INT. J. PROD. RES., 2002, VOL. 40, NO. 3, 601–624

Buffer capacity for accommodating machine downtime in serial production lines

EMRE ENGINARLAR†, JINGSHAN LI‡, SEMYON M. MEERKOV†* and RACHEL Q. ZHANG§

This paper investigates the smallest level of buffering (LB), necessary to ensure the desired production rate in serial lines with unreliable machines. The reliability of machines is assumed to obey either exponential, or Erlang, or Rayleigh models. The LB is measured in units of the average downtime, T_{down}. The dependence of LB on the reliability model, the number of machines, M, the average uptime, T_{up}, and the efficiency, $e = T_{\mathrm{up}}/(T_{\mathrm{up}} + T_{\mathrm{down}})$ is analysed. It is shown that reliability models with larger coefficient of variation require larger LB, and an empirical law that connects LB of the exponential model with those for other reliability models is established. It is shown that LB is an increasing function of M, but with an exponentially decreasing rate, saturating at around $M = 10$. Also, it is shown that LB does not depend explicitly on T_{up} and is a decreasing function of e. Based on these results, rules-of-thumb are provided for selecting buffer capacity, which guarantee sufficiently high line efficiency.

1. Motivation

Machine downtime may, and often does, lead to loss of throughput in manufacturing systems. To minimize this loss, in-process buffers are used. If the capacity of these buffers is very large, the machines are practically decoupled, and the system production rate, PR (i.e. the average number of parts produced by the last machine per unit of time), is maximized. We denote this production rate as PR_∞, indicating that it is attained when the capacity of the buffers is infinite. Obviously, large buffers may lead to excessive inventory, long part-in-process time, low quality and other manufacturing ills. Therefore, it is of interest to determine the buffer capacity, which, on one hand, decouples the machines so that PR is sufficiently close to PR_∞ and, on the other hand, is as small as possible. The goal of this paper is to characterize such a capacity and, on this basis, establish 'rules-of-thumb' for selecting the buffers so that PR is 95, or 90, or 85% of PR_∞.

In practice, both up- and downtime of the machines are random variables, even in the case of the so-called scheduled downtime (for instance, tool change time). In this paper, we model these random variables by three types of distributions: exponential, Erlang and Rayleigh. The exponential distribution is intended to

Revision received July 2001.

† Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA.

‡ Enterprise Systems Laboratory, GM Research and Development Center, Warren, MI 48090-9055, USA.

§ Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, USA.

* To whom correspondence should be addressed. e-mail: smm@eecs.umich.edu

International Journal of Production Research ISSN 0020–7543 print/ISSN 1366–588X online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540110091703

model up- and downtime with a relatively large variability (measured by the coefficient of variation, i.e. the ratio of the standard deviation to the average). The coefficient of variation of Erlang distribution depends on the number of exponential stages involved and is a decreasing function of the number of stages. Rayleigh distribution is selected to model random variables other than exponential or Erlang.

Obviously, the capacity of buffers, which accommodate downtime, depends on the average value of this downtime (denoted as T_{down}). Therefore, the rules-of-thumb, sought in this paper, give the capacity of the buffers in units of T_{down}. For instance, 'k-downtime buffer' denotes the capacity of a buffer that is capable of storing the number of parts produced or consumed during k downtimes. The number k is referred to as the level of buffering (LB). Note that to accommodate tool change time, production line designers usually select the buffers of one-downtime capacity. It will be shown in section 6 that in many practical situations, this may lead to 30% loss of system throughput.

Intuitively, LB may depend not only on the average downtime, T_{down}, but also on the average uptime, T_{up}, and on the machine efficiency, e, defined by:

$$e = \frac{T_{\text{up}}}{T_{\text{up}} + T_{\text{down}}} = \frac{1}{1 + T_{\text{down}}/T_{\text{up}}}.$$

Finally, it may depend also on the number of machines, M, in the system.

Therefore, this paper is intended to provide a characterization of the LB, necessary to accommodate machine downtime, in terms of the following function:

LB $= F$ (type of up- and downtime distribution; machine efficiency, e, or $T_{\text{up}}/T_{\text{down}}$; number of machines in the system, M; average uptime, T_{up}).

Roughly speaking, the results obtained in this paper are as follows.

(1) Production lines with machines having larger coefficient of variation of the downtime require a larger LB. This implies that exponential machines need larger LB than other distributions considered.

(2) LB for lines with Erlang and Rayleigh distributions is related to LB for lines with exponential machines according to the following empirical law:

$$\widehat{k^A} = \text{CV}^A_{\text{down}} \cdot k^{\text{ex}} = \frac{\sigma^A_{\text{down}}}{T_{\text{down}}} \cdot k^{\text{ex}},$$

where A is the distribution of the downtime, k^{ex} is the LB for exponentially distributed downtime, $\widehat{k^A}$ is an estimate of the LB for downtime distributed according to A and $\text{CV}^A_{\text{down}}$ is the coefficient of variation of distribution A defined by:

$$\text{CV}^A_{\text{down}} = \frac{\sigma^A_{\text{down}}}{T_{\text{down}}}.$$

(3) Larger machine efficiency, e, or larger $\alpha = T_{\text{up}}/T_{\text{down}}$ requires a smaller level of buffering. For example, in 10-machine lines with exponential machines, this relationship is characterized by:

$$k^{\text{ex}}_{0.9}(\alpha) = -0.0015\alpha^3 + 0.068\alpha^2 - 1.156\alpha + 9.8504, \tag{1}$$

where $k_{0.9}^{\mathrm{ex}}$ is the smallest level of buffering necessary to achieve $\mathrm{PR} = 0.9\mathrm{PR}_\infty$ in serial lines with exponential machines.

(4) LB does not depend explicitly on T_{up} (i.e. it depends only on the ratio of T_{up} and T_{down}, akin the machine efficiency, e).

(5) LB is an increasing function of M. However, this increase is exponentially decaying saturating at about $M = 10$. For instance, if $e = 0.9$, this relationship is described by the following exponential approximation:

$$k_{0.9}^{\mathrm{ex}}(M) = 0.045 + 4.365\left(1 - \exp\left[-\frac{M-2}{3.5}\right]\right), \qquad M \in [2, 3, \ldots). \tag{2}$$

Justification of these conclusions and additional results are given in sections 6 and 7.

The results reported here are obtained using analytical calculations based on aggregation (for exponential distribution), Markov chain analysis (for Erlang with $M = 2$), and discrete event simulations (for Erlang with $M > 2$ and Rayleigh distributions).

The outline of the paper is as follows: section 2 describes the relevant literature. In section 3, the model of the serial production line under consideration is introduced. The problem formulation is given in section 4. Methods of analysis used throughout are outlined in section 5. The results obtained are described in sections 6 and 7. Conclusions are given in section 8. Notation is given in appendix 1.

2. Related literature

Buffer capacity allocation in production lines has been studied quantitatively for over 50 years and hundreds of publications are available. The remarks below are intended to place the current paper in the framework of this literature rather than to provide a comprehensive review.

With respect to the machines, production lines can be classified into two groups: unreliable machines with fixed cycle time and reliable machines with random processing time. This work addresses the first group.

With respect to the machine efficiency, production lines with unreliable machines can be further divided into two groups: balanced (i.e. the machines have identical up- and downtime distributions) and unbalanced. This work first addresses the balanced case and then extends the results to the unbalanced one.

Buffer capacity allocation in production lines, similar to those addressed here, has been first considered in the classic papers by Vladzievskii (1950, 1951), Sevastyanov (1962), and Buzacott (1967). A review of the early work in this area is given by Buzacott and Hanifin (1978). In particular, Buzacott (1967) showed that the coefficient of variation of the downtime strongly affects the efficacy of buffering. In addition, Buzacott associated the buffer capacity allocation with the average downtime and stated that buffering beyond five-downtime can hardly be justified. These results are confirmed and further quantified in the present work.

Conway *et al.* (1988) also connected buffer allocation with downtime. They showed that one-downtime buffering was sufficient to regain about 50% of production losses if the downtime was constant (deterministic). They suggested that random (exponential) downtime may require a twice larger capacity to result in comparable gains. This suggestion is further explored in the current paper.

Another line of research on buffer capacity allocation is related to the so-called storage bowl phenomenon (Hillier and So 1991). According to this phenomenon, more buffering should be assigned to middle machines in balanced lines. It can be shown, however, that unbalancing the buffering in lines with downtime coefficient of variation < 1 results in only 1–3% of throughput improvement, if at all. (For further details, see Jacobs and Meerkov 1995b where it is proved that optimal buffers are of equal capacity, if the work is distributed according to the optimal bowl.) Since this improvement is quite small, the present paper does not consider bowl-type storage allocation and assigns equal capacity to all buffers in balanced lines and appropriately selected unequal buffering in unbalanced ones.

