Journal of the Operational Research Society (2000) 51, 1170-1178 0160-5682/00 $15.00

www.stockton-press.co.uk/jors

On the distribution of throughput of transfer lines

C Dinçer[1] and B Deler[2]

[1]*Bilkent University, Turkey and* [2]*Northwestern University, USA*

Transfer lines simply characterise the interrelationship of manufacturing stages with their buffers and they are used to model the key features of such manufacturing environments with simplifying assumptions. There is a vast literature on these systems, however, little has been done on the transient analysis of the transfer lines by making use of the higher order moments of their performance measures due to the difficulty in determining the evolution of the stochastic processes under consideration. This paper examines the transient behaviour of relatively short transfer lines and derives the distribution of the performance measures of interest. An experiment is designed in order to compare the results of this study with the state-space representations and the simulation. Furthermore, extensions are briefly discussed and directions for future research are suggested.

Keywords: manufacturing; transfer lines; stochastic process evolution; transient analysis; steady-state analysis;

Introduction

A transfer line is a manufacturing system with a very special structure. It is a linear network of service stations or machines ($M_1, M_2, \ldots, M_n$) separated by buffer storages ($B_0, B_1, \ldots, B_n$). Material flows from outside the system to B_0, then to M_1, then to B_1, and so forth until it reaches B_n after which it leaves. Figure 1 depicts a two-machine transfer line. The squares represent machines and the triangles represent buffers.

There are two major reasons for studying the transfer lines. Firstly that they are of economic importance as they are generally used in high volume production. Secondly that they transfer lines simply characterise the interrelationship of manufacturing stages with their buffers, which are used to model the key features of such manufacturing environments with simplifying assumptions.

There is a vast literature on modelling and analysis of the transfer lines. However, most of the results are for steady-state operation. The literature emphasises this type of analysis because the equations involved are considerably simplified in the limit and relatively straightforward techniques such as balance equations can then be used. However, such steady-state analyses are inappropriate in many applied situations since the time horizon of operation naturally terminates, steady-state measures of system simply do not make sense. However, transient results can be quite difficult to obtain, tend to be rather complicated, and are available only for a fairly restricted class of models.

Almost all the methods in the literature deal with steady-state average production rates and steady-state average buffer levels though the variance of the production and the buffer levels during a time period is also important. Gershwin[1] reports based on the simulation experimentation and factory observations that the standard deviation of weekly production can be over 10% of the mean which means that it is very likely that customer requirements cannot be met on time in most of the time. It is striking that this inherent characteristic of the manufacturing systems, variability, is so little appreciated by researchers. However, today it has received a significant amount of importance as compared to the the past due to the emphasis on the just-in-time production.

This article shows how to calculate the mean and the variance of the throughput rate, which we define as the number of parts produced by a transfer line with buffer inventories per unit time, and then allows to calculate interval estimates for the throughput. These interval estimates provide an operational guide for the production manager. More importantly, we are now able to examine the transient behaviour of relatively short transfer lines and derive the distribution of the throughput, the number of parts leaving the system of interest at an arbitrary instant in time. This also leads to the calculation of the steady-state mean and variance of the throughput rate. Since transfer lines with high efficiencies and low variances are generally preferred, our results can be used to help design economically feasible transfer lines.

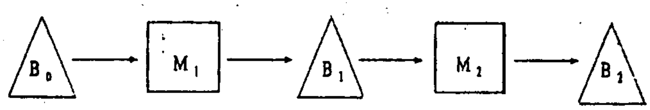

Figure 1 Two-machine serial line production system.

Correspondence: Dr C Dinçer, Department of IE, Bilkent University, Turkey.
E-mail: dincer@bilkent.edu.tr

Organisation of the remaining part of this paper is as follows. In Section 2, an overview of the transfer line literature, that emphasises the studies on the transient and steady-state analyses in which the first and second order moments of performance measures of interest are calculated, is given. Section 3 shows how the proposed model is developed. Numerical results and discussions are given in Section 4 and finally conclusions and new directions for future research are given in Section 5. The evolutions of the stochastic process of throughput for two- and three-machine transfer lines with finite buffer inventories as well as the corresponding representations are available in Dincer and Deler.[2]

Literature survey

In this section, we present a classification, which is based on the methodology followed, for the studies on the transient and steady-state analyses of the serial line production systems. Two major methodologies, that are analytical approach (exact *vs* approximate analysis) and simulation based approach, are under consideration. There is some cross fertilisation among these approaches, for example, some analytical approaches use simulation to generate or evaluate alternative initial states. Furthermore, we divide the models into two: (1) models for determining the first and second order moments of the performance measures in the steady-state; and (2) models for analysing the transient behavior of the transfer lines. The classification of the related literature is presented in Table 1.

