Adv. Appl. Prob. **21**, 629–660 (1989)
Printed in N. Ireland

THE FORK-JOIN QUEUE AND RELATED SYSTEMS WITH SYNCHRONIZATION CONSTRAINTS: STOCHASTIC ORDERING AND COMPUTABLE BOUNDS

FRANÇOIS BACCELLI,* *INRIA*
ARMAND M. MAKOWSKI,** *University of Maryland*
ADAM SHWARTZ,*** *Technion–Israel Institute of Technology*

Abstract

A simple queueing system, known as the fork-join queue, is considered with basic performance measure defined as the delay between the fork and join dates. Simple lower and upper bounds are derived for some of the statistics of this quantity. They are obtained, in both transient and steady-state regimes, by stochastically comparing the original system to other queueing systems with a structure simpler than the original system, yet with identical stability characteristics. In steady-state, under renewal assumptions, the computation reduces to standard $GI/GI/1$ calculations and the bounds constitute a first sizing-up of system performance. These bounds can also be used to show that for homogeneous fork-join queue system under assumptions, the moments of the system response time grow logarithmically in the number of parallel processors provided the service time distribution has rational Laplace–Stieltjes transform. The bounding arguments combine ideas from the theory of stochastic ordering with the notion of associated random variables, and are of independent interest to study various other queueing systems with synchronization constraints. The paper is an abridged version of a more complete report on the matter [6].

STOCHASTIC BOUNDS; INCREASING CONVEX ORDERING; ASSOCIATED RANDOM VARIABLES

1. Introduction

A K-dimensional fork-join (FJ) queue is a queueing system operated by K parallel servers with *synchronized* arrival and departure streams. Each server is attended by a buffer of infinite capacity and individually operates according to the FIFO discipline. Customers arrive into the system in batches of size not larger than K and are processed according to the following discipline.

Received 23 January 1987; revision received 26 May 1988.

* Postal address: INRIA—Centre de Sophia Antipolis, Avenue E. Hughes, 06565, Valbonne Cedex, France.

The work of this author was supported partially through a grant from AT & T Bell Laboratories and partially through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland, College Park, MD 20742, USA.

** Postal address: Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742, USA.

The work of this author was supported partially through ONR Grant N00014-84-K-0614, partially through NSF Grant ECS-83-51836 and partially through a grant from AT & T Bell Laboratories.

*** Postal address: Electrical Engineering Department, Technion–Israel Institute of Technology, Haifa 32000, Israel.

The work of this author was supported through a grant from AT & T Bell Laboratories.

Upon arrival, a batch of size $S \leqq K$, bringing customers to S of the K servers, is immediately split so that each one of its S customers is allocated to exactly one server (the so-called *fork primitive*).

As soon as all the S customers constituting a batch have been serviced, the batch is immediately recomposed (the so-called *join primitive*) and leaves the system at once. This second synchronization constraint is achieved by parking already serviced customers in an auxiliary buffer of infinite capacity, where they await being reunited to serviced customers of the same batch whose service has not been completed yet.

Such queueing models arise in many application areas, including flexible manufacturing and parallel processing, with a wide variety of interpretations. In the context of production systems, a batch customer can be interpreted as a customer's order with several components, each component or suborder being attended by a separate production device. An example very similar to this one is obtained by considering the production of multipart items. In computer systems with parallel architecture, a batch customer can be viewed as a program composed of several subroutines, each one to be executed on a different processor (e.g. the cobegin and coend structures in concurrent languages like Concurrent Pascal, CSP, etc).

For this type of application, the determination of batch response time (defined as the delay between the fork and the join dates) is of crucial importance in quantifying system performance. In two dimensions and for constant batch sizes, i.e., $K = S = 2$, the stationary joint distribution of the number of customers in the two queues was determined by Flatto and Hahn [11] under Markovian assumptions. A somewhat more general problem with Poisson arrivals and general service times was also analyzed by Baccelli in [1]. However, in more dimensions (i.e., $K > 2$) and/or for more general interarrival distributions, the problem still remains open. The difficulty stems from the *statistical dependence* between the individual response times of different queues, which is due to the common arrival process.

In this paper, simple lower and upper bounds are derived for various statistics of this response time, including its moments. The bounds are obtained by a direct stochastic comparison of the queueing system to several other systems with K parallel servers, for which the statistics of interest can be computed *explicitly*. Moreover, the stability conditions for these 'bounding' systems are identical to those for the original system. Transient as well as steady-state bounds are obtained.

In order to carry out this program, it is convenient (and natural) to consider a somewhat larger class of queueing systems with K parallel servers operating under synchronization constraints. This enlarged class is obtained by removing the synchronization constraint on the arrival streams and by allowing more general loading patterns. This class of models includes the simplest FJ queue studied earlier by the authors [2], [5], and is described in detail in Section 2 where the performance measures of interest are defined.

Two basic bounding methodologies are developed and used throughout this work. The first approach relies on the *convex increasing* order for probability distributions [18], [21], [22] and is found most useful for establishing ordering results on the waiting times which are generated through Lindley's equation. In Section 3 a basic bounding result is obtained for such a recursion by comparing the waiting times when Lindley's recursion is driven by different arrival, service and/or loading processes. The ideas are then applied in Section 4 to the queueing system of interest, to obtain several simpler queueing systems whose performance bounds that of the original system. Preliminary versions of the results discussed here can be found in the conference papers [2], [5] for the FJ queue under renewal type assumptions.

The second methodology makes use of the notion of *associated* random variables [4], [9] and is especially useful in establishing comparison results for the *maximum* of positively correlated random variables. The necessary steps are developed in Section 5, where increasing the number of independent components (by taking independent versions of some of the arrival and loading processes) is shown to *stochastically* increase the system response time. The resulting bounds are given in terms of the waiting times and other performance measures of (partially) decoupled queueing systems. This technique was already used by Nelson and Tantawi [16] who obtained only first-moment information in the special case of Poisson arrivals and exponential servers.

In Section 6, these bounds are specialized to the FJ queue with synchronized arrivals under a set of renewal assumptions. The discussion emphasizes the *computability* of the steady-state bounds, in the sense that their evaluation reduces to analyzing the statistics of K *independent* $GI/GI/1$ queueing systems, in contrast with the initial problem which is also K-dimensional but with *strongly coupled* components. The calculations are carried out for the case of *exponentially* distributed service times, and simple explicit expressions are given for *homogeneous* FJ queues, i.e., when the loading patterns and service requirements are identical for all processors. The reader is referred to the conference papers [2], [5] for simulation results on the quality of some of the bounds obtained here.

A further use of these bounds is discussed in Section 7 where homogeneous FJ queues are considered. Asymptotics on the system response time statistics are obtained as the number K of servers grows large. It is shown under standard renewal assumptions that the moments of the system response time grow *logarithmically* in the number of parallel servers, provided the Laplace–Stieltjes transform of the service time distribution is *rational*. This result generalizes a similar result obtained by Nelson and Tantawi [16] for the special case of Poisson arrivals and exponential servers. The asymptotics are developed by making use of a result of Lai and Robbins [14] on the maximum of identically distributed random variables. The derivation provides for an explicit estimate of the tightness of the asymptotic bounds.

2. The model

The queueing model of interest in this paper is now presented, together with the notation and some of the basic assumptions enforced throughout. Although the proposed queueing model is more general than the synchronized FJ queue model which motivated the work reported here [2], [5], its usefulness will soon become apparent to the reader in the forthcoming sections.

2.1. *The basic random variables.* Emphasis is put on sample path representations for the quantities of interest, and as further developments will demonstrate, this approach is quite fruitful in establishing bounds. To that end, an underlying probability triple $(\Omega, \mathcal{F}, P)$ is postulated on which all the random variables (RV) mentioned in this paper are defined. A positive integer K is given and held fixed hereafter. As a convention, the kth component RV of any $\mathbb{R}^K$-valued RV is denoted by the same symbol as this RV but superscripted by k; a similar convention is adopted to denote the components of any vector in $\mathbb{R}^K$. This probability triple $(\Omega, \mathcal{F}, P)$ is assumed to simultaneously carry the sequences $\{\tau_n\}_1^\infty$, $\{\sigma_n\}_0^\infty$ and $\{u_n\}_0^\infty$ of $\mathbb{R}_+^K$-valued RV's, together with an $\mathbb{R}_+^K$-valued RV W.

The queueing system generated by the constituting sequence

$$\langle W, \sigma_n, u_n, \tau_{n+1}, n = 0, 1, \cdots \rangle \tag{2.1}$$

is defined as a queueing system composed of K *parallel* servers with the following features: each one of these servers has its own buffer of *infinite* capacity and operates according to the FIFO discipline. The RVs $\{\tau_n^k\}_1^\infty$ model the interarrival times as experienced by the kth server, so that arrivals to its queue are taking place along the time sequence $\{A_n^k\}_0^\infty$ defined by

$$A_n^k = \sum_{m=0}^{n-1} \tau_{m+1}^k, \qquad n = 1, 2, \cdots \tag{2.2}$$

with $A_0^k = 0$. The customers arriving at the kth queue at such times are called type k customers hereafter. The nth customer of type k brings an amount of processing time σ_n^k to be executed by the kth server, whereas the RV u_n^k represents a *loading* factor in that the actual work to be processed by the kth server due to the nth arrival of type k is $u_n^k \cdot \sigma_n^k$. A customer is assumed to arrive at time $t = 0$, at which time an initial load is already awaiting service in the various buffer areas, and the RV W^k thus represents the amount of time required by the kth server to clear this initial load from its buffer. In the discussion, it will be convenient to aggregate the nth customers of all types into a single entity referred to as the nth *composite* customer.

Two particular cases are especially relevant for what follows.

(C.1) The RVs $\{u_n\}_0^\infty$ take their vaues in $\{0, 1\}^K$, in which case, each RV u_n determines a (random) subset I_n of $\{1, 2, \cdots, K\}$ given by

$$I_n := \{k, 1 \leqq k \leqq K : u_n^k = 1\}, \qquad n = 0, 1, \cdots \tag{2.3}$$

and the system is one fed by K arrival streams where the nth composite customer brings work only to the servers with index in I_n, possibly not all at the same time.

(C2) The RVs $\{u_n\}_0^\infty$ are *synchronized* in the sense that

$$u_n^1 = u_n^2 = \cdots = u_n^K = v_n, \qquad n = 0, 1, \cdots. \tag{2.4}$$

In this system, the amounts of actual work brought by the nth customers to the K parallel queues are positively correlated through a common scaling factor v_n. Such a correlation seems quite natural in all the practical interpretations discussed earlier (e.g., the sizes of parallel subprograms in a program).

2.2. *The performance measures.* In order to define reasonable performance measures, consider the sequence of $\mathbb{R}_+^K$-valued RVs $\{W_n\}_0^\infty$ generated componentwise by the Lindley recursions

$$W_{n+1}^k = [W_n^k + u_n^k \,.\, \sigma_n^k - \tau_{n+1}^k]^+, \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots \tag{2.5}$$

with $W_0 = W$, and the standard notation $x^+ = \max(x, 0)$ is used for all x in $\mathbb{R}$. The RV W_n^k represents the *waiting time* of the nth customer of type k, whereas the quantities R_n^k and S_n^k, akin to response times in the queueing system attended by the kth server, are defined by

$$R_n^k := u_n^k \,.\, (W_n^k + \sigma_n^k), \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots \tag{2.6}$$

and

$$S_n^k := W_n^k + u_n^k \,.\, \sigma_n^k, \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots \tag{2.7}$$

respectively. For reasons that will become clear later on, two quantities can be naturally defined as the *system response time* for the nth composite customer. These are denoted T_n^1 and T_n^2, and are given by

$$T_n^1 := \max_{1 \leqq k \leqq K} R_n^k = \max_{1 \leqq k \leqq K} u_n^k \,.\, (W_n^k + \sigma_n^k), \qquad n = 0, 1, \cdots \tag{2.8}$$

and

$$T_n^2 := \max_{1 \leqq k \leqq K} S_n^k = \max_{1 \leqq k \leqq K} W_n^k + u_n^k \,.\, \sigma_n^k, \qquad n = 0, 1, \cdots \tag{2.9}$$

respectively.

The definitions (2.8)–(2.9) attempt to provide meaningful measures of system performance for as large a class of models as possible. To get a better feel for the meaning of these definitions, consider the situation where the arrivals are *synchronized* in the sense that

$$\tau_{n+1}^1 = \tau_{n+1}^2 = \cdots = \tau_{n+1}^K, \qquad n = 0, 1, \cdots. \tag{2.10}$$

This corresponds to the situation where the nth customers of all types arrive into the system at the same time, in which case the arrival stream *common* to all K queues is still denoted by $\{\tau_{n+1}\}_0^\infty$. In the particular case (C.1), T_n^1 now reads

$$T_n^1 = \max_{k \in I_n} R_n^k, \qquad n = 0, 1, \cdots \tag{2.11}$$

with I_n given by (2.2), and is exactly the time that elapses between the fork and join dates of the nth composite customer. On the other hand, when the loading sequence has the form (2.4) as in (C.2), it is more natural to define the system response time of the nth composite customer by (2.9).

