International Journal of Production Research,
Vol. 44, No. 7, 1 April 2006, 1375–1398

Taylor & Francis
Taylor & Francis Group

Comparisons of two-machine line models in throughput analysis

JINGSHAN LI*, DENNIS E. BLUMENFELD
and JEFFREY M. ALDEN

Manufacturing Systems Research Laboratory, General Motors Research &
Development Center, Mail Code 480-106-359, 30500 Mound Road,
Warren, MI 48090-9055, USA

(*Received September 2005*)

Throughput analysis is important for the design, operation and management of manufacturing systems. Due to the large state space, exact analytical results only exist for two-machine serial lines. To analyse longer lines or more complex systems, the two-machine line model becomes the building block for further development. Therefore, many two-machine line models have been presented in the literature. A thorough understanding of the nature and differences of the various two-machine line models is critical for future analysis and development. In this paper, we study eight different two-machine models, categorized by synchronous and asynchronous lines with both time-dependent and operation-dependent failures, respectively. For each model, we introduce and compare the assumptions, calculation formulae, and their performance in system throughput. The results show that all the models exhibit similar performance with small differences, comparable to (or less than) the accuracy of data collection on the factory floor.

Keywords: Throughput analysis; Two-machine line models

1. Introduction

1.1 *Motivation*

To design, improve and manage production systems, an accurate estimate of system performance is necessary and important. Typically, the system performance in production volume is characterized by line throughput, i.e., the average number of parts produced by the last machine per unit of time (also known as production rate). Extensive studies of line throughput have been conducted during the last 50 years (see reviews by Dallery and Gershwin (1992) and Papadopoulos and Heavey (1996), and monographs by Buzacott and Shanthikumar (1993), Gershwin (1994) and Altiok (1997)). However, due to the large state space, exact analytical results only exist for the simplest system, the two machine–one buffer system. To analyse the performance of longer lines, various decomposition and aggregation methods are proposed to approximate the system throughput (see, for instance, Buzacott (1967, 1971), Gershwin (1987), De Koster (1987), Dallery *et al.* (1988), Lim *et al.* (1990), Jacobs

*Corresponding author. Email: jingshan.li@gm.com

International Journal of Production Research
ISSN 0020–7543 print/ISSN 1366–588X online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540500371980

and Meerkov (1995), Chiang *et al.* (2000, 2001), Le Bihan and Dallery (2000) and Li and Meerkov (2003)). Longer lines are either decomposed into several 'equivalent two-machine lines' or are formed by aggregating the two-machine line into an 'equivalent single machine' recursively, and the results from two-machine models are used as a building block to estimate the system throughput. In addition, the analysis of more complex systems (for instance, assembly lines, parallel lines, rework loops, reentrant lines, etc.) also relies on the results of two-machine line model. Therefore, the two-machine line model plays a key role in throughput analysis.

In the production system literature, various two-machine models have been presented and then used for longer line (or more complex system) analysis. However, how these models are applicable to different systems, what are the major differences among them and what are the impacts, how do they differ in system performance, what is the common nature of all the models, etc., all these questions have not been answered in the current literature. Therefore, in order to answer these questions and provide a solid base for further analysis and development, a thorough understanding and comparison of the available results on two-machine line models are necessary and important. In addition, it is helpful for us to investigate the nature of production systems. The goal of this paper is to provide a brief description and comparison of several two-machine line models.

1.2 *System and model description*

In this paper, we limit our study to the system with deterministic processing time (i.e., cycle time), unreliable machines and finite buffers. Many large volume manufacturing systems fall into this category, in particular systems with automatic operations. If the machines have the same cycle time and the states of the machines change only at the beginning (or the end) of the cycle, then the line is referred to as a synchronous line in this paper, otherwise it is referred to as an asynchronous line. (In some literature, synchronous and asynchronous line models are also referred to as discrete (or deterministic) and continuous models, respectively.) Typically, in systems with automated material handling (e.g., transfer line), synchronous lines are common. In systems with machining operations, most lines are not synchronized. In this paper, we examine both synchronous and asynchronous line models.

Machine breakdowns can be characterized by two failure models: time-dependent failure (TDF) and operation-dependent failure (ODF). In the time-dependent failure model, a machine breakdown depends only on the amount of time since the last repair, i.e., a machine can be down when it is blocked or starved; whereas in the operation-dependent failure model, a machine breakdown depends only on the number of operations that have been performed since the last repair, in other words, a machine will not fail if it is idle. Empirical studies show that, of all the machine breakdowns, operation-dependent failure occurs more often than time-dependent failure. However, both models are applicable (see Buzacott and Hanifin (1978) and Buzacott and Shanthikumar (1993) for a comparison between the TDF and ODF models based on a comprehensive study of downtime data). For instance, power failures are time dependent, while tool breaks are operation dependent.

By assuming the TDF model, the analysis is simplified. Closed form expressions can be obtained and some system theoretic properties can be proved. In addition, with closed form expressions, it is easier to provide fundamental properties of production systems with longer lines and guidelines for continuous improvements. For instance, through forward and backward aggregations (Jacobs 1993, Jacobs and Meerkov 1995, Chiang *et al*. 2000, 2001, Li and Meerkov 2003), the convergence of the iteration procedure and some system-theoretic properties, such as reversibility, monotonicity, improvability, bowl phenomenon, etc., can be proved analytically. Therefore, the study of throughput analysis with the TDF model is also necessary and important. In this paper, we discuss both the TDF and ODF models and compare their differences.

Machine uptimes and downtimes are assumed to follow a certain type of distribution. In the continuous model (asynchronous lines), an exponential distribution is a common assumption due to its 'memoryless' property (see Inman 1999) for an empirical evaluation of the exponential assumptions in queueing models of manufacturing systems). Correspondingly, the geometric distribution is assumed in discrete models (synchronous lines). When machine downtime is short and comparable to its cycle time, the Bernoulli model can be applied. In this paper, we will concentrate on the study of systems with exponential, geometric, and Bernoulli types of machine reliability.

In addition to these models, other reliability models exist, for instance machines with Erlang, phase type, or coxian type up- and downtimes (e.g., Altiok (1985) and Buzacott and Shanthikumar (1993)). However, they are difficult to apply to the study of longer lines. As an alternative, empirical laws or simulations are used to evaluate system performance (e.g., Enginarlar *et al*. (2002) and Li and Meerkov (2005)). These models are not discussed in this paper.

We limit our discussion to the models illustrated in figure 1. Eight two-machine line models are classified by synchronous and asynchronous lines with operation-dependent and time-dependent failures. In particular, synchronous line models are studied by Buzacott and Shanthikumar (TDF and ODF), Gershwin (ODF), Jacobs and Meerkov (TDF) and Li and Meerkov (TDF), whereas asynchronous line models are investigated by Alden (ODF), Gershwin (ODF) and Jacobs and Meerkov (TDF).

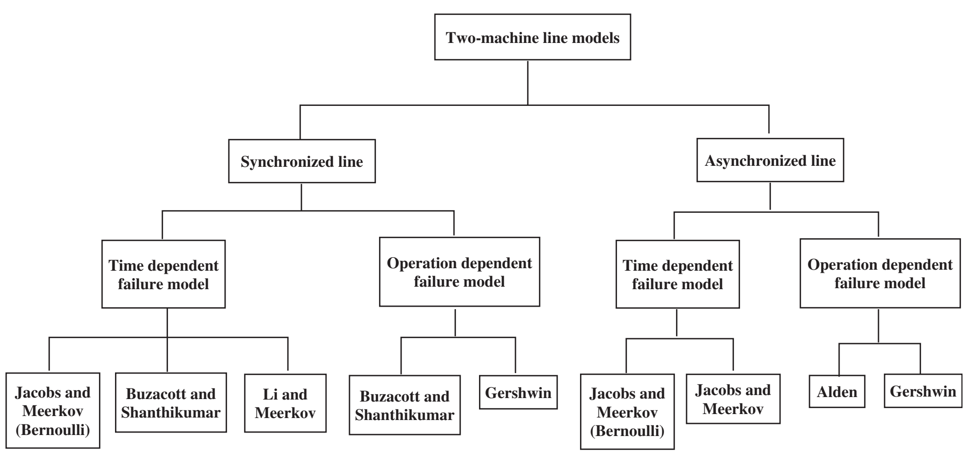

Figure 1. Taxonomy of two-machine line analysis models.