If the relationships that compose the models are simple enough, it is possible to use mathematical methods (such as algebra, calculus, or probability theory) to obtain exact

Table 1 Classification from the methodology viewpoint

Methodology	*Publications*
Analytical methods	
Exact analysis	
Steady-state behaviour	Lavenberg[3]; Miltenburg[4]; Ou and Gershwin[5]; Hendricks[6]; Heavey, Papadopoulos and Browne[7]; Papadopoulos[8]; Papadopoulos and O'Kelly[9]; Tan[10]
Transient behaviour	Kelton and Law[11]; Klutke and Seiford[12]; Gopalan and Dinesh Kumar[13,14]
Approximate analysis	
Steady-state behaviour	Hong, Glassey and Seong[15]; Gershwin[16]; Glassey and Hong[17]; Springer[18]; Wu[19]; Papadopoulos[20]; Frein, Commault and Dallery[21]
Simulation-based methods	
Steady-state behaviour	Hendricks and McClain[22]
Transient behaviour	Lin and Cochran[23]

information on questions of interest. However, only special systems have exact solutions due to the complexities introduced by the buffers in the serial line production systems: Hendricks[6] developed a technique to analytically describe the output process of a serial production line of n machines with exponential processing time distributions and finite buffer capacities. Papadopoulos[8] considered the throughput of multi-station reliable production lines with no intermediate buffers and exponentially distributed processing times and, more importantly, this study provides the distribution function of the holding time at the stations. Heavey *et al*[7] examined multi-station series production lines, but with unreliable machines and Erlang-type distributed processing times with the number of phases allowed to vary for each station. The methodology they propose generates the associated set of linear equations that are solved for comparison purposes against approximate results which exist in the literature. In another study, Papadopoulos and O'Kelly[9] developed an exact procedure for the analysis of a production line consisting of single machines linked in series with no intermediate buffers between them. Their exact algorithm gives the marginal probability distribution of the number of parts processed by each machine, the mean queue length and the critical input rate, that is the throughput rate of the line.

In this review, we also describe the most important models and results of transient behavior of transfer line literature: Kelton and Law[11] considered the one-station system with parallel servers, exponential inter-arrival and service times, and also an arbitrary number of parts present at time zero. They obtain probabilities in a relatively simple closed form that can be used to evaluate exactly several measures of system performance, including the expected delay in queue of each arriving part. In another study, Klutke and Seiford[12] obtained an exact analytical expression for the expected output time of the ith part by using a recursive equation to compute the transient expected output times. Finally, a merge production system which has two parallel stations in the first stage followed by a single station in the second stage is analysed to study its transient behaviour in Gopalan and Dinesh Kumar[13,14] by using the concept of semi-regenerative phenomena.

It appears from the preceding studies that exact solutions of relatively short serial lines are available for a wide range of models. However, it seems hopeless to expect to obtain exact solutions of serial lines with more machines even when more powerful computers are available. In such cases, the use of approximate solutions is the only viable alternative: Hong *et al*[15] developed an efficient analytical method for the analysis of a n-machine production line with unreliable machines and random processing times. The behaviour of the n-machine line is approximated by the behaviours of the $(n-1)$ two-machine lines based on the decomposition technique proposed by Gershwin.[24] In their next study, Glassey and Hong[17] extend Hong *et al*[15] serial model by making use of the steady-state output of machine M_j in the n-machine line rather than the flow rate-idle time relationship. Springer[18] also proposes a new approximation for estimating the throughput rate and work-in-process inventory of finite-buffer exponential queues in series. Papadopoulos[20] derives another approximate analytical formula, using the holding time model method, for calculating the average throughput rate of an n-station production line with exponential service times, manufacturing blocking, and no intermediate buffers between adjacent stations. Finally Frein *et al*[21] approximate the behaviour of a closed-loop production line with unreliable machines and finite buffers by a continuous flow model that is analysed with a decomposition technique, which is similar to that used for (open) production lines.

Simulation is certainly more tractable than analytical formulations of production system problems. Moreover, there is no concern about feasibility since there is no need to make any simplifying assumptions. Therefore, the simulation model can be built as close to reality as one needs to: Hendricks and McClain[22] examined the output process of a serial production line of n machines with general processing time distributions and finite buffer capacities via the use of simulation. In another noteworthy study, Lin and Cochran[23] studied the transient behaviour of $GI/G/n$ queueing system for the often encountered dynamic event of machine breakdown by computer simulation.

The performance measures of almost all the studies correspond to the steady-state average production rates and steady-state average buffer levels. However, the essence of transfer lines is their variability and this issue has been mostly neglected. As far as we are aware of, there are only two published papers that deal with the variance of the performance measures of a transfer line over a limited period: Lavenberg[3] derives an expression for the Laplace–Stieltjes transform of the steady-state distribution of the queueing time for the $M/G/1$ finite capacity queue that can be differentiated readily in order to obtain higher order moments of the steady-state queuing time and Miltenburg[4] presents a procedure for calculating the variance of the number of units produced by two-machine transfer line.

There are also four working papers on the variance calculation of the performance measures of interest: Ou and Gershwin[5] derive a closed-form expression for the variance of the lead time distribution of a two-machine transfer line with a finite buffer and Gershwin[16] analyses the variance of a tandem production system. Wu[19] develops algorithms to calculate the variance of the number of departures at fixed time intervals from both tandem queuing network models and discrete-time models with breakdowns and repairs. Finally, Tan[10] determines analytically the variance of the throughput rate of an n-machine production line with no intermediate buffers and time-dependent failures.