The system response times defined through (2.8) and (2.9) coincide when the loading sequence has the simplified form

$$u_1^k = \cdots = u_n^K = 1, \qquad n = 0, 1, \cdots \tag{2.12}$$

as in the simplest FJ queue system studied earlier by the authors [2], [5]. The difficulty in analyzing these queueing systems with synchronization constraints is already apparent in the simple situation defined by (2.10) and (2.12), say under standard renewal assumptions [2], [5]. Indeed, in such a case, the single-server queueing system associated with each server embedded in the FJ queue operates like a standard $GI/GI/1$ queueing system. However, these K parallel $GI/GI/1$ systems are *not* independent in general since they have *identical* inputs owing to (2.10). It is precisely the presence of this correlation between the input streams to the various queues that makes the computation of the statistics of the RVs $\{T_n^i\}_0^\infty$, $i = 1, 2$, hard. In view of these difficulties, it seems relevant to seek ways of generating *bounds* and *approximations* to the various statistics of $\{T_n^i\}_0^\infty$, $i = 1, 2$. The results on stochastic ordering presented in this paper readily lead to the derivation of *simple computable* bounds to these statistics in the context of the FJ queue as well as for the more general queueing systems described earlier in this section. The comparison results discussed in this paper are derived through direct bounding arguments that explicitly exploit the sample path nature of the recursions (2.5). The basic idea consists in directly constructing from the RVs that define the original system, a new queueing system of the same type but with different constituting sequence (2.1).

2.3. *Notation.* Although more general situations will be covered during the discussion, the results will often be specialized to various models of interest. In particular, it will be convenient to consider the system under the assumption of *independence* (I) (respectively (Ib)) if the conditions below are satisfied, namely

(I) The $\mathbb{R}_+^K$-valued RVs $\{W, \sigma_n, u_n, \tau_{n+1}, n = 0, 1, \cdots\}$ are *mutually independent*; and

(Ib) The $\mathbb{R}_+$-valued RVs $\{W^k, \sigma_n^k, u_n^k, \tau_{n+1}^k, \ 1 \leqq k \leqq K, \ n = 0, 1, \cdots\}$ are *mutually independent.*

This section closes with a word on the notation used throughout the paper. The Laplace–Stieltjes transform of the probability distribution function $F(\cdot)$ is denoted by $F^*(\cdot)$. The conditional expectation of any $\mathbb{R}$-valued RV X with respect to any sub-σ-field $\mathscr{D}$ of $\mathscr{F}$ is often denoted by $X^{\mathscr{D}}$, with

$$X^{\mathscr{D}} := E[X \mid \mathscr{D}], \tag{2.13}$$

whenever meaningful. The notation $\sigma_n^{k,\mathcal{D}}$, $W_n^{k,\mathcal{D}}$, $R_n^{k,\mathcal{D}}$, $S_n^{k,\mathcal{D}}$ and $T_n^{i,\mathcal{D}}$ is thus used with this meaning for all $n = 0, 1, \cdots, 1 \leqq k \leqq K$, and $i = 1, 2$.

3. Bounds in the convex increasing ordering

The first bounding methodology is based on a simple result that provides for a *direct stochastic* comparison between different queueing systems of the type described earlier. This elementary result uses in an essential way the structure of the Lindley's recursions (2.5), and allows for a *unified* treatment of many of the bounds presented here. The discussion that follows finds its origin in a folk theorem of queueing theory stating that *determinism minimizes waiting* (or *response*) *times* in many queueing systems. For $G/G/1$ systems, such results have been established under a variety of assumptions by a number of authors, including Hajek [12], Humblet [13], Rogozin [17] and Whitt [22] to name a few.

3.1. *A basic bounding methodology.* To set the stage for the discussion, consider a sequence $\{O_n\}_1^\infty$ of $\mathbb{R}$-valued RVs and an $\mathbb{R}_+$-valued RV V defined on the probability triple $(\Omega, \mathcal{F}, P)$. These RVs are assumed to satisfy the *finite* mean condition

$$E[V] < \infty \quad \text{and} \quad E[O_n] < \infty, \qquad n = 1, 2, \cdots. \tag{3.1}$$

For any σ-field $\mathcal{D}$ of events contained in $\mathcal{F}$, the $\mathbb{R}$-valued RVs $\{O_n^{\mathcal{D}}\}_1^\infty$ are defined as (any one version of) the conditional expectations

$$O_n^{\mathcal{D}} := E[O_n \mid \mathcal{D}], \qquad n = 1, 2, \cdots \tag{3.2}$$

in agreement with the convention (2.13). The main result of this section is then the following.

Theorem 3.1. *Let* $\{V_n\}_0^\infty$ *and* $\{V_n(\mathcal{D})\}_0^\infty$ *be the sequences of* $\mathbb{R}_+$*-valued* RVs *defined through the recursions*

$$V_{n+1} = [V_n + O_{n+1}]^+, \qquad n = 0, 1, \cdots \tag{3.3}$$

and

$$V_{n+1}(\mathcal{D}) = [V_n(\mathcal{D}) + O_{n+1}^{\mathcal{D}}]^+, \qquad n = 0, 1, \cdots \tag{3.4}$$

with $V_0 = V_0(\mathcal{D}) = V$. *Under the assumptions made, whenever the* RV V *is* $\mathcal{D}$*-measurable, the inequalities*

$$V_n(\mathcal{D}) \leqq E[V_n \mid \mathcal{D}] \quad \text{a.s.}, \qquad n = 0, 1, \cdots \tag{3.5}$$

hold.

Proof. The proof proceeds by induction. Since the RV V is $\mathcal{D}$-measurable, (3.5) trivially holds for $n = 0$ since $V_0 = V_0(\mathcal{D}) = V$.

Take as induction hypothesis that (3.5) holds true for *some* $n = m \geqq 0$. For such m, Jensen's inequality gives

$$E[V_{m+1} \mid \mathscr{D}] \geqq [E[V_m \mid \mathscr{D}] + E[O_{m+1} \mid \mathscr{D}]]^+ \tag{3.6}$$

since the function $\mathbb{R} \to \mathbb{R} : x \to x^+$ is convex monotone non-decreasing. Substitution of (3.2) into (3.6) and use of the induction hypothesis lead to almost sure relations

$$E[V_{m+1} \mid \mathscr{D}] \geqq [V_m(\mathscr{D}) + O^{\mathscr{D}}_{m+1}]^+ = V_{m+1}(\mathscr{D}), \tag{3.7}$$

where the equality follows from (3.4). This shows that (3.5) holds for $n = m + 1$ and since it holds for $n = 0$, it holds by induction for all $n = 0, 1, \cdots$.

The following corollary is an easy consequence of Theorem 3.1, and its proof is available in [6].

Corollary 3.2. *If* $\mathscr{D}_1$ *and* $\mathscr{D}_2$ *are two σ-fields of events such that*

$$\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \mathscr{F}, \tag{3.8}$$

then, with the notation of Theorem 3.1, *almost sure inequalities*

$$V_n(\mathscr{D}_1) \leqq E[V_n(\mathscr{D}_2) \mid \mathscr{D}_1] \leqq E[V_n \mid \mathscr{D}_1] \qquad n = 0, 1, \cdots \tag{3.9}$$

hold provided $V_0 = V_0(\mathscr{D}_1) = V_0(\mathscr{D}_2) = V$ *and* V *is* $\mathscr{D}_1$*-measurable.*

3.2. *Transient analysis.* The basic results of the previous section are now applied to the K-server queueing system generated by the constituting sequence

$$\langle W, \sigma_n, u_n, \tau_{n+1}, n = 0, 1, \cdots \rangle \tag{3.10}$$

as described in Section 2. The discussion is given under the assumptions (A.1)–(A.2), where

(A.1) For $1 \leqq k \leqq K$ and $n = 0, 1, \cdots$, the RVs σ^k_n, u^k_n and τ^k_{n+1} all have *finite* means.

(A.2) There exists a sub σ-field of events $\mathscr{D}$ of $\mathscr{F}$ with the property that W is $\mathscr{D}$-measurable and for each $n = 0, 1, \cdots$, the RV u_n is *conditionally independent* of the σ-field $\mathscr{G}_n$ *given* $\mathscr{D}$, with

$$\mathscr{G}_n := \sigma\{W, \sigma_n\} \vee \sigma\{\sigma_m, u_m, \tau_{m+1}, 0 \leqq m < n\}, \qquad n = 0, 1, \cdots. \tag{3.11}$$

Note that (A.2) is automatically satisfied for any σ-field $\mathscr{D}$ when the constituting loading sequence has the simple form (2.12) as in [2], [5].

Let $\mathscr{D}$ be any σ-field of events contained in $\mathscr{F}$. With a notation consistent with that introduced in Theorem 3.1, let the RVs $\{W_n(\mathscr{D})\}_0^\infty$, $\{R_n(\mathscr{D})\}_0^\infty$, $\{S_n(\mathscr{D})\}_0^\infty$ and $\{T^i_n(\mathscr{D})\}_0^\infty$, $i = 1, 2$, be defined through (2.5)–(2.9), but for the K-server queueing system generated by the constituting sequence

$$\langle W, \sigma^{\mathscr{D}}_n, u^{\mathscr{D}}_n, \tau^{\mathscr{D}}_{n+1}, n = 0, 1, \cdots \rangle. \tag{3.12}$$

Theorem 3.3. *Under the assumptions* (A.1)–(A.2), *for all* $1 \leqq k \leqq K$, *the almost sure inequalities*

$$W_n^k(\mathscr{D}) \leqq W_n^{k,\mathscr{D}}, \qquad n = 0, 1, \cdots \tag{3.13}$$

and

$$R_n^k(\mathscr{D}) \leqq R_n^{k,\mathscr{D}}, \qquad S_n^k(\mathscr{D}) \leqq S_n^{k,\mathscr{D}}, \qquad n = 0, 1, \cdots \tag{3.14}$$

hold, whence

$$T_n^i(\mathscr{D}) \leqq T_n^{i,\mathscr{D}}, \qquad i = 1, 2, \qquad n = 0, 1, \cdots. \tag{3.15}$$

Proof. For all $n = 0, 1, \cdots$, the RV W_n is $\mathscr{G}_n$-measurable, and by assumption (A.2), the RV u_n is thus conditionally independent of the RVs $\{\sigma_n, W_n\}$ given the σ-field $\mathscr{D}$, whence

$$E[u_n^k \,.\, \sigma_n^k \mid \mathscr{D}] = u_n^{k,\mathscr{D}} \,.\, \sigma_n^{k,\mathscr{D}} \quad \text{a.s.} \qquad n = 0, 1, \cdots \tag{3.16a}$$

and

$$E[u_n^k \,.\, W_n^k \mid \mathscr{D}] = u_n^{k,\mathscr{D}} \,.\, W_n^{k,\mathscr{D}} \quad \text{a.s.} \qquad n = 0, 1, \cdots \tag{3.16b}$$

for all $1 \leqq k \leqq K$.

With the $\mathbb{R}_+^K$-valued RVs $\{O_{n+1}\}_0^\infty$ defined componentwise by

$$O_{n+1}^k = u_n^k \,.\, \sigma_n^k - \tau_{n+1}^k \qquad n = 0, 1, \cdots \tag{3.17}$$

for all $1 \leqq k \leqq K$, it is plain from (3.16a) that

$$O_{n+1}^{k,\mathscr{D}} = u_n^{k,\mathscr{D}} \,.\, \sigma_n^{k,\mathscr{D}} - \tau_{n+1}^{k,\mathscr{D}} \qquad n = 0, 1, \cdots \tag{3.18}$$

for all $1 \leqq k \leqq K$. The inequality (3.13) now follows readily from Theorem 3.1 applied to the recursions (2.5)–(2.9), whereas (3.14) are direct consequences of (3.13) and (3.16). It is also plain that

$$E[T_n^1 \mid \mathscr{D}] \geqq \max_{1 \leqq k \leqq K} E[R_n^k \mid \mathscr{D}] \geqq \max_{1 \leqq k \leqq K} R_n^k(\mathscr{D}) \quad \text{a.s.,} \qquad n = 0, 1, \cdots \tag{3.19}$$

where (3.14) was used in the last inequality, and the validity of (3.15) for the case $i = 1$ is now established. The proof for the case $i = 2$ is similar and is therefore omitted.

The following corollary is an easy by-product of Theorem 3.3.