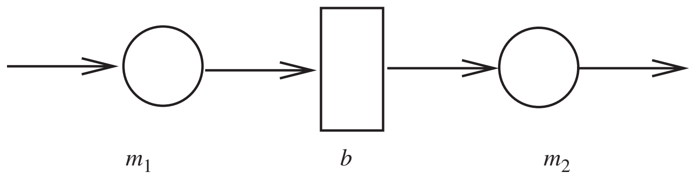

Figure 2. Two-machine line with two machines m_1 and m_2 separated by a buffer b.

The main conclusion obtained from this paper is that all the models discussed here exhibit similar system performance. When the buffer is small, the differences in system throughput are of the same order as (or less than) the accuracy of data collection on the factory floor. When the buffer is large, the differences are indistinguishable.

The remainder of this paper is structured as follows. Subsection 1.3 introduces the common assumptions of all models discussed in this paper. Sections 2 and 3 investigate synchronous and asynchronous line models, respectively. A comparison of the models discussed is presented in section 4. Section 5 formulates the conclusions.

Due to space limitation, we concentrate only on comparing their assumptions, calculation formulae, and throughput estimates. Details of each model can be found in the corresponding literature referred to in this paper.

1.3 *General assumptions*

The assumptions defined below are common to all the models discussed in this paper. Additional specific assumptions for each particular model are introduced in subsequent sections.

1. The system consists of two machines arranged serially and a buffer separating the machines (see figure 2, where circles represent the machines and the rectangle is the buffer).
2. Each machine (m_1 and m_2) has two states: up and down. When up, the machine is capable of producing parts; when down, no production takes place. Machine uptimes and downtimes are independent random variables.
3. Each machine requires a fixed unit of time to process a part. This unit is referred to as the cycle time.
4. The buffer is characterized by its capacity N, $0 < N < \infty$.
5. Machine m_1 is never starved; machine m_2 is never blocked.
6. If a machine fails while processing a part, the part remains at the machine, and repair will include any necessary operations needed to complete processing of the part. No scrap is assumed.
7. The buffer does not delay job movement, i.e. the buffer transition time is zero.

2. Synchronous lines

In this section, the synchronous line models of Buzacott and Shanthikumar, Gershwin, Jacobs and Meerkov, and Li and Meerkov are introduced

(Buzacott and Shanthikumar 1993, Jacobs 1993, Gershwin 1994, Jacobs and Meerkov 1995, Li 2000, Li and Meerkov 2003). We begin with the simplest, the Bernoulli reliability model, and then extend to Markovian reliability models.

In synchronous line models, the time axis is slotted with the slot duration equal to the cycle time, and all machines begin their operations at the beginning of each cycle. The up- and downtimes are assumed to be either Bernoulli or geometric. Specifically,

(i) all machines have identical cycle times to process a part (one part per cycle),
(ii) machine m_2 is starved during the cycle if the buffer is empty, and
(iii) in the time-dependent failure model, the machine can experience failures even if it is blocked or starved, whereas in the operation-dependent failure model, the machine cannot fail if it is idle.

Below, synchronous line models with both time-dependent and operation-dependent failures are introduced.

2.1 *Time-dependent failure models*

2.1.1 Bernoulli model by Jacobs and Meerkov. When the machine downtime is quite short and comparable to the machine cycle time, it can be viewed that, at each cycle, the machine has an independent probability of being up or down, analogous to a Bernoulli trial. Therefore, Jacobs and Meerkov introduced a simple machine reliability model, the so-called Bernoulli reliability model (Jacobs 1993, Jacobs and Meerkov 1995).

The Bernoulli model is easier for analysis, and it enables us to investigate the nature of production systems (see, for instance, Jacobs (1993), Jacobs and Meerkov (1995), Kuo (1996) and Li (2000)). It is appropriate for production lines where the duration of the 'breakdowns' is of the order of the cycle time and, therefore, it can be viewed that each cycle has a probability of producing a part or not. In such cases, the typical breakdowns are push-button stops, pallet jams, etc.; to fix such problems, only a short period of time is required. In many assembly lines these are the majority of the breakdowns. In these situations, the Bernoulli reliability model is appropriate. In other cases, the Markovian (geometric or exponential) models are more applicable.

In addition to the general assumptions, the following assumptions for the Bernoulli model are introduced.

- The state of the machine in each cycle time is determined by the process of Bernoulli trials. In other words, it is assumed that, during each slot, machine m_i, $i = 1, 2$, is up with probability p_i and down with probability $1 - p_i$; the state of the machine is determined at the beginning of each cycle, independent of the state of this machine in the previous cycle.
- If the buffer is full at the beginning of the time slot and m_2 does not take a part from the buffer at the beginning of this slot, then m_1 is blocked during this time slot.

Under these assumptions, the system throughput (TP) can be calculated as

$$TP = p_1[1 - Q(p_2, p_1, N)] = p_2[1 - Q(p_1, p_2, N)], \tag{1}$$

where

$$Q(p_1, p_2, N) = \begin{cases} \dfrac{(1-p_1)(1-\alpha)}{1-(p_1/p_2)\alpha^N}, & \text{if } p_1 \neq p_2, \\ \dfrac{1-p}{N+1-p}, & \text{if } p_1 = p_2 = p, \end{cases} \tag{2}$$

$$\alpha = \frac{p_1(1-p_2)}{p_2(1-p_1)}. \tag{3}$$

When both machines are identical, i.e., $p_1 = p_2 = p$,

$$TP = \frac{Np}{N+1-p}. \tag{4}$$

Remark 1: It is observed that the buffer capacity in the Bernoulli model is approximately equivalent to the number of machine downtimes that the buffer can accommodate.

Remark 2: Function $Q(\cdot)$ in the Bernoulli model (and similar Q functions in geometric and exponential TDF models) exhibits monotonic properties and plays an important role in longer line (and more complex system) analysis to prove the procedure convergence, system monotonicity, reversibility, improvability and other properties. Moreover, most of the properties proved in the Bernoulli model can apply to other models. Therefore, the Bernoulli model helps us to understand the nature of production systems.

2.1.2 Geometric model by Buzacott and Shanthikumar. The Buzacott–Shanthikumar model (Buzacott 1967, 1971, Buzacott and Shanthikumar 1993) is based on the use of the Markov chain. It was first presented in the 1960s and has been widely applied.

The following specific assumptions address the Buzacott–Shanthikumar model.

- Parameter λ_i is the probability of failure of machine i in a cycle, and μ_i is the probability of completing the repair of machine i in a cycle, $i = 1, 2$. The up- and downtimes are assumed to be geometric.
- If the buffer is full at the beginning of the time slot, then m_1 is blocked during this time slot.
- The change in inventory level is determined by the state at the beginning of the interval.

Under the above assumptions, the system throughput TP is

$$TP = \begin{cases} \dfrac{1-A\sigma^N}{1+x_1-(1+x_2)A\sigma^N}, & \text{if } \lambda_1\mu_2 \neq \lambda_2\mu_1, \\ \dfrac{\lambda_1+\lambda_2+\mu_1+\mu_2-(\lambda_1+\mu_1)(\lambda_2+\mu_2)+(\lambda_1+\mu_1)(\lambda_2+\mu_2)N(1+x_1)}{\left\{ \begin{array}{c} (\lambda_1+\lambda_2+\mu_1+\mu_2)(1+x_1)+(\lambda_1+\mu_1)(\lambda_2+\mu_2) \\ \times\{[(\mu_1+\mu_2)x_1/(\mu_1\mu_2)]+(N-1)(1+x_1)^2\} \end{array} \right\}}, & \text{if } \lambda_1\mu_2 = \lambda_2\mu_1, \end{cases} \tag{5}$$

where

$$
\begin{aligned}
A &= \frac{\mu_1\lambda_2(1+x_1)}{\lambda_1\mu_2(1+x_2)},\\
x_i &= \frac{\lambda_i}{\mu_i}, \quad i = 1, 2,\\
\sigma &= \frac{\alpha_2\beta_1}{\alpha_1\beta_2},\\
\alpha_1 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_1\lambda_2,\\
\alpha_2 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_2\lambda_1,\\
\beta_1 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_1\mu_2,\\
\beta_2 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_2\mu_1.
\end{aligned}
\tag{6}
$$

If both machines are identical, i.e., $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$, then

$$
TP = \frac{2 - \mu(1+x) + \mu N(1+x)^2}{2(1+x)^2 + \mu(N-1)(1+x)^3}, \tag{7}
$$

where

$$
x = \frac{\lambda}{\mu}.
$$

This notation is used in subsequent sections of this paper.