A detailed description and discussion of the models, which have been developed till 1992, in the transfer line literature can be found in Dallery and Gershwin[25] and Buzacott and Shanthikumar.[25] Another noteworthy review is by Papadopoulos and Heavey[27] in which a bibliography of material (from 1992 to early 1995) concerned with modelling of production and transfer lines using queuing networks is provided. A similar literature survey on the transient behaviour of transfer line literature is also available in Deler and Dinçer.[28]

The model

The model assumptions and notation

This section is devoted to the verbal description of the transfer lines with buffer storages. The assumptions listed in Table 2 describe the mostly encountered production line in the literature,[25] and also, the basic notation used throughout this paper is introduced Table 3.

Modelling

As stated in the previous section, the general system under consideration is a n-machine-$(n+1)$-buffer line. However, we specialise on three systems: (1) The atomic model; (2) two-machine-one-buffer system; and (3) three-machine-two-buffer system. This section gives the verbal and graphical descriptions of the transient behavior of these

Table 2 The assumptions

1. The production line is a serial arrangement of a finite number of n machines. Each machine can operate on one unit of product at a time and has internal storage capacity for that unit.
2. The arrival process is assumed to be Poisson. Therefore, the inter-arrival time for part i, $T^i_\lambda (i = 1, \ldots)$, is an independent identically distributed exponential random variable with density function $f_i(t) = \lambda e^{-\lambda t}$, $\lambda > 0$, $t \geqslant 0$.
3. The machines $M_j (j = 1, \ldots, n)$ have mutually independent processing times that are also exponentially distributed with density function $f_j(t) = \mu_j e^{-\mu_j t}$, $\mu_j > 0$, $t \geqslant 0$.
4. The first buffer of the line is assumed to have zero capacity (the parts arriving the system while the first machine is busy are lost) and the last buffer is considered to have infinite capacity (last machine never gets blocked).
5. The buffers between the machines of the line have finite storage capacities. There are no overflows or lost parts. If a machine has finished working on a part and the next downstream buffer is full, that machine becomes blocked and stops processing parts until a buffer slot becomes available (blocking-after-service policy).
6. All machines are reliable and produce no bad parts.
7. No batching and no setup times (single product) are considered.
8. The output process is not necessarily stationary (a steady-state distribution for the output of the system under consideration may not exist).
9. The production line assumes idle and empty initial conditions.

Table 3 The notation

$N_j(t)$:	random variable of the number of parts that have left machine j up to time t in an n-machine transfer line, $j = 1, \ldots n$
l:	number of parts leaving the system at an instance in time
k:	index of the time frame in which the system is
$p^{c,k}$:	number of parts arriving the system under Case c in time frame k
n:	number of machines in the system
λ:	arrival rate to the system
μ_j:	service rate of machine $j, j = 1, \ldots, n$
T_λ^i:	inter-arrival time for part i
$T_{\mu_j}^i$:	service time for part i on machine $j, j = 1, \ldots, n$
b_m:	size of the mth buffer in the n-machine transfer line, $m = 0, 1, \ldots, n$
$f_{\lambda,\mu_1}(t)$:	probability density function of the random variable $T_\lambda + T_{\mu_1}$
$F_{\lambda,\mu_1}(t)$:	cumulative distribution function of the random variable $T_\lambda + T_{\mu_1}$
$f_{\lambda,\mu_1}^l(t)$:	lth convolution of the probability density function for the random variable $T_\lambda + T_{\mu_1}$
$F_{\lambda,\mu_1}^l(t)$:	lth convolution of the cumulative distribution function for the random variable $T_\lambda + T_{\mu_1}$
$E(N_n(t))$:	mean of the throughput, the number of parts leaving the n-machine system at time t
$V(N_n(t))$:	variance of the throughput, number of parts leaving the n-machine system at time t
$E(t)$:	mean of the throughput rate at time t
$V(t)$:	variance of the throughput rate at time t
E:	steady-state mean of the throughput rate
V:	steady-state variance of the throughput rate

systems. The corresponding descriptions help us to write down the representations for the number of parts leaving the system at an arbitrary instant in time. Finally, the relevant representations are utilised to obtain the distribution of throughput by which the mean and variance of the number of parts leaving the system are calculated. The approach is explained in detail after presentation of the stochastic process evolutions of the related systems.

The atomic model

The atomic model corresponds to one-machine system as depicted in Figure 2.

The evolution of the stochastic process of the random variable, $N_1(t)$, is presented on the following page.