Finally, there exists a large body of literature on numerical algorithms that calculate the optimal buffer allocation (e.g. Ho *et al.* 1979, Jacobs and Meerkov 1995a, Glasserman and Yao 1996, Gershwin and Schor 2000). The current work does not address this issue.

Thus, the present paper follows Buzacott (1967) and Conway *et al.* (1988) and provides additional results on rules-of-thumb for buffer capacity allocation necessary to accommodate downtime and achieve the desired efficiency of serial production lines with unreliable machines.

3. Model

The block-diagram of the production system considered here is shown in figure 1, where the circles are the machines and the rectangles are buffers. The following are the assumptions concerning the machines, buffers and interactions among them (i.e. blockages and starvations).

3.1. *Machines*

(1) Each machine, m, has two states: up and down. When up, the machine is capable of producing one part per unit of time (*machine cycle time*); when down, no production takes place.

(2) The up- and downtime of each machine are random variables distributed according to either of the following distributions

(a) Exponential:

$$\left.\begin{aligned} f_{\rm up}^{\rm ex}(t) &= p_{\rm ex} e^{-p_{\rm ex} t}, \qquad t \geq 0, \\ f_{\rm down}^{\rm ex}(t) &= r_{\rm ex} e^{-r_{\rm ex} t}, \qquad t \geq 0; \end{aligned}\right\} \tag{3}$$

(b) Erlang:

$$\left.\begin{aligned} f_{\rm up}^{{\rm Er}_P}(t) &= p_{\rm Er} e^{-p_{\rm Er} t} \frac{(p_{\rm Er} t)^{P-1}}{(P-1)!}, \qquad t \geq 0, \\ f_{\rm down}^{{\rm Er}_R}(t) &= r_{\rm Er} e^{-r_{\rm Er} t} \frac{(r_{\rm Er} t)^{R-1}}{(R-1)!}, \qquad t \geq 0; \end{aligned}\right\} \tag{4}$$

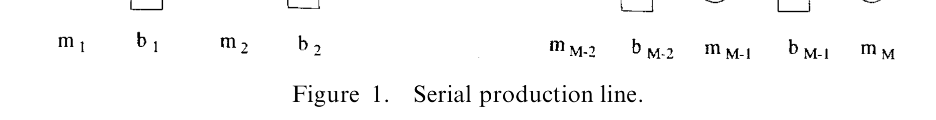


Figure 1. Serial production line.

(c) Rayleigh:

$$\left.\begin{aligned} f_{\mathrm{up}}^{\mathrm{Ra}}(t) &= p_{\mathrm{Ra}}^2 t e^{-p_{\mathrm{Ra}}^2 t^2/2}, \qquad t \geq 0, \\ f_{\mathrm{down}}^{\mathrm{Ra}}(t) &= r_{\mathrm{Ra}}^2 t e^{-r_{\mathrm{Ra}}^2 t^2/2}, \qquad t \geq 0. \end{aligned}\right\} \tag{5}$$

The expected T_{up} and T_{down}, the variances σ_{up}^2 and σ_{down}^2, and the coefficients of variation $\mathrm{CV}_{\mathrm{up}}$ and $\mathrm{CV}_{\mathrm{down}}$, of each of these distributions are given in table 1.

(3) The parameters of distributions (3)–(5) are selected so that machine efficiencies, e, and, moreover, T_{up} and T_{down} are identical for all reliability models, i.e.

$$\begin{aligned} T_{\mathrm{up}} &= \frac{1}{p_{\mathrm{ex}}} \quad \text{(exponential)} \\ &= \frac{P}{p_{\mathrm{Er}}} \quad \text{(Erlang with } P \text{ stages)} \\ &= \frac{1.2533}{p_{\mathrm{Ra}}} \quad \text{(Rayleigh);} \end{aligned}$$

$$\begin{aligned} T_{\mathrm{down}} &= \frac{1}{r_{\mathrm{ex}}} \quad \text{(exponential)} \\ &= \frac{R}{r_{\mathrm{Er}}} \quad \text{(Erlang with } R \text{ stages)} \\ &= \frac{1.2533}{r_{\mathrm{Ra}}} \quad \text{(Rayleigh).} \end{aligned}$$

3.2. *Buffers*

(4) Each buffer has the capacity N defined by

$$N = \lceil kT_{\mathrm{down}} \rceil,$$

where $\lceil x \rceil$ is the smallest integer $> x$ and T_{down} is measured in units of the cycle time. Coefficient $k \in R_+$ is referred to as the *level of buffering*.

3.3. *Interactions among the machines and buffers*

(5) Machine m_i, $i = 2, \ldots, M$, is starved at time t if b_{i-1} is empty and m_{i-1} fails to put a part into b_{i-1} at time t. Machine m_1 is never starved.

(6) Machine m_i, $i = 1, \ldots, M-1$, is blocked at time t if buffer b_i is full and m_{i+1} fails to take a part from b_i at time t. Machine m_M is never blocked.

Distribution	T_{up}	T_{down}	σ_{up}^2	σ_{down}^2	$\mathrm{CV}_{\mathrm{up}}$	$\mathrm{CV}_{\mathrm{down}}$
Exponential	$1/p_{\mathrm{ex}}$	$1/r_{\mathrm{ex}}$	$1/p_{\mathrm{ex}}^2$	$1/r_{\mathrm{ex}}^2$	1	1
Erlang	P/p_{er}	R/r_{Er}	P/p_{Er}^2	R/r_{Er}^2	$1/\sqrt{P}$	$1/\sqrt{R}$
Rayleigh	$1.2533/p_{\mathrm{Ra}}$	$1.2533/r_{\mathrm{Ra}}$	$0.4292/p_{\mathrm{Ra}}^2$	$0.4292/r_{\mathrm{ra}}^2$	0.5227	0.5227

Table 1. Expected value, variance and coefficient of variation of up- and downtime distributions considered.

Assumptions (1)–(6) define the production system considered in sections 4–6. In section 7, an unbalanced version of this system is analysed.

4. Problem formulation

The production rate, PR, of the serial line (1)–(6) is the average number of parts produced by the last machine, m_M, during a cycle time (in the steady state of system operation). When the capacity of the buffers is infinite, the production rate of the line, PR_∞, is equal to

$$\mathrm{PR}_\infty = \min(e_1, \ldots, e_M).$$

When the buffers are finite and selected according to assumption (4), PR is, obviously, smaller. Let E denote the *efficiency of the line*, defined by

$$E = \frac{\mathrm{PR}}{\mathrm{PR}_\infty}, \quad E \in \left[\prod_{i=1}^{M} e_i, 1\right),$$

and k_E be the smallest level of buffering necessary to attain line efficiency E.

The following are the problems analysed in this work: Given the production system defined by (1)–(6).

(α) Analyse the properties of k_E in production lines with identical machines. In particular, investigate the dependence of k_E on the machine reliability model, e, M, and T_{up} and, on this basis, provide rules-of-thumb for selecting k_E for $E = 0.95$, or 0.9 or 0.85.

(β) Extend results obtained to production systems with non-identical machines.

(γ) Investigate production losses, measured by $(\mathrm{PR}_\infty - \mathrm{PR}_k)/\mathrm{PR}_\infty$, when $k = 1$, where PR_k denotes the production rate of the system (1)–(6) with $\mathrm{LB} = k$. This case is intended to model the current industrial practice used in the design of modern production systems, where buffer capacity is selected using the 'one-downtime' rule.

Solutions of problems (α) and (γ) are given in section 6. Problem (β) is discussed in section 7.

5. Methods of analysis

5.1. *Exponential machines*

For $M = 2$ and exponential reliability model with parameters p_i and r_i, $i = 1, 2$, PR of the serial line defined by (1)–(6) is calculated by Jacobs (1993) to be

$$\mathrm{PR} = \begin{cases} \dfrac{r_1 r_2}{(p_1 + r_1)(p_2 + r_2)} \left[\dfrac{p_1(p_2 + r_2) - p_2(p_1 + r_1)e^{-\beta N}}{p_1 r_2 - p_2 r_1 e^{-\beta N}} \right], & \text{if } \dfrac{p_1}{r_1} \neq \dfrac{p_2}{r_2}, \\[2ex] \dfrac{r_2^2(r_1 + r_2) + N r_1 r_2 (p_2 + r_2)^2}{(p_2 + r_2)^2 [r_1 + r_2 + N r_1 (p_2 + r_2)]}, & \text{if } \dfrac{p_1}{r_1} = \dfrac{p_2}{r_2}, \end{cases} \tag{6}$$

where

$$\beta = \frac{(r_1 + r_2 + p_1 + p_2)(p_1 r_2 - p_2 r_1)}{(r_1 + r_2)(p_1 + p_2)}.$$