Remark 3: In this model, machine m_2 can still produce a part during a time slot if both machines are up and the buffer is empty at the beginning of the time slot. In other words, if the buffer is empty, machine m_2 is up (or from down to up) and machine m_1 is from down to up at the beginning of the time slot, however both machines m_1 and m_2 are still able to produce a part during this time slot. This assumption may lead to an overestimate of system throughput. However, when the buffer is large, the impact of this assumption in system throughput is negligible.

2.1.3 Geometric model by Li and Meerkov. To slightly modify the assumption on the empty buffer case (see remark 3) in the Buzacott–Shanthikumar model, Li and Meerkov introduced the geometric model (Li 2000, Li and Meerkov 2003). In particular, it is assumed that:

- the uptime and downtime of machine i are distributed geometrically with parameters λ_i and μ_i, respectively;
- if the buffer is full at the beginning of a time slot and m_2 does not take a part from the buffer at the beginning of this slot, then m_1 is blocked during this time slot;
- the status of the machine is determined at the beginning of each cycle and the change of the buffer level determined at the end of each cycle.

The system throughput can be calculated as

$$
TP = E_2[1 - Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N)] = E_1[1 - Q(\lambda_2, \mu_2, \lambda_1, \mu_1, N)], \tag{8}
$$

and

$$E_i = \frac{\mu_i}{\lambda_i + \mu_i}, \quad i = 1, 2,$$

$$Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N) = \begin{cases} \dfrac{\lambda_1 \beta_2}{(\mu_1 + \mu_2 - \mu_1\mu_2)(\mu_1 + \lambda_1)}, & \text{if } N = 1, \\ \dfrac{\lambda_1 \alpha_1 \alpha_2 \beta_2^2 (\mu_2 + \lambda_2)}{\Phi + \Psi + \Gamma + \Delta}, & \text{if } N > 1, \end{cases} \tag{9}$$

where

$$\alpha_1 = \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_1\lambda_2,$$

$$\alpha_2 = \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_2\lambda_1,$$

$$\beta_1 = \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_1\mu_2,$$

$$\beta_2 = \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_2\mu_1,$$

$$\sigma = \frac{\alpha_2\beta_1}{\alpha_1\beta_2},$$

$$\Phi = \lambda_1\mu_2\alpha_1\alpha_2\beta_2(\lambda_2 + \beta_2),$$

$$\Psi = \lambda_1\mu_1\mu_2\alpha_2\left[\beta_2^2 + \lambda_2(\alpha_1 + \beta_2)(\alpha_2 + 2\beta_2)\right],$$

$$\Gamma = \sum_{k=2}^{N-1} \lambda_1\lambda_2\mu_1\mu_2(\alpha_2 + \beta_2)^3 \sigma^{k-1},$$

$$\Delta = \lambda_2\mu_1\alpha_1\beta_2[\mu_2(\alpha_1 + \beta_1) + \alpha_2(\lambda_1 + \mu_1)]\sigma^{N-1}. \tag{10}$$

In the identical machine case, i.e., $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$, for any $N > 0$,

$$TP = \frac{2 - \mu + \mu x^2 + \mu(N-1)(1+x)^2}{(1+x)^2[2 - \mu + \mu(N-1)(1+x)]}. \tag{11}$$

Remark 4: In this model, machine m_2 is starved during a time slot if both machines are up and the buffer is empty at the beginning of this time slot. This assumption leads to a lower estimate of system throughput compared with the Buzacott–Shanthikumar model, especially when $N = 1$. When $N > 1$, the difference is very small.

Remark 5: When $\lambda + \mu = 1$, this model is equivalent to the Bernoulli model with $p = \mu$.

2.2 *Operation-dependent models*

2.2.1 Geometric model by Buzacott and Shanthikumar. With a slight modification of the failure mode of the Buzacott–Shanthikumar TDF model, the ODF model for synchronous lines is also available (Buzacott and Shanthikumar 1993).

The assumptions are kept the same as in the Buzacott–Shanthikumar TDF model except that now machines follow the operation-dependent failure mode, i.e., machines cannot fail when starved or blocked. Now the system throughput is

$$TP = \begin{cases} \dfrac{1 - A\sigma^N}{1 + x_1 - (1 + x_2)A\sigma^N}, & \text{if } \lambda_1\mu_2 \neq \lambda_2\mu_1, \\ \dfrac{\mu_1 + \mu_2 - \mu_1\mu_2 + \mu_1\mu_2 N(1 + x_1)}{(\mu_1 + \mu_2)(1 + 2x_1) + \mu_1\mu_2(N - 1)(1 + x_1)^2}, & \text{if } \lambda_1\mu_2 = \lambda_2\mu_1, \end{cases} \tag{12}$$

where

$$\begin{aligned} A &= \frac{\lambda_2\mu_1\beta_1}{\lambda_1\mu_2\beta_2}, \\ x_i &= \frac{\lambda_i}{\mu_i}, \quad i = 1, 2, \\ \sigma &= \frac{\alpha_2\beta_1}{\alpha_1\beta_2}, \\ \alpha_1 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_1\lambda_2, \\ \alpha_2 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_2\lambda_1, \\ \beta_1 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_1\mu_2, \\ \beta_2 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \lambda_2\mu_1. \end{aligned} \tag{13}$$

When both machines are identical, i.e. $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$, then

$$TP = \frac{2 - \mu + N\mu(1 + x)}{2(1 + 2x) + \mu(N - 1)(1 + x)^2}. \tag{14}$$

Remark 6: It is shown in Buzacott and Shanthikumar (1993) that the ODF model results in a slightly higher throughput estimate than the TDF model. However, the difference is very small (less than 3% when $N = 1$ and indistinguishable if N is large). Similar results are obtained in section 4.

2.2.2 Geometric model by Gershwin. Another operation-dependent failure synchronous model is presented by Gershwin (1994), and is also referred to as the deterministic model. In particular, it assumes:

- Machines are assumed to have geometrically distributed times between failures and times to repair. Thus, the mean operating time between failures is $1/\lambda_i$ and the mean time to repair is $1/\mu_i$.
- Machine m_1 is blocked if the buffer is full.
- Repair and failures occur at the beginning of time units, and a change in buffer levels take place at the end of the time units.

The system throughput is

$$TP = \frac{\mu_2}{\lambda_2 + \mu_2} - CZ\frac{\mu_1 + \mu_2 - \mu_1\mu_2 - \mu_1\lambda_2}{\mu_1\lambda_2}, \tag{15}$$

where

$$\begin{aligned}
\alpha_1 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \mu_1\lambda_2,\\
\alpha_2 &= \mu_1 + \mu_2 - \mu_1\mu_2 - \mu_2\lambda_1,\\
\beta_1 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_1\lambda_2,\\
\beta_2 &= \lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \mu_2\lambda_1,\\
Y_1 &= \frac{\alpha_1}{\beta_2},\\
Y_2 &= \frac{\alpha_2}{\beta_1},\\
Z &= \frac{Y_2}{Y_1},
\end{aligned} \tag{16}$$

and C, Y_1 and Y_2 are calculated from normalization and balance equations (see Gershwin (1994) for details).