The memorylessness property of the inter-arrival times (the Poisson arrival process) and the zero capacity of the first buffer in the line facilitate the representation of the

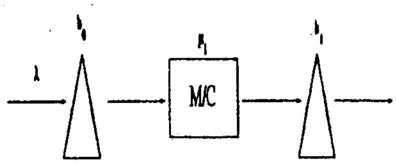

Figure 2 $T_\lambda^i \sim \exp(\lambda)$, $T_{\mu_1}^i \sim \exp(\mu_1)$ $\forall i$; $b_0 = 0$ and $b_1 = \infty$.

stochastic process evolution for the atomic model. Whenever a part leaves the system, say part $(i-1)$, the next one arrives T_λ^i time units later. Hence, the number of parts leaving the system at time t can be written as

$$N_1(t) = \begin{cases} 0 & \text{if } 0 < t < T_\lambda^1 + T_{\mu_1}^1 \\ 1 & \text{if } T_\lambda^1 + T_{\mu_1}^1 < t < \sum_{i=1}^{2}(T_\lambda^i + T_{\mu_1}^i) \\ 2 & \text{if } \sum_{i=1}^{2}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{3}(T_\lambda^i + T_{\mu_1}^i) \\ \vdots & \quad \vdots \qquad\qquad \vdots \\ l-1 & \text{if } \sum_{i=1}^{l-1}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{l}(T_\lambda^i + T_{\mu_1}^i) \\ l & \text{if } \sum_{i=1}^{l}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{l+1}(T_\lambda^i + T_{\mu_1}^i) \\ l+1 & \text{if } \sum_{i=1}^{l+1}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{l+2}(T_\lambda^i + T_{\mu_1}^i) \\ \vdots & \quad \vdots \qquad\qquad \vdots \end{cases} \quad (1)$$

Two-machine-one-buffer system

The system under consideration is illustrated in Figure 4.

Now, there occurs an increase in the number of the sources of variability (T_λ^i, $T_{\mu_1}^i$, and $T_{\mu_2}^i$, $i \geqslant 1$) due to the more number of machines and buffers in the system of interest and this leads to the existence of two mutually exclusive and collectively exhaustive events that describe the behaviour of the system: (i) the part leaving M_1 finds M_2^* busy; and (ii) the part leaving M_1 finds M_2^* starved. While M_1 refers to the first machine itself, M_2^* is not considered only as the second machine of the line, but as the rest of the system consisting of the second buffer (B_1) and the second machine (M_2) of the line. In the former event, the second buffer of the line is at its full capacity and the second machine is processing a part at the time the part on the first machine is ready to leave. Hence, the system gets blocked (blocking-after-service policy). In the latter event, the part leaving the first machine enters the queue and depending on the state of the second machine (idle or busy), it is delayed in the queue for either zero or more number of units. The effect of the buffer capacity on the throughput behaviour is implicitly characterised via the consideration of this case. The two-machine transfer line will be in either of these mutually exclusive and collectively exhaustive cases at different instants during the evolution of the relevant stochastic process. This is why we recommend to make use of the time frames in which the mutually exclusive and collectively exhaustive descriptions of the system behaviour take place for particular number of parts arriving the system in order to keep a better track of the evolution of the throughput. Each time frame is labeled with index k in which the first and second cases are assumed to hold consecutively for $p^{1,k}$ and $p^{2,k}$ parts,

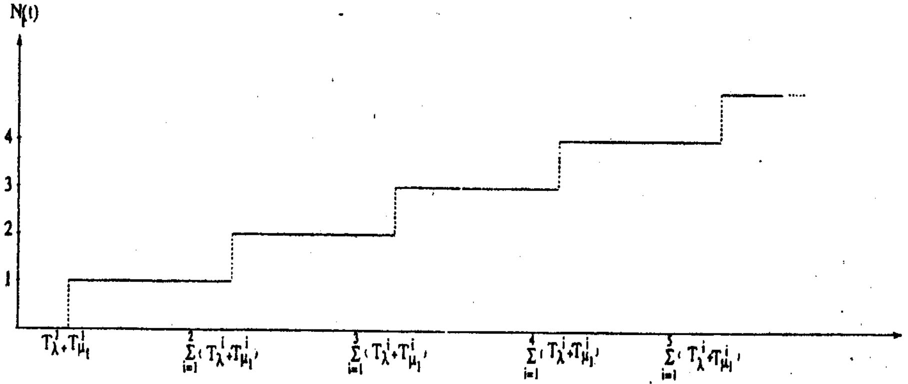

Figure 3 Evolution of the stochastic process, $N_1(t)$.

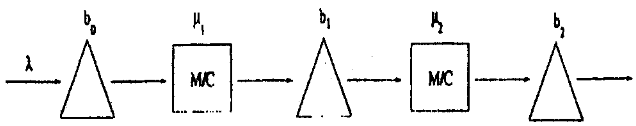

Figure 4 $T_\lambda^i \sim \exp(\lambda)$, $T_{\mu_j}^i \sim \exp(\mu_j)$ $\forall i, j = 1, 2$; $b_0 = 0$, $b_1 \geqslant 0$ and $b_2 = \infty$.

respectively. What happens in time frames helps us to describe the system behaviour and obtain the analytical derivations for the number of parts leaving the system at different instants in time. We must note that the $p^{1,k}$ and $p^{2,k}$ variables are considered to be known in advance by the production manager and, accordingly, the analytical derivations for the number of parts leaving the system, $N_2(t)$, can be modified at any instance in time. The evolution of the stochastic processes and the corresponding analytical derivations for $N_2(t)$ under Case 1, Case 2, and the aggregate case, under which the first case holds for $p^{1,k}$ parts and the second case for $p^{2,k}$ parts in time frame k, for the two-machine transfer line with finite buffer storage, are available in Dinçer and Deler.[2]

Three-machine-two-buffer system

The system under consideration is illustrated in Figure 5.