Corollary 3.4. *Under the assumptions* (A.1)–(A.2), *the inequalities*

$$E[T_n^i(\mathscr{D})] \leqq E[T_n^i], \qquad i = 1, 2, \qquad n = 0, 1, \cdots, \tag{3.20}$$

hold; more generally, for all convex monotone non-decreasing functions $\phi : \mathbb{R} \to \mathbb{R}$,

$$E[\phi(T_n^i(\mathscr{D}))] \leqq E[\phi(T_n^i)], \qquad i = 1, 2, \qquad n = 0, 1, \cdots, \tag{3.21}$$

provided both expectations exist.

Proof. It suffices to establish (3.21) since (3.20) follows from it by taking $\phi(x)=x$ for all x in $\mathbb{R}$. Theorem 3.3 yields

$$T_n^i(\mathcal{D}) \leqq E[T_n^i \mid \mathcal{D}], \qquad i=1, 2, \qquad n=0, 1, \cdots \tag{3.22}$$

and Jensen's inequality thus implies the inequalities

$$\phi(T_n^i(\mathcal{D})) \leqq E[\phi(T_n^i) \mid \mathcal{D}], \qquad i=1, 2, \qquad n=0, 1, \cdots. \tag{3.23}$$

The conclusion (3.21) is now immediate from (3.23) upon taking the mathematical expectation on both sides of these inequalities.

The comparison results of Theorem 3.3 and of its Corollary 3.4 are really statements on the *convex increasing ordering* between waiting times and response times in two different queueing systems, as understood by Ross [18], Stoyan [21] and many other authors. More precisely, let X and Y be any two $\mathbb{R}^K$-valued RVs. The (distribution of the) RV X is said to be *greater* than the (distribution of the) RV Y in the (*stochastic*) *convex increasing order* if and only if $E[\phi(Y)] \leqq E[\phi(X)]$ for all *convex monotone non-increasing* mapping $\phi: \mathbb{R}^K \to \mathbb{R}$ for which the expectations exist; this is denoted in short by $X \leqq_{\text{icx}} Y$.

With this notation, Corollary 3.3 can be restated simply as saying that

$$T_n^i(\mathcal{D}) \leqq_{\text{icx}} T_n^i, \qquad i=1, 2, \qquad n=0, 1, \cdots. \tag{3.24}$$

In fact, an argument identical to the one made in Corollary 3.4, when applied to the inequalities (3.13)–(3.14), yields the following stronger stochastic ordering in *vector* form.

Corollary 3.5. *Under the assumptions* (A.1)–(A.2), *the inequalities*

$$W_n(\mathcal{D}) \leqq_{\text{icx}} W_n, \qquad n=0, 1, \cdots \tag{3.25}$$

and

$$R_n(\mathcal{D}) \leqq_{\text{icx}} R_n, \qquad S_n(\mathcal{D}) \leqq_{\text{icx}} S_n, \qquad n=0, 1, \cdots \tag{3.26}$$

hold.

It is plain that the inequalities (3.24) are immediate consequences of (3.26) since the mapping $\mathbb{R}^K \to \mathbb{R}: x \to \max_{1 \leqq k \leqq K} x^k$ is convex monotone increasing. The next corollary parallels Corollary 3.2 to the present set-up and shows how some of the bounds could possibly be improved.

Corollary 3.6. *Let* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *be two σ-fields such that* $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \mathcal{F}$, *which both satisfy* (A.2). *Under the finite mean assumption* (A.1), *the chains of inequalities*

$$W_n(\mathcal{D}_1) \leqq_{\text{icx}} W_n(\mathcal{D}_2) \leqq_{\text{icx}} W_n, \qquad n=0, 1, \cdots \tag{3.27}$$

and

$$S_n(\mathcal{D}_1) \leqq_{\text{icx}} S_n(\mathcal{D}_2) \leqq_{\text{icx}} S_n, \qquad n=0, 1, \cdots \tag{3.28}$$

hold true.

Sketch of the proof. Details of the proof can be found in [6]. The right-hand inequalities are already contained in Corollary 3.5 (with $\mathscr{D} = \mathscr{D}_2$). To prove the left-hand ones, fix $n = 0, 1, \cdots$ and consider any integrable $\mathbb{R}$-valued RV Y which is *conditionally independent* of the RV u_n given each one of the σ-fields $\mathscr{D}_1$ and $\mathscr{D}_2$. This condition implies

$$(3.29) \qquad E[u_n^k \,.\, Y \mid \mathscr{D}_i] = E[u_n^k \mid \mathscr{D}_i] \,.\, E[Y \mid \mathscr{D}_i], \qquad i = 1, 2, \qquad n = 0, 1, \cdots$$

for all $1 \leqq k \leqq K$. Any integrable $\mathscr{G}_n$-measurable RV is such an RV by virtue of the enforced condition (A.2) on the σ-fields $\mathscr{D}_1$ and $\mathscr{D}_2$, where $\mathscr{G}_n$ is the σ-field entering (A.2). The inclusion $\mathscr{D}_1 \subseteq \mathscr{D}_2$ now implies

$$(3.30) \qquad (u_n^k \,.\, Y)^{\mathscr{D}_1} = ((u_n^k \,.\, Y)^{\mathscr{D}_2})^{\mathscr{D}_1} = (u_n^{k,\mathscr{D}_2} \,.\, Y^{\mathscr{D}_2})^{\mathscr{D}_1}, \qquad n = 0, 1, \cdots$$

by the smoothing for conditional expectations, and therefore

$$(3.31) \qquad E[u_n^k \mid \mathscr{D}_1] \,.\, E[Y \mid \mathscr{D}_1] = (u_n^{k,\mathscr{D}_2} \,.\, Y^{\mathscr{D}_2})^{\mathscr{D}_1}, \qquad n = 0, 1, \cdots$$

upon comparing (3.30) to (3.29) (with $i = 1$). This last relation can be used exactly as (3.16a) was used in the proof of Theorem 3.3 to derive the leftmost inequalities. Indeed, observe that in the proof of Theorem 3.3, only the consequences (3.16a) of the assumption (A.2) enforced on the σ-field $\mathscr{D}$ are needed to carry out the arguments in order to get (3.13).

3.3. *Steady-state analysis.* The bounds obtained thus far are transient in nature but extend readily to statistical (or steady-state) equilibrium when appropriate. To that end, consider the additional conditions (A.3) and (A.3b) on the constituting sequences (3.10) and (3.12) (for some σ-field $\mathscr{D}$).

(A.3) The RVs $\{(\sigma_n, u_n, \tau_{n+1})\}_0^\infty$ form a *stationary ergodic* sequence.

(A.3b) The RVs $\{(\sigma_n^{\mathscr{D}}, u_n^{\mathscr{D}}, \tau_{n+1}^{\mathscr{D}})\}_0^\infty$ form a *stationary ergodic* sequence.

The next proposition is obtained through a vector version of the increasing scheme technique of Loynes [15].

Theorem 3.7. *Under the assumptions* (A.1) *and* (A.3), (*or* (A.1) *and* (A.3b), *respectively*), *the condition*

$$(3.32) \qquad E[u_n^k \,.\, \sigma_n^k] < E[\tau_{n+1}^k], \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots$$

$$(3.33) \qquad (\text{or } E[u_n^{k,\mathscr{D}} \,.\, \sigma_n^{k,\mathscr{D}}] < E[\tau_{n+1}^{k,\mathscr{D}}], \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots)$$

guarantees the weak convergence of the sequence of waiting times $\{W_n\}_0^\infty$ (*or* $\{W_n(\mathscr{D})\}_0^\infty$, *respectively*) *to some non-defective probability distribution function on* $\mathbb{R}_+^K$.

Proof. The proof proceeds in two steps which are specified through the value of the initial workload W. To indicate the dependence on this initial workload, denote by $\{{}^W W_n\}_0^\infty$ the sequence of waiting times which are defined componentwise through the recursions (2.5) when the initial workload is W, i.e., for all $1 \leqq k \leqq K$,

$$ {}^{W}W_{n+1}^{k} = [{}^{W}W_{n}^{k} + u_{n}^{k} \cdot \sigma_{n}^{k} - \tau_{n+1}^{k}]^{+}, \qquad n = 0, 1, \cdots \tag{3.34} $$

with ${}^{W}W_0 = W$.

The first step assumes the initial workload W to be 0, in which case iterating (3.34) yields the well-known representation

$$ {}^{0}W_{n}^{k} = \max\{0, O_{n}^{k}, O_{n}^{k} + O_{n-1}^{k}, \cdots, O_{n}^{k} + \cdots + O_{1}^{k}\}, \qquad n = 1, 2, \cdots \tag{3.35} $$

with the RVs $\{O_n^k\}_0^\infty$ as defined in (3.17).

Following Loynes [15], it is convenient to embed the sequence $\{O_n\}_1^\infty$ into a bi-infinite *stationary ergodic* sequence, say $\{O_n\}_{-\infty}^{+\infty}$. Such an extension is possible owing to the enforced assumption (A.3). The $\mathbb{R}_+^K$-valued RVs $\{V_n\}_0^\infty$ are now defined componentwise by

$$ V_{n}^{k} := \max\{0, O_{-1}^{k}, O_{-1}^{k} + O_{-2}^{k}, \cdots, O_{-1}^{k} + \cdots + O_{-n}^{k}\}, \qquad n = 1, 2, \cdots \tag{3.36} $$

with $V_0 = 0$, and equivalence in law being denoted by $=_{\mathrm{st}}$, it is clear that

$$ {}^{0}W_n =_{\mathrm{st}} V_n \qquad n = 0, 1, \cdots. \tag{3.37} $$

The ergodicity assumption now entails

$$ \lim_n \frac{1}{n} \sum_{i=1}^{n} O_{-i}^{k} = E[O_0^k] < 0 \quad \text{a.s.} \tag{3.38} $$

where the last inequality follows from the stability condition (3.32). Consequently, the convergence

$$ \lim_n \sum_{i=1}^{n} O_{-i}^{k} = -\infty \quad \text{a.s.} \tag{3.39} $$

takes place for all $1 \leqq k \leqq K$, and implies the existence of an almost surely *finite* integer N_k with the property that for all $n > N_k$,

$$ \sum_{i=1}^{n} O_{-i}^{k} < 0 \quad \text{a.s.} \tag{3.40} $$

The reader will easily check from the defining relation (3.36) that the sequence of RVs $\{V_n\}_0^\infty$ is componentwise *monotone non-decreasing*, i.e., $0 \leqq V_n^k \leqq V_{n+1}^k$ for all $1 \leqq k \leqq K$ and $n = 0, 1, \cdots$, and the RV V_∞ whose kth component is given by $V_\infty^k := \lim_n V_n^k$, $1 \leqq k \leqq K$, is thus well defined. Owing to (3.40), this RV V_∞^k is readily interpreted as the maximum of an almost surely *finite* number of RVs, whence it is almost surely finite, and the sequence of $\mathbb{R}_+^K$-valued RVs $\{V_n\}_0^\infty$ converges almost surely to the almost surely *finite* RV V_∞. Consequently, owing to (3.37), the $\mathbb{R}_+^K$-valued RVs $\{{}^{0}W_n\}_0^\infty$ necessarily converge *weakly* to an almost surely finite $\mathbb{R}_+^K$-valued RV, say W_∞, which is identical in law with the non-defective RV V_∞.

For an arbitrary initial workload $W \geqq 0$, an easy induction argument shows that ${}^{W}W_n^k \geqq {}^{0}W_n^k$, for all $n = 0, 1, \cdots$. In addition, as pointed by Loynes [15], the RV ν_k

given by

$$\nu_k := \inf\{n \geqq 0 : {}^W W_n^k = 0\}, \tag{3.41}$$

is almost surely *finite* under the stability condition (3.32). Therefore, ${}^W W_{\nu_k}^k = {}^0 W_{\nu_k}^k = 0$, and ${}^W W_n^k = {}^0 W_n^k$ necessarily for all $n \geqq \nu_k$ by virtue of (3.34). The RV ν given by

$$\nu = \max_{1 \leqq k \leqq K} \nu_k \tag{3.42}$$

is thus almost surely *finite* and has the property that ${}^W W_n = {}^0 W_n$ for all $n \geqq \nu$.

The first part of the proof now implies that the sequence of $\mathbb{R}_+^K$-valued RVs $\{{}^W W_n\}_0^\infty$ converges *weakly* to the almost surely finite (and thus non-defective) RV W_∞, i.e., the theorem is indeed obtained for the sequence $\{W_n\}_0^\infty$. It is clear from the discussion that the limiting distribution is independent of the initial workload distribution. The corresponding result for the sequence of $\mathbb{R}_+^K$-valued RVs $\{W_n(\mathscr{D})\}_0^\infty$ is obtained by similar arguments which are omitted for the sake of brevity.