For the identical machine case (i.e., $\mu_1 = \mu_2 = \mu$, $\lambda_1 = \lambda_2 = \lambda$), we can show that

$$TP = \frac{2(1 + x - x^2) + \mu(N - 3 + \mu)(1 + x)^2}{(1 + x)\left[2(1 + 2x) + \mu(N - 3)(1 + x)^2\right]}. \tag{17}$$

Remark 7: Unlike the other synchronous models discussed previously, this model does not have a closed form solution for system throughput, and only exists for lines in which $N \geq 4$.

3. Asynchronous models

Synchronous models are only applicable to well-balanced transfer lines. However, many production lines are unsynchronized. Therefore, the study of asynchronous models is important. Among all the asynchronous models, continuous models are assumed such that parts flow into the system (except for the Bernoulli model where asynchronous lines are studied through the M–B transformation; see subsection 3.1.2 for details). In particular, it is typically assumed that:

(i) When machine m_i is up, it is capable of producing parts at rate S_i parts per cycle (i.e., cycle time $1/S_i$); when m_i is down, no production takes place.

(ii) The uptime and downtime of each machine m_i are random variables distributed exponentially with parameters λ_i and μ_i, respectively. In other words, machine up and down can occur during the time interval $(t, t + \delta t)$ with probability approximately $\mu_i\delta t$ or $\lambda_i\delta t$, respectively, for small δt.

(iii) The machine cannot fail if it is blocked or starved in the operation-dependent failure model, whereas no such restriction exists in the time-dependent failure model.

Below we discuss both the time-dependent failure and operation-dependent failure models.

3.1 *Time-dependent failure models*

3.1.1 Exponential model by Jacobs and Meerkov. The following assumptions define the machines and the interactions between machines and buffers (Jacobs 1993, Chiang *et al.* 2000, 2001):

- machine m_2 is starved at time t if the buffer is empty and m_1 fails to put a part into the buffer at time t;
- machine m_1 is blocked at time t if the buffer is full and m_2 fails to take a part from the buffer at time t.

The system throughput can be calculated as follows.

Case 1: $S_1 = S_2 = S$,

$$TP = E_2[1 - Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N)]S = E_1[1 - Q(\lambda_2, \mu_2, \lambda_1, \mu_1, N)]S, \tag{18}$$

where

$$Q(\lambda_1, \mu_1, \lambda_2, \mu_2, N) = \begin{cases} \dfrac{(1-E_1)(1-\phi)}{1-\phi e^{-\beta N}}, & \text{if } \dfrac{\lambda_1}{\mu_1} \neq \dfrac{\lambda_2}{\mu_2}, \\ \dfrac{\lambda_1(\lambda_1+\lambda_2)(\mu_1+\mu_2)}{(\lambda_1+\mu_1)[(\lambda_1+\lambda_2)(\mu_1+\mu_2)+\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)N]}, & \text{if } \dfrac{\lambda_1}{\mu_1} = \dfrac{\lambda_2}{\mu_2}, \end{cases}$$

$$E_i = \frac{\mu_i}{\lambda_i + \mu_i}, \quad i = 1, 2, \tag{19}$$

and

$$\phi = \frac{E_1(1-E_2)}{E_2(1-E_1)},$$
$$\beta = \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_1\mu_2 - \lambda_2\mu_1)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}. \tag{20}$$

Case 2: $S_1 < S_2$,

$$TP = \frac{S_2E_2\alpha e^{k_1N} + S_1E_1\beta e^{k_2N} + S_1E_1\gamma e^{-k_2N}}{\alpha e^{k_1N} + \beta e^{k_2N} + \gamma e^{-k_2N}}, \tag{21}$$

where

$$E_i = \frac{\mu_i}{\lambda_i+\mu_i}, \quad i = 1, 2,$$
$$k_1 = \frac{1}{2S_1S_2(\mu_1+\mu_2)(S_1-S_2)}\Big(\mu_1S_1^2(\mu_1+\mu_2+\lambda_2) - S_1S_2[(\mu_1+\mu_2)^2 + (\mu_1+\mu_2)(\lambda_1+\lambda_2) - (\mu_1\lambda_2+\mu_2\lambda_1)] + \mu_2S_2^2(\mu_1+\mu_2+\lambda_1)\Big),$$
$$k_2 = \frac{(S_1\mu_1+S_2\mu_2)\delta}{2S_1S_2(\mu_1+\mu_2)(S_2-S_1)},$$
$$\delta = \sqrt{[S_1(\mu_1+\mu_2+\lambda_2) - S_2(\mu_1+\mu_2+\lambda_1)]^2 + 4S_1S_2\lambda_1\lambda_2},$$
$$\alpha = \mu_1\delta^2 + \mu_1\delta[S_1(\mu_1+\mu_2+\lambda_2) - S_2(\mu_1+\mu_2+\lambda_1)],$$
$$\beta = \mu_2\lambda_1S_2[(S_1-S_2)(\mu_1-\mu_2) - (S_2\lambda_1+S_1\lambda_2) - \delta],$$
$$\gamma = \frac{E_2(S_2-S_1E_1)\alpha + S_1E_1(1-E_2)\beta}{S_1E_1(E_2-1)}. \tag{22}$$

Case 3: $S_1 > S_2$,

$$TP = \frac{S_1 E_1 \alpha e^{k_1 N} + S_2 E_2 \beta e^{k_2 N} + S_2 E_2 \gamma e^{-k_2 N}}{\alpha e^{k_1 N} + \beta e^{k_2 N} + \gamma e^{-k_2 N}}, \tag{23}$$

where

$$\begin{aligned}
E_i &= \frac{\mu_i}{\lambda_i + \mu_i}, \quad i = 1, 2, \\
k_1 &= \frac{1}{2S_1 S_2(\mu_1 + \mu_2)(S_2 - S_1)} \Big(\mu_1 S_1^2(\mu_1 + \mu_2 + \lambda_2) - S_1 S_2 \Big[(\mu_1 + \mu_2)^2 \\
&\quad + (\mu_1 + \mu_2)(\lambda_1 + \lambda_2) - (\mu_1 \lambda_2 + \mu_2 \lambda_1) \Big] + \mu_2 S_2^2(\mu_1 + \mu_2 + \lambda_1) \Big] \Big), \\
k_2 &= \frac{(S_1 \mu_1 + S_2 \mu_2)\delta}{2S_1 S_2(\mu_1 + \mu_2)(S_2 - S_1)}, \\
\delta &= \sqrt{[S_1(\mu_1 + \mu_2 + \lambda_2) - S_2(\mu_1 + \mu_2 + \lambda_1)]^2 + 4S_1 S_2 \lambda_1 \lambda_2}, \\
\alpha &= \mu_1 \delta^2 + \mu_1 \delta [S_1(\mu_1 + \mu_2 + \lambda_2) - S_2(\mu_1 + \mu_2 + \lambda_1)], \\
\beta &= \mu_1 \lambda_2 S_1 [(S_1 - S_2)(\mu_1 - \mu_2) - (S_2 \lambda_1 + S_1 \lambda_2) + \delta], \\
\gamma &= \frac{E_1(S_1 - S_2 E_2)\alpha + S_2 E_2(1 - E_1)\beta}{S_2 E_2(E_1 - 1)}.
\end{aligned} \tag{24}$$

Remark 8: It is shown by Chiang *et al*. (2000, 2001) that the system throughput is monotonically increasing with respect to μ_i, $i = 1, 2$, and N, and decreasing with respect to λ_i.

If both machines are identical, i.e., $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$, $S_1 = S_2 = S$, then

$$TP = \frac{S[2 + N\mu(1 + x)^2]}{(1 + x)^2[2 + N\mu(1 + x)]}. \tag{25}$$

3.1.2 Bernoulli model by Jacobs and Meerkov. Most production lines involve Markovian (exponential) machines. In order to apply the Bernoulli model, the original Markovian model can be transformed into an appropriate Bernoulli model (M–B transformation) so that the system production rate is 'preserved' (Kuo 1996). In order to predict the performance of the resulting Markovian system, the results need to be transformed back from the Bernoulli to the Markovian model. Numerical evidence indicates that such transformations have acceptable precision (see Kuo (1996) for detailed description and justification). The M–B transformation is defined as follows:

$$\begin{aligned}
p_i^{\text{Ber}} &= \frac{\mu_i S_i}{(\lambda_i + \mu_i) \max\{S_1, \ldots, S_M\}}, \quad i = 1, \ldots, M, \\
N_i^{\text{Ber}} &= \min\left\{ \frac{N_i^{\text{Mar}}}{S_{i+1}} \mu_i, \frac{N_i^{\text{Mar}}}{S_i} \mu_{i+1} \right\} + 1, \quad i = 1, \ldots, M.
\end{aligned} \tag{26}$$

Then the production rate calculated using the Bernoulli model is transformed back into the Markovian system.