In this system, there arises four mutually exclusive and collectively exhaustive events that describe the behaviour of the line: (i) the part leaving M_1 finds M_2^* busy and the part leaving M_2^* finds M_3^* busy; (ii) the part leaving M_1 finds M_2^* busy and the part leaving M_2^* finds M_3^* starved; (iii) the part leaving M_1 finds M_2^* starved and the part leaving M_2^* finds M_3^* busy; and (iv) the part leaving M_1 finds M_2^* starved and the part leaving M_2^* finds M_3^* starved. Similarly, M_2^* corresponds to the second buffer (B_1) and the second machine of the line (M_2) while M_3^* represents the third buffer (B_2) and the last machine (M_3) of the line. The behaviour of the first part of the system, which is composed of M_1 and M_2^*, implicitly describes the effects of μ_1, b_1, and μ_2 on the throughput of the line while the consideration of M_3^* helps to characterise the effects of the b_2 and μ_3 on the transient behaviour of the system behaviour. However, none of these mutually exclusive and collectively exhaustive events uniquely represent the true system behaviour by itself. Hence, an aggregate case is considered under which each condition is allowed to be valid for particular number of parts, assumed to be known in advance, in each time frame.

The corresponding analytical derivation for $N_3(t)$ under the aggregate case, in which Case c holds for $p^{c,k}$ parts in each time frame labeled with $k, c = 1, 2, 3, 4$, for three-machine transfer line with finite buffer storages, is also available in Dinçer and Deler.[2]

Approach

In this section, we explain how the analytical derivations for the number of parts leaving the system at an instance in time can be readily utilised to obtain the various moments of the throughput rate. This is shown on the atomic model due to the simplicity in deriving the first and second order moments of the throughput rate of the system.

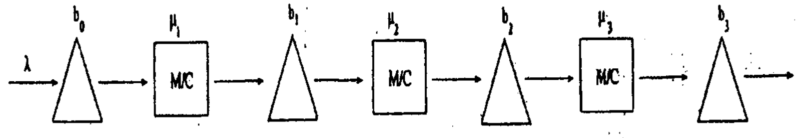

Figure 5 $T_\lambda^i \sim \exp(\lambda)$, $T_{\mu_j}^i \sim \exp(\mu_j)$ $\forall i, j = 1, 2, 3$; $b_0 = 0$, $b_1 \geqslant 0$, $b_2 \geqslant 0$ and $b_3 = \infty$.

The number of parts leaving the system, the atomic model, at time t has been obtained in (1). It is deduced from this statement that,

$$P(0 < t < T_\lambda^1 + T_{\mu_1}^1) = P(N_1(t) = 0), \quad (2)$$

and

$$P\left(\sum_{i=1}^{l}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{l+1}(T_\lambda^i + T_{\mu_1}^i)\right) = P(N_1(t) = l), \quad l = 1, 2, \ldots \quad (3)$$

By definition of the expectation,

$$E(N_1(t)) = \sum_{l=1}^{\infty} l \times P\left(\sum_{i=1}^{l}(T_\lambda^i + T_{\mu_1}^i) < t < \sum_{i=1}^{l+1}(T_\lambda^i + T_{\mu_1}^i)\right). \quad (4)$$

Simultaneously, $P(N_1(t) = l)$, $l = 1, 2, \ldots$ can be written as the difference of the convolution of distribution functions,

$$P(N_1(t) = l) = F_{\lambda,\mu_1}^l(t) - F_{\lambda,\mu_1}^{l+1}(t), \quad (5)$$

If this is substituted into the expression (4), then we determine the mean and the variance of the number of parts leaving the system at time t,

$$E[N_1(t)] = \sum_{l=1}^{\infty} l \times (F_{\lambda,\mu_1}^l(t) - F_{\lambda,\mu_1}^{l+1}(t)), \quad (6)$$

$$E[N_1(t) \times (N_1(t) - 1)] = \sum_{l=2}^{\infty} l \times (l-1) \times (F_{\lambda,\mu_1}^l(t) - F_{\lambda,\mu_1}^{l+1}(t)), \quad (7)$$

and

$$V[N_1(t)] = E[N_1(t) \times (N_1(t) - 1)] + E[N_1(t)] - E[N_1(t)]^2. \quad (8)$$

Then, the mean and the variance of the throughput rate over a length of time t, can be written as,

$$E(t) = E\left[\frac{N_1(t)}{t}\right] = \frac{E[N_1(t)]}{t}, \quad (9)$$

and

$$V(t) = V\left[\frac{N_1(t)}{t}\right] = \frac{V[N_1(t)]}{t^2}. \quad (10)$$

The mean and the variance of the throughput rate in the steady-state can also be calculated as