The reader will readily check from (3.16a) that under the assumptions (A.1)–(A.3b), the queueing systems generated by the constituting sequences (3.10) and (3.12) exhibit the *same* stability condition, to wit (3.32). In what follows, the $\mathbb{R}_+^K$-valued RV W_∞ (or $W_\infty(\mathscr{D})$) will be any RV on $(\Omega, \mathscr{F}, P)$ distributed according to the limiting distribution of $\{W_n\}_0^\infty$ (or $\{W_n(\mathscr{D})\}_0^\infty$), whose existence is guaranteed by Theorem 3.7 under the stability condition (3.32). Similar interpretations are given for R_∞ (or $R_\infty(\mathscr{D})$), S_∞ (or $S_\infty(\mathscr{D})$) and T_∞^i (or $T_\infty^i(\mathscr{D})$, respectively), $i = 1, 2$. This section concludes with the following steady-state version of Corollaries 3.5 and 3.6.

Theorem 3.8. *Under the assumptions* (A.1)–(A.3b), *whenever the stability condition* (3.32) *is satisfied, the inequalities*

$$W_\infty(\mathscr{D}) \leqq_{\text{icx}} W_\infty \tag{3.43}$$

$$R_\infty(\mathscr{D}) \leqq_{\text{icx}} R_\infty, \qquad S_\infty(\mathscr{D}) \leqq_{\text{icx}} S_\infty \tag{3.44}$$

and

$$T_\infty^i(\mathscr{D}) \leqq_{\text{icx}} T_\infty^i, \qquad i = 1, 2 \tag{3.45}$$

hold true.

Proof. Let ϕ be any convex non-decreasing function $\phi : \mathbb{R}^K \to \mathbb{R}_+$ with the property

$$E[\phi(W_\infty)] < \infty \quad \text{and} \quad E[\phi(W_\infty(\mathscr{D}))] < \infty. \tag{3.46}$$

The RVs $\{V_n\}_0^\infty$ defined by the pathwise scheme (3.36) satisfy the inequalities $V_n^k \leqq V_{n+1}^k \leqq V_\infty^k$ for all $1 \leqq k \leqq K$ and $n = 0, 1, \cdots$. The monotone convergence theorem thus yields

$$\begin{aligned} E[\phi({}^0 W_n)] &= E[\phi(V_n)] \\ &\leqq \lim_n E[\phi(V_n)] = E[\phi(V_\infty)] = E[\phi(W_\infty)], \qquad n = 0, 1, \cdots \end{aligned} \tag{3.47}$$

upon making use of some of the facts noted in the proof of Theorem 3.7. A similar conclusion is obtained for the sequence $\{{}^0W_n(\mathscr{D})\}_0^\infty$.

Using obvious notation, the RVs $\phi(V_n)$ and $\phi(V_n(\mathscr{D}))$ are thus *integrable* as a result of (3.46). Furthermore, Corollary 3.5 and (3.37) yield the relation

$$E[\phi(V_n(\mathscr{D}))] \leqq E[\phi(V_n)], \qquad n = 0, 1, \cdots. \tag{3.48}$$

The proof of (3.43) is now completed by letting n go to infinity in (3.48) and by making use of (3.47) (for both $\{{}^0W_n\}_0^\infty$ and $\{{}^0W_n(\mathscr{D})\}_0^\infty$). The inequalities (3.44)–(3.45) are now immediate under the enforced assumptions.

4. The first family of bounds

Consider the K-dimensional FJ queue described in Section 2 under the synchronization condition (2.10) on the arrivals, with assumption (A.1) enforced throughout. The discussion given in Section 3 clearly indicates how bounds can be obtained on the statistics of the RVs $\{W_n\}_0^\infty$, $\{R_n\}_0^\infty$ $\{S_n\}_0^\infty$ and $\{T_n^i\}_0^\infty$, $i = 1, 2$. This is done by appropriately selecting several sub σ-fields $\mathscr{D}$ of $\mathscr{F}$, with a view towards giving a different system interpretation in each case. The material of this section being illustrative of the methodology of Section 3, no attempts are made to give the strongest possible results.

4.1. *The FJ queue with deterministic arrival and loading patterns.* Consider the sub σ-algebra $\mathscr{D}_1$ of $\mathscr{F}$ given by

$$\mathscr{D}_1 = \sigma\{W, \sigma_n, n = 0, 1, \cdots\} \tag{4.1}$$

under the conditions (A.2) and (A.4).

(A.4) The σ-field $\mathscr{D}_1$ is *independent* of the σ-field $\sigma\{u_n, \tau_{n+1}, n = 0, 1, \cdots\}$.

This set of conditions will be satisfied for instance if the independence assumption (I) holds. Moreover, note also that *under* (A.4), condition (A.2) for $\mathscr{D}_1$ is equivalent to the RV u_n being independent of the RVs $\{u_k, \tau_{k+1}, 0 \leqq k < n\}$ for all $n = 0, 1, \cdots$.

With the RVs $\{O_{n+1}\}_0^\infty$ defined by (3.17), condition (A.4) implies

$$O_{n+1}^{k,\mathscr{D}_1} = E[u_n^k] \,.\, \sigma_n^k - E[\tau_{n+1}], \qquad n = 0, 1, \cdots \tag{4.2}$$

for all $1 \leqq k \leqq K$. A direct application of Theorem 3.3 then shows that the FJ queue with both *deterministic* arrivals and loading patterns constitutes a *lower bound* system to the original one (in the convex increasing order sense). Interest in this FJ queueing system with deterministic components becomes apparent whenever the families of component RVs $\{W^k, \sigma_n^k, n = 0, 1, \cdots\}$, $1 \leqq k \leqq K$, are *mutually independent*, so that the families of RVs $\{R_n^k(\mathscr{D}_1)\}_0^\infty$, $1 \leqq k \leqq K$, turn out to be *mutually independent*. A similar statement holds for the quantities $\{S_n^k(\mathscr{D}_1)\}_0^\infty$, $1 \leqq k \leqq K$.

This lower bound system receives a different interpretation depending on the situation. In case (C.1), it is clear that

$$E[u_n^k] = P[k \in I_n], \qquad n = 0, 1, \cdots \tag{4.3}$$

for all $1 \leqq k \leqq K$, while in case (C.2), $E[u_n^k]$ represents the mean value of a common scaling factor, namely.

$$E[u_n^1] = \cdots = E[u_n^K] = E[v_n], \qquad n = 0, 1, \cdots. \tag{4.4}$$

4.2. *The FJ queue with deterministic service times.* Consider the σ-field $\mathscr{D}_2$ defined by

$$\mathscr{D}_2 = \sigma\{W, \tau_{n+1}, n = 0, 1, \cdots\} \tag{4.5}$$

under the conditions (A.2) and (A.5).

(A.5) The σ-field $\mathscr{D}_2$ is *independent* of the σ-field $\sigma\{\sigma_n, u_n, n = 0, 1, \cdots\}$.

Here again, this set of conditions will be satisfied under assumption (I), whereas *under* (A.5), the condition (A.2) for $\mathscr{D}_2$ is equivalent to the RV u_n being independent of the RVs σ_n and $\{u_k, \sigma_k, 0 \leqq k < n\}$ for all $n = 0, 1, \cdots$. Under the assumptions (A.1)–(A.2) and (A.5), it is clear that

$$O_{n+1}^{k, \mathscr{D}_2} = E[u_n^k] \,.\, E[\sigma_n^k] - \tau_{n+1}, \qquad n = 0, 1, \cdots \tag{4.6}$$

and Theorem 3.3 shows that the FJ queue with both *deterministic* service times and loading factors also constitutes a lower bound system to the original system. If the system is initially empty (i.e., $W = 0$) and the homogeneity condition

$$E[u_n^k] \,.\, E[\sigma_n^k] = s^k, \qquad n = 0, 1, \cdots \tag{4.7}$$

holds for all $1 \leqq k \leqq K$, then the system response times $\{T_n^2(\mathscr{D}_2)\}_0^\infty$ can be expressed as the response times in a $G/D/1$ system with interarrival stream $\{\tau_{n+1}\}_0^\infty$ and constant service requirements $\max_{1 \leqq k \leqq K} s^k$.

4.3. *The FJ queue with divisible statistics.* Assume the synchronization condition (2.10) to hold, and the interarrival times $\{\tau_{n+1}\}_0^\infty$ to be *K-divisible* in the sense that the following conditions (D.1)–(D.3) are satisfied.

(D.1) There exists a sequence $\{\tilde{\tau}_{n+1}\}_0^\infty$ of *mutually independent* $\mathbb{R}_+^K$-valued RVs such that

$$\tau_{n+1} = \frac{1}{K} \sum_{k=1}^{K} \tilde{\tau}_{n+1}^k, \qquad n = 0, 1, \cdots. \tag{4.8}$$

(D.2) For all $n = 0, 1, \cdots$, the $\mathbb{R}_+$-valued RVs $\{\tilde{\tau}_{n+1}^1, \cdots, \tilde{\tau}_{n+1}^K\}$ are *exchangeable.*

(D.3) The families of RVs $\{\tilde{\tau}_{n+1}\}_0^\infty$ and $\{W, \sigma_n, u_n, n = 0, 1, \cdots\}$ are *mutually independent.*

In the renewal case discussed in Section 6, the conditions (D.1)–(D.3) reduce to the condition that the common distribution function of the RVs $\{\tau_{n+1}\}_0^\infty$ be K-divisible in the classical sense. Moreover, if the σ-field $\mathcal{T}$ is defined by $\mathcal{T} := \sigma\{\tau_{n+1},\ n=0, 1, \cdots\}$, then it is easy to check from (D.1) and (D.2) that

$$E[\tilde{\tau}^1_{n+1} \mid \mathcal{T}] = \cdots = E[\tilde{\tau}^K_{n+1} \mid \mathcal{T}] = \tau_{n+1}, \qquad n = 0, 1, \cdots, \tag{4.9}$$

with the last equation following from (4.8).

Consider now the *non*-synchronized queueing system generated by the constituting sequence

$$\langle W, \sigma_n, u_n, \tilde{\tau}_{n+1}, n = 0, 1, \cdots \rangle, \tag{4.10}$$

and observe that under the assumptions (D.1)–(D.3), the σ-field $\mathcal{D}_3$ defined by

$$\mathcal{D}_3 := \sigma\{W, \sigma_n, u_n, \tau_{n+1}, n = 0, 1, \cdots\} \tag{4.11}$$

trivially satisfies the assumption (A.2) with respect to (4.10). In complete analogy with (3.17), if the RVs $\{\tilde{O}_{n+1}\}_0^\infty$ are defined componentwise by

$$\tilde{O}^k_{n+1} = u^k_n \cdot \sigma^k_n - \tilde{\tau}^k_{n+1}, \qquad n = 0, 1, \cdots \tag{4.12}$$

for all $1 \leqq k \leqq K$, then the equality (4.9) and the independence assumption (D.3) readily yield

$$\tilde{O}^{k, \mathcal{D}_2}_{n+1} = u^k_n \cdot \sigma^k_n - \tau_{n+1}, \qquad n = 0, 1, \cdots \tag{4.13}$$

for all $1 \leqq k \leqq K$.

From the results of Section 3, the system generated by (4.10) thus provides an *upper bound* to the original system. Interest in the system generated by such a divisible input stream becomes clear when the families of component RVs $\{W^k, \sigma^k_n, u^k_n, \tilde{\tau}^k_n, n = 0, 1, \cdots\}$, $1 \leqq k \leqq K$, are mutually independent, since in that case the system (4.10) provides a *decoupled* upper bound system to the original FJ queue in the sense defined in Section 3. The interested reader is invited to consult [2], [5] for a specific example under renewal assumptions.

Several further extensions and refinements can be obtained from this bounding technique, but will not be discussed here for the sake of brevity; the interested reader is referred to [6] for a detailed discussion on the matter.

The upper bounds obtained in this section can be improved by the method discussed in Section 5. There, statements will be obtained in a stochastic sense stronger than the convex increasing ordering used here, but only for the statistics of the system response times $\{T^i_n\}_0^\infty$, $i = 1, 2$. As pointed out by the authors in [4], bounding systems such as (4.10) are the only ones for which the strong *vector* ordering statement of Corollary 3.5 holds.

4.4. *A counterexample.* At this point in the discussion, the reader may entertain other possible extensions to the previous results. A natural idea consists in making deterministic only some of the K arrival streams, say L $(1 < L \leqq K)$ of them, while keeping the remaining $K-L$ streams unchanged. Under standard independence and renewal assumptions, such an approach would lead to a partially decoupled system composed of an FJ queue of smaller dimension fed by the initial arrival process and of an independent FJ system with deterministic arrivals (and thus composed of independent channels). The exact solution being available for two-dimensional FJ queues (at least in some particular cases [1], [11]), the system response time statistics would also be computable for such a composite system when $L=K-2$. The reader might hope at first that such a partially decoupled system still provides a lower bound for the initial system since its constituting sequence is less variable.