Remark 9: The M–B transformation works with high accuracy when the speeds of the machines do not differ by too much. However, when the discrepancy between the speeds is larger, the accuracy of the transformation decreases.

Moreover, consider the case of identical machines. From (26), probability p in the Bernoulli model can be rewritten in terms of x:

$$p = \frac{1}{1+x}.$$

In addition,

$$N^{\text{Ber}} = \frac{N^{\text{Mar}}}{S}\mu + 1.$$

Then (4) can be rewritten as

$$TP = \frac{S + N^{\text{Mar}}\mu}{Sx + (N^{\text{Mar}}\mu + S)(1+x)}. \tag{27}$$

3.2 *Operation-dependent failure models*

Although TDF models for asynchronous lines provide closed formulae to calculate system throughput, the majority of the systems may obey operation-dependent failure. Therefore, the following ODF models with exponential machines are introduced.

3.2.1 Exponential model by Gershwin. In this model, the interactions between the machine and the buffers are assumed to be (Gershwin 1994):

- When the buffer is neither empty nor full and both machines are up, m_1 processes $S_1\delta t$ material during an interval of length δt, and m_2 processes $S_2\delta t$.
- When the buffer is empty and both machines are up, both machines process $S_1\delta t$ during an interval of length δt if $S_1 < S_2$. When the buffer is full and both machines are up, both machines process $S_2\delta t$ during an interval of length δt if $S_1 > S_2$.
- Machine m_2 is starved and cannot fail if the buffer is empty. Machine m_1 is blocked and cannot fail if the buffer is full.

The system throughput (TP) can be evaluated from

$$TP = \frac{\mu_2 S_2}{\lambda_1}(1 - p_s), \tag{28}$$

where

$$p_s = p(0,0,1) + \left(1 - \frac{S_1}{S_2}\right)p(0,1,1), \tag{29}$$

and $p(x, \alpha_1, \alpha_2)$ denotes the probability that machines m_1 and m_2 are in states α_1 and α_2, respectively, with buffer occupancy between x and $x + \delta x$, and machine up and down are denoted as 1 and 0, respectively. Such probabilities $p(0,0,1)$ and $p(0,1,1)$

are obtained through solving the parametric equations which describe the steady-state behaviour of internal storage (see Gershwin (1994) for details).

When both machines are identical, $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$ and $S_1 = S_2 = S$, after some derivation we finally obtain

$$TP = \frac{S[2S + N\mu(1 + x)]}{2S(1 + 2x) + N\mu(1 + x)^2}. \tag{30}$$

Remark 10: As we can see, closed formulae are not available in Gershwin ODF models, which makes it impossible to prove monotonicity and other system properties analytically. Such properties may still hold, but can only be verified through numerical examples (see Gershwin (1994) for details).

3.2.2 Exponential model by Alden. In addition to the general assumption, it is also specifically assumed (Alden 2002) that:

- No simultaneous downtime, i.e. at most one machine is down at any time. However, to account for possible simultaneous downtime, when one machine is down, the other machine's speed is reduced to its normal speed times its efficiency.
- Failures of machine i form a Poisson process with rates:
  - λ_i during periods with no special conditions, i.e. both machines are up, and machine i is not the faster machine during a speed blocking or speed starving period;
  - $\lambda_i f_i^{ss}$, during speed starving periods where machine i is the faster machine;
  - $\lambda_i f_i^{sb}$, during speed blocking periods where machine i is the faster machine;
  - 0, during periods when any machine is down

(in the above, f_i^{ss} and f_i^{sb} are the respective fractions of a speed blocking period and a speed starving period that machine i actually spends processing jobs).

The system throughput can then be calculated from the balance equations (see Alden (2002) for details):

$$TP = \begin{cases} S_2(P\{U\} + P\{SB\}) + \hat{S}_2 P(E), & \text{if } S_1 > S_2, \\ S_2 P\{U\} + \hat{S}_2 P\{E\}, & \text{if } S_1 = S_2, \\ S_2(P\{U\} + P\{SS\}) + \hat{S}_2 P\{E\}, & \text{if } S_1 < S_2, \end{cases} \tag{31}$$

where

$P\{U\}$: the probability that both machines are processing,

$P\{SB\}$: the probability of speed blocking,

$P\{SS\}$: the probability of speed starving,

$P\{E\}$: the probability of buffer emptying , i.e., m_1 down and m_2 processing,

$$\hat{S}_2 = \frac{\mu_2}{\lambda_2 + \mu_2}.$$

When two machines are identical, i.e., $\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$ and $S_1 = S_2 = S$, the system throughput simplifies to (Blumenfeld and Li 2005)

$$TP = \frac{S[2S + \mu N(1 + 2x)]}{(1 + 2x)[2S + \mu N(1 + x)]}. \tag{32}$$

4. Comparisons

In this section, comparisons between synchronous and asynchronous line models, time-dependent and operation-dependent failure models, are presented. In particular, table 1 provides a simple comparison of all the models discussed in this report for the case of identical machines. It turns out that they have a similar form and the differences are small. This is also verified in numerical examples.

Three typical systems are investigated (see table 2): two identical machines, non-identical machines with equal cycle times and non-identical machines with unequal cycle times. Similar results are observed in other examples with different machine parameters (for instance, machines with longer up- and downtimes). In the following subsections, comparisons between synchronous and asynchronous, TDF and ODF models are presented. More comparison examples can be found in Li *et al*. (2003).

4.1 *Synchronous line models*

A comparison of synchronous line models (both TDF and ODF) for systems with identical machines, and non-identical machines with equal cycle times is shown in figures 3 and 4, respectively.

Examining these figures, we obtain the following.

- Comparing the ODF and TDF models, both by Buzacott and Shanthikumar, the ODF model always has a higher estimate than the TDF model. However, the difference is very small and reaches a maximum with a small buffer (about 3–4% for $N=1$). When the buffer capacity becomes large (for instance, $N \geq 10$), the difference diminishes. Similar results are also shown in Buzacott and Shanthikumar (1993). In addition, the difference in throughput between the Buzacott–Shanthikumar TDF model and the Li–Meerkov model (where the former is slightly higher) may be due to the assumption for the case of an empty buffer in the Buzacott–Shanthikumar model addressed in remark .
- Using the M–B transformation, the Bernoulli model gives estimates that are very close to the Buzacott–Shanthikumar ODF model, which implies that the accuracy of the transformations is acceptable.
- When the buffer is small (for instance, $4 \leq N < 10$), the Gershwin geometric model provides an estimate similar to, but slightly lower than, that of the Li–Meerkov model. When the buffer is large, there is no difference at all.
- In summary, although the largest difference in throughput estimate is slightly higher (up to 7% if $N=1$ between the Buzacott–Shanthikumar and Li–Meerkov models)) in the case of a small buffer, in most cases the results are almost indistinguishable. It is comparable to the accuracy of data collection (about 10% error). Therefore, the difference in throughput calculation among all the models is insignificant, in particular when the buffer is large.

4.2 *Asynchronous line models*

Comparisons between asynchronous line models (TDF and ODF) for systems with identical machines, and non-identical machines with equal cycle times, are shown

Table 1. Throughput model of two identical machine lines ($x = \lambda/\mu$).