$$\lim_{t\to\infty} E(t) = E, \quad (11)$$

and

$$\lim_{t\to\infty} V(t) = V. \quad (12)$$

The major difficulty is due to the development of the closed-form expressions for the mean and the variance of $N_1(t)$, which are expressed as in (6) and (8). Therefore, the essential step turns out to be the determination of l-fold convolution of distribution function, $F_{\lambda,\mu_1}^l(t)$. In fact, this is the l-fold convolution of the random variable $T_\lambda^i + T_{\mu_1}^i$, which is hypoexponentially distributed with parameters λ and μ_1, respectively, in case $\lambda \neq \mu_1$. Otherwise, the l-fold convolution of the random variable $T_\lambda^i + T_{\mu_1}^i$ has an Erlang-type distribution. In other words, $T_\lambda^i + T_{\mu_1}^i \sim \text{hypo}(\lambda, \mu_1)$ if $\lambda \neq \mu_1$ and $T_\lambda^i + T_{\mu_1}^i \sim \text{Erlang}_2(\mu_1)$ if $\lambda = \mu_1$. In case the arrival rate is equal to the service rate of the server, $\lambda = \mu_1$, the closed-form expressions for the probability density function and the distribution function can be readily obtained by the definition of the Erlang distribution function.[29] However, although the sum of hypoexponentially distributed random variables is also hypoexponentially distributed with the requirement that all the random variables are independent with different parameters,[30] there is no concrete information about whether the sum of hypoexponentially distributed random variables is also hypoexponential in case the parameters of the random variables are identical. At this point, we recommend to make use of the convolution method by which the closed-form expression of $F_{\lambda,\mu_1}^l(t)$ can be obtained.[29] In this way, the closed-form expressions for the probability density function and the distribution function are obtained as in case $\lambda \neq \mu_1$:

$$f_{\lambda,\mu_1}^l(t) = \begin{cases} -\dfrac{\lambda \times \mu_1 \times (e^{-t\mu_1} - e^{-t\lambda})}{\mu_1 - \lambda} & \text{if } l = 0 \\ \dfrac{\lambda^1 \times \mu_1^l \times e^{-t\mu_1} \times \sum_{i=1}^{l}(-1)^{i+1} \times A_i \times t^{l-i} \times (\lambda - \mu_1)^{l-i}}{(l-1)! \times (\lambda - \mu_1)^{2l-1}} \\ + \dfrac{\lambda^l \times \mu_1^l \times e^{-t\lambda} \times \sum_{i=1}^{l}(-1)^{i+1} \times A_i \times t^{l-i} \times (\mu_1 - \lambda)^{l-i}}{(l-1)! \times (\mu_1 - \lambda)^{2l-1}} & \text{if } l \geqslant 1 \end{cases} \quad (13)$$

and

$$F_{\lambda,\mu_1}^l(t) = \begin{cases} \dfrac{\lambda \times e^{-t\mu_1}}{\mu_1 - \lambda} + \dfrac{\mu_1 \times e^{-t\lambda}}{\lambda - \mu_1} + 1 & \text{if } l = 0 \\ \dfrac{\lambda^l \times \mu_1^l \times \sum_{i=1}^{l}(-1)^{i+1} \times A_i \times g_{l-i}(t) \times (\lambda - \mu_1)^{l-i}}{(l-1)! \times (\lambda - \mu_1)^{2l-1}} \\ + \dfrac{\lambda^l \times \mu_1^l \times \sum_{i=1}^{l}(-1)^{i+1} \times A_i \times h_{l-i}(t) \times (\mu_1 - \lambda)^{l-i}}{(l-1)! \times (\mu_1 - \lambda)^{2l-1}} & \text{if } l \geqslant 1 \end{cases} \quad (14)$$

where $g_{l-i}(t)$ and $h_{l-i}(t)$ are incomplete gamma functions that can be defined as $g_{l-i}(t) = \int_0^t x^{l-i} \times e^{-x\mu_1}\,dx$ and $h_{l-i}(t) = \int_0^t x^{l-i} \times e^{-x\lambda}\,dx$, and also the coefficients A_i are numerically calculated in the Maple V environment by convoluting the relevant density functions.

The substitution of expression (14) in (6) and (7) leads to the determination of the exact values of the mean and the variance of number of parts leaving the system at time t.

Finally, the mean and the variance of the throughput rate at time t and the steady-state mean and variance of the throughput rate by letting time, t, approach infinity, are numerically obtained.

Validation

After the development of the model, an experiment is designed in order to determine whether the model can capture the true transient and steady-state behaviours of the corresponding systems.

State-space representation

For each system, the mean and the variance of the throughput rate obtained by the analytical method are compared with the ones given by the state-space representations developed under exactly the same assumptions for the analytical models (Table 4). The state-space representations correspond to the Markov chain models that are developed via the use of balance equations and that are used in order to obtain the steady-state performance measures of the systems.

The design used for the experiments is the paired comparison design in which the precision can be greatly improved by making comparisons within pairs of experimental results at a significance level of 95%. The results of the experimental design show that there is no evidence to indicate that the two approaches produce statistically significant difference in the estimation of the performance measures, which are the mean and the variance of the throughput rate.