Although the simulation results of [5] might lend credence to such a conjecture, this is unfortunately not true as shown by the following counterexample. Consider a two-queue FJ system with *deterministic* loading and service sequences satisfying $\sigma_n^1=\sigma_n^2=\sigma$, $u_n^1=u_n^2=1$ and with *non*-deterministic synchronized interarrival times $\tau_{n+1}^1=\tau_{n+1}^2=\tau_{n+1}$ for all $n=0,1,\cdots$. Let $\{W_n\}_0^\infty$ be the sequence of waiting times in this system where $W_0=W=0$ and let $\{\hat{W}_n\}_0^\infty$ be the corresponding sequence when the arrival sequences $\{\hat{\tau}_{n+1}\}_0^\infty$ are given by $\hat{\tau}_{n+1}^1=E[\tau_{n+1}]$ and $\hat{\tau}_{n+1}^2=\tau_{n+1}$ for all $n=0,1,\cdots$.

With an obvious notation, the system response time for the first incoming customer is given by $T_1=\sigma+(\sigma-\tau_1)^+$ and $\hat{T}_1=\sigma+(\sigma-\min(\tau_1,E[\tau_1]))^+$, respectively. Hence $\hat{T}_1\geqq T_1$ and the event $[\hat{T}_1>T_1]$ has positive probability thus proving that $\hat{T}_1\geqq_{\mathrm{st}}T_1$ (where $\leqq_{\mathrm{st}}$ denotes the strong order on distribution functions [18], [21] as defined in Section 5), so that the ordering $\hat{T}_1\geqq_{\mathrm{icx}}T_1$ holds, i.e., the new system is not a lower bound to the original system.

4.5. *Steady-state analysis.* Assume the constituting sequence of the original FJ queue to satisfy the condition (A.3). For sub σ-fields considered in Sections 4.1, 4.2 and 4.4, condition (A.3b) is an immediate consequence of condition (A.3), and, owing to Theorem 3.8, all the transient bounds derived there can be given a steady-state version provided the initial constituting sequence fulfills the stability condition of Theorem 3.7. For the FJ queue with divisible statistics, additional conditions typically need to be imposed on the initial constituting sequence for validating the passage to steady-state.

4.6. *Ross's conjecture.* It was conjectured by Ross and proved in [4], [19] [20] under various statistical assumptions that the response times in a $G/G/1$ queue with a server speed modulated by an independent stochastic process are always larger for the convex ordering than the corresponding response times in a $G/G/1$ queue where the environment is averaged out. This result can also be established for the generalized response times considered here. A proof of such a result can be obtained by combining the results of Section 4 to the methods of [4]; the proof is omitted for the sake of brevity.

5. A family of upper bounds

A second bounding methodology is developed in this section. Its application leads very naturally to the definition of a family of queueing systems that provide upper bounds on the performance measures of interest.

5.1. *Associated* RVs. This second bounding methodology is based in an essential way on the notion of associated RVs [9]. The $\mathbb{R}$-valued RVs $\{X_1, \cdots, X_K\}$ are *associated* if and only if, with the notation $X := (X_1, \cdots, X_K)$, the inequality

$$E[f(X)g(X)] \geqq E[f(X)]E[g(X)] \tag{5.1}$$

holds for all pairs of *monotone non-decreasing* mappings $f, g: \mathbb{R}^K \to \mathbb{R}$ for which the expectations $E[f(X)]$, $E[g(X)]$ and $E[f(X)g(X)]$ exist.

In order to explain the usefulness of this concept in the present context, it will be convenient to say that the $\mathbb{R}$-valued RVs $\{\bar{X}_1, \cdots, \bar{X}_K\}$ form *independent versions* of the RVs $\{X_1, \cdots, X_K\}$ if

(i) the RVs $\{\bar{X}_1, \cdots, \bar{X}_K\}$ are *mutually independent*;

(ii) for every $1 \leqq k \leqq K$, the RVs X_k and $\bar{X}_k$ have the *same* probability distribution.

The following proposition [10] is an easy consequence of the definition (5.1).

Theorem 5.1. *If the* RVs $\{X_1, \cdots, X_K\}$ *are associated, then the inequality*

$$P\Big[\max_{1 \leqq k \leqq K} X_k \leqq x\Big] \geqq P\Big[\max_{1 \leqq k \leqq K} \bar{X}_k \leqq x\Big] \tag{5.2}$$

holds true for all x in $\mathbb{R}$.

This result can be viewed as a statement on the stochastic ordering between the maximum of the RVs $\{X_1, \cdots, X_K\}$ and the corresponding quantity for the independent version. More precisely, if X and Y are two $\mathbb{R}$-valued RVs, then the (distribution of the) RV X is said to be *greater* than the (distribution of the) RV Y in the *stochastic order* if and only if

$$P[Y > t] \leqq P[X > t] \tag{5.3}$$

for all t in $\mathbb{R}$; this is denoted in short by $Y \leqq_{\text{st}} X$. As well known [18], [21], (5.3) is equivalent to the statement that

$$E[\phi(Y)] \leqq E[\phi(X)] \tag{5.4}$$

for all *monotone non-decreasing* mappings $\phi: \mathbb{R} \to \mathbb{R}$ for which the expectations exist. With this notion, Theorem 5.1 can be restated simply as saying that

$$\max_{1 \leqq k \leqq K} X_k \leqq_{\text{st}} \max_{1 \leqq k \leqq K} \bar{X}_k. \tag{5.5}$$

The elements of a 'calculus' for associated RVs are provided in [10], pp. 29–31. Some of these facts, which are often used in the discussion, have been collected in the next lemma for easy reference.

Lemma 5.2.

(i) *Independent* RVs *are associated.*

(ii) *The union of independent collections of associated* RVs *forms a set of associated* RVs.

(iii) *Any subset of a family of associated* RVs *forms a set of associated* RVs.

(iv) *Any monotone non-decreasing function of associated* RVs *generates a set of associated* RVs.

The basic bounding result is given in the next proposition. Let $\{\bar{X}_1, \cdots, \bar{X}_K\}$ be *independent versions* of some $\mathbb{R}$-valued RVs $\{X_1, \cdots, X_K\}$. For any subset I of the index set $\{1, \cdots, K\}$, define the $\mathbb{R}^K$-valued RV X^I by posing

$$X^{I,k} = \begin{cases} \bar{X}_k & \text{if} \quad k \in I; \\ X_k & \text{if} \quad k \notin I. \end{cases} \tag{5.6}$$

Note that $X^I = X$ when $I = \emptyset$ and that $X^I = \bar{X}$ when $I = \{1, \cdots, K\}$. The next result constitutes a strengthening of Theorem 5.1. Its proof can be found in [6].

Theorem 5.3. *Assume the* RVs $\{X_1, \cdots, X_K\}$ *and their independent versions* $\{\bar{X}_1, \cdots, \bar{X}_K\}$ *to be mutually independent. If the* RVs $\{X_1, \cdots, X_K\}$ *are associated, then*

$$\max_{1 \leqq k \leqq K} X^{I,k} \leqq_{\text{st}} \max_{1 \leqq k \leqq K} X^{J,k}, \tag{5.7}$$

for any pair I and J of subsets of $\{1, \cdots, K\}$ *such that* $I \subseteq J$.

5.2. *The upper bound systems.* The results of Section 5.1, especially Theorem 5.3, suggest ways of obtaining upper bounds on the response time statistics for the queueing system generated by the constituting system

$$\langle W, \sigma_n, u_n, \tau_{n+1}, n = 0, 1, \cdots \rangle \tag{5.8}$$

provided certain independence assumptions hold. This is done by introducing a family of queueing systems parametrized by the subsets of the index set $\{1, \cdots, K\}$. Let $\bar{W}$, $\{\bar{\sigma}_n\}_0^\infty$, $\{\bar{u}_n\}_0^\infty$ and $\{\bar{\tau}_{n+1}\}_0^\infty$ be sequences of $\mathbb{R}_+^K$-valued RVs defined on the probability triple $(\Omega, \mathcal{F}, P)$ under the following assumptions (A.6)–(A.8).

(A.6) The collections of RVs $\{W, (\sigma_n, u_n, \tau_{n+1}), \ n = 0, 1, \cdots\}$ and $\{\bar{W}, (\bar{\sigma}_n, \bar{u}_n, \bar{\tau}_{n+1}), n = 0, 1, \cdots\}$ are *mutually independent.*

(A.7) The sequences of RVs $\{\bar{W}^k, (\bar{\sigma}_n^k, \bar{u}_n^k, \bar{\tau}_{n+1}^k), n = 0, 1, \cdots\}$, $1 \leqq k \leqq K$, are *mutually independent*; and

(A.8) For each $1 \leqq k \leqq K$, the sequence of RVs $\{\bar{W}^k, (\bar{\sigma}_n^k, \bar{u}_n^k, \bar{\tau}_{n+1}^k), n = 0, 1, \cdots\}$ is *statistically indistinguishable* from the original sequence $\{W^k, (\sigma_n^k, u_n^k, \tau_{n+1}^k), n = 0, 1, \cdots\}$.

With the notation introduced in (5.6), for any subset I of the index set $\{1, \cdots, K\}$, consider the K-server queueing system generated by the constituting sequence

$$(5.9) \qquad \langle W^I, \sigma_n^I, u_n^I, \tau_{n+1}^I, n = 0, 1, \cdots \rangle.$$

All the quantities of interest for this system have an index I. In particular, the corresponding waiting times and response times form sequences of $\mathbb{R}_+^K$-valued RVs $\{W_n^I\}_0^\infty$, $\{R_n^I\}_0^\infty$ and $\{S_n^I\}_0^\infty$; the former is generated componentwise by the recursive scheme

$$(5.10) \qquad W_{n+1}^{I,k} = [W_n^{I,k} + u_n^{I,k} \cdot \sigma_n^{I,k} - \tau_{n+1}^{I,k}]^+, \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots$$

with $W_0^I = W^I$, whereas the latter are defined by

$$(5.11) \qquad R_n^{I,k} = u_n^{I,k} \cdot (W_n^{I,k} + \sigma_n^{I,k}) \quad \text{and} \quad S_n^{I,k} = W_n^{I,k} + u_n^{I,k} \cdot \sigma_n^{I,k}, \qquad n = 0, 1, \cdots$$

for all $1 \leqq k \leqq K$. The system response times $\{T_n^{i,I}\}_0^\infty$, $i = 1, 2$, are then given simply by

$$(5.12) \qquad T_n^{1,I} = \max_{1 \leqq k \leqq K} R_n^{I,k} \quad \text{and} \quad T_n^{2,I} = \max_{1 \leqq k \leqq K} S_n^{I,k}, \qquad n = 0, 1, \cdots.$$

This system clearly reduces to the original system (with constituting system (5.8)) when $I = \varnothing$.

The definitions (5.10)–(5.11) of the RVs W_n^I, R_n^I and S_n^I are consistent with the definition (5.6) where the RVs $\bar{W}_n$, $\bar{R}_n$ and $\bar{S}_n$ are defined through (5.10)–(5.11) with $I = \{1, \cdots, K\}$. More specifically, the RVs $\{\bar{W}_n\}_0^\infty$ are generated componentwise by the recursive scheme

$$(5.13) \qquad \bar{W}_{n+1}^k = [\bar{W}_n^k + \bar{u}_n^k \cdot \bar{\sigma}_n^k - \bar{\tau}_{n+1}^k]^+, \qquad 1 \leqq k \leqq K, \qquad n = 0, 1, \cdots$$

with $W_0 = \bar{W}$, whereas the RVs $\{\bar{R}_n\}_0^\infty$ and $\{\bar{S}_n\}_0^\infty$ are defined by

$$(5.14) \qquad \bar{R}_n^k = \bar{u}_n^k \cdot (\bar{W}_n^k + \bar{\sigma}_n^k) \quad \text{and} \quad \bar{S}_n^k = \bar{W}_n^k + \bar{u}_n^k \cdot \bar{\sigma}_n^k, \qquad n = 0, 1, \cdots$$

for all $1 \leqq k \leqq K$.

5.3. *The main result in the stochastic ordering* $\leqq_{\text{st}}$. The material of the Sections 5.1 and 5.2 is now combined to obtain the following basic result for the systems of interest here.

Theorem 5.4. *Assume both sets of* RVs $\{R_n^1, \cdots, R_n^K\}$ *and* $\{S_n^1, \cdots, S_n^K\}$ *to form sets of associated* RVs *for all* $n = 0, 1, \cdots$. *Under the assumptions* (A.6)–(A.8), *for any two subsets* I *and* J *of the index set* $\{1, \cdots, K\}$ *such that* $I \subseteq J$, *the inequalities*

$$(5.15) \qquad T_n^i \leqq_{\text{st}} T_n^{i,I} \leqq_{\text{st}} T_n^{i,J}, \qquad i = 1, 2, \qquad n = 0, 1, \cdots$$

hold.