Model	Line type	Reliability distribution	Failure type	Throughput calculation
Jacobs–Meerkov	Synchronous	Bernoulli	TDF	$TP = \dfrac{1 + N\mu}{x + (N\mu + 1)(1 + x)}$
Buzacott–Shanthikumar	Synchronous	Geometric	TDF	$TP = \dfrac{2 - \mu(1 + x) + \mu N(1 + x)^2}{2(1 + x)^2 + \mu(N - 1)(1 + x)^3}$
Li–Meerkov	Synchronous	Geometric	TDF	$TP = \dfrac{2 - \mu + \mu x^2 + \mu(N - 1)(1 + x)^2}{(1 + x)^2[2 - \mu + \mu(N - 1)(1 + x)]}$
Buzacott–Shanthikumar	Synchronous	Geometric	ODF	$TP = \dfrac{2 - \mu + N\mu(1 + x)}{2(1 + 2x) + \mu(N - 1)(1 + x)^2}$
Gershwin	Synchronous	Geometric	ODF	$TP = \dfrac{2 + \mu x + (N - 3)\mu(1 + x)}{2(1 + 2x) + (N - 3)\mu(1 + x)^2}$
Jacobs–Meerkov	Asynchronous	Bernoulli	TDF	$TP = \dfrac{S + N\mu}{Sx + (N\mu + S)(1 + x)}$
Jacobs–Meerkov	Asynchronous	Exponential	TDF	$TP = \dfrac{S[2 + \mu N(1 + x)^2]}{(1 + x)^2[2 + \mu N(1 + x)]}$
Gershwin	Asynchronous	Exponential	ODF	$TP = \dfrac{S[2S + \mu N(1 + x)]}{2S(1 + 2x) + \mu N(1 + x)^2}$
Alden	Asynchronous	Exponential	ODF	$TP = \dfrac{S[2S + \mu N(1 + 2x)]}{(1 + 2x)[2S + \mu N(1 + x)]}$

Table 2. System parameters.

	λ_1	μ_1	λ_2	μ_2	S_1	S_2
System 1	0.1	0.4	0.1	0.4	1	1
System 2	0.1	0.8	0.08	0.7	1	1
System 3	0.08	0.3	0.06	0.55	3	2.5

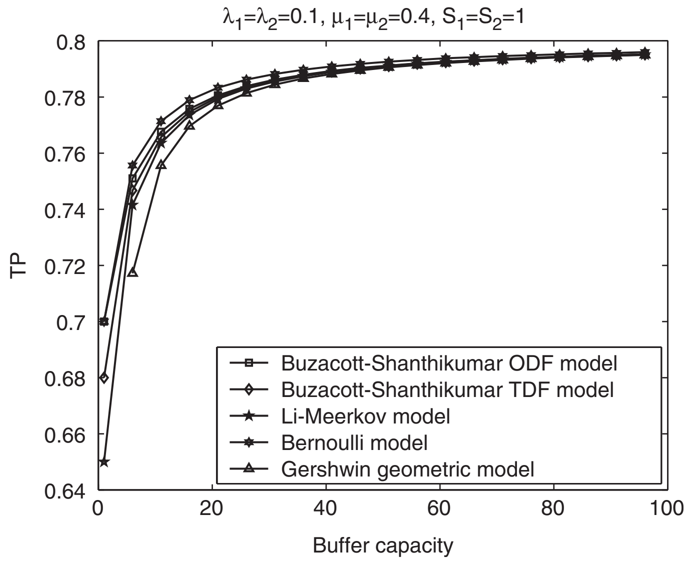

Figure 3. Synchronous line models with identical machines.

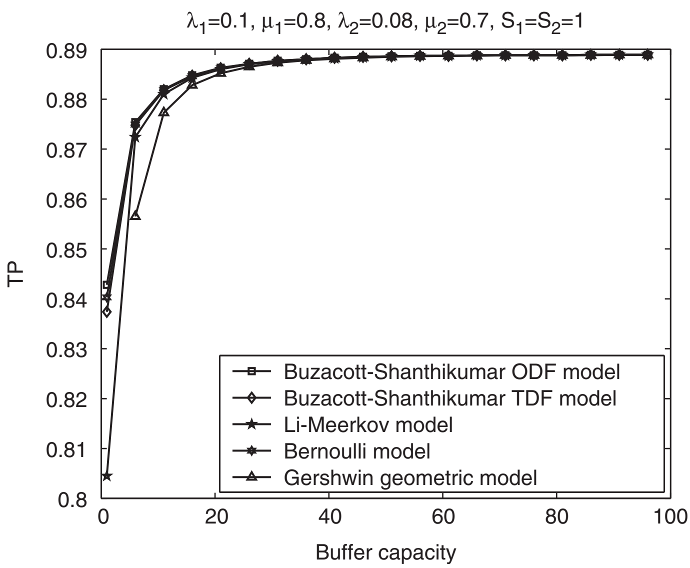

Figure 4. Synchronous line models with non-identical machines.

in figures 5 and 6, respectively (the comparison for a system of non-identical machines with unequal cycle times is shown in section 4.5).

Similar to the synchronous line models, there is almost no difference among the Alden, Gershwin exponential and Jacobs–Meerkov models except for very small buffers (for instance, $N < 5$). The difference between the Bernoulli model and other models is also small. The combined results from synchronous lines show

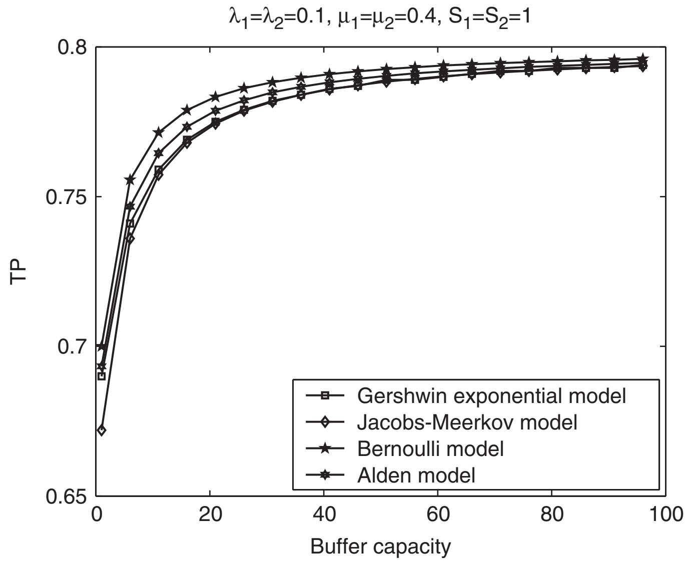

Figure 5. Asynchronous line models with identical machines.

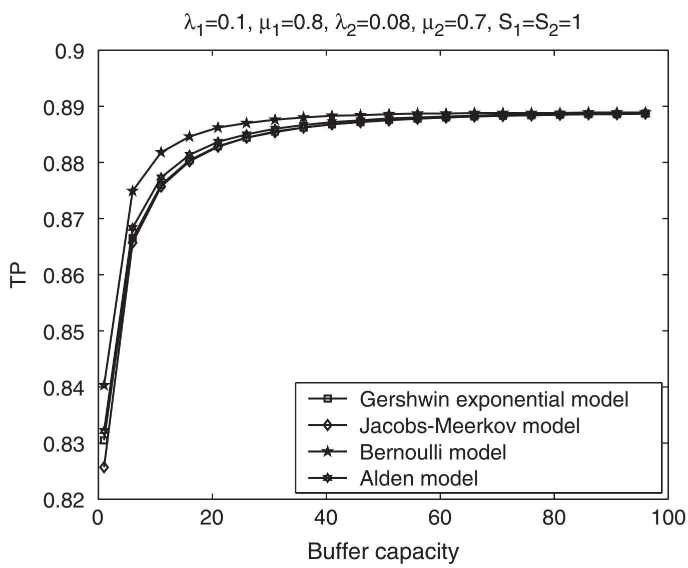

Figure 6. Asynchronous line models with non-identical machines and identical cycle times.