Table 4 Analytical results *vs* state-space representations

System parameters $(b_1, \ldots, b_{n-1}, \lambda, \mu_1, \ldots, \mu_n)$	*Analytical method*		*Markov chain*	
	Mean	*Variance*	*Mean*	*Variance*
The atomic model				
(3,3)	3.000	9.000	3.000	9.000
(1,0,5)	0.499	0.249	0.500	0.250
(0,5,1)	0.499	0.249	0.500	0.250
Two-machine-one-buffer system				
(0,1,3,3)	0.727	1.650	0.727	1.653
(0,1,7,3)	0.834	1.804	0.833	1.804
(2,1,3,3)	0.748	1.686	0.749	1.687
(2,1,7,3)	0.871	1.854	0.873	1.857
(5,1,3,3)	0.748	1.686	0.750	1.687
(5,1,7,3)	0.874	1.856	0.875	1.859
Three-machine-two-buffer system				
(0,0,1,3,3,3)	0.239	1.832	0.241	1.833
(0,0,1,6,4,3)	0.274	2.000	0.275	1.999
(2,2,1,3,3,3)	0.282	0.128	0.285	0.131
(2,2,1,6,4,3)	0.392	0.164	0.388	0.163
(2,5,1,6,4,3)	0.401	0.207	0.398	0.205
(5,2,1,6,4,3)	0.408	0.209	0.409	0.211
(5,5,1,6,4,3)	0.412	0.214	0.412	0.213

Simulation

The simulation analysis helps to determine whether the model operates appropriately in the transient state. The codes to simulate the systems developed under exactly the same assumptions for the analytical models, are written in the SIMAN simulation language.[31] While obtaining the numerical results for the mean and variance of throughput at arbitrary instants in time via the use of simulation models, the following formulas are used:

$$E(N_n(t)) = \sum_{y=1}^{500} \frac{N_n^y(t)}{500}$$

and

$$\mathrm{Var}(N_n(t)) = \sum_{y=1}^{500} \frac{N_n^y(t) - E(X))^2}{499},$$

where $N_n^y(t)$ corresponds to the number of parts leaving the n-machine system at time t in the yth replication. These formulas are basically the definitions of the sample mean and the sample variance that are available in the reference of Montgomery.[32] Then, the mean and variance of throughput rate of the n-machine transfer line can be numerically calculated by

$$E(t) = E\left[\frac{N_n(t)}{t}\right] = \frac{E[N_n(t)]}{t}$$

and

$$V(t) = V\left[\frac{N_n(t)}{t}\right] = \frac{V[N_n(t)]}{t^2},$$

respectively, at time t.

It is observed that the pattern which analytical results follow is smoother than the one of the simulation results (Table 5). This is due to the random number generator that is one of the main mechanisms of the simulation software.

The simulation analysis indicates that the proposed method works fairly well in reflecting the transient behaviour of the systems under consideration (Table 5). This is confirmed by the paired comparison design that is done at a significance level of 95% with the result that the analytical and simulation based approaches do not produce statistically significant difference in the estimation of mean and variance of the throughput rate.

Conclusions and new directions for future research

In this paper we propose an analytical method for estimating the mean and variance of the throughput of a serial line production system with reliable machines and finite buffers. The transient and steady-state behaviours of the system are determined by using the evolution of the stochastic processes under consideration. The analysis of the evolution of stochastic processes enabled us to derive the

Table 5 Analytical results *vs* simulation

System parameters $(b_1, \ldots, b_{n-1}, \lambda, \mu_1, \ldots, \mu_n, t)$	*Analytical method*		*Simulation*	
	Mean	*Variance*	*Mean*	*Variance*
The atomic model				
(2,20,2)	1.802	1.422	1.809	1.431
(2,20,5)	1.816	1.434	1.824	1.438
(2,20,10)	1.882	1.496	1.885	1.501
(2,20,25)	1.844	1.450	1.846	1.451
(2,20,50)	1.837	1.451	1.838	1.449
(2,20,100)	1.828	1.449	1.827	1.449
Two-machine-one-buffer system				
(5,1,3,3,2)	0.869	0.318	0.872	0.320
(5,1,3,3,5)	0.801	0.267	0.794	0.270
(5,1,3,3,10)	0.772	0.242	0.776	0.249
(5,1,3,3,25)	0.761	0.238	0.762	0.239
(5,1,3,3,50)	0.755	0.235	0.752	0.238
(5,1,3,3,100)	0.753	0.237	0.752	0.236
Three-machine-two-buffer system				
(2,5,20,6,4,3,2)	2.012	1.364	1.767	1.361
(2,5,20,6,4,3,5)	2.296	1.492	2.287	1.467
(2,5,20,6,4,3,10)	2.511	1.668	2.509	1.666
(2,5,20,6,4,3,25)	2.674	1.692	2.666	1.678
(2,5,20,6,4,3,50)	2.703	1.715	2.699	1.716
(2,5,20,6,4,3,100)	2.713	1.728	2.713	1.727

distribution of the throughput. This distribution is then utilised to find the higher order moments of the throughput rate for measuring the performance of the system. The method based on the analytical derivation of the distribution provides correct results for typical transfer line models encountered in real applications. Moreover, the iterative algorithms coded in Maple V environment seem to be efficient: in all examples we test, it always converges and, in general, very rapidly. Furthermore, an experiment is designed in order to compare the analytical models with the state-space representations and simulation models in which the results verified the accuracy of the methodology developed in this paper.