For $M > 2$, no closed formula for PR is available. However, several approximation techniques have been developed (Gershwin 1987, Dallery *et al.* 1989, Chiang *et al.* 2000). We use here the one developed in Chiang *et al.*, since it is directly applicable to model (1)–(6). It consists of the so-called forward and backward aggregation. In the forward aggregation, using expression (6), the first two machines are aggregated in a single machine, m_2^f, defined by parameters p_2^f and r_2^f. Then m_2^f is aggregated with m_3 to result in m_3^f,which is then aggregated with m_4 to give m_4^f, and so on until all machines are aggregated in m_M^f. In the backward aggregation, m_{M-1}^f is aggregated with m_M to produce m_{M-1}^b, which is then aggregated with m_{M-2}^f to result in m_{M-2}^b, and so on until all machines are aggregated in m_1^b. Then the process is repeated anew. Formally, this recursive procedure has the following form:

$$\left.\begin{aligned}
p_i^b(s+1) &= \frac{p}{1-Q(p_{i+1}^b(s+1), r_{i+1}^b(s+1), p_i^f(s), r_i^f(s), N_i)}, \quad 1 \le i \le M-1,\\
r_i^b(s+1) &= \frac{1}{\dfrac{Q(p_{i+1}^b(s+1), r_{i+1}^b(s+1), p_i^f(s), r_i^f(s), N_i)}{p} + \dfrac{1}{r}}, \quad 1 \le i \le M-1,\\
p_i^f(s+1) &= \frac{p}{1-Q(p_{i-1}^f(s+1), r_{i-1}^f(s+1), p_i^b(s+1), r_i^b(s+1), N_{i-1})}, \quad 2 \le i \le M,\\
r_i^f(s+1) &= \frac{1}{\dfrac{Q(p_{i-1}^f(s+1), r_{i-1}^f(s+1), p_i^b(s+1), r_i^b(s+1), N_{i-1})}{p} + \dfrac{1}{r}}, \quad 2 \le i \le M,
\end{aligned}\right\} \tag{7}$$

with boundary conditions

$$\left.\begin{aligned}
p_1^f(s) = p, \quad r_1^f(s) = r,\\
p_M^b(s) = p, \quad r_M^b(s) = r,\\
s = 0, 1, 2, \ldots,
\end{aligned}\right\} \tag{8}$$

and initial conditions

$$p_i^f(0) = p, \quad r_i^f(0) = r, \quad i = 2, \ldots, M-1, \tag{9}$$

where function Q is given by

$$Q(p_1, r_1, p_2, r_2, N_1) = \begin{cases} \dfrac{(1-e_1)(1-\phi)}{1-\phi e^{-\beta N_1}}, & \text{if } \dfrac{p_1}{r_1} \ne \dfrac{p_2}{r_2},\\[2ex] \dfrac{p_1(p_1+p_2)(r_1+r_2)}{(p_1+r_1)[(p_1+p_2)(r_1+r_2)+p_2 r_1(p_1+p_2+r_1+r_2)N_1]}, & \text{if } \dfrac{p_1}{r_1} = \dfrac{p_2}{r_2}, \end{cases}$$

$$e_i = \frac{r_i}{p_i + r_i}, \quad i = 1, 2,$$

$$\phi = \frac{e_1(1-e_2)}{e_2(1-e_1)},$$

$$\beta = \frac{(r_1 + r_2 + p_1 + p_2)(p_1 r_2 - p_2 r_1)}{(r_1 + r_2)(p_1 + p_2)}.$$

It is shown by Chiang *et al.* that this procedure is convergent, and the following limits exist:

$$\lim_{s\to\infty} p_i^f(s) =: p_i^f, \quad \lim_{s\to\infty} p_i^b(s) =: p_i^b,$$

$$\lim_{s\to\infty} r_i^f(s) =: r_i^f, \quad \lim_{s\to\infty} r_i^b(s) =: r_i^b, \quad i = 1, \ldots, M.$$

Since the last machine is never blocked and the first machine is never starved, the estimate of PR, denoted as $\widehat{\mathrm{PR}}$, is defined as:

$$\widehat{\mathrm{PR}}(p_1, r_1, p_2, r_2, \ldots, p_M, r_M, N_1, N_2, \ldots, N_{M-1}) = \frac{r_M^f}{p_M^f + r_M^f} = \frac{r_1^b}{p_1^b + r_1^b}. \tag{10}$$

It is shown by Chiang *et al.* that this estimate results in sufficiently high precision.

5.2. *Erlang machines*

For $M = 2$ and Erlang reliability, PR can be calculated using the method developed by Altiok (1985). According to this method, each stage of the Erlang distribution is treated as a state (along with all other states, defined by the occupancy of the buffer). Since the residence time in each stage is distributed exponentially, a standard Markov process description applies. To simplify calculations, a discrete time approximation of the continuous time Markov process is utilized. Thus, according to this method, the performance analysis of system (1)–(6) reduces to the calculation of the stationary probability distribution of a discrete time Markov chain. Once this probability distribution is found, the production rate, PR, is calculated by summing up the probabilities of the states where m_2 is up and not starved (Altiok 1985).

It should be pointed out that, due to the increase of dimensionality, this method is practical only when the buffer capacity is not too large (<100). For systems with larger buffers or with more than two machines, discrete event simulations seem to be faster than the method described above, even if the Erlang distribution with only two stages is considered.

5.3. *Simulations*

Unfortunately, no analytical calculation methods exist for PR evaluation in systems with Rayleigh machines. For production lines with Erlang machines and $M > 2$, the PR calculations are prohibitively time consuming. Therefore, we analyse these systems using discrete event simulations. It should be pointed out that the analytical calculations are many orders of magnitude faster than the discrete event simulation; for instance, calculation of $\widehat{\mathrm{PR}}$ on a PC for a line with 10 exponential machines, using (7)–(10), takes about 3.5 sec, whereas discrete event simulation takes >2 h.

The simulations have been carried out as follows: a discrete event model of line (1)–(6) has been constructed. Zero initial conditions for all buffers were assumed and the states of all machines at the initial time moment have been selected to be ‘up’. First, 100 000 time slots of warm-up period were carried out and the next 1 000 000 slots of stationary operation were used to evaluate the production rate statistically.

The 95% confidence intervals, calculated as explained in Law and Kelton (1991), were <0.0005 when each simulation was carried out 10 times.

5.4. *Calculation of* k_E

The level of buffering, k_E, which ensures the desired line efficiency E ($E = 0.95$, or 0.9 or 0.85), has been determined as follows.

For each model of machine reliability, PR of line (1)–(6) was evaluated, first for $N = 0$, then for $N = 1$ and so on until PR reached the level of $E \cdot \mathrm{PR}_\infty$. This buffer capacity, N_E, was then divided by T_{down} (in units of the cycle time). This provided the desired level of buffering k_E. Results of these calculations are described below.

6. Results: identical machines

Here, we assume that all machines obey the same reliability model and the average uptime (respectively, downtime) of all machines is the same. Non-identical machines are addressed in section 7.

6.1. *Two-machine case*

6.1.1. *Exponential machines*

Expression (6), under the assumption of $p_1 = p_2 = p$ and $r_1 = r_2 = r$, leads to a closed form expression for k_E^{ex}. Indeed, assuming that $\mathrm{PR} = E \cdot \mathrm{PR}_\infty = E \cdot r/(r+p)$, from (6) it follows that the buffer capacity, N_E, which results in this production rate, is defined by

$$N_E = \begin{cases} \left\lceil \dfrac{2(1-e)(E-e)}{p(1-E)} \right\rceil, & \text{if } E > e, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where, as before, $\lceil x \rceil$ is the smallest integer $> x$. Therefore, for two-machine lines with exponential machines LB is given by

$$k_E^{\text{ex}} = \frac{N_E}{T_{\text{down}}} = \begin{cases} \dfrac{2e(E-e)}{1-E}, & \text{if } E > e, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

As follows from (12), LB depends explicitly only on machine efficiency e and is independent of T_{up}. Also, (12) shows that k_E^{ex} is decreasing as a function of e for $e > 0.5E$ and increasing for $e < 0.5E$. Since in most practical situations $e > 0.5$, we consider throughout this paper only the machines with $e \geq 0.5 > 0.5E$.

The behaviour of k_E^{ex} for $E = 0.95$, 0.9, and 0.85 as a function of $T_{\text{up}}/T_{\text{down}}$ (or e) is illustrated in figure 2.

6.1.2. *Erlang and Rayleigh machines*

Using Markov chain analysis for the Erlang case (with $P = R = 2$ and $P = R = 5$) and discrete event simulations for the Rayleigh case, we calculate $k_E^{\text{Er}_P}$ and k_E^{Ra} for $E = 0.95$, 0.9 and 0.85. The results are shown in figure 3 (for $T_{\text{up}} = 30$) and figure 4 (for $T_{\text{up}} = 60$), where the exponential case is also included for comparison.

Results shown in figures 2–4 lead to the following conclusions.

- LB for the Erlang and Rayleigh machines, akin the exponential case, is independent of T_{up} (since figures 3 and 4 are practically identical).