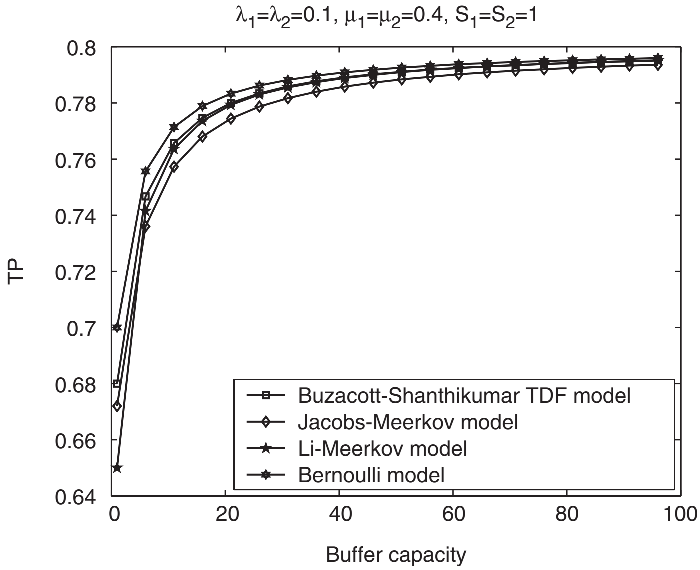

Figure 7. Time-dependent failure models with identical machines.

that the assumptions on time-dependent or operation-dependent failures do not make a significant difference in throughput analysis.

4.3 *Time-dependent failure models*

Figures 7 and 8 illustrate comparisons among all TDF models (synchronous and asynchronous lines) for systems with identical machines and non-identical machines with equal cycle times. From figures 7 and 8 we conclude the following.

- As before, the Buzacott–Shanthikumar model has a higher estimate of system throughput than the Li–Meerkov and Jacobs–Meerkov models.
- The Li–Meerkov model has a lower estimate when the buffer capacity is small (especially when $N=1$), however the estimates nearly agree with the Buzacott–Shanthikumar model for large buffer capacity (e.g., $N > 15$), and the Jacobs–Meerkov model provides the smallest estimate in this range.
- The estimate of the Bernoulli model through M–B transformations is slightly higher than for all the other models, but still within an acceptable range.
- Again, the differences among all models are very small and all the differences diminish with larger buffers.

4.4 *Operation-dependent failure models*

Figures 9 and 10 illustrate comparisons among all ODF models (synchronous and asynchronous) for systems with identical machines and non-identical machines with equal cycle times. Again, it can be seen that the Buzacott–Shanthikumar model gives a slightly higher estimate of system throughput than the other models. The Gershwin geometric model provides a lower estimate when the buffer capacity is small, whereas the Gershwin exponential model provides a lower estimate when the buffer capacity is large.

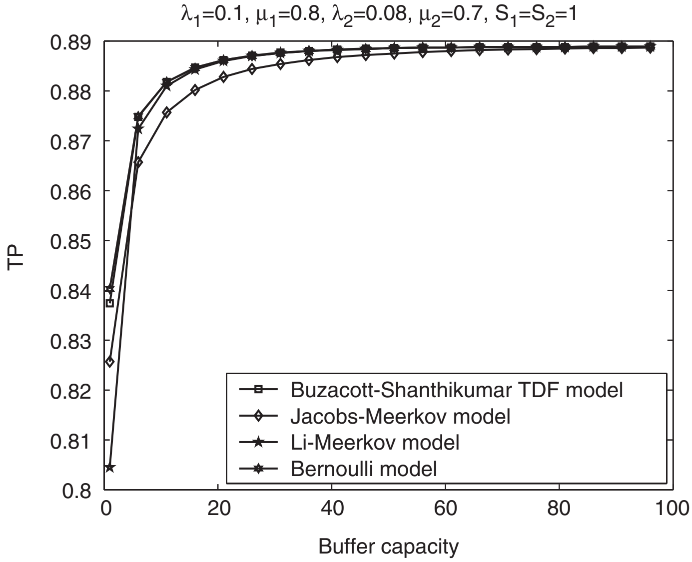

Figure 8. Time-dependent failure models with non-identical machines and identical cycle times.

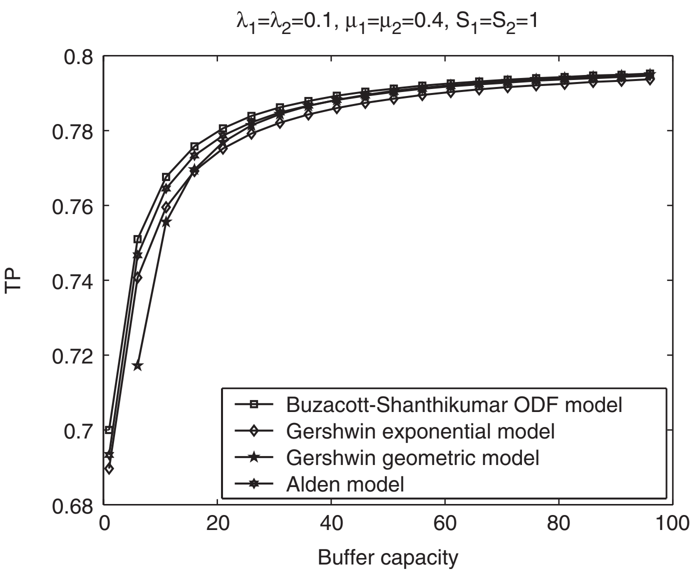

Figure 9. Operation-dependent failure models with identical machines.

4.5 *Comparisons among all models*

A comparison of all the models for systems with identical machines, and non-identical machines with equal and unequal cycle times are illustrated in figures 11–13, respectively.

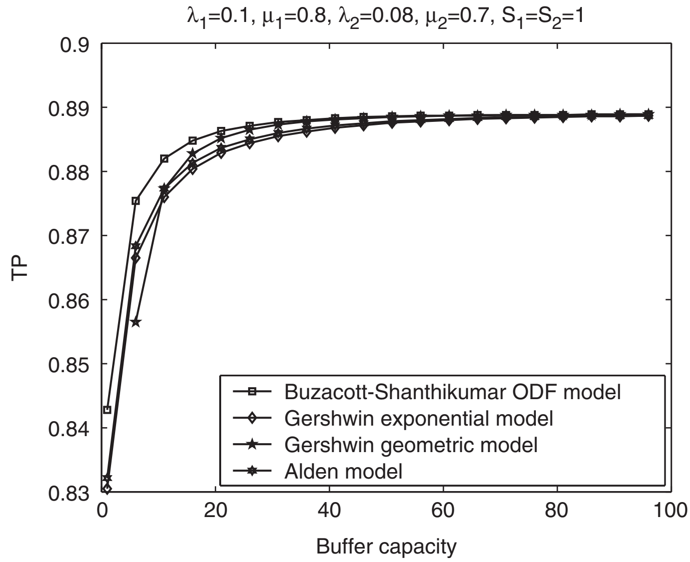

Figure 10. Operation-dependent failure models with non-identical machines and identical cycle times.

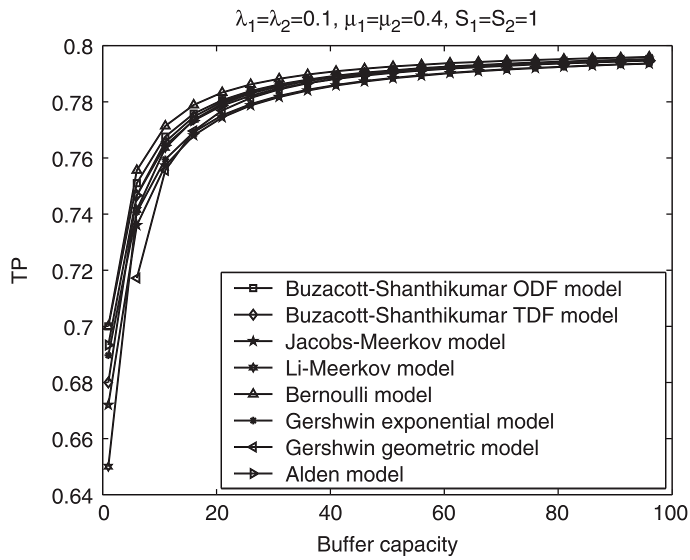

Figure 11. All models with identical machines.