This research should be extended to develop analytical methods for the analysis of more complicated systems such as longer transfer lines with non-exponential distributions and multiple-part types. Furthermore, line design issues, pull-type systems, and general networks can be studied by making use of this method.

References

1 Gershwin SB (1994). *Manufacturing Systems Engineering*. Prentice-Hall: New Jersey.

2 Dinçer C and Deler B (1997). *On the distribution of the throughput rate of transfer lines*. Research Report IEOR-9713, August, Bilkent University, Ankara.

3 Lavenyerg SS (1975). The steady-state queueing-time distribution for the $M/G/1$ finite capacity queue. *Mgmt Sci* **21**: 501–506.

4 Miltenburg GJ (1987). Variance of the number of units produced on a transfer line with buffer inventories during a period of length t. *Naval Res Logist* **34**: 811–822.

5 Ou J and Gershwin SB (1989). *The variance of the lead time distribution of a two-machine transfer line with a finite buffer*. Technical Report LMP-89-028. MIT Laboratory for Manufacturing and Productivity, USA.

6 Hendricks KB (1992). The output processes of serial production lines of exponential machines with finite buffers. *Opns Res* **40**: 1139–1147.

7 Heavey JM, Papadopoulos HT and Browne J (1993). The throughput rate of multi-station unreliable production lines. *Eur J Opl Res* **68**: 69–89.

8 Papadopoulos HT (1993). The throughput of multi-station production lines with no intermediate buffers. *Opns Res* **43**: 712–715.

9 Papadopoulos HT and O'Kelly MEJ (1993). Exact analysis of production lines with no intermediate buffers. *Eur J Opl Res* **65**: 118–137.

10 Tan B (1996). *Variance of the throughput of a n-station production line with no intermediate buffers and time-dependent failures*. Internal Report, Graduate School of Business, Koc University.

11 Kelton WD and Law AM (1985). The transient behavior of the $M/M/n$ queue, with implications for steady-state simulation. *Opns Res* **33**: 378–396.

12 Klutke GA and Seiford M (1991). Transient behavior of finite capacity tandem queues with blocking. *Int J Sys Sci* **22**: 2205–2215.

13 Gopalan MN and Dinesh Kumar U (1992). Stochastic analysis of a two-stage production system with n parallel stations in the first stage. *Int J Mgmt Sys* **8**: 263–275.

14 Gopalan MN and Dinesh Kumar U (1994). On the transient behavior of a merge production system with an end buffer. *Int J Prod Econ* **34**: 157–165.

15 Hong Y, Glassey CR and Seong D (1992). The analysis of a production line with unreliable machines and random processing times. *IIE Trans* **24**: 77–83.

16 Gershwin SB (1992). Variance of output of a tandem production system. In: Onvural R and Akyildiz I (eds). *Proceedings of the Second International Conference on Queuing Networks with Finite Capacity*, Elsevier: Amsterdam.

17 Glassey CR and Hong Y (1993). Analysis of behavior of an unreliable n-stage transfer line with $(n-1)$ inter-stage storage buffers. *Int J Prod Res* **31**: 519–530.

18 Springer MC (1994). A decomposition approximation for finite-buffered flow lines of exponential queues. *Eur J Opl Res* **74**: 95–110.

19 Wu Q (1994). *Variance of output of transfer lines with finite buffer inventories*. Department of Operations Research, Stanford University.

20 Papadopoulos HT (1996). An analytic formula for the mean throughput of n-station production lines with no intermediate buffers. *Eur J Opl Res* **91**: 481–494.

21 Frein Y, Commault C and Dallery Y (1996). Modeling and analysis of closed-loop production lines with unreliable machines and finite buffers. *IIE Trans*. **28**: 545–554.

22 Hendricks KB and McClain JO (1993). The output processes of serial production lines of general machines with finite buffers. *Mgmt Sci* **39**: 1194–1201.

23 Lin L and Cochran JK (1990). Meta-models of production line transient behavior for sudden machine breakdowns. *Int J Prod Res* **28**: 1791–1806.

24 Gershwin SB (1987). An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Opns Res* **35**: 291–305.

25 Dallery Y and Gershwin SB (1992). Manufacturing flow line systems: A review of models and analytical results. *Queueing Sys* **12**: 3–94.
26 Buzacott JA and Shanthikumar JG (1992). Design of manufacturing systems using queueing models. *Queueing Sys* **12**: 135–214.
27 Papadopoulos HT and Heavey JM (1996). Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines. *Eur J Opl Res* **92**: 1–27.
28 Deler B and Dinçer C. (1997). *Transient analysis of serial line production systems: a critical review.* Research Report IEOR-9712, August, Bilkent University, Ankara.
29 Rohatgi VK (1976). *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons: New York.
30 Trivedi KS (1982). *Probability and Statistics with Reliability, Queueing and Computer Science Applications*. Prentice Hall, Inc: Englewood Cliffs, NJ.
31 Pegden CD, Shannon RE and Sadowski RP (1996). *Introduction to Simulation using SIMAN*. McGraw-Hill: New York.
32 Montgomery DC (1991). *Design and Analysis of Experiments*. John Wiley & Sons: New York.

Received September 1997; accepted July 1999 after two revisions