As k increases this conditional cdf approaches the original unconditional cdf $F(x)$. Thus, we can apply the Lindeberg-Feller central limit theorem for triangular arrays to conclude that the conditional distribution of $[S_k - M_k - k]/\sqrt{k}$ given M_k converges in distribution to $N(0, \sigma^2)$, independently of the value of M_k, and we obtain the desired asymptotic independence.

From (8.1),

$$\frac{D(k, n_k) - k - n_k a_k}{\sqrt{k}} = \frac{(n_k - 1) M_k - n_k a_k + S_k - k}{\sqrt{k}} ,$$

so that we obtain (a), (b) and (c) from (8.3) and standard arguments, i.e., Theorems 4.1 and 5.5 of

Billingsley (1968). ■

Corollary 8.3. If $a_k M_k \Rightarrow m$ with $a_k \to 0$ as $k \to \infty$ then

$$a_{\lfloor xn \rfloor} \frac{D(\lfloor xn \rfloor, n)}{n} \Rightarrow m$$

for any $x > 0$.

Proof. From (8.1), note that

$$a_{\lfloor xn \rfloor} \frac{D(\lfloor xn \rfloor, n)}{n} = a_{\lfloor xn \rfloor} \frac{(n-1)}{n} M_{\lfloor xn \rfloor} + a_{\lfloor xn \rfloor} \frac{S_n}{n} ;$$

then apply the assumed limit for M_n and the law of large numbers for S_n. ■

We conclude this section by illustrating Corollaries 8.2 and 8.3 with the exponential case.

Corollary 8.4. If $V(i, 1)$ is exponential for all i, then

$$[D(k, n_k) - k]\sqrt{k} \Rightarrow N(m, 1) \text{ as } k \to \infty$$

provided that $k^{-1/2} n_k \log k \to m$ as $k \to \infty$ and

$$D(\lfloor xn \rfloor, n)/n \log (xn) \Rightarrow 1 \text{ as } n \to \infty .$$

Proof. By Example 1.7.2, p. 20, of Leadbetter et al. (1983), $M_k/\log k \Rightarrow 1$ as $k \to \infty$. ■

Note that $D(n, n)$ is of order $n \log n$ here as opposed to of order n in § 6.

9. Concluding Remarks

There are several stones left unturned. First, it would be nice to identify the hydrodynamic limit $\gamma(x)$ in (6.1) for non-exponential service-time distributions and determine how this limit depends on the service-time distribution. We conjecture that the limit depends on the service-time distribution beyond its first two moments. (This is confirmed by simulation.) It would also be nice to establish a refined distributional limit, i.e., a weak convergence limit for $n^{-\beta}(D(\lfloor xn \rfloor, n) - \gamma(x)n)$ as $n \to \infty$ for some $\beta > 0$.

Second, it would be useful to know more about the limit process $f(\hat{B})$ in § 3. In Remark 3.3 we noted that $\hat{\Delta}_1 = \hat{D}_2 - \hat{D}_1$ is a reflecting BM, so that $E[\hat{\Delta}_1(1)] = 2\sigma/\sqrt{\pi}.$ Moreover, by Theorem 5.2, $\hat{\Delta}_k(1)$ is stochastically decreasing as k increases. However, it would be nice to know the joint distribution, or at least the means, of $(\hat{\Delta}_1(1), \ldots, \hat{\Delta}_k(1))$. Moreover, it would be nice to know the constant α in (7.1) and (7.2). (Simulation suggests that $\alpha = 2$.)

Finally, an old open problem involves the limiting behavior of the stationary departure process from n queues as $n \rightarrow \infty$. Here we assume that the service times $V(i, j)$ are i.i.d. for $i \geq 1$ and $j \geq 2$ while the service times $V(i, 1)$ are i.i.d. (or just stationary and ergodic) with $E\, V(1, 1) > E\, V(1, 2)$, so that the departure process from the first queue corresponds to an external arrival process with mean interarrival time greater than the subsequent mean service times. For the case in which $V(1, 2)$ is exponentially distributed, but $V(1, 1)$ is not, it is widely believed that the stationary departure process from queue n is asymptotically Poisson as $n \rightarrow \infty$. A corresponding result for infinite-server queues was established by Vere-Jones (1968).