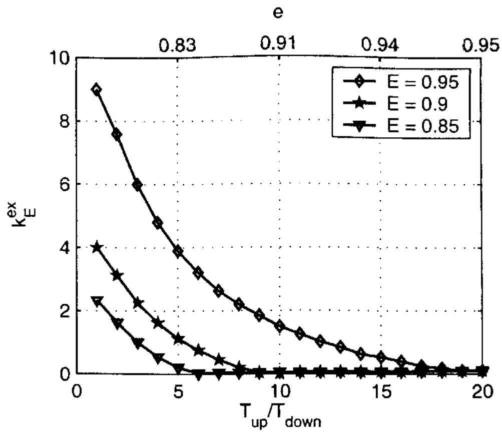
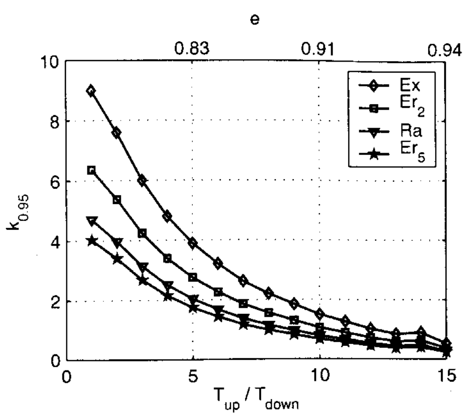
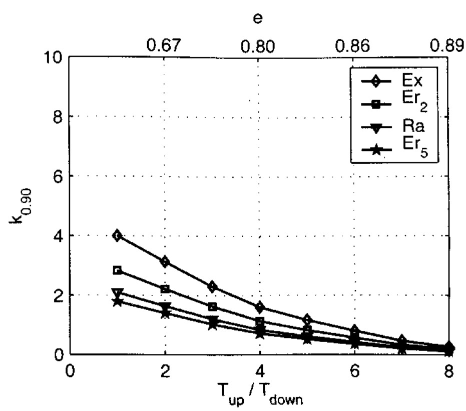


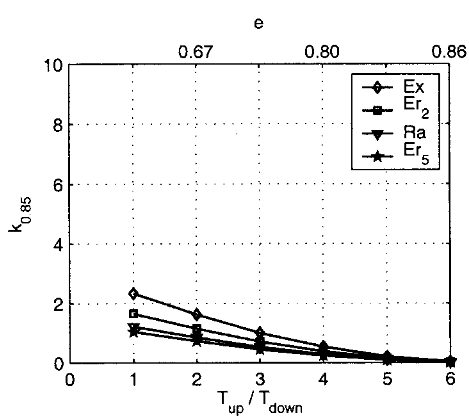
Figure 2. Level of buffering for exponential machines ($M = 2$, $T_{up} = 200$).



(a) $PR = 0.95PR_\infty$

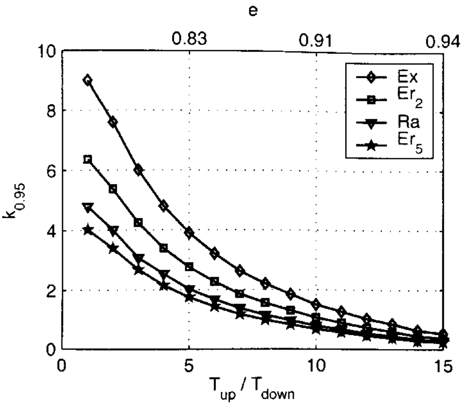


(b) $PR = 0.9PR_\infty$

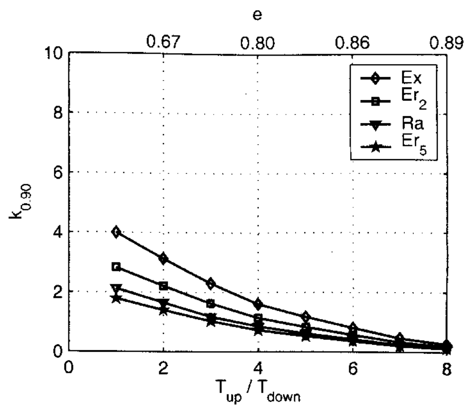


(c) $PR = 0.85PR_\infty$

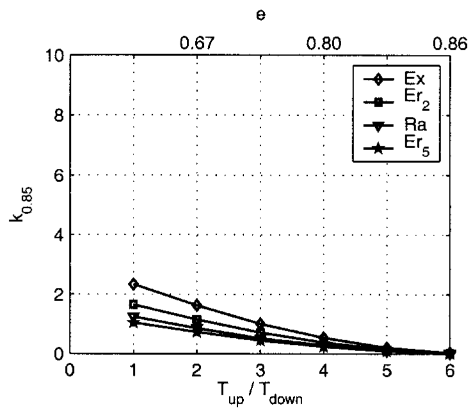
Figure 3. Level of buffering for Erlang, Rayleigh and exponential machines ($M = 2$, $T_{up} = 30$).



(a) PR $= 0.95\mathrm{PR}_\infty$



(b) PR $= 0.9\mathrm{PR}_\infty$



(c) PR $= 0.85\mathrm{PR}_\infty$

Figure 4. Level of buffering for Erlang, Rayleigh and exponential machines ($M = 2$, $T_{\text{up}} = 60$).

- Smaller variability of up- and downtime distributions of the machines leads to smaller level of buffering, LB (since $CV_{\text{ex}} > CV_{\text{Er}_2} > CV_{\text{Ra}} > CV_{\text{Er}_5}$ and the curves are related as shown in figures 3 and 4).
- Smaller machine efficiency, e, requires larger buffering, k_E, to attain the same line efficiency, E.
- Rules-of-thumb for two-machine lines:
  - If PR $= 0.95\mathrm{PR}_\infty$ is desired:

    (α) three-downtime buffer is sufficient for all reliability models if $e \approx 0.85$; and

    (β) zero LB is acceptable if $e \geq 0.94$.
  - If PR $= 0.9\mathrm{PR}_\infty$ is desired:

    (α) one-downtime buffer is sufficient for all reliability models if $e \approx 0.85$; and

(β) zero LB is acceptable if $e \geq 0.88$.

- If $\mathrm{PR} = 0.85\mathrm{PR}_\infty$ is desired, zero LB is acceptable for all reliability models if $e \geq 0.85$.

6.1.3. *Empirical law*

As pointed out above, calculation of k_E is fast and simple for exponential machines and requires lengthy discrete event simulations for Erlang and Rayleigh machines. It would be desirable to have an 'empirical law' that could provide k_E for Erlang and Rayleigh reliability models as a function of k_E for exponential machines. From the data of figures 2–4, one can conclude that such a law can be formulated as follows:

$$\widehat{k_E^A} = \mathrm{CV}_{\mathrm{down}}^A k_E^{\mathrm{ex}}, \tag{13}$$

where A is the reliability model (i.e. the distribution of the downtime — either Erlang or Rayleigh), $\mathrm{CV}_{\mathrm{down}}^A$ is the coefficient of variation of the downtime, $\widehat{k_E^A}$ is the estimate of LB for reliability model A, and k_E^{ex} is the LB for exponential machines defined by (12).

The quality of approximation (13) is illustrated in figure 5 and table 2, where the accuracy of (13) is evaluated in terms of the error, μ_E^A, defined by

$$\mu_E^A = \frac{\widehat{k_E^A} - k_E^A}{k_E^A}. \tag{14}$$

As it follows from these data, $\widehat{k_E^A}$ approximates k_E^A with sufficiently high precision. Moreover, since $\widehat{k_E^A} > k_E^A$, selection of LB according to $\widehat{k_E^A}$ does not lead to a loss of performance.

Empirical law (13) will be used below for the case $M > 2$ as well.

6.2. *M-machine case, $M > 2$*

6.2.1. *Level of buffering as a function of machine efficiency*

For various M, the level of buffering, k_E^{ex}, as a function of $\alpha = T_{\mathrm{up}}/T_{\mathrm{down}}$ or e is shown in figure 6 for $E = 0.95$, 0.9, and 0.85.