Proof. Only the rightmost inequalities in (5.15) need to be established since the leftmost ones follow immediately from them upon taking I and J to be $\varnothing$ and I, respectively. As a result of the assumptions (A.6)–(A.8), for each $n=0, 1, \cdots$, the RVs $\{\bar{R}_n^1, \cdots, \bar{R}_n^K\}$ and $\{\bar{S}_n^1, \cdots, \bar{S}_n^K\}$ constitute *independent versions* for the RVs $\{R_n^1, \cdots, R_n^K\}$ and $\{S_n^1, \cdots, S_n^K\}$, respectively. Moreover, by virtue of (A.6), the RVs $\{R_n^1, \cdots, R_n^K\}$ and $\{\bar{R}_n^1, \cdots, \bar{R}_n^K\}$ are independent, and so are $\{S_n^1, \cdots, S_n^K\}$ and $\{\bar{S}_n^1, \cdots, \bar{S}_n^K\}$. The hypotheses of Theorem 5.3 are thus satisfied and the result (5.15) follows upon immediate identification.

Sufficient conditions are now given to ensure that the collections of RVs $\{R_n^1, \cdots, R_n^K\}$ and $\{S_n^1, \cdots, S_n^K\}$ form sets of associated RVs for all $n=0, 1, \cdots$. To that end, consider the assumptions (A.9)–(A.10).

(A.9) The RVs $\{W, (\sigma_n, u_n, \tau_{n+1}), n=0, 1, \cdots\}$ are *mutually independent.*

(A.10) Each one of the collections of RVs $\{W^1, \cdots, W^K\}$ and $\{\sigma_n^1, u_n^1, -\tau_{n+1}^1, \cdots, \sigma_n^K, u_n^K, -\tau_{n+1}^K\}$, $n=0, 1, \cdots$, form a set of *associated* RVs.

Theorem 5.5. *Under the assumptions* (A.9)–(A.10), *for all* $n=0, 1, \cdots$, *each one of the three collections of* RVs $\{W_n^1, \cdots, W_n^K\}$, $\{R_n^1, \cdots, R_n^K\}$ *and* $\{S_n^1, \cdots, S_n^K\}$ *forms a set of associated* RVs.

Proof. Take as induction hypothesis that the RVs $\{W_m^1, \cdots, W_m^K\}$ are *associated* for *some* $n=m\geqq 0$. By virtue of (A.9), the RV W_m is independent of the RV $(\sigma_m, u_m, -\tau_{m+1})$, and therefore the induction hypothesis and (A.10) imply that the RVs $\{W_m^1, \cdots, W_m^K, \sigma_m^1, u_m^1, -\tau_{m+1}^1, \cdots, \sigma_m^K, u_m^K, -\tau_{m+1}^K\}$ are associated, upon applying Lemma 5.2(ii). Part (iv) of Lemma 5.2 now gives the conclusion that the sets of RVs $\{W_{m+1}^1, \cdots, W_{m+1}^K\}$, $\{R_{m+1}^1, \cdots, R_{m+1}^K\}$ and $\{S_{m+1}^1, \cdots, S_{m+1}^K\}$ are three sets of associated RVs. In passing, this shows that the induction hypothesis holds for $n=m+1$, and since it holds for $n=0$, by virtue of assumption (A.10), it holds for all $n=0, 1, \cdots$.

The main result of this section is now obtained upon combining Theorems 5.4 and 5.5.

Theorem 5.6. *Under the assumptions* (A.6)–(A.10), *for any pair of subsets I and J of the index sets* $\{1, \cdots, K\}$ *such that* $I\subseteq J$, *the inequalities*

$$T_n^i \leqq_{\text{st}} T_n^{i,I} \leqq_{\text{st}} T_n^{i,J}, \qquad i=1, 2, \qquad n=0, 1, \cdots \tag{5.16}$$

hold true.

The following corollary follows by a simple application of (5.4) and (5.16) with $\phi(x)=x$ for all x in $\mathbb{R}$.

Corollary 5.7. *Under the assumptions* (A.6)–(A.10), *the inequalities*

$$E[T_n^i] \leqq E[T_n^{i,I}] \leqq E[T_n^{i,J}], \qquad i=1, 2, \qquad n=0, 1, \cdots \tag{5.17}$$

hold true.

Theorem 5.5 was already obtained by Nelson and Tantawi [16] under more restrictive assumptions and only for the case $I=\emptyset$. Their proof, used here, is an inductive one and is applicable to the more general situation with minor modifications. In the case $u_n^1=\cdots=u_n^K=1$ for all $n=0,1,\cdots$, Nelson and Tantawi [16] gave (5.17) with $I=J=\{1,\cdots,K\}$. Theorem 5.6 and its Corollary 5.7 thus provide a strengthened version of these earlier results.

5.4. *A special case.* Consider the FJ queue system of Section 2 with synchronized arrivals (2.10) and synchronized loading factors (2.3) of model (C.2), and assume the constituting sequence (5.8) to satisfy the *strong independence* assumption (Ib). Conditions (A.9)–(A.10) are then both satisfied, the latter as a result of Lemma 5.2(i), and the FJ queue system with constituting sequence (5.9) and $I=\{1,\cdots,K\}$ is constituted by K *independent* components, each one having the same statistics as the corresponding component in the original system. By Theorem 5.5, this fully decoupled system provides a *computable* upper bound to the initial system, in the sense that

$$T_n^2 \leqq_{\text{st}} T_n^{2,\{1,\cdots,K\}}, \qquad n=0,1,\cdots \tag{5.18}$$

a relation obtained by specializing the leftmost inequality of (5.16) to the case $I=\{1,\cdots,K\}$.

Theorems 5.4 and 5.5 can also be used to generate better upper bounds on the performance measures of interest, which are still computable. To see this, assume the response time distribution to be computable for the synchronized model under consideration in L dimensions for some $2\leqq L\leqq K$. Then for any subset I of $\{1,\cdots,K\}$ with cardinality $K-L$, (5.16) gives the bound

$$T_n^2 \leqq_{\text{st}} T_n^{2,I} \leqq_{\text{st}} T_n^{2,\{1,\cdots,K\}}, \qquad n=0,1,\cdots \tag{5.19}$$

which improves on (5.18). The statistics of the response times $\{T_n^{2,I}\}_0^\infty$ are computable since the corresponding system has $L+1$ *independent* components, namely, an L-dimensional synchronized FJ system and $K-L$ independent unsynchronized single server queues.

Theorem 5.4 can also be used directly to generate yet better upper bounds as follows. Let $\{I_1,\cdots,I_p\}$ be a partition of $\{1,\cdots,K\}$, i.e., $I_j\cap I_l=\emptyset$ for $1\leqq j<l\leqq p$ and $\bigcup_{j=1}^p I_j=\{1,\cdots,K\}$, such that $|I_j|\leqq L$ for $1\leqq j\leqq p$. Define the $\mathbb{R}$-valued RVs $\{X_n^j\}_0^\infty$, $1\leqq j\leqq p$, by

$$X_n^j := \max_{k\in I_j} S_n^k, \qquad n=0,1,\cdots \tag{5.20}$$

for all $1\leqq j\leqq p$.

For each $n=0,1,\cdots$, the RVs $\{S_n^1,\cdots,S_n^K\}$ are associated by Theorem 5.5 and so are the RVs $\{X_n^1,\cdots,X_n^p\}$ by virtue of Lemma 5.2(iv), whence

$$T_n^2=\max_{1\leqq j\leqq p} X_n^j \leqq_{\text{st}} \max_{1\leqq j\leqq p} \bar{X}_n^j, \qquad n=0,1,\cdots \tag{5.21}$$

where the RVs $\{\bar{X}_n^1, \cdots, \bar{X}_n^p\}$ form independent versions of $\{X_n^1, \cdots, X_n^p\}$. With this notation, it is also clear that

$$\bar{X}_n^j \leqq_{\text{st}} \max_{k \in I_j} \bar{S}_n^k, \qquad n = 0, 1, \cdots \tag{5.22}$$

for all $1 \leqq j \leqq p$, by a direct application of Theorem 5.6 to the FJ queue system made up of servers whose indices are in I_j. If the RVs $\{\Theta_n^{I_1, \cdots, I_p}\}_0^\infty$ are now given by

$$\Theta_n^{I_1, \cdots, I_p} := \max_{1 \leqq j \leqq p} \bar{X}_n^j, \qquad n = 0, 1, \cdots \tag{5.23}$$

then the inequalities

$$T_n^2 \leqq_{\text{st}} \Theta_n^{I_1, \cdots, I_p} \leqq_{\text{st}} T_n^{2, \{1, \cdots, K\}}, \qquad n = 0, 1, \cdots \tag{5.24}$$

are obtained by combining (5.21) and (5.22). The system with response times $\{\Theta_n^{I_1, \cdots, I_p}\}_0^\infty$ thus provides a refinement on the upper bound (5.18). It is also computable since it is composed of p *independent* FJ systems, all of dimension no greater than L.

A similar discussion can be carried out in the context of model (C.1) under the independence assumption (I) provided the independent RVs σ_n, u_n and $-\tau_{n+1}$ have *associated* components for all $n = 0, 1, \cdots$.

5.5. *Steady-state analysis.* Assume the constituting RVs (5.8) that define the original system to satisfy *both* the conditions (A.3) and (A.9), in which case the RVs $\{(\sigma_n, u_n, \tau_{n+1})\}_0^\infty$ form a sequence of i.i.d. $\mathbb{R}_+^{3K}$-valued RVs. It is now easy to see under conditions (A.6)–(A.8) that the constituting sequence (5.9) also satisfies (A.3) for every subset I of the index set $\{1, \cdots, K\}$, i.e., both (5.8) and (5.9) satisfy the stability condition (3.32) at the same time. Therefore, if the stability condition (3.32) is enforced, Theorem 3.7 ensures that the sequence of RVs $\{W_n^I\}_0^\infty$ (for each subset I of $\{1, \cdots, K\}$) converges weakly to some non-defective distribution function on $\mathbb{R}_+^K$. Again, generic RVs which are distributed according to the limiting distribution functions of $\{W_n^I\}_0^\infty$, $\{R_n^I\}_0^\infty$, $\{S_n^I\}_0^\infty$ and $\{T_n^{i,I}\}_0^\infty$, $i = 1, 2$, are denoted simply by W_∞^I, R_∞^I, S_∞^I and $T_\infty^{i,I}$, respectively. From the stability of the stochastic ordering $\leqq_{\text{st}}$ under weak limits ([21], Proposition 1.2.3, p. 6), the transient bounds of Theorem 5.6 also hold in statistical equilibrium.

Theorem 5.8. *Under the assumptions* (A.3) *and* (A.6)–(A.10), *whenever the stability condition* (3.32) *holds, the inequalities*

$$T_\infty^i \leqq_{\text{st}} T_\infty^{i,I} \leqq_{\text{st}} T_\infty^{i,J}, \qquad i = 1, 2 \tag{5.25}$$

hold for every pair of subsets I and J of the index set $\{1, \cdots, K\}$ such that $I \subseteq J$.

6. The renewal case—computable bounds

This section is devoted to explicit calculations for some of the bounds obtained so far. More specifically, the discussion is carried out when the arrivals are *synchronized* in the sense of (2.10), under the following set of *renewal* assumptions (R-1)–(R-6),

(R.1) The RV W and the sequences of RVs $\{\sigma_n\}_0^\infty$, $\{u_n\}_0^\infty$ and $\{\tau_{n+1}\}_0^\infty$ are *mutually independent.*

(R.2) The RVs $\{\tau_{n+1}\}_0^\infty$ form a *renewal* sequence with common probability distribution function $A(\cdot)$.

(R.3) The sequences $\{\sigma_n^k\}_0^\infty$, $1 \leqq k \leqq K$, are *mutually independent.*

(R.4) For each $1 \leqq k \leqq K$, the RVs $\{\sigma_n^k\}_0^\infty$ form a *renewal* sequence with common probability distribution function $B_k(\cdot)$.

(R.5) The RVs $\{u_n\}_0^\infty$ are *mutually independent.*

(R.6) For each $1 \leqq k \leqq K$, the RVs $\{u_n^k\}_0^\infty$ form a sequence of *i.i.d.* RVs with common probability distribution function $H_k(\cdot)$.

Under these assumptions, the RVs $\{u_n^1, \cdots, u_n^K\}$ are not necessarily independent. Moreover, it should also be clear that the assumptions (R.1)–(R.6) imply the independence assumption (I). To fix the notation, set

$$E[u_n^k] := \nu_k, \qquad n = 0, 1, \cdots \tag{6.1}$$

for all $1 \leqq k \leqq K$, and as usual, define the arrival and service rates through the relations

$$\frac{1}{\lambda} := E[\tau_{n+1}] = \int_0^\infty t\, dA(t), \qquad n = 0, 1, \cdots \tag{6.2}$$

and

$$\frac{1}{\mu_k} := E[\sigma_n^k] = \int_0^\infty t\, dB_k(t), \qquad n = 0, 1, \cdots \tag{6.3}$$

for all $1 \leqq k \leqq K$, respectively.