Studying these results, it can be seen that the differences in terms of system throughput for all the models discussed in this paper are small. The differences decrease with respect to buffer capacity. For instance, the maximum difference is about 7% when $N = 1$ and decreases to about 4% if $N = 2$. When the buffer capacity is over 20, all the differences are less than 1%. Similar results have been obtained

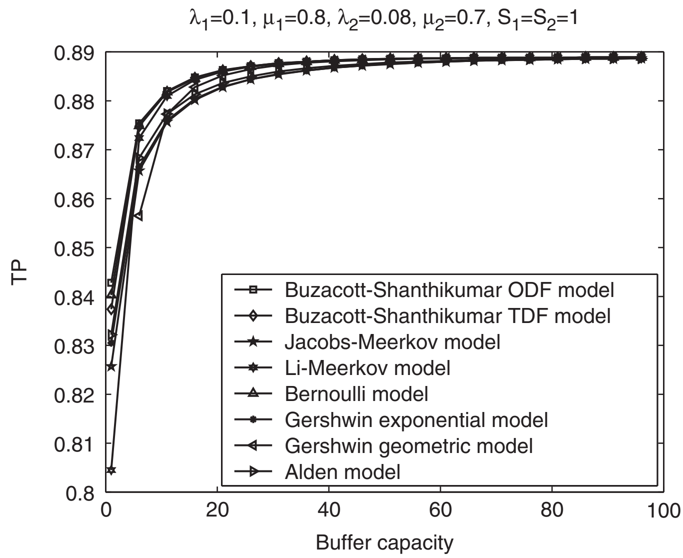

Figure 12. All models with non-identical machines and identical cycle times.

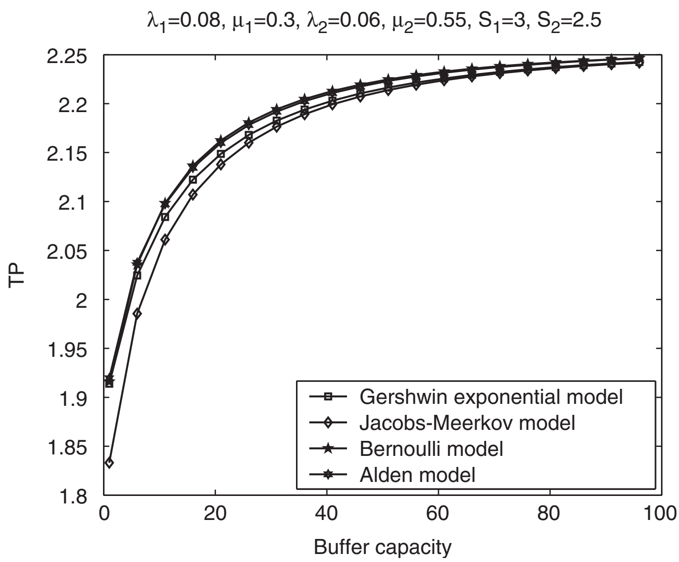

Figure 13. All models with non-identical machines and unequal cycle times.

in other examples with a wide range of parameters (see Li *et al.* (2003) for more examples).

5. Conclusions

Models of two machine–one buffer systems are the basic building blocks in throughput analysis. In this paper, eight different two-machine line models for operation-dependent failure, time-dependent failure, and synchronous and asynchronous lines are discussed. Comparisons of their assumptions and performances are presented.

It is shown that all the models discussed here exhibit similar performance in system throughput. Compared with the accuracy of data collection, the differences in throughput estimates among all the models are small and can be neglected. The differences in assumptions do not lead to a significant change in estimating system performance.

References

Alden, J.M., Estimating performance of two workstations in series with downtime and unequal speeds. General Motors Research & Development Center, Report R&D-9434, Warren, MI, 2002.

Altiok, T., Production lines with phase-type operation and repair times and finite buffers. *Int. J. Prod. Res.*, 1985, **23**, 489–498.

Altiok, T., *Performance Analysis of Manufacturing Systems*, 1997 (Springer: New York).

Blumenfeld, D.E. and Li, J., An analytical formula for throughput of a production line with identical stations and random failures. *Math. Prob. Eng.*, 2005, 293—308.

Buzacott, J.A., Automatic transfer lines with buffer stocks. *Int. J. Prod. Res.*, 1967, **5**, 183–200.

Buzacott, J.A., Methods of reliability analysis of production systems subject to breakdowns. In *Operations Research and Reliability*, edited by F. Grouchko, pp. 211–232, 1971 (Gordon and Breach: New York).

Buzacott, J.A. and Hanifin, L.E., Models of automatic transfer lines with inventory banks: a review and comparison. *IIE Trans.*, 1978, **10**, 197–207.

Buzacott, J.A. and Shanthikumar, J.G., *Stochastic Models of Manufacturing Systems*, 1993 (Prentice Hall: Englewood Cliffs, NJ).

Chiang, S.-Y., Kuo, C.-T. and Meerkov, S.M., DT-bottleneck in serial production lines: theory and application. *IEEE T. Robotic. Autom.*, 2000, **16**, 567–580.

Chiang, S.-Y., Kuo, C.-T. and Meerkov, S.M., *c*-Bottleneck in serial production lines: identification and application. *Math. Prob. Eng.*, 2001, **7**, 543–578.

Dallery, Y., David, R. and Xie, X.-L., An efficient algorithm for analysis of transfer lines with unreliable machines and finite buffers. *IIE Trans.*, 1988, **20**, 280–283.

Dallery, Y. and Gershwin, S.B., Manufacturing flow line systems: a review of models and analytical results. *Queuing Syst.*, 1992, **12**, 3–94.

De Koster, M.B.M., Estimation of the line efficiency by aggregation. *Int. J. Prod. Res.*, 1987, **25**, 615–626.

Enginarlar, E., Li, J., Meerkov, S.M. and Zhang, R.Q., Buffer capacity for accommodating machine downtime in serial production lines. *Int. J. Prod. Res.*, 2002, **40**, 601–624.

Gershwin, S.B., An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Oper. Res.*, 1987, **35**, 291–305.

Gershwin, S.B., *Manufacturing Systems Engineering*, 1994 (Prentice Hall: Englewood Cliffs, NJ).

Inman, R.R., Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. *Prod. Oper. Manage.*, 1999, **8**, 409–432.

Jacobs, D.A., Improvability in production systems: theory and case studies. PhD thesis, Department of EECS, University of Michigan, Ann Arbor, MI, 1993.

Jacobs, D.A. and Meerkov, S.M., A system-theoretic property of serial production lines: improvability. *Int. J. Sys. Sci.*, 1995, **26**, 95–137.

Kuo, C.-T., Bottlenecks in production systems: a systems approach. PhD thesis, Department of EECS, University of Michigan, Ann Arbor, MI, 1996.

Le Bihan, H. and Dallery, Y., A robust decomposition method for the analysis of production lines with unreliable machines and finite buffers. *Ann. Oper. Res.*, 2000, **93**, 265–297.

Li, J., Production variability in manufacturing systems: a systems approach. PhD thesis, Department of EECS, Univsity of Michigan, Ann Arbor, MI, 2000.

Li, J. and Meerkov, S.M., Due-time performance in production systems with Markovian machines. In *Analysis of Manufacturing Systems*, edited by S.B. Gershwin *et al.*, pp. 221–253, 2003 (Kluwer Academic: Boston, MA).

Li, J., Blumenfeld, D.E. and Alden, J.M., Throughput analysis of serial production line with two machines: summary and comparisons. General Motors Research & Development Center, Report R&D-9532, Warren, MI, 2003.

Li, J. and Meerkov, S.M., Evaluation of throughput in serial production lines with non-exponential machines. In *Analysis, Control and Optimization of Complex Dynamic Systems*, edited by E.K. Boukas and R. Malhame, pp. 55–82, 2005 (Springer: New York).

Lim, J.T., Meerkov, S.M. and Top, F., Homogeneous, asymtotically reliable serial production lines: theory and a case study. *IEEE T. Automat. Contr.*, 1990, **35**, 524–534.

Papadopoulos, H.T. and Heavey, C., Queueing theory in manufacturing systems analysis and design: a classification of models for production and transfer lines. *Eur. J. Oper. Res.*, 1996, **92**, 1–27.