Polynomial approximations of these functions for $M = 10$ can be given as follows:

$$\widehat{k}_{0.95}(\alpha) = -0.0035\alpha^3 + 0.1607\alpha^2 - 2.6492\alpha + 20.7627,$$

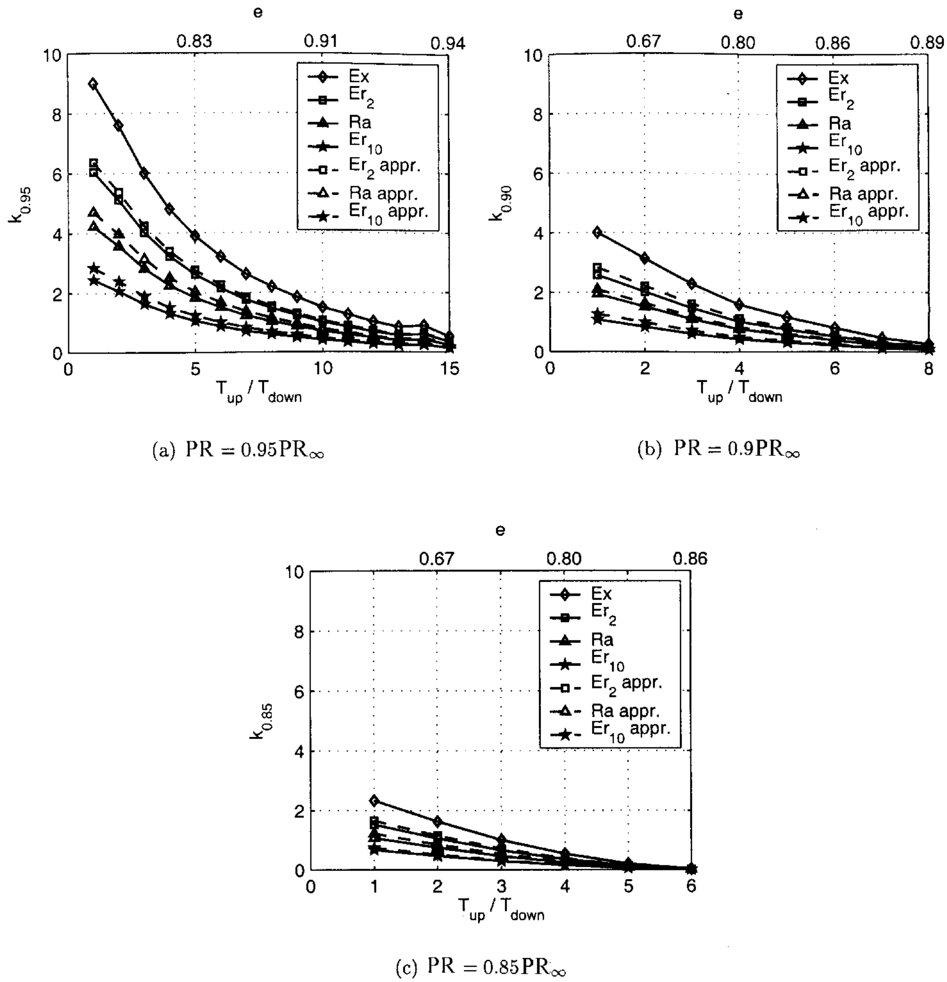
$$\widehat{k}_{0.9}(\alpha) = -0.0015\alpha^3 + 0.068\alpha^2 - 1.156\alpha + 9.8504,$$

$$\widehat{k}_{0.85}(\alpha) = -0.0007\alpha^3 + 0.0361\alpha^2 - 0.6635\alpha + 6.2102.$$

For $M = 10$, graphs for $k_E^{\mathrm{Er}_P}$ and k_E^{Ra} and their approximations according to (13) are illustrated in figure 7 and table 3. These data indicate that the empirical law results in acceptable precision for $M > 2$ as well.

Based on the data of figures 6 and 7, we conclude the following.

- Longer lines require larger level of buffering between each two machines.
- As before, larger machine efficiency requires less buffering.
- Rules-of-thumb for 10-machine lines with exponential machines:



(a) PR = 0.95PR$_\infty$

(b) PR = 0.9PR$_\infty$

(c) PR = 0.85PR$_\infty$

Figure 5. Level of buffering, k_E, for Erlang, Rayleigh and exponential machines and approximation, $\hat{k}_E$, using empirical law (13) ($M = 2$, $T_{\text{up}} = 30$).

- If PR $= 0.95$PR$_\infty$ is desired and machine efficiency $e \approx 0.9$, seven-downtime buffers are required for exponential machines and 4.5-downtime for Erlang ($P \geq 2$) and Rayleigh machines.
- If PR $= 0.9$PR$_\infty$ is desired and machine efficiency $e \approx 0.9$, four-downtime buffer is required for exponential machines and about 3-downtime for Erlang ($P \geq 2$) and Rayleigh machines.
- If PR $= 0.85$PR$_\infty$ is desired and machine efficiency $e \approx 0.9$, 2.5-downtime buffer is required for exponential machines and about 2-downtime for Erlang ($P \geq 2$) and Rayleigh machines.
- Zero LB is not acceptable, even if e is as high as 0.95 and E as low as 0.85.

6.2.2. *Level of buffering as a function of the average uptime*

For $M = 2$, expression (12) states that k_E is independent of T_{up}. No analytical result of this type is available for $M > 2$. Therefore, we verify this property using

(a) $E = 0.95$

Distribution	$e = 0.8$	$e = 0.85$	$e = 0.9$	$r = 0.94$
Rayleigh	0.08	0.07	0.08	0.09
Erlang 2	0.06	0.05	0.05	0.06
Erlang 10	0.14	0.12	0.11	0.15

(b) $E = 0.9$

	$e = 0.7$	$e = 0.75$	$e = 0.8$	$e = 0.85$
Rayleigh	0.11	0.06	0.06	0.05
Erlang 2	0.07	0.09	0.08	0.10
Erlang 10	0.14	0.12	0.11	0.12

(c) $E = 0.85$

	$e = 0.65$	$e = 0.7$	$e = 0.75$	$e = 0.8$
Rayleigh	0.08	0.09	0.05	0.09
Erlang 2	0.09	0.08	0.11	0.07
Erlang 10	0.12	0.15	0.09	0.12

Table 2. Accuracy, μ_E^A, of empirical law (13) as a function of e $(M = 2, T_{\mathrm{up}} = 30)$.

the aggregation procedure of Subsection 5.1. Calculations have been carried out for ten-machine lines with $T_{\mathrm{up}} = 200$ and $T_{\mathrm{up}} = 400$, $T_{\mathrm{up}}/T_{\mathrm{down}} \in \{1, \ldots, 20\}$, and $E \in \{0.85, 0.9, 0.95\}$. As it turned out, k_E for $T_{\mathrm{up}} = 200$ and $T_{\mathrm{up}} = 400$ differ at most by 0.1%. Therefore, we conclude that k_E^{ex} for $M > 2$ does not depend on T_{up} either.

6.2.3. *Level of buffering as a function of the number of machines*

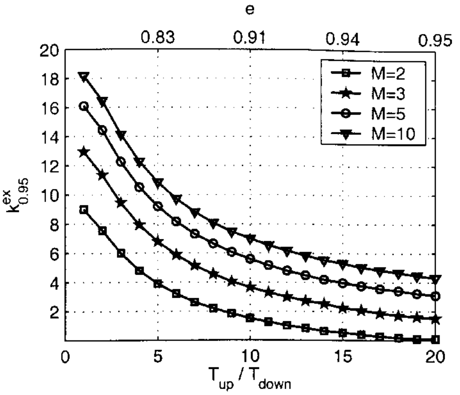
From figures 6 and 7, it is clear that k_E is an increasing function of M. To investigate further this dependency, we calculated k_E as a function of M. The results are shown in figure 8.

Clearly, although k_E^{ex} is an increasing function of M, the rate of increase is exponentially decreasing and saturates at about $M = 10$. This happens, perhaps, due to the fact that the machines, separated by nine appropriately selected buffers become, to a large degree, decoupled.

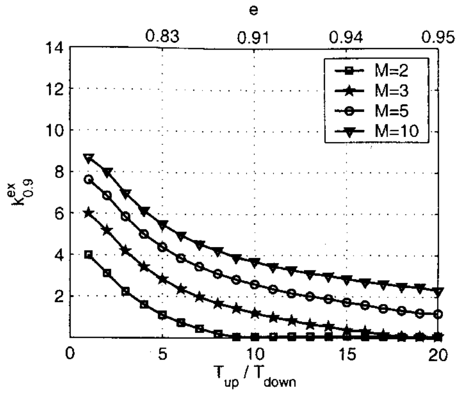
The curves shown in figure 8 have a convenient exponential approximation. For instance, if $e = 0.9$, these approximations are:

$$k_{0.95}^{\mathrm{ex}}(M) = 1.8 + 6.255\left(1 - \exp\left(-\frac{(M-2)}{3}\right)\right),$$

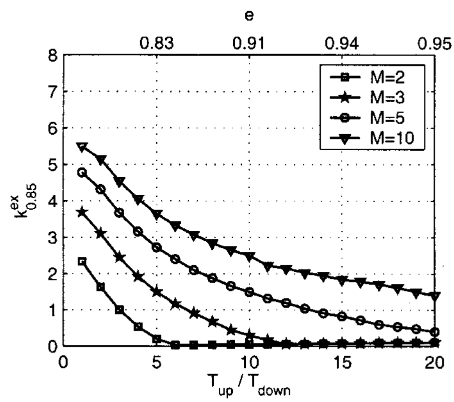
$$k_{0.90}^{\mathrm{ex}}(M) = 0.045 + 4.365\left(1 - \exp\left(-\frac{(M-2)}{3.5}\right)\right),$$



(a) PR = 0.95PR$_\infty$



(b) PR = 0.9PR$_\infty$



(c) PR = 0.85PR$_\infty$

Figure 6. Level of buffering for exponential machines with various M.