It will be convenient to refer to the *homogeneous* situation as the one where the probability distribution functions $\{B_k(\cdot)\}_1^K$ and $\{H_k(\cdot)\}_1^K$ all coincide with some probability distribution functions $B(\cdot)$ and $H(\cdot)$, respectively, in which case, set $\mu_1 = \cdots = \mu_K =: \mu$ and $\nu_1 = \cdots = \nu_K =: \nu$.

6.1. *Computable bounds.* As indicated in earlier sections, *computable* bounds are obtained whenever *statistical decoupling* takes place between the various components of the corresponding bounding systems. For the first moments, this follows from well-known elementary facts, namely that for any set of *independent* $\mathbb{R}_+$-valued RVs $\{X_1, \cdots, X_K\}$,

$$E\left[\max_{1 \leqq k \leqq K} X_k\right] = \int_0^\infty \left(1 - \prod_{k=1}^K P[X_k \leqq x]\right) dx. \tag{6.4}$$

The first example is obtained from the results on convex ordering given in Section

4.1. Under the assumptions (R.1)–(R.5), the σ-field $\mathcal{D}_1$ defined by (4.1) satisfies both assumptions (A.2) and (A.4). Here, the definition (4.2) of the RVs $\{O_{n+1}^{k,\mathcal{D}_1}\}_0^\infty$ reduces to

$$(6.5) \qquad O_{n+1}^{k,\mathcal{D}_1} = v_k \, . \, \sigma_n^k - \frac{1}{\lambda}, \qquad n = 0, 1, \cdots$$

for all $1 \leqq k \leqq K$. The RVs $\{S_n^1(\mathcal{D}_1), \cdots, S_n^K(\mathcal{D}_1)\}$ are *mutually independent*, and the remark (6.4) thus yields the expression

$$(6.6) \qquad E[T_n^2(\mathcal{D}_1)] = \int_0^\infty \left(1 - \prod_{k=1}^K P[S_n^k(\mathcal{D}_1) \leqq x]\right) dx, \qquad n = 0, 1, \cdots.$$

Note that for every $1 \leqq k \leqq K$, the RVs $\{S_n^k(\mathcal{D}_1)\}_0^\infty$ are the successive response times in a standard $GI/GI/1$ queue with deterministic arrival times $\{n/\lambda\}_0^\infty$ and service requirements $\{v_k \, . \, \sigma_n^k\}_0^\infty$. Similar comments can be made concerning the computation of the statistics of the system response times $\{T_n^1(\mathcal{D}_1)\}_0^\infty$.

Consider now the upper bounds derived in Section 5. Take $I = \{1, \cdots, K\}$ and assume the RVs $\{u_n^1, \cdots, u_n^K\}$ to be *associated*. It is easy to check that the assumptions (A.9)–(A.10) are satisfied. By construction, the bounding system with constituting sequence (5.9) described by (5.13)–(5.14) exhibits *independent* components, and therefore

$$(6.7) \quad E[T_n^{1,\{1,\cdots,K\}}] = \int_0^\infty \left(1 - \prod_{k=1}^K P[u_n^k \, . \, (W_n^k + \sigma_n^k) \leqq x]\right) dx, \qquad n = 0, 1, \cdots.$$

Here the RVs $\{W_n^k\}_0^\infty$ are defined for the original FJ queueing system through (2.5), and correspond to the successive waiting times in a $GI/GI/1$ system with interarrival stream $\{\tau_{n+1}\}_0^\infty$ and service requirements $\{u_n^k \, . \, \sigma_n^k\}_0^\infty$. This last expression (6.7) simplifies somewhat when each loading sequence $\{u_n^k\}_0^\infty$ is a $\{0, 1\}$-valued *Bernoulli* sequence, since then

$$(6.8) \quad P[u_n^k \, . \, (W_n^k + \sigma_n^k) \leqq x] = \pi_k \, . \, P[W_n^k + \sigma_n^k \leqq x] + (1 - \pi_k), \qquad n = 0, 1, \cdots$$

for all $x \geqq 0$, with the notation

$$(6.9) \qquad \pi_k = P[u_n^k = 1], \qquad n = 0, 1, \cdots$$

for all $1 \leqq k \leqq K$.

Note that the bounding systems discussed above all exhibit the same stability condition, namely

$$(6.10) \qquad \max_{1 \leqq k \leqq K} \frac{\lambda \, . \, v_k}{\mu_k} < 1.$$

Under this condition, the formulas (6.6) and (6.7) readily extend to statistical equilibrium, with n replaced everywhere by ∞.

Under the renewal and independence assumptions stated above, more explicit expressions can be obtained for the steady-state versions of these bounds when the service time distributions $B_k(\cdot)$, $1 \leqq k \leqq K$, are all *exponential.* In the next two sections, the calculations are carried out for this special case, with (6.6) and (6.7) as point of departure for the lower and upper bounds, respectively.

6.2. *Lower bounds—exponential servers.* As pointed out in the remark following (6.6), the evaluation of $E[T^2_\infty(\mathscr{D}_1)]$ amounts to computing the equilibrium response time distribution for K independent single-server systems; the kth such system is a $D/M/1$ queue with arrival times $\{n/\lambda\}_0^\infty$ and exponential service times with parameter $\gamma_k = \mu_k/\nu_k$, $1 \leqq k \leqq K$. It is well known that the response time in such a $D/M/1$ system is exponentially distributed. More precisely, for all $1 \leqq k \leqq K$,

$$P[S^k_\infty(\mathscr{D}_1) > x] = \exp(-\delta_k x), \qquad x \geqq 0 \tag{6.11}$$

with

$$\delta_k := \gamma_k \cdot (1 - \beta_k) = \frac{\mu_k}{\nu_k} \cdot (1 - \beta_k), \tag{6.12}$$

where β_k is the smallest positive solution to the equation

$$\beta = \exp\left(-\frac{\gamma_k(1-\beta)}{\lambda}\right), \qquad \beta \geqq 0. \tag{6.13}$$

The expression (6.6) in statistical equilibrium now becomes

$$E[T^2_\infty(\mathscr{D}_1)] = \int_0^\infty \left(1 - \prod_{k=1}^K (1 - \exp(-\delta_k x))\right) dx. \tag{6.14}$$

Elementary calculations show that

$$1 - \prod_{k=1}^K (1 - \exp(-\delta_k x)) = \sum_{k=1}^K (-1)^{k+1} \sum_{T \in I_k} \exp\left(-\sum_{k \in I} \delta_k x\right), \qquad x \geqq 0 \tag{6.15}$$

where the simplifying notation

$$\mathscr{I}_k := \{I \subseteq \{1, \cdots, K\} : |I| = k\}, \qquad 1 \leqq k \leqq K \tag{6.16}$$

is used. For any non-empty subset I of $\{1, \cdots, K\}$, it is plain that

$$\int_0^\infty \exp\left(-\sum_{k \in I} \delta_k x\right) dx = \left(\sum_{k \in I} \delta_k\right)^{-1} \tag{6.17}$$

and direct substitution of (6.15)–(6.17) into (6.14) readily yields

$$E[T^2_\infty(\mathscr{D}_1)] = \sum_{k=1}^K (-1)^{k+1} \sum_{I \in \mathscr{I}_k} \left(\sum_{k \in I} \delta_k\right)^{-1}. \tag{6.18}$$

In the *homogeneous* case, $\delta_k = \delta$ for all $1 \leqq k \leqq K$ and the easy identity

$$\sum_{k=1}^K \frac{(-1)^{k+1}}{k} \binom{K}{k} = \sum_{k=1}^K \frac{1}{k} \tag{6.19}$$

allows a rewriting of (6.18) in the following simpler form

$$E[T_\infty^2(\mathscr{D}_1)] = \frac{1}{\delta} \cdot \sum_{k=1}^{K} \frac{1}{k} \tag{6.20}$$

since $|\mathscr{I}_k| = \binom{K}{k}$ for all $1 \leqq k \leqq K$.

6.3. *Upper bounds—exponential servers.* Under the foregoing assumptions, the computation of (6.9) in equilibrium passes through the calculation of the equilibrium time distribution for K independent $GI/GI/1$ systems. When the service times are exponential and the loading sequences are Bernoulli, explicit expressions are available for the distribution of the stationary waiting times for this bounding system of independent queues [7], and will be used below in deriving the explicit upper bounds (6.29). The relevant expressions are summarized in the next proposition, where the notation

$$\tilde{A}_k^*(s) = \frac{\pi_k \,.\, A^*(s)}{1 - (1 - \pi_k) \,.\, A^*(s)}, \qquad s \geqq 0 \tag{6.21}$$

is used for all $1 \leqq k \leqq K$.

Lemma 6.1. *Under the foregoing assumptions* (R.1)–(R.6), *the stability condition* (6.10) *reads* $\pi_k \lambda < \mu_k$, $1 \leqq k \leqq K$, *in which case the sequence of waiting times* $\{\bar{W}_n\}_0^\infty$ *converges weakly to some non-defective distribution* $F(\cdot)$ *on* $\mathbb{R}_+^K$. *This limiting distribution* $F(\cdot)$ *is independent of the initial waiting time distribution and is of the form* $F(x) = \prod_{k=1}^K F_k(x^k)$ *for all* $x = (x^1, \cdots, x^K)$ *in* $\mathbb{R}_+^K$, *with*

$$F_k(x^k) = 1 - \alpha_k \exp(-\theta_k x^k), \qquad x^k \geqq 0 \tag{6.22}$$

where α_k *is the smallest positive solution to the equation*

$$x = \tilde{A}_k^*(\mu_k(1 - x)), \qquad x \geqq 0 \tag{6.23a}$$

and

$$\theta_k := \mu_k(1 - \alpha_k). \tag{6.23b}$$

The corresponding Laplace–Stieltjes transform $F_k^*(\cdot)$ is then given by

$$F_k^*(s) = (1 - \alpha_k) + \alpha_k \,.\, \frac{\theta_k}{\theta_k + s}, \qquad s \geqq 0 \tag{6.24}$$

and by elementary calculations, it follows that

$$E[\exp(-sW_\infty^k)]B^*(s) = \left[\frac{\theta_k}{\theta_k + s}\right], \qquad s \geqq 0 \tag{6.25}$$

as expected. It is now plain from (6.8) and (6.25) that in statistical equilibrium, the expression (6.7) becomes

$$E[T_\infty^{1,\{1,\cdots,K\}}] = \int_0^\infty \left(1 - \prod_{k=1}^K [\pi_k . (1 - \exp(-\theta_k x)) + (1 - \pi_k)]\right) dx. \tag{6.26}$$

In the *homogeneous case*, $\theta_k = \theta$ and $\pi_k = \pi$ for all $1 \leqq k \leqq K$, and (6.26) reduces to

$$E[T_\infty^{1,\{1,\cdots,K\}}] = \int_0^\infty (1 - (1 - \pi \exp(-\theta x))^K)\, dx, \tag{6.27}$$

whence

$$E[T_\infty^{1,\{1,\cdots,K\}}] = \frac{1}{\theta} \cdot \sum_{k=1}^K (-1)^{k+1} \binom{K}{k} \frac{\pi^k}{k} \tag{6.28}$$

by elementary calculations and (6.17). From this last expression and from the fact that $\int_0^\pi x^{k-1}\, dx = \pi^k/k$ for all $1 \leqq k \leqq K$, the reader can readily check that

$$E[T_\infty^{1,\{1,\cdots,K\}}] = \frac{1}{\theta} \sum_{k=1}^K \frac{1 - (1 - \pi)^k}{k} \tag{6.29}$$

upon making use of elementary properties of geometric series.

7. Asymptotic analysis for homogeneous FJ queues

The derivation of asymptotics is now considered for a class of *homogeneous* FJ queues as the number of servers, K, grows large. For sake of simplicity, the discussion is carried in statistical equilibrium, under the renewal assumptions (R.1)–(R.6) and the additional assumptions (R.7)–(R.8), where

(R.7) The RVs $\{\sigma_n^k,\ 1 \leqq k \leqq K,\ n = 0, 1, \cdots\}$ form a collection of *i.i.d.* RVs whose common distribution $B(\cdot)$ has a *rational* Laplace–Stieltjes transform.