Acknowledgment. We thank Raj Srinivasan for helping motivate this paper and Bill Massey, Marty Reiman, Scott Shenker and Alan Weiss for helping to improve it.

References

Andjel, E. D. (1982). Invariant measures for the zero-range process. *Ann. Probab.* 10, 525-547.

Andjel, E. D. and C. Kipnis (1984). Derivation of the hydrodynamical equation for the zero-range interaction process. *Ann. Probab.* 12, 325-334.

Barlow, R. E. and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing,* Holt, Reinhart and Winston, New York.

Benassi, A. and J. P. Fouque (1987). Hydrodynamical limit for the asymmetric simple exclusion process. *Ann. Probab.* 15, 546-560.

Billingsley, P. (1968). *Convergence of Probability Measures,* Wiley, New York.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* 23, 494-507.

Cramér, H. (1938). On a new limit theorem in the theory of probability. In *Colloquium on the Theory of Probability,* Hermann, Paris.

Csörgó′, M., P. Deheuvels and L. Horvath (1987). An approximation of stopped sums with application in queueing theory. *Adv. Appl. Prob.* 19, 674-690.

Csörgó′, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics,* Academic Press, New York.

Disney, R. L. and P. C. Kiessler (1987). *Traffic Processes in Queueing Networks,* The John Hopkins University Press, Baltimore.

Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes: Characterization and Convergence,* Wiley, New York.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications,* Vol. I, 3rd edition, Wiley, New York.

Glynn, P. W. and W. Whitt (1991). A new view of the heavy-traffic limit theorem for infinite-server queues. *Adv. Appl. Prob.* 23, to appear.

Greenberg, A., O. Schlunk and W. Whitt (1990). Departures from Many Queues in Series: A Simulation Study. AT&T Bell Laboratories, Murray Hill, NJ.

Hanqin, Z., H. Guanghui and W. Rongxin. (1990). Strong approximations for multichannel queues in heavy traffic. *J. Appl. Prob.* 27, 658-670.

Harrison, J. M. (1973). The heavy traffic approximation for single server queues in series. *J. Appl. Prob.* 10, 613-629.

Harrison, J. M. (1978). The diffusion approximation for tandem queues in heavy traffic. *Adv. Appl. Prob.* 10, 886-905.

Harrison, J. M. (1985). *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.

Harrison, J. M. and M. I. Reiman (1981a). Reflected Brownian motion on an orthant. *Ann. Probab.* 9, 302-308.

Harrison, J. M. and M. I. Reiman (1981b). On the distribution of multidimensional reflected Brownian motion. *SIAM J. Appl. Math* 41, 345-361.

Harrison, J. M. and R. J. Williams (1987a). Multidimensional reflected Brownian motions having exponential stationary distributions. *Ann. Probab.* 15, 115-137.

Harrison, J. M. and R. J. Williams (1987b). Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* 22, 77-115.

Horvath, L. (1990). Strong approximations of open queueing networks. Department of

Mathematics, University of Utah.

Iglehart, D. L. and W. Whitt (1970). Multiple channel queues in heavy traffic, II: sequences, networks and batches. *Adv. Appl. Prob.* 2, 355-369.

Kamae, T., U. Krengel and G. L. O'Brien (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* 5, 899-912.

Kelly, F. P. (1982). The throughput of a series of buffers. *Adv. Appl. Prob.* 14, 633-653.

Kelly, F. P. (1984). An asymptotic analysis of blocking. *Modeling and Performance Evaluation Methodology* eds. F. Baccelli and G. Fayolle, Springer-Verlag, New York, 3-20.

Kipnis, C. (1986). Central limit theorems for infinite series of queues and applications to simple exclusion. *Ann. Probab.* 14, 397-408.

Komlós, J., P. Major and G. Tusnády (1975). An approximation of partial sums of independent R.V.'s and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32, 111-131.

Komlós, J., P. Major and G. Tusnády (1976). An approximation of partial sums of independent R.V.'s and the sample DF. II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 34, 33-58.

Leadbetter, M. R., G. Lindgren and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes,* Springer-Verlag, New York.

Liggett, T. M. (1985). *Interacting Particle Systems,* Springer-Verlag, New York.

Muth, E. G. (1979). The reversibility property of production lines. *Management Sci.* 25, 152-158.

Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability*, Holden Day, San Francisco.

Pinedo, M. and R. W. Wolff (1982) A comparison between tandem queues with dependent and independent service times. *Oper. Res.* 30, 464-479.

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

Resnick, S. I. (1986). Point processes, regular variation and weak convergence. *Adv. Appl. Prob.* 18, 66-138.

Rost, H. (1981). Non-equilibrium behavior of a many particle process: density profile and local equilibria. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 58, 41-53.

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* 1, 261-290.

Skorohod, A. V. (1961). Stochastic differential equations for a bounded region. *Theor. Probability Appl.* 6, 264-274.

Srinivasan, R. (1989). Queues in series via interacting particle systems. Department of Mathematics, University of Saskatchewan.

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*, Wiley, Chichester.

Suresh, S. and W. Whitt (1990). The heavy-traffic bottleneck phenomenon for open queueing networks. *Operations Res. Letters,* to appear.

Tembe, S. V. and R. W. Wolff (1974). The optimal order of service in tandem queues. *Operations Res.* 24, 824-832.

Vanderbei, R. J. and A. Weiss (1988). Large deviations and their applications to computer and communications systems, Part I. AT&T Bell Laboratories, Murray Hill, NJ 07974.

Varadhan, S. R. S. (1984). *Large Deviations and Applications,* SIAM, Philadelphia.

Vere-Jones, D. (1968). Some applications of probability generating functionals to the study of input-output streams. *J. Roy. Statist. Soc.,* Ser. B, 30, 321-333.

Whitt, W. (1974). Preservation of rates of convergence under mappings. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 29, 39-44.

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

Wolff, R. W. (1982) Tandem queues with dependent service times in light traffic. *Oper. Res.,* 30, 619-635.

We first show, under appropriate conditions, that the departure process from the n^{th} queue obeys an invariance principle or functional central limit theorem (FCLT). The FCLT supports approximating the beginning of the departure process, after appropriate normalization, by an infinite-dimensional reflected Brownian motion (RBM) on the infinite-dimensional orthant $[0, \infty)^{\infty}$. This infinite-dimensional RBM is the natural extension of finite-dimensional RBMs considered by Harrison (1978), Harrison and Reiman (1981a,b), Reiman (1984) and Harrison and Williams (1987a,b).

We are primarily interested in the special case in which the service times of all the customers at all the queues are i.i.d. Then the invariance principle implies that the approximation depends on the service-time distribution only through its mean and variance. Moreover, the mean and variance play a relatively trivial role. In particular, the mean service time only determines the deterministic rate customers flow through the queues; without loss of generality, we can let the mean service time be one. The service-time variance only appears (via its square root) as a constant multiplicative factor in front of the multivariate RBM associated with service-time variance 1. Hence, just as with the familiar one-dimensional Brownian motion (BM) approximation for partial sums of i.i.d. real-valued random variables, there is essentially only one fundamental limit process for this system for all service-time distributions. We call this limit process the *departure RBM.*

The model we consider has no external arrival process, but the same model can be interpreted as starting out empty with an external arrival process. Simply interpret the departure process from the first queue as the external arrival process. Of course, the assumption that the service times be all i.i.d. implies that the interarrival-time distribution must then be exactly the same as each service-time distribution. However, this is not required. The limiting behavior remains unchanged if the service-time distributions at an initial finite set of queues are different. (The stated results cover this generalization.)