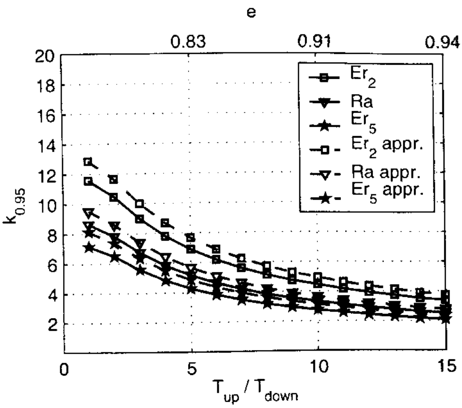
$$k^{\mathrm{ex}}_{0.85}(M) = 0.045 + 3.061\left(1 - \exp\left(-\frac{(M-2)}{3.75}\right)\right),$$
$$M \in [2, 3, \ldots).$$

The quality of this approximation is illustrated in figure 9.

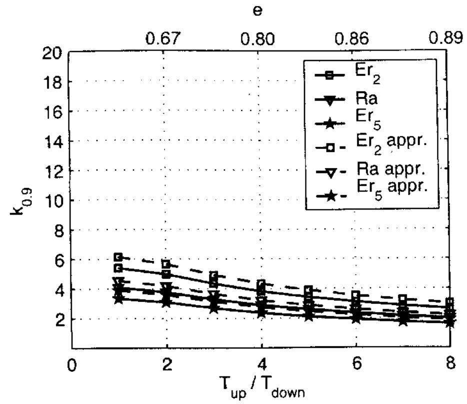
Figures 8 and 9 characterize $k^{\mathrm{ex}}_E(M)$ for the exponential machines. Empirical law (13) can be invoked to evaluate $k_E(M)$ for Erlang and Rayleigh machines as well. The behaviour of $k^{\mathrm{Er}_P}_E(M)$ and $k^{\mathrm{Ra}}_E(M)$ obtained by simulation, and $\hat{k}^{\mathrm{Er}_P}_E(M)$ and $\hat{k}^{\mathrm{Ra}}_E(M)$ obtained from (13) is shown in figure 10; its accuracy (14) is characterized in table 4. The conclusion is that empirical law (13) results in acceptable precision for $M > 2$.

Based on the above results, we arrive at the following conclusions.

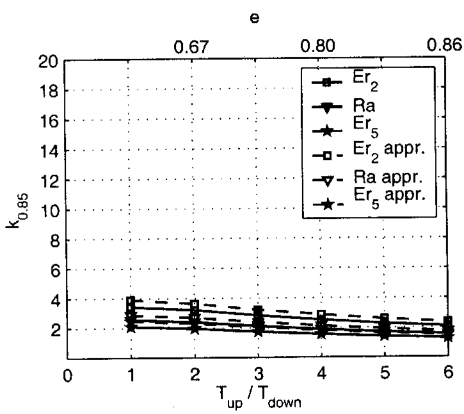
- Although longer lines require larger level of buffering, the increase is exponentially decreasing as a function of M.



(a) $PR = 0.95PR_\infty$



(b) $PR = 0.9PR_\infty$



(c) $PR = 0.85PR_\infty$

Figure 7. Level of buffering , k_E, for Erlang and Rayleigh machines and approximation, $\widehat{k_E}$, using empirical law (13) ($M = 10$, $T_{\mathrm{up}} = 30$).

- Roughly speaking, buffering necessary for $M = 10$ is sufficient to accommodate downtime in all lines with $M > 10$.
- Rules-of-thumb established in Subsection 6.2.1 remain valid for Erlang and Rayleigh machines as well if the level of buffering is modified by the coefficient of variation of the downtime.

6.3. *Production losses for* $k = 1$

As it was pointed out above, one-downtime rule is often used by production line designers. Performance of 10-machine lines with this buffer allocation is characterized in figure 11. As it follows from this figure, if $e = 0.9$, throughput losses are about 30% of PR_∞ if machine reliability is exponential and about 25% if it is Er_2. Thus, the 'one-downtime' rule may not be advisable if high line efficiency is pursued.

(a) $E = 0.95$

Distribution	$e = 0.8$	$e = 0.85$	$e = 0.9$	$r = 0.94$
Rayleigh	0.09	0.05	0.06	0.08
Erlang 2	0.11	0.12	0.14	0.13
Erlang 10	0.12	0.10	0.09	0.11

(b) $E = 0.9$

	$e = 0.7$	$e = 0.75$	$e = 0.8$	$e = 0.85$
Rayleigh	0.07	0.08	0.09	0.05
Erlang 2	0.14	0.11	0.08	0.16
Erlang 10	0.09	0.10	0.12	0.11

(c) $E = 0.85$

	$e = 0.65$	$e = 0.7$	$e = 0.75$	$e = 0.8$
Rayleigh	0.08	0.12	0.11	0.10
Erlang 2	0.12	0.12	0.15	0.16
Erlang 10	0.08	0.12	0.12	0.09

Table 3. Accuracy, μ_E^A, of empirical law (13) as a function of e ($M = 2, T_{\text{up}} = 30$).

7. Extension: non-identical machines

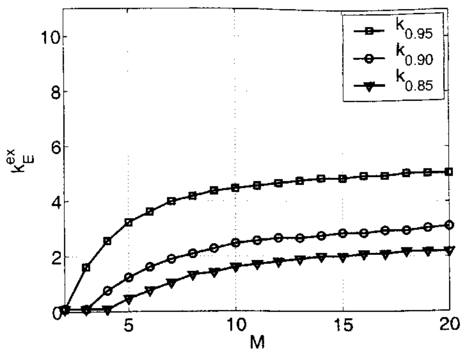
7.1. *Description of machines*

Identical machines imply that up- and downtime obey the same reliability model and the average uptime (respectively, downtime) of all machines is the same. Non-identical machines mean that either or both of these assumptions is violated. The goal of this section is to extend the results of section 6 to non-identical machines assuming, however, that the efficiency, e, of all machines is the same. This assumption is made to account for the fact that in most practical cases all machines of a production line are roughly of the same efficiency. To simplify the presentation, we consider only two-machine lines here.

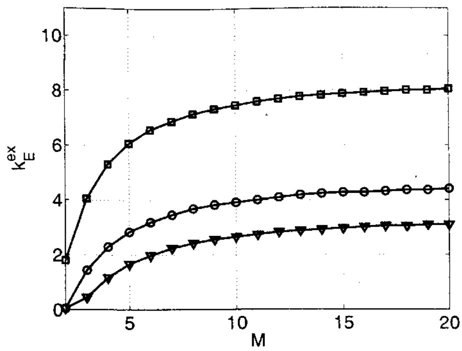
In this section, each machine, m_i, $i = 1, 2$, is denoted by a pair $\{A(p_i), B(r_i)\}$, where the first symbol, $A(p_i)$ (respectively, the second symbol $B(r_i)$) denotes the distribution of the uptime (respectively, downtime) defined by parameter p_i (respectively, r_i); the subscript i indicates whether the first or second machine is addressed. For instance, $\{\text{Er}_5(p_2), \text{Ex}(r_2)\}$ denotes the second machine of a two-machine line with the uptime being distributed according to the Erlang distribution with five stages, defined by parameter p_2, and the downtime distributed according to the exponential distribution, defined by parameter r_2. Obviously, in this case the average up- and downtime of the second machine are $5/p_2$ and $1/r_2$, respectively. Note that in these notations, the systems considered in section 6 consist of machines $\{A(p_i), A(r_i)\}$, $p_i = p$, $r_i = r$, $\forall i = 1, \ldots, M$.

7.2. *Cases analysed*

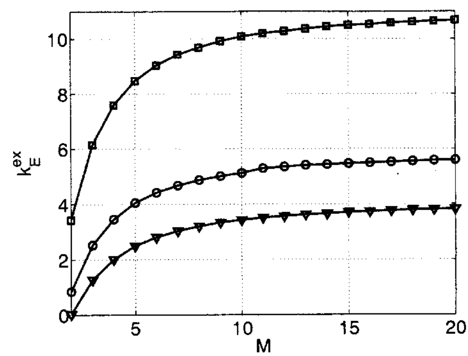
To investigate the properties of LB in production lines with non-identical machines, the following five cases have been analysed:



(a) $e = 0.95$



(b) $e = 0.9$



(c) $e = 0.85$

Figure 8. Level of buffering k_E^{ex} as a function of M.