(R.8) The loading RVs $\{u_n\}_0^\infty$ are given by

$$u_n^1 = \cdots = u_n^K = 1, \qquad n = 0, 1, \cdots. \tag{7.1}$$

7.1. *Asymptotics for GI/GI/1 systems.* Consider a *stable GI/GI/1* queueing system where $A(\cdot)$ and $B(\cdot)$ denote the probability distribution functions of the interarrival and service times, respectively. The Laplace–Stieltjes transform $B^*(\cdot)$ is assumed *rational* so that the function $s \to f(s)$ which is initially defined for $\mathcal{R}(s) = 0$ by

$$f(s) = A^*(s) B^*(-s), \tag{7.2}$$

can be continued in the region $\mathcal{R}(s) \geqq 0$. Define μ^+ as

$$\mu^+ := \inf \{s \in \mathbb{R}_+ : f(s) < \infty\}. \tag{7.3}$$

It is plain under the enforced assumptions that $\mu^+ > 0$ and $f(\mu^+) = \infty$. Consequently, the queueing system being stable, $f'(0) < 0$ and convexity of $f(\cdot)$ implies the existence of a unique real number q in $(0, \mu^+)$ such that

$$f(q) = 1. \tag{7.4}$$

Let W, R and I be generic RVs which are distributed respectively according to the stationary waiting time, response time and idle period distributions of the $GI/GI/1$ queue under consideration. The following result is available in the monograph by Borovkov ([10], Theorem 11, p. 129) and is given here in appropriate form for handy reference.

Lemma 7.1. *The Laplace–Stieltjes transform of W is given by the expression*

$$E[e^{-sW}] = \frac{1 - P[W > 0]}{1 - \phi(s)}, \tag{7.5}$$

where $\phi(\cdot)$ is a completely monotone function which is analytic in the region $\{s : \mathcal{R}(s) > -(q + \varepsilon)\}$ for some $\varepsilon > 0$ and which satisfies the conditions

$$\phi(-q) = 1 \quad \text{and} \quad \phi'(-q) = \frac{f'(q)}{E[e^{-qI}] - 1}. \tag{7.6}$$

The Laplace–Stieltjes transform of the response time R is thus given by

$$E[e^{-sR}] = B^*(s) E[e^{-sW}] \tag{7.7}$$

and the function $x \to P[R > x]$, the so-called *complementary* distribution function of R, has Laplace transform given by

$$\int_0^\infty P[R > x] e^{-sx}\, dx = \frac{1}{s}\left[1 - B^*(s)\frac{1 - P[W > 0]}{1 - \phi(s)}\right]. \tag{7.8}$$

This transform function is analytic in the region $\mathcal{R}(s) > -(q + \varepsilon)$ for $\varepsilon > 0$ sufficiently small but for a pole of order 1 at $s = -q$, owing to the fact that $q < \mu^+$. The residue C associated to this pole is therefore given by

$$C = \frac{B^*(-q)(1 - E[e^{-qI}])}{qf'(q)}(1 - P[W > 0]) \tag{7.9}$$

and classical results on the first left singularity of Laplace transforms then yield the following estimate.

Lemma 7.2. *Under the foregoing assumptions,*

$$P[R > x] = Ce^{-qx}(1 + o(1)) \tag{7.10}$$

when x goes to infinity, with q and C given by (7.4) *and* (7.9), *respectively.*

7.2. *Maximum of identically distributed* RVs. The derivation of asymptotic bounds relies on the stochastic ordering theorems of the preceding sections and on results of Lai and Robbins [14] on the asymptotic behavior of the maximum of i.i.d RVs. Let $\{Y_k\}_1^\infty$ be a family of i.i.d. $\mathbb{R}_+$-valued RVs with common probability distribution function $G(\cdot)$, and introduce the RVs $\{M_K\}_1^\infty$ defined by

$$M_K := \max\{Y_1, \cdots, Y_K\}, \qquad K = 1, 2, \cdots. \tag{7.11}$$

Also, for ease of notation, define the $\mathbb{R}_+$-valued sequence $\{m_K\}_1^\infty$ by

$$m_K := \inf\left\{x \geqq 0 : 1 - G(x) \leqq \frac{1}{K}\right\}, \qquad K = 1, 2, \cdots. \tag{7.12}$$

The following result is a simplified version of Theorem 5(ii) given by Lai and Robbins ([14], p. 103) adapted to the present set-up.

Theorem 7.3. *Let* $\{Y_k\}_1^\infty$ *be a family of i.i.d* $\mathbb{R}_+$*-valued* RVs *whose common distribution function* $G(\cdot)$ *satisfies the conditions*

$$G(x) < 1 \quad \text{for all} \quad x \geqq 0 \tag{7.13a}$$

and

$$\lim_{x \to +\infty} \frac{1 - G(cx)}{1 - G(x)} = 0 \quad \text{for all} \quad c > 1. \tag{7.13b}$$

Under these conditions, the convergence

$$\lim_{K \to \infty} E\left[\left|\frac{M_K}{m_K} - 1\right|\right] = 0 \tag{7.14}$$

takes place and the asymptotics

$$E[M_K] = m_K(1 + o(1)), \qquad K = 1, 2, \cdots \tag{7.15}$$

hold true with K going to infinity.

In view of Lemma 7.2, it is now natural to consider probability distribution functions $G(\cdot)$ with tail behavior

$$P[Y_1 > x] = 1 - G(x) = Ce^{-qx}(1 + o(1)), \qquad x \geqq 0 \tag{7.16}$$

for some $q > 0$ and $C > 0$. The next proposition summarizes the asymptotic properties associated with this tail behavior (7.16). To simplify the exposition, for every $r \geqq 1$, denote by $G_r(\cdot)$ the probability distribution function of the rth power of any $\mathbb{R}_+$-valued RV distributed according to $G(\cdot)$, and in complete analogy with (7.12), define the real numbers $\{m_{K,r}\}_1^\infty$ by

$$m_{K,r} = \inf\left\{x \geqq 0 : 1 - G_r(x) \leqq \frac{1}{K}\right\}, \qquad K = 1, 2, \cdots. \tag{7.17}$$

It is plain that

$$P[|Y_1|^r > x] = 1 - G_r(x) = C \exp(-qx^{1/r})(1 + o(1)), \qquad x \geqq 0 \tag{7.18}$$

with the identifications $G_r(\cdot) = G(\cdot)$ and $m_{K,r} = m_K$ taking place for $r = 1$. The main result of this section can now be given.

Theorem 7.4. *Let* $\{Y_k\}_1^\infty$ *be a family of i.i.d* $\mathbb{R}_+$*-valued* RVs *whose common distribution function* $G(\cdot)$ *exhibits the tail behavior* (7.16). *In that case, for all* $r \geqq 1$,

(i) *the probability distribution* $G_r(\cdot)$ *satisfies the conditions* (7.13);

(ii) *the asymptotic equivalence*

$$\lim_{K\to\infty}\left[m_{K,r}\cdot\left[\frac{q}{\log K}\right]^r\right]=1 \tag{7.19}$$

holds true;

(iii) *the asymptotics*

$$E[|M_K|^r]=\left[\frac{\log K}{q}\right]^r.(1+o(1)),\qquad K=1,2,\cdots \tag{7.20}$$

hold true with K going to infinity.

Proof. Part (i) is readily checked by direct calculations which are omitted for sake of brevity. To show the equivalence (7.19), fix $r \geqq 1$ and observe from (7.18) that for every $0<\varepsilon<1$, there exists $x^*(\varepsilon)>0$ with the property that

$$1-G_r^{-\varepsilon}(x)\leqq 1-G_r(x)\leqq 1-G_r^{+\varepsilon}(x) \tag{7.21}$$

for all $x>x^*(\varepsilon)$, with

$$G_r^{\pm\varepsilon}(x):=1-C\exp(-qx^{1/r})(1\pm\varepsilon),\qquad x\geqq 0. \tag{7.22}$$

If the sequences $\{m_{K,r}^{\pm\varepsilon}\}_1^\infty$ are defined through (7.17) but with $G_r^{\pm\varepsilon}(\cdot)$ instead of $G_r(\cdot)$, then simple computations show that

$$m_{K,r}^{\pm\varepsilon}=\left[\frac{\log(K\cdot C(1\pm\varepsilon))}{q}\right]^r,\qquad K=1,2,\cdots \tag{7.23}$$

and the inequalities

$$m_{K,r}^{-\varepsilon}\leqq\left[\frac{\log(K.C)}{q}\right]^r\leqq m_{K,r}^{+\varepsilon},\qquad K=1,2,\cdots \tag{7.24}$$

are obtained. It also follows from (7.21) that for $K\geqq K_\varepsilon$ with some integer K_ε,

$$m_{K,r}^{-\varepsilon}\leqq m_{K,r}\leqq m_{K,r}^{+\varepsilon} \tag{7.25}$$

and upon combining (7.24)–(7.25), the asymptotic equivalence (7.19) is obtained since obviously

$$\lim_{K\to\infty}\left[\frac{m_{K,r}^{+\varepsilon}}{m_{K,r}^{-\varepsilon}}\right]=1. \tag{7.26}$$

The asymptotics (7.15) of Theorem 7.3 and (7.19), when combined with the obvious relation

$$|M_K|^r=\max\{|Y_1|^r,\cdots,|Y_K|^r\}, \tag{7.27}$$

readily yield (7.20).

References

[1] BACCELLI, F. (1985) Two parallel queues created by arrivals with two demands. Rapport de Recherche **426,** INRIA—Rocquencourt France.

[2] BACCELLI, F. AND MAKOWSKI, A. M. (1985) Simple computable bounds for the fork-join queue. *Proc. 19th Annual Conf. Information Sciences and Systems,* The Johns Hopkins University, Baltimore, MD, March 1985, 436–441.

[3] BACCELLI, F. AND MAKOWSKI, A. M. (1986) Stability and bounds for single server queue in random environment. *Stoch. Models* **2,** 281–292.

[4] BACCELLI, F. AND MAKOWSKI, A. M. (1989) Multidimensional stochastic ordering and associated random variables. *Operat. Res.* **37.**

[5] BACCELLI, F., MAKOWSKI, A. M. AND SHWARTZ, A. (1986) Simple computable bounds and approximations for the fork-join queue. *Intern. Workshop on Computer Performance Evaluation,* Tokyo, September 1985, pp. 437–450.

[6] BACCELLI, F., MAKOWSKI, A. M. AND SHWARTZ, A. (1987) The fork-join queue and related systems with synchronization constraints: stochastic ordering, approximations and computable bounds. Technical Research Report **TR-87-01,** System Research Center, University of Maryland, College Park, MD, January 1987 and Rapport de Recherche **687,** INRIA–Rocquencourt (France), June 1987.

[7] BACCELLI, F., MAKOWSKI, A. M. AND SHWARTZ, A. (1988) Computations for synchronized queueing networks. In preparation.

[8] BACCELLI, F., MASSEY, W. A. AND TOWSLEY, A. (1989) Acyclic fork-join queueing network. *J. Assoc. Comput. Mach.* **36.**

[9] BARLOW, R. E. AND PROSCHAN, F. (1975) *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston, Reading, MA.

[10] BOROVKOV, A. A. (1976) *Stochastic Processes in Queueing Theory* (English translation). Springer-Verlag, New York.

[11] FLATTO, L. AND HAHN, S. (1984) Two parallel queues created by arrivals with two demands I. *SIAM J. Appl. Math.* **44,** 1041–1053.

[12] HAJEK, B. (1983) The proof of a folk theorem on queueing delay with applications to routing in networks. *J. Assoc. Comput. Mach.* **30,** 834–851.

[13] HUMBLET, P. (1982) Determinism minimizes waiting times in queues. Technical Report LIDS—Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

[14] LAI, T. L. AND ROBBINS, H. (1978) A class of dependent random variables and their maxima. *Z. Wahrscheinlichkeitsch.* **42,** 89–111.

[15] LOYNES, R. M. (1962) The stability of a queue with non-independent interarrival and service times. *Proc. Camb. Phil. Soc.* **5,** 497–520.

[16] NELSON, R. AND TANTAWI, A. N. (1988) Approximate analysis of fork/join synchronization in parallel queues. *IEEE Trans Comput.* **37,** 739–743.

[17] ROGOZIN, B. A. (1966) Some extremal problems in queueing theory. *Theory Prob. Appl.* **11,** 144–151.

[18] ROSS, S. (1986) *Stochastic Processes.* Wiley, New York.

[19] ROLSKI, T. (1981) Queues with non-stationary input streams: Ross' conjecture. *J. Appl. Prob.* **13,** 603–68.

[20] ROLSKI, T. (1984) Comparison theorems for queues with dependent inter-arrival times. In *Modelling and Performance Evaluation Methodology,* Paris (France), January 1983. Lecture Notes in Control and Information Sciences **60,** Springer-Verlag, New York.

[21] STOYAN, D. (1984) *Comparison Methods for Queues and Other Stochastic Models* (English translation, ed. D. J. Daley). Wiley, New York.

[22] WHITT, W. (1984) Minimizing delays in the *GI/GI/*1 queue. *Operat. Res.* **32,** 41–51.