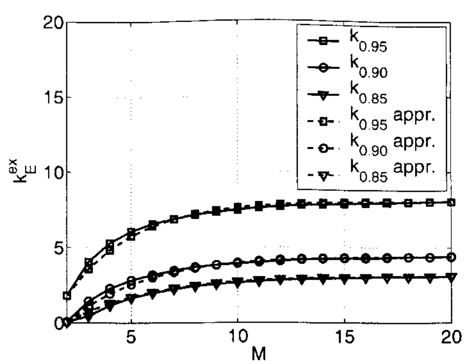
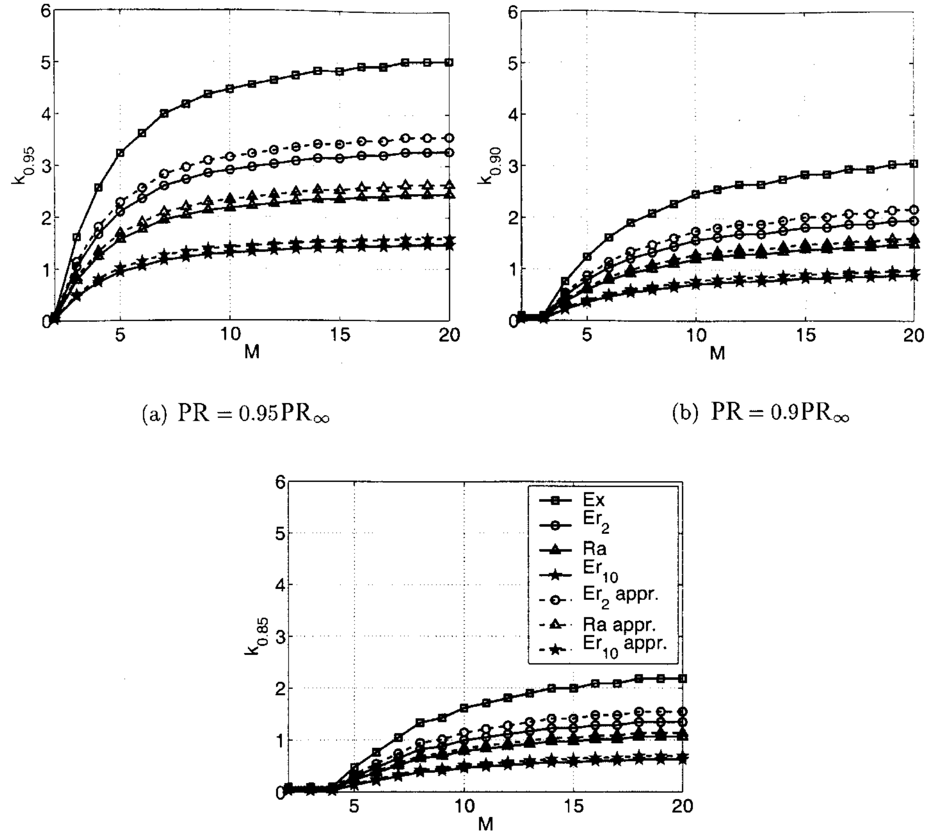


Figure 9. Approximations of k_E^{ex} for $e = 0.9$.



(a) $\mathrm{PR} = 0.95\mathrm{PR}_\infty$

(b) $\mathrm{PR} = 0.9\mathrm{PR}_\infty$

(c) $\mathrm{PR} = 0.85\mathrm{PR}_\infty$

Figure 10. Levels of buffering $k_E^{\mathrm{Er}_P}$, k_E^{Ra}, and their approximations according to (13) as a function of M ($T_{\mathrm{up}} = 30$, $e = 0.95$).

Case 1: Non-identical T_{up} and T_{down}. Specific systems analysed were:

$$\{\mathrm{Ex}(p_1), \mathrm{Ex}(r_1)\}, \quad \{\mathrm{Ex}(p_2), \mathrm{Ex}(r_2)\},$$

$$\{\mathrm{Ra}(p_1), \mathrm{Ra}(r_1)\}, \quad \{\mathrm{Ra}(p_2), \mathrm{Ra}(r_2)\},$$

$$\{\mathrm{Er}_2(p_1), \mathrm{Er}_2(r_1)\}, \quad \{\mathrm{Er}_2(p_2), \mathrm{Er}_2(r_2)\},$$

$$\{\mathrm{Er}_5(p_1), \mathrm{Er}_5(r_1)\}, \quad \{\mathrm{Er}_5(p_2), \mathrm{Er}_5(r_2)\},$$

$$p_2 = 2p_1, r_2 = 2r_1.$$

Case 2: Non-identical up- and downtime distribution laws. Systems considered here were:

$$\{\mathrm{Ex}(p_1), \mathrm{Ra}(r_1)\}, \quad \{\mathrm{Ex}(p_2), \mathrm{Ra}(r_2)\},$$

$$\{\mathrm{Er}_5(p_1), \mathrm{Er}_2(r_1)\}, \quad \{\mathrm{Er}_5(p_2), \mathrm{Er}_2(r_2)\},$$

(a) $E = 0.95$

Distribution	$M = 5$	$M = 10$	$M = 15$	$M = 20$
Rayleigh	0.09	0.09	0.07	0.11
Erlang 2	0.10	0.08	0.07	0.10
Erlang 10	0.11	0.09	0.12	0.12

(b) $E = 0.9$

	$M = 5$	$M = 10$	$M = 15$	$M = 20$
Rayleigh	0.12	0.11	0.15	0.08
Erlang 2	0.14	0.08	0.09	0.11
Erlang 10	0.09	0.09	0.10	0.06

(c) $E = 0.85$

	$M = 5$	$M = 10$	$M = 15$	$M = 20$
Rayleigh	0.10	0.07	0.09	0.14
Erlang 2	0.14	0.11	0.09	0.16
Erlang 10	0.16	0.12	0.14	0.15

Table 4. Accuracy, μ_E^A, of empirical law (13) as a function of M ($e = 2, T_{\text{up}} = 30$).

$$\{\text{Er}_2(p_1), \text{Er}_5(r_1)\}, \quad \{\text{Er}_2(p_2), \text{Er}_5(r_2)\},$$

$$p_2 = p_1, \quad r_2 = r_1.$$

Case 3: Non-identical up- and downtime distribution laws, non-identical T_{up} and T_{down}. Systems studied here were:

$$\{\text{Ex}(p_1), \text{Ra}(r_1)\}, \quad \{\text{Ex}(p_2), \text{Ra}(r_2)\},$$

$$\{\text{Er}_5(p_1), \text{Er}_2(r_1)\}, \quad \{\text{Er}_5(p_2), \text{Er}_2(r_2)\},$$

$$\{\text{Er}_2(p_1), \text{Er}_5(r_1)\}, \quad \{\text{Er}_2(p_2), \text{Er}_5(r_2)\},$$

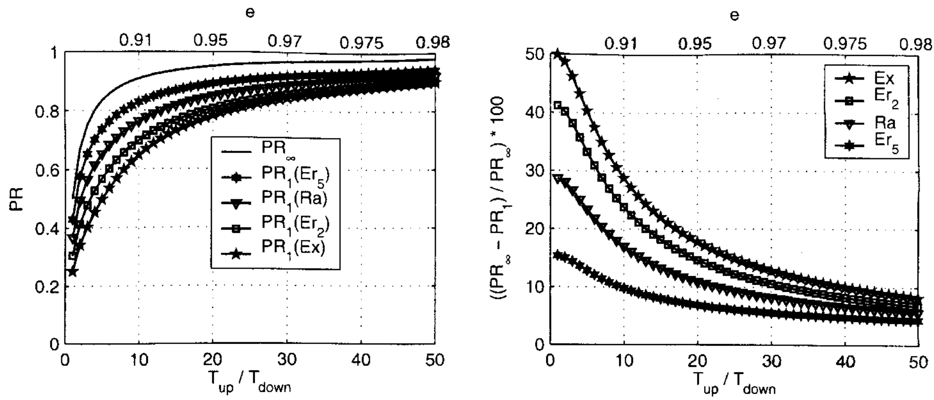
$$p_2 = 2p_1, \quad r_2 = 2r_1.$$

Case 4: Non-identical uptime distribution laws, non-identical downtime distribution laws. The systems analysed were:

$$\{\text{Ex}(p_1), \text{Ra}(r_1)\}, \quad \{\text{Er}_5(p_2), \text{Ex}(r_2)\}, \quad \frac{1.2533}{r_1} = \frac{1}{r_2},$$

$$\{\text{Er}_2(p_1), \text{Er}_5(r_1)\}, \quad \{\text{Er}_5(p_2), \text{Er}_2(r_2)\}, \quad \frac{5}{r_1} = \frac{2}{r_2},$$

$$\{\text{Ra}(p_1), \text{Ex}(r_1)\}, \quad \{\text{Er}_3(p_2), \text{Ra}(r_2)\}, \quad \frac{1}{r_1} = \frac{1.2533}{r_2}.$$



(a) PR_∞ and PR_1 for various models

(b) Losses: $\frac{\mathrm{PR}_\infty - \mathrm{PR}_1}{\mathrm{PR}_\infty}$

Figure 11. Performance of 10-machine lines with $k_E^{\mathrm{ex}} = 1$ as a function of e.

Case 5: Non-identical uptime distribution laws, non-identical downtime distribution laws and non-identical T_{up} and T_{down}. The systems investigated here were:

$$\{\mathrm{Ex}(p_1), \mathrm{Ra}(r_1)\}, \quad \{\mathrm{Er}_5(p_2), \mathrm{Ex}(r_2)\}, \quad \frac{1.2533}{r_1} = \frac{2}{r_2},$$

$$\{\mathrm{Er}_2(p_1), \mathrm{Er}_5(r_1)\}, \quad \{\mathrm{Er}_5(p_2), \mathrm{Er}_2(r_2)\}, \quad \frac{5}{r_1} = \frac{4}{r_2},$$

$$\{\mathrm{Ra}(p_1), \mathrm{Ex}(r_1)\}, \quad \{\mathrm{Er}_3(p_2), \mathrm{Ra}(r_2)\}, \quad \frac{1}{r_1} = \frac{2.5066}{r_2}.$$

7.3. *Results obtained*

We provide here only the summary of the results obtained. More details can be found in Enginarlar *et al.* (2000).

The main result can be formulated as follows. The selection of LB for a two-machine line with non-identical machines can be reduced to the selection of LB for a two-machine line with identical machines, provided that the latter is defined appropriately. Specifically, consider the production line $\{A(p_1), B(r_1)\}$, $\{C(p_2), D(r_2)\}$. Without loss of generality, assume that the first machine has the largest average downtime, i.e. $T_{\mathrm{down}_1} > T_{\mathrm{down}_2}$, and the second machine has the largest coefficient of variation of the downtime, i.e. $\mathrm{CV}_{\mathrm{down}_1} < \mathrm{CV}_{\mathrm{down}_2}$. Assume that the LB sought is in units of the largest average downtime, i.e.

$$k_E = \frac{N_E}{T_{\mathrm{down}_1}}.$$

Then, the level of buffering for the line

$$\{A(p_1), B(r_1)\}, \{C(p_2), D(r_2)\}$$

can be selected as the level of buffering of the following production line with identical machines

$$\{D(p), D(r)\}, \{D(p), D(r)\},$$

where p and r are selected as follows:

$$E_{D(r)}\{t_{\text{down}}\} = E_{B(r_1)}\{t_{\text{down}}\}$$
$$E_{D(p)}\{t_{\text{up}}\} = E_{A(p_1)}\{t_{\text{up}}\}.$$

Here $E_{R(v)}\{t\}$ denotes the expected value of random variable t distributed according to distribution R defined by parameter v.

Thus, selecting LB for two-machine lines with non-identical machines is reduced to the problem of selecting LB for lines with identical machines, the solution of which is given in subsection 6.1.

8. Conclusions

Based on this study, the following rules-of-thumb for selecting the level of buffering, k_E, in serial production lines as a function of machine efficiency, e, line efficiency, E, number of machines, M, and the downtime coefficient of variation, CV_{down}, can be provided.

(1) If all machines are identical and obey the exponential reliability model, k_E^{ex} can be selected as indicated in table 5. If the number of machines in the system is substantially less than 10, the level of buffering can be reduced by using the data of figure 8.

(2) If the machines are identical but not exponential, all k_E^{ex} from table 5 should be multiplied by the coefficient of variation of the downtime, CV_{down}. For machines with Erlang and Rayleigh reliability models, this leads to about 50% reduction of buffer capacity. This might justify the effort for evaluating not only the average value of the downtime but also its variance.

(3) If the machines are not identical, the capacity of the buffer between each pair of consecutive machines can be chosen according to

$$N_i = \lceil k_E^{\text{ex}} \cdot \max\{\text{CV}_{\text{down}_{i-1}}, \text{CV}_{\text{down}_i}\} \cdot \max\{T_{\text{down}_{i-1}}, T_{\text{down}_i}\}\rceil, \quad i = 1, \ldots, M-1,$$

where k_E^{ex} is selected from table 5.

It should be pointed out that this paper does not address the issue of which line efficiency should be pursued – 0.95, 0.90 or 0.85. However, given the data of table 5, it is reasonable to conclude that $E = 0.95$ might require too much buffering, as far as practical considerations are concerned (unless the downtime variability is very small). Efficiency $E = 0.85$ might be too low for many industrial situations. Therefore, it seems reasonable that the second column of table 5 provides the

e	$E = 0.85$	$E = 0.90$	$E = 0.95$
0.85	3.5	5	10
0.90	2.5	4	7
0.95	1.5	2.5	4.5

Table 5. Level of buffering, k_E^{ex}, as a function of machine and line efficiency.

most important practical information. This information defines how 'lean' a production line could be to result in a reasonable performance.

Acknowledgements

The authors are grateful to Professor J. A. Buzacott for valuable advice in connection with his paper (1967). The helpful comments of anonymous reviewers are also acknowledged. The work was supported by NSF Grant No. DMI-9820580.

Appendix 1: Notation

b_i	ith buffer,
CV	coefficient of variation,
e	machine efficiency,
ex	exponential distribution,
E	production line efficiency,
$E_{R(v)}\{t\}$	expected random variable t distributed according to $R(v)$,
Er	Erlang distribution,
k_E	smallest level of buffering necessary to achieve line efficiency E,
LB	level of buffering,
m_i	ith machine,
M	number of machines in the line,
N	buffer capacity,
N_E	buffer capacity necessary to achieve line efficiency E,
p	parameter of the uptime distribution,
PR	production rate,
PR_∞	production rate when the capacity of all buffers is infinite,
PR_k	production rate when the level of buffering is k,
Q	function defining the aggregation procedure,
r	parameter of the downtime distribution,
Ra	Rayleigh distribution,
s	step of the aggregation procedure,
T_{up}	average machine uptime,
T_{down}	average machine downtime,
σ	standard deviation, and
μ	accuracy of the empirical law.

References

Altiok, T., 1985, Production lines with phase – type operation and repair times and finite buffers, *International Journal of Production Research,* **23**, 489–498.

Buzacott, J. A., 1967, Automatic transfer lines with buffer stocks. *International Journal of Production Research*, **5**, 183–200.

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

Chiang, S.-Y., Kuo, C.-T. and Meerkov, S. M., 2000, DT-bottlenecks in serial production lines: theory and application. *IEEE Transactions on Robotics and Automation*, **16**, 567–580.

Conway, R., Maxwell, W., McClain, J. O. and Thomas, L. J., 1988, The role of work-in-process inventory in serial production lines. *Operations Research*, **36**, 229–241.

Dallery, Y., David, R. and Xie, X., L., 1989, Approximate analysis of transfer lines with unreliable machines and finite buffers. *IEEE Transactions on Automatic Control*, **34**, 943–953.

Enginarlar, E., Li, J., Meerkov, S. M. and Zhang, R., 2000, Buffer capacity for accommodating machine downtime in serial production lines. Control Group Report No. CGR-00-07, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor.

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

Gershwin, S. B. and Schor, J. E., 2000, Efficient algorithms for buffer space allocation. *Annals of Operations Research*, **93**, 117–144.

Glasserman, P. and Yao, D. D., 1996, Structured buffer-allocation problems. *Discrete Event Dynamic Systems*, **6**, 9–41.

Hillier, F. S. and So, K. C., 1991, The effect of machine breakdowns and internal storage on the performance of production line systems. *International Journal of Production Research*, **29**, 2043–2055.

Ho, Y. C., Eyler, M. A. and Chien, T. T., 1979, A gradient technique for general buffer storage design in a production line. *International Journal of Production Research*, **7**, 557–580.

Jacobs, D. A., 1993, Improvability of production systems: theory and case studies. PhD thesis, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor.

Jacobs, D. and Meerkov, S. M., 1995a, A system-theoretic property of serial production lines: improvability. *International Journal of Systems Science*, **26**, 755–785.

Jacobs, D. and Meerkov, S. M., 1995b, Mathematical theory of improvability for production systems. *Mathematical Problems in Engineering*, **1**, 99–137.

Law, A. M. and Kelton, W. D., 1991, *Simulation, Modeling and Analysis* (New York: McGraw-Hill).

Sevastyanov, B. A., 1962, Influence of storage bin capacity on the average standstill time of a production line. *Theory of Probability and Its Applications*, **7**, 429–438.

Vladzievskii, A. P., 1950, The theory of internal stocks and their influence on the output of automatic lines. *Stanki i Instrumenty*, **21**, 4–7.

Vladzievskii, A. P., 1951, The theory of internal stocks and their influence on the output of automatic lines. *Stanki i Instrumenty*, **22**, 16